UNIVERSITY OF LEEDS

BIOMOLECULAR RECOGNITION AND PROTEIN DYNAMICS IN AMYLOID AGGREGATION

Theodoros Karamanos

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

The University of Leeds
Astbury Centre for Structural Molecular Biology

September 2013

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.

"In order to understand the world,
one has to turn away from it on occasion"

~Albert Camus

ABSTRACT

The deposition of proteinaceous aggregates known as amyloid fibrils is related to the onset of numerous human pathologies termed protein folding diseases. Even though the amyloid fold is proposed to lie in an energy minimum, the process of amyloid aggregation is remarkably slow both *in vivo* and *in vitro*. This arises from the necessity to overcome high energy barriers that lead to the formation of intermolecular interactions that promote further polymerisation and/or the generation of aggregation-prone intermediate species. Structural characterisation of protein association in the early stages of amyloid assembly, and most importantly, its link to protein dynamics and/or conformational changes remains poorly characterised due to the heterogeneity and the transient nature of the interactions involved. In this thesis, using NMR techniques sensitive enough to detect species that are lowly populated in solution, the interaction surfaces that lead to inhibition, promotion or nucleation of amyloid assembly by the protein β_2-microglobulin (β_2m) were mapped in atomic detail. As visualised by paramagnetic relaxation enhancement and analysis of chemical shift perturbation data using human β_2m and its murine homologue, inhibition and promotion of assembly involve similar interfaces but the interactions differ in details sufficient to determine the fate of polymerisation. Inhibition occurs by the formation of specific, kinetically trapped off-pathway species, the accumulation of which results in prolongation of the onset of amyloid, while promotion involves more diffuse, weaker protein-protein interactions that cause destabilisation/conformational changes in both interacting partners. By contrast, nucleation progresses through a more complicated multiple-site binding, reminiscent of oligomer formation. Overall, the results reveal an atomic level description of the fine interplay between structure, affinity and dynamics that dictates the course of amyloid formation by proteins that are very similar both in terms of sequence and structure and provide a general framework of the mechanism by which β_2m (hetero)polymerises.

ACKNOWLEDGEMENTS

There are numerous people that, one way or another, contributed to this work over the years, to which all I would like to express my gratitude.

My main supervisor, Sheena Radford for her support along the way. Her interest and commitment to good science has been a constant motivation throughout the years.

Special thanks to Arnout Kalverda and Gary Thompson for their invaluable help on NMR and their general knowledge about science, which was critical for the outcome of this thesis. Steve Homans and Jean Baum for helpful advice on NMR experiments, Timo Eichner for introducing me to the world of the Radford lab and Isabella Felli and her team in Florence for performing the direct carbon detection experiments.

I would like to thank Claire Sarell, Rebecca Thompson, Maya Pandya, Clare Pashley, Kevin Tipping, Alessandro Sicorello, Nasir Khan and Girish Tampi. All of you contributed to the outcome of this research and/or my time in Leeds.

Finally, I would like to thank my family and Maria for their love and support over the years.

CONTENTS

II. THEORY, MATERIALS & METHODS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

AFM	Atomic Force Microscopy
ApoE	Apolipoprotein-E
ASA	Accessible Surface Area
AUC	Analytical Ultracentrifugation
Aβ	Amyloid-beta peptide
BSA	Buried Surface Area
CD	Circular Dichroism
CPMD rd	Carr-Pursell-Meiboom-Gill relaxation dispersion
CSP	Chemical Shift Perturbation
DNAse-I	Deoxy-ribonuclease-I
DRA	Dialysis-Related Amyloidosis
EM	Electron Microscopy
ESI-MS	Electron-Spray-Ionisation Mass Spectrometry
FTIR	Fourier-Transform Infrared
H/D exchange	Hydrogen-Deuterium Exchange
HIV	Human Immunodeficiency Virus
HSQC	Heteronuclear Singe Quantum Coherence
Hβ_2m	Human β_2m
IDP	Inrinsically Disordered Protein
IPTG	Isopropyl β-D-1-thiogalactopyranoside
MHC-I	Major Histocompatibility complex-I
MTSL	S-(2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl-oxyl)methyl methanesulfonothioate
Mβ_2m	Murine β_2m
NMR	Nuclear Magnetic Resonance
nOe	nuclear Overhauser effect
PBD	Protein Data Bank
PMSF	Phenylmethanesulfonyl fluoride
ppm	parts per million
PRE	Paramagnetic Relaxation Enhancement
PSC	Pseudo Contact Shift
PTI	Pancreatic Trypsin Inhibitor

PTS	PhosphoTransferase System
RDC	Residual Dipolar Coupling
RMSD	Root Mean Square Deviation
SEM	Standard Error of the Mean
SH3	Src-homology domain 3
SOD-1	Superoxide Dismutase-1
SSB proteins	Single Strand Binding proteins
ssNMR	Solid state NMR
TEM	Transmission Electron Microscopy
TTR	Transthyretin
β_2m	Beta-2 microglobulin
$\Delta\delta\omega$	Chemical shift difference

CHAPTER I

INTRODUCTION

1.1 A historical perspective

This thesis builds up on the work of numerous previous researchers that inspired generations of scientists and developed the field of Protein Science to the post-genomic era that we are currently experiencing. The study of protein molecules started 223 years ago, but the advances made this relatively short period of time are truly extraordinary.

It was in 1838 when Jons Jacob Berzelius, in a letter to Gerardus Johannes Mulder, introduced for the first time the term protein to describe a new distinct class of biomolecules. As its name denotes (greek: πρωτείος, superior) this class of molecules was first thought to be "the primitive or principal substance of animal nutrition' (Jons Jacob Berzelius, 1838). However, the first types of proteins were identified back in 1789 by Antoine Francois, who was successful in characterising albumin, fibrin and gelatin among others[1]. By 1902, proteins were known to be linear polymers, consisting of building blocks connected to each other (Emil Fisher and Franz Homeister). However, the pursuit of identifying these 20 blocks took almost 130 years[2]. In 1809 Mulder managed to determine an almost correct molecular weight (131Da) for leucine[3] that had already been identified as a protein degradation product by the French pharmacist Henry Braconnet, while threonine was the last to be identified in 1936[4]. It was only in 1949 that Frederick Sanger managed to correctly determine the primary sequence of insulin, showing undoubtedly that proteins consist of amino acids and not other branched chains or colloids[5]. For this achievement, he won the Nobel prize in 1958.

In early 1930s, it was William Astbury who proposed the first models for the structure of proteins in 3D space. While in Leeds, based on his X-ray diffraction work on fibrous protein molecules he suggested that proteins can adopt two forms: the α-form, which is helical and a more extended β-form[6,7]. Astbury's models (although incorrect in details but accurate in essence), inspired Linus

Pauling in 1951 to propose for the first time the α-helix and the β-sheet as the secondary structure elements of all proteins, keeping Astbury's nomenclature[8-10]. Pauling's victory came in 1958 when John Kendrew and Max Perutz 10 years later solved the first crystal structures of myoglobin[11] and hemoglobin respectively[12] (sharing the Nobel prize in 1962). Following these discoveries, an outburst of X-ray crystallography with protagonists Max Perutz, William and Laurence Bragg, William Astbury and John Desmont Bernal, led to what now (May 2013) is 90611 biomolecular structures in the protein data bank repository (PDB)[13] (determined also by other methods such as nuclear magnetic resonance spectroscopy-NMR).

The blossoming of X-ray crystallography in the last century reinforced Fischer's idea of a 'lock and key' mechanism for enzymatic substrate binding[14,15]. All proteins seemed to be static. Even one of the first structures of a complex, that of hen-egg lysozyme with a trisaccharide (1966) showed a remarkable alignment between the side-chains of residues in the active site and the ligand[16-18]. In opposition to common belief at the time, Monod, Wyman and Changeux, as early as in 1965, proposed that proteins can adopt numerous conformations in solution, and that ligand binding shifts the equilibrium to the active state (conformational selection theory)[19]. It was only in 1984, when advances in NMR spectroscopy allowed Kurt Wuthrich to determine the first structure of a protein in solution (that of bull seminal protease inhibitor – BUSI)[20,21] showing that proteins are inherently dynamic entities and shifting the 'lock and key' mechanism to include protein dynamics[22].

Even though the structure of haemoglobin revealed four protein subunits that collaborate to achieve their function, the role of protein-protein interactions in expressing protein functionality, or even establishing new protein roles, was (until very recently) underestimated[23]. The domination of the dogma 'one gene/one enzyme/one function' proposed by Beadle and Tatum[24] after their analysis of a series of auxotrophic mutants in *Neurospora Crassa* misled the field to believe that a linear combination between genes and phenotypes is sufficient to explain nature's complexity in terms of protein function. Protein-protein interactions were thought to be artifacts and were not studied until the

1970s. However, key molecular processes such as DNA replication, transcription and translation[25] were soon identified to involve more than one partner and gathered scientific interest[26].

Soon after, questions such as how biomolecules form stable interactions, what is the origin of their affinity and their specificity, what are the advantages over single subunit systems and which physical properties govern their formation started to emerge. Back in 1940, with virtually no structural or biochemical data available, Linus Pauling along with Max Delbruck had another very good guess[27]. They had foreseen that apart from the formation of covalent bonds, non-bonded interactions such as van der Waals attraction or repulsion, electrostatic interactions and hydrogen bond formation could also be involved in processes such as protein synthesis and folding, highlighting at the same time the importance of surface complementarity. 75 years later, the outline of this statement remains unchanged with the addition of the hydrophobic effect to the list. An overview of the history of protein research is given in Figure 1.1.

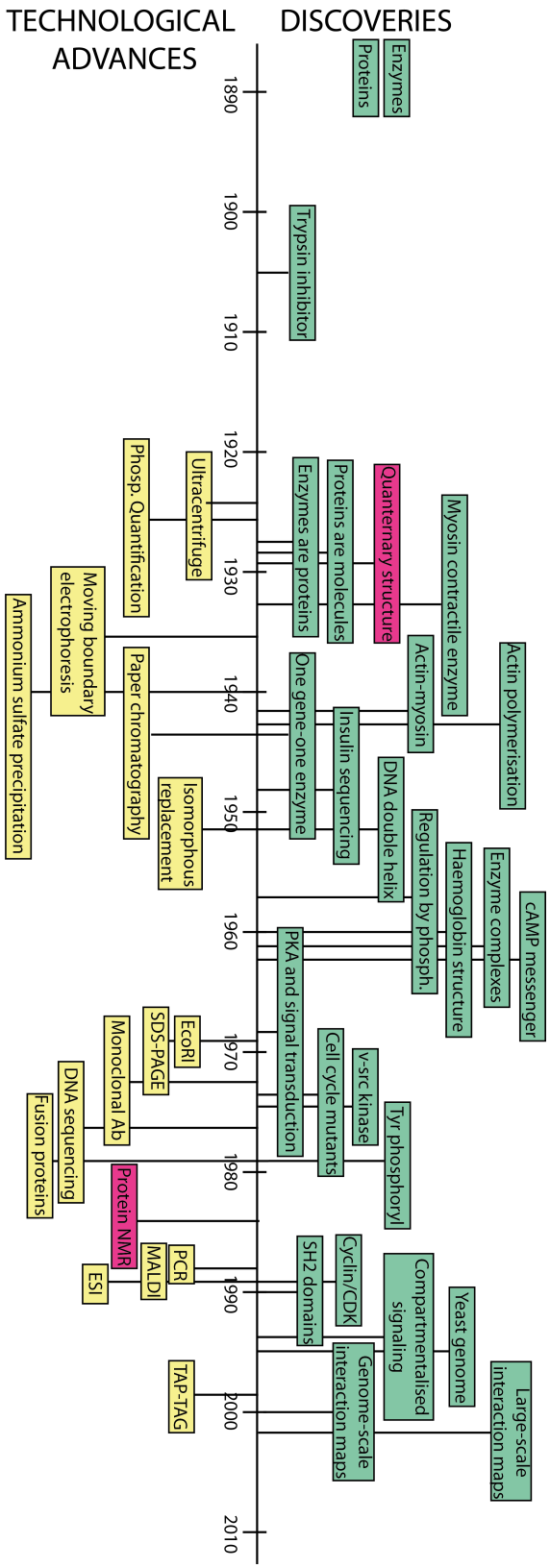

Figure 1.1: Timetable of protein history.
In the top panel key discoveries in the field of Protein Science are shown, accompanied with emerging technological break-throughs in the lower panel. Adapted from[26].

Today with more than 28000[28] structures of protein complexes available and numerous studies reporting on the affinity and energetics of protein binding, our knowledge of protein recognition is much enhanced. However, we are still far from understanding and even more important, predicting how proteins interact with other biomolecules.

In this thesis I will try to connect protein-protein recognition and protein dynamics to protein assembly *in vitro* but using an *in vivo* relevant system under close to physiological conditions. While both binding and dynamics are now considered essential aspects in any biological/cellular process, the link between them and how they affect each other fall into an area of research that is still developing. In protein aggregation, which constitutes the general focus of this thesis, thermodynamic and kinetic stability of the protein precursor, as assessed by its molecular motions are well known to modulate the amyloid propensity[29-31]. The elucidation of the structural properties of the monomeric precursor that relate to its increased amyloidogenicity as well as the characterisation of the late steps in aggregation that usually involve a (cytotoxic) inter-conversion of oligomeric species to fibrils, have gathered much scientific interest. However, the events in the early stages in this process that require the association of low molecular weight species remain elusive. Yet, all aggregation proceeds through these steps and understanding the nature of the interactions involved will help prevent the onset of the disease and the appearance of the cytotoxic entity that is more likely to happen later in the reaction.

Chapter I starts by giving an overview of protein-protein interactions and their physiological importance focusing on large macromolecular assemblies that fulfil crucial roles inside the cell. The next section moves on to describe the essential physicochemical properties that formulate the basis of macromolecular recognition and the scientific nomenclature used in the field. In contrast with the specific/obligate interactions their transient counterparts often represent hidden/invisible states on pathway to the formation of a specific complex[32]. Due to their transient nature, these interactions depend highly on, and/or influence the dynamical properties of the partners involved

and are inaccessible to standard biophysical tools. Thus, an introduction to the importance of transient macromolecular associations, the way they affect protein dynamics and newly developed methods to get insights into them is given in the following sections. To put all this information in the context of amyloid aggregation which is the general framework of this thesis, the relevance of transient protein-protein interactions in the early stages of fibril formation is then discussed by giving examples of the recent literature on the structural and kinetic properties of oligomeric species that act as precursors to amyloid. Finally an introduction to β_2m, the model system of this study is given in the last part of the opening Chapter.

1.2 The value of molecular machines

Klotz et al, in their 1970 paper[33] described four significant advantages of multi domain proteins: first, it is much more efficient and cost-effective to have monomeric copies of a protein associating with each other rather than having a big multimeric moiety with individual domains covalently linked between them. For instance it would be much more demanding for microtubules to assemble and achieve their function if they were comprised by a single tubulin–like polyprotein, rather than having individual tubulin domains coming together. In a similar manner, the same point stands even in the level of DNA. It is much more economic to have one copy of a gene that encodes for a protein that interacts with others to fulfill different roles in the cell rather than having multiple copies of the same gene for all the different protein functions it may encode. Additionally, translation of large proteins is prone to translation errors, which are vastly reduced if only translation of smaller subunits is required. Third, the relationship between protein synthesis and protein function is much more efficient for macromolecular assemblies rather than larger entities. This strategy allows for synthesis in one position in the cell and assembly at a different one, thus compartmentalisation of protein function is achieved much more readily. Finally, homo-multimeric proteins are much more evolutionary favoured, since a single mutation in e.g. the active site of a homodimeric enzyme, affects both partners, increasing (or decreasing) the rate of the enzymatic reaction two-fold rather than one-fold. Furthermore, it is

much easier for a multimeric protein to change its subunit composition in order to change activity or modulate the cellular response. Examples include haemoglobin that changes its composition from α2γ2 in its fetal form to α2β2, to form the adult version of the protein[34], the multiple interactions of transcription factors with RNA polymerase[35] and many more[36,37].

After all this, it is not surprising that some of the key cellular processes are accomplished by complex, elaborate intermolecular assemblies, known as molecular machineries[23]. DNA replication for instance, even though it was widely believed in the 1960s that is achieved by the action of a single enzyme, DNA polymerase, proved to be the outcome of the coordinated actions of numerous proteins involved (DNA helicase, DNA primase, single strand binding [SSB] proteins and others)[25]. Nowadays, it is widely appreciated that 'cell function in general is undertaken by sets of protein complexes rather than by proteins that act individually as isolated species' (Alberts, 1993). A few recent examples will be given in the following paragraphs.

Probably one of the best examples of a molecular machine is the ribosome, not only because of its physiological importance, but also because it illustrates the evolution in the field of structural molecular biology in the last century. Even though it was known since the 1960s that protein translation occurs via a large intermolecular assembly[38], its size and complexity didn't allow detailed structural characterisation until very recently[39,40]. The first structural insights on the ribosomal architecture were provided in 1960 by George Palades who used electron microscopy to capture the so-called Palades particles.[41] Since then the field has progressed from reconstructions of negative stained electron micrographs[42] to low resolution models[43] and very recently the high resolution structure of essentially the complete 70S prokariotic subunit[44,45]. This gigantic 2.5MDa protein-RNA assembly contains in its small 30S subunit more than 1500 nucleotides of ribosomal RNA (rRNA) and 21 proteins, while the larger 50S subunit has ~3000 nucleotides of rRNA and 34 proteins (Figure 1.2). To highlight the importance of molecular recognition, only for the elongation phase of protein translation and for the recognition of the mRNA codon by the tRNA anti-codon, contacts between the 30S ribosomal subunit, the mRNA and

the tRNA have to be made. If the correct tRNA is bound, these contacts cause A1493, A1492 and G530 of the 30S subunit to change their conformation sending the signal that translation can proceed[46,47]. These local conformational changes are then translated into more global changes in structure, with the 'head' of the small subunit moving towards the large subunit activating GTP hydrolysis and subsequent sliding of the ribosome toward the 3'-end of mRNA. Binding and conformational changes are strongly coupled in the ribosome and their specificity is of crucial importance for cell survival (Figure 1.2A)[48].

The F-ATPase is another elaborate system where the collaboration of numerous protein subunits results in energy production essential for cell survival. In this case, the individual protein subunits assemble to form two rotary motors of which one is driven by electric current (F_o) and the other one by chemical gradients (F_1) connected by a rotary shaft and an eccentric stator[49]. The soluble F_1 subunit is comprised of an α/β symmetric hexamer which retains the ATP-synthase activity, with the γ subunit inserted in the middle and the δ, ε subunits attached at the top and bottom of the α/β hexamer respectively-$\delta(\alpha\beta)_3\gamma\varepsilon$, while the transmembrane F_0 subunit which is responsible for proton uptake, consists of 10 to 15 copies of the c subunit, two b and one a subunits. Proton translocation through the membrane via F_0 causes a rotation of the c subunits which is then transferred to the central rotary beam (γ-subunit)[50] and eventually causes conformational changes to the catalytic hexamer $(\alpha\beta)_3$ in F_1. In this sophisticated molecular nanorotor, protein-protein interactions not only contribute to the generation of force (torque) that is essential for ATP production, but are also responsible for the formation of new binding sites: the ATP bindings sites are formed in the interface between the α and β subunits, while the proton translocation is taking place in the interface of the a and b subunits in F_o (Figure 1.2B)[51].

In order to maintain and regulate the levels of a protein the cell has to apply a very tight control on protein synthesis and proteolytic degradation. This degradation function is known to be achieved by the proteasome, a key player in maintaining cell proteostasis[52]. Over the last decade, our knowledge of cell's proteolytic machine has significantly increased. The 26S eukaryotic

proteasome (>1.5MDa) is responsible for the degradation of proteins that are labelled with polyubiquitin chains in an ATP-dependent manner has a cylindrical morphology and consists of two subunits (20S and 19S)[53,54]. The cylindrical appearance of the 20S subunit is the outcome of the axial stacking of four heteroheptameric rings, each one comprised of structurally similar α or β subunits. In this case the α-rings form the gate by which substrates insert into the proteolytic chamber made by the β-rings. The 19S proteasome, also known as (19S regulatory particle - RP[55]) on the other hand, forms the 'lid' and the 'base' of the 20S cylinder, is comprised of 19 subunits in total and is responsible for carrying out the ATPase activity. The mechanism of gate opening for the insertion of the protein or peptide substrate in the catalytic β-rings is the outcome of a synchronised 'dance' of many different protein subunits. Normally the N-termini of the seven gate-keeper α-subunits (20S proteasome) interact with each other, keeping the gate of the proteolytic chamber closed[56]. The interaction of the 19S-RP with the α-ring results in the insertion of the 19S C-terminal regions in between the α-subunits and promotes the interaction of 19S activation loops with the reverse loops of the α subunits. This interaction causes the rotation of the N-terminal segments and the opening of the gate (Figure 1.2C-D)[57].

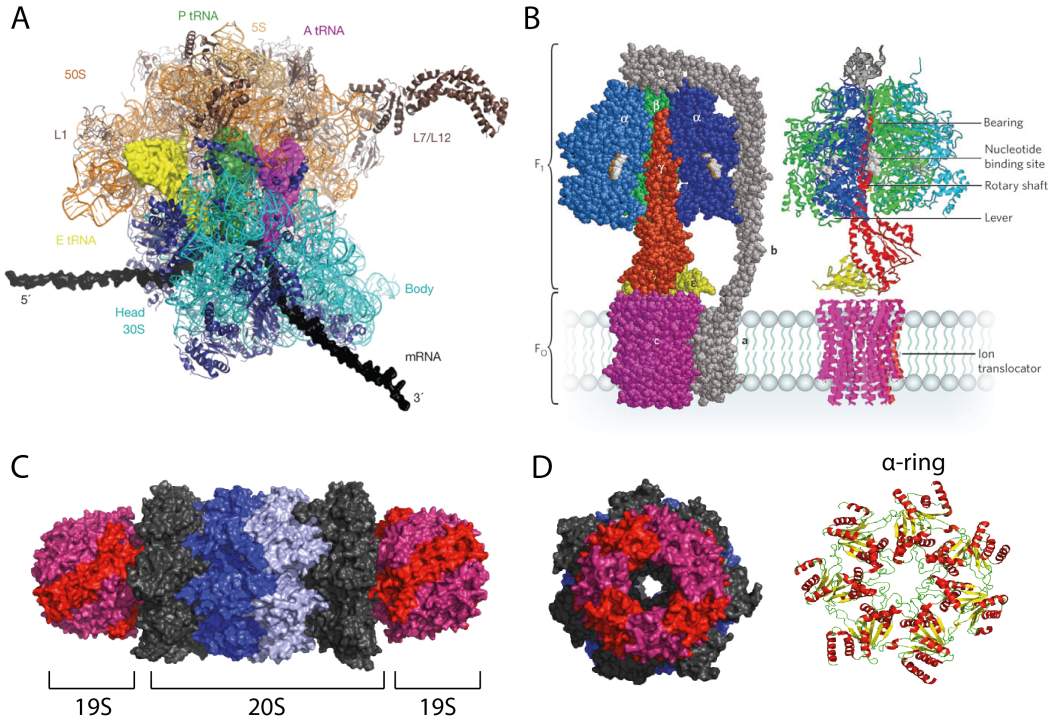

Figure 1.2: Molecular machines in biology.

(A) Top view of the high resolution structure of the almost full 70S ribosome in complex with mRNA and tRNA. Mobile parts of the complex (L7/L12) were modelled in the structure. Adapted from[40]. (B) Side view of the F_0F_1 – ATPase complex shown as a surface (left) and as cartoon representation (right) – adapted from[49]. (C) Side view of the 26S, proteasome. α- and β- rings are coloured black and light/dark blue respectively, while the two regulatory 19S particles are shown in red. (D) Top view of the 20S proteasome as shown in (C) - left. Right - cartoon representation of the α-ring conformation in complex with the 19S regulatory particle (pdbID:1FNT)[58].

Apart from the three examples of molecular machines given here, the list extends to viral capsids[59], motor proteins such as kinesin[60], proteins that carry nucleotide tri-phosphate hydrolysis activity (NTPases)[61] and many more. Their detailed description is beyond the scope of this thesis, but all together these examples highlight the importance of macromolecular recognition inside the cell.

1.3 Scientific formalism and principles of macromolecular recognition

1.3.1 Definition of affinity at equilibrium

The two most important parameters that characterise macromolecular interactions are the stoichiometry and the binding affinity. In most cases a 2-state association is assumed, while more complex systems can be broken down to a series of one-to-one interactions. Therefore, the reaction can be written as:

$$[A] + [B] \underset{k_{off}}{\overset{k_{on}}{\Leftrightarrow}} [AB] \qquad \text{scheme (1)}$$

where [A] and [B] are the concentrations of the reactants and [AB] the concentration of the complex product, k_{on} is the second–order rate constant for the unimolecular association and k_{off} the first-rate dissociation constant. Here and using the law of mass action, the equilibrium constant K_d is defined as the ratio of the dissociation and the association constants:

$$K_d = \frac{[A_f][B_f]}{[AB]} = \frac{k_{off}}{k_{on}} \qquad \text{equation (1)}$$

Where, according to the law for mass conservation:

$$[A_f] = [A_t] - [AB] \qquad \text{equation (2)}$$

and

$$[B_f] = [B_t] - [AB] \qquad \text{equation (3)}$$

Where A_t and B_t are the total concentrations of reactants A and B respectively, k_{on} is measured in $M^{-1}s^{-1}$ and thus concentration dependent, while k_{off} is expressed in s^{-1}, therefore K_d has a dimension of concentration.

If we introduce eq. 2 and eq.3 in eq. 1 we get:

$$K_d = \frac{([A_t] - [AB])([B_t] - [AB])}{[AB]} \qquad \text{equation (4)}$$

or

$$\frac{[AB]}{[A]} = \frac{[B_f]}{K_d + [B_f]} \qquad \text{equation (5)}$$

The equation of fractional saturation (eq. 5) describes the portion of protein B that is saturated with protein A and a plot of $[AB]/[A_t]$ vs $[B_f]$ will result in a binding rectangular hyperbola that is commonly seen in experiments for the determination of K_d (Figure 1.3). The linear form of eq. 5 results in the commonly used Scatchard plot of $[AB]/[A_t]$ vs $\log[B_f]$.

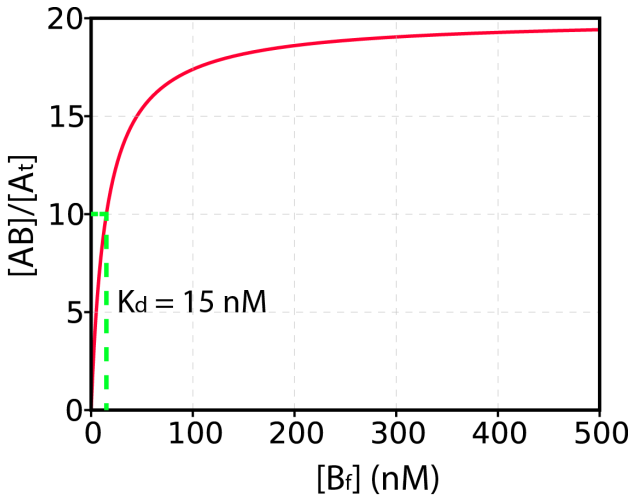

Figure 1.3: Simulated binding hyperbola.
A 1:1 stoichiometry and a K_d of 15nM is assumed. The fractional saturation on the Y axis could be quantified by any observable that changes upon binding.

The equilibrium constant (K_d) is related to the Gibbs free energy of dissociation by the formula:

$$\Delta G_d^o = -RT \ln \frac{K_d}{c^o} = \Delta H_d^o - T\Delta S_d^o \qquad \text{equation (6)}$$

where R is the gas constant (8.3144 $JK^{-1}mol^{-1}$), T the temperature in Kelvin, c^o is the standard state concentration (for interactions that take place in solution c^o is assumed to be 1M, but the situation might be different in the gas phase), ΔH_d^o and ΔS_d^o the changes in enthalpy and the entropy of binding. Even if ΔG_d^o is known as the free energy of binding, it is not a proper energy, but rather describes the affinity between the two partners and hence the stability of the complex. However and based on eq. 5, a high affinity implies that the K_d is much smaller than [B], meaning that the affinity depends not only on the K_d but also on the concentration of the ligand and it is not a inherent property of the complex. In reality, a K_d of 1μM for a hormone receptor might be considered not functional since the concentration of hormones in blood is well below the μM range. However the same relatively high K_d of 1μM describes the dissociation of haemoglobin dimers, but because haemoglobin is present at >mM concentrations in the blood, the presence of the non-functional dimers can be considered negligible.

1.3.2 Structural properties that define affinity

Macromolecular recognition is the outcome of the balance between factors that favour binding and others that oppose it. Bond formation between the interacting subunits (enthalpic binding) or contribution from the solvation entropy (water molecules that get excluded from the interface show increased mobility) are promoting interactions, whereas the loss of translational and rotational degrees of freedom of interfacial side chains or mobile parts that get stabilised upon binding result to an unfavourable entropy of binding. In general the entropic contributions in binding are difficult to assess even with newly developed techniques[36,62,63] and thus correlations between the binding affinity and structural features of the bound complex were made over years.

Some of the main descriptors of binding will be described in the following paragraphs.

1. Buried surface area - Hydrophobicity

Proteins in their native state normally exist in a water-based solvent. The concept of accessible surface area (ASA) describes the ability and the extent to which protein atoms can form contacts with water and is normally measured by rolling a solvent probe of radius X Å (e.g 1.4Å) and measuring the contact area to the surface of the protein without penetrating the structure[64,65]. On the contrary, buried surface area (BSA) describes the loss of ASA for both partners upon binding (Figure 1.4). BSA is a very good descriptor of the hydrophobic interactions that participate in complex formation and are often considered the driving force of macromolecular recognition[66]. An empirical estimation of an energy gain of 25 cal per $Å^2$ of BSA has been made although this value can vary between cases. A very good, almost linear correlation between binding affinity and BSA is observed for high affinity complexes that do not undergo conformational changes upon binding[66-68], whereas this is not the case for more transient interactions or interactions that involve conformational changes[68], suggesting that hydrophobicity is not the only determinant of binding.

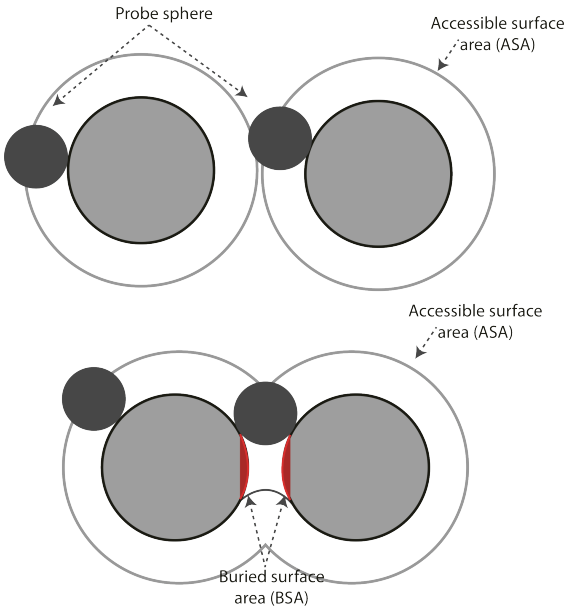

Figure 1.4: Accessible surface area (ASA) and buried surface area (BSA). In the case of two non-interacting proteins (top) the probe sphere is able to freely sample the entire molecular surface (the proteins are assumed perfectly spherical for simplicity). However when the molecules make contacts with

each other (bottom) some of the surface is excluded from sampling (red), forming the BSA.

2. Electrostatics

The formation of hydrogen bonds and attractive van der Waals energies in the interface also contribute significantly to binding energetics. Hydrogen bonds made in the interface are generally more favourable than those made with solvent and therefore contribute significantly to the free energy of binding. On the other hand, van der Waals forces are normally not that strong in the interface (and not more preferable than interactions with the solvent) but because the number of interfacial atoms is large, their contribution cannot be considered negligible. Furthermore, there is recent evidence that this kind of interactions are very important in the first stages of biomolecular association before the formation of the stereo-specific complex (see Section 1.5.1)[69-71].

4. Shape complementarity

Shape complementarity plays an important role in interactions where the protein partners interact as rigid bodies[72]. Normally, it is the outcome of both electrostatics and the shape of the proteins in the interface. However, and as it was mentioned for BSA, this descriptor of binding fails to provide useful information when the interactions is weak and involves changes in the structure of the monomers involved[73].

5. Hot spots or anchor residues

In most cases the BSA at the interface extends to more than 1000Å, involving numerous residues from both partners[74]. Thus, one would imagine that single point mutations, even at interfacial residues, would not yield any significant changes on biomolecular recognition. However, this is not the case. Nowadays, it is widely appreciated that even single side chains can contribute large parts of the binding free energy[75,76], an observation that was first made by Wells et al, who used alanine scanning mutagenesis to probe the interface of human growth hormone and its receptor[77,78]. In general, hot spot residues seem to obey three basic rules[74]:

- They are more conserved than other parts of the protein

- They differ in amino-acid distribution
- And they are protected from bulk water

Hot spot residues can either affect the kinetics of protein association (k_{on}) or the stability of the specific complex (k_{off}) showing that binding is not only due to the hydrophobic effect as it was believed in the past.

As mentioned above, the entropic contributions to binding are usually very difficult to study as they report on changes on very fast protein dynamics that are not accessible to most biophysical tools. However, in a recent study Kalodimos et al. used NMR methods to show that binding to DNA is dictated by changes in protein conformational dynamics (which are often considered negligible)[79,80], showing that the structural studies normally carried out to characterise binding might not be sufficient in all cases.

1.3.3 Properties of protein-protein interfaces from a structural view

In the literature, two main categories of protein complexes have been distinguished: obligate (or permanent) complexes and non-obligate complexes[81,82]. In the first class fall complexes that only exist in their bound form and the involved protomers are never encountered free in solution, once the complex is formed. Most complexes in this class are often symmetric homodimers, such as human immunodeficiency virus (HIV) reverse trancriptase[83] which encompasses an active site formed by residues of both monomers, and the permanent homodimer of interleukin 8 cytokine(IL-8)[84]. On the other hand, non-obligate complexes can exist in both the bound and unbound forms and include numerous associations such as antibody-antigen interactions[85], interactions that take place in the biogenesis of the cytoskeleton (e.g actin filament assembly) and many others[32] (Figure 1.5).

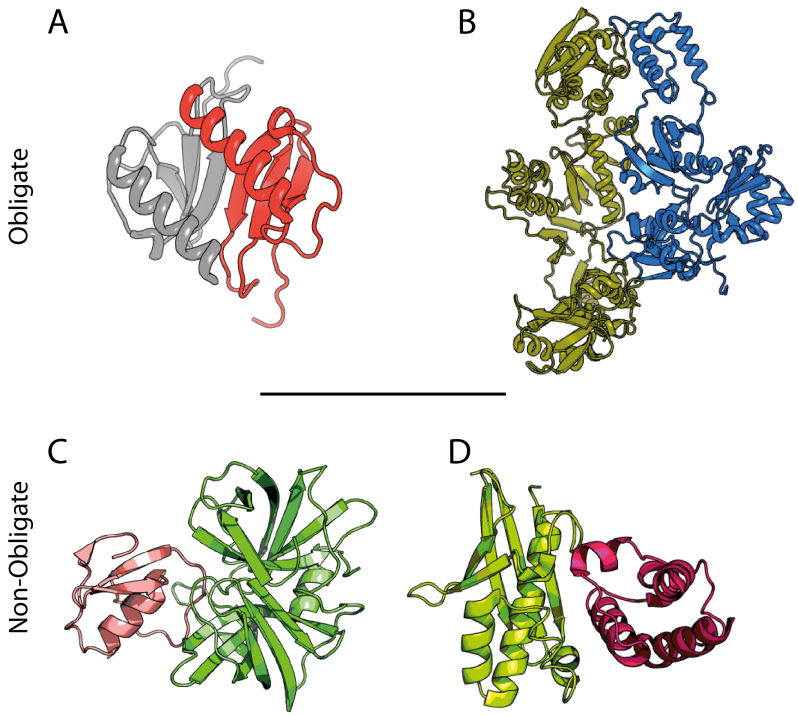

Figure 1.5: Obligate and non-obligate complexes.
(A) Crystal structure of the interleukin kinase 8 (IL-8), pdbID: 1IL8[84] and (B) Human immunodeficiency virus (HIV) reverse transcriptase, pdbID:3HVT[83], both belonging in the family of obligate complexes. (C) Protease B (*Streptomyces griseus*) in complex with the turkey ovomucoid inhibitor, pdbID:3SGB[86]. (D) Structure of cellulase in complex with cellulolyticum, pdbID: 1FBO_A:E[87].

A comparison of the properties of the interface of obligate and non obligate complexes performed by the Thornton group already in 1996 (and with only 58 non-homologous complexes available) shows clear differences between the two classes. In this study[82], non-obligate complexes were found to have smaller, less well packed and less planar interface than their permanent counterparts. However, permanent interfaces seem to be more hydrophobic, containing on average fewer hydrogen bonds than non obligate complexes (Table 1.1)[88].

Interface feature	Obligate complexes		Non-obligate complexes
BSA (Å^2)	Mean	1772	804
	σ	1985	147
Planarity (Å)	Mean	3.56	2.57
	σ	1.65	0.52
Gap index (Å)	Mean	2.12	2.66
	σ	0.94	0.87

Table 1.1: Interfacial properties of obligate and non-obligate complexes. Planarity: a measure of the interface curvature. Root-mean-square deviation of the atoms in the interface from the least square plane through all these atoms. Gap index: the ratio of gap volume to the BSA at the interface. Adapted from[74].

The differences in the structural features of the interface in permanent and non-obligate complexes show that apart from evolutionary constraints, protein-protein interactions differ significantly in their conformational properties depending on their affinity. Controlling protein function by modulating interface properties and thus affinities seems to be crucial for the maintenance of cell viability (see Section1.4). However, there is not a clear correlation between the size or the physicochemical properties of the interface and binding affinity, especially for dynamic binders that undergo conformational changes upon binding. Even though some differences can be seen between the properties of the interface in obligate and non-obligate complexes, it is still difficult to distinguish different types of protein-protein interactions based on structural analysis of the interface[68].

1.4 Classification of macromolecular complexes based on their affinities

Based on their physiological role, non-obligate protein-protein interactions span a wide range of affinities *in vivo*. Evolution has produced numerous methods to ensure the functionality of a complex and to control its affinity, with protein co-localisation being one of the most important ones. Since the protomers involved in obligate interactions are unstable on their own, they are often co-localised with their partners and thus the interaction can take place instantaneously after protein translation[89]. This strategy is easy to imagine for obligate homodimers, but even in the case of hetero-complexes, the monomeric subunits may share the same promoter (e.g. the subunits of cathepsin D)[90,91]. Using the same philosophy, non-obligate interactions can also be controlled by

protein co-localisation. Such systems usually perform a physiological function for which a relatively high affinity is crucial for the cell, with examples including antibody-antigen associations and proteases with their inhibitors. In such cases, the very slow dissociation rate is often the parameter that determines high affinity, suggesting that the interaction once formed is often irreversible. In other cases however, a modulation of binding affinity is necessary, depending on cell conditions. In cell signalling for example[92], the association of the α subunit of protein G with the βγ heterodimer exhibits an almost 1000 fold increase/decrease in its binding affinity depending on the presence of GTP vs GDP[93].

To highlight the elaborate manner with which affinity is related with protein function, a few examples of complexes with high (fM-nM), moderate(nM-μM) and low (μM-mM) affinity will be given in the next paragraphs.

Often during cell life cycle, there arises the need of DNA breakage, DNA repair or recombination or RNA maturation. All require hydrolysis of the phosphodiester bond, performed by enzymes known as nucleases[94,95]. However, errors in these procedures are most likely to be deleterious for the host cell and hence tight mechanisms have evolved to protect against them. The control of specific nucleases (such as restriction enzymes) is often achieved by methylation of the target sequence which prevents access of the enzyme to DNA. In the case of more generic/non-specific nucleases however, such a strategy is not applicable. Instead, the undesirable action of these enzymes can be prevented through the action of specific nuclease inhibitors[74,96]. The interactions between nucleases and their inhibitors provide examples of some of the highest affinity interactions that are known in the literature ($K_d \leq 10^{-14}$M)[68]. Numerous studies reporting on the free and bound states, the kinetics and the thermodynamics of the association provide an almost complete picture of the binding event[97-100].

One of the best-characterised interactions of this type is the interaction of barnase with its inhibitor barstar. Barnase is an extracellular RNAse secreted by *Bacilus amyloliquefaciens,* while the inhibitor barstar is only found

intracellularly. Even though the proteins are not co-localised and thus not expected to interact, occasional errors in targeting barnase to the secretion pathway raises the need for the organism to have a potent inhibitor of its activity within the cytosol. Structural studies on the complex revealed that barstar occupies the active site of barnase in the complex, resulting in a highly complementary interface and a loss of 1590Å^2 of ASA upon binding[101]. All active residues of barnase make contacts with barstar by forming either salt bridges or hydrogen bonds. Another feature of barstar that enhances its inhibitory function is its ability to mimic phosphorus P1, one of the key features of the RNA substrate[102]. To achieve this, D39 makes electrostatic interactions with the enzyme that are similar to those mediated by P1 (Figure 1.6). All these structural features are translated into a very high affinity complex (K_d = 10^{-14} M), making barnase-barstar one of the best model systems to study binding energetics, entropy-enthalpy compensation and protein engineering at the interface.

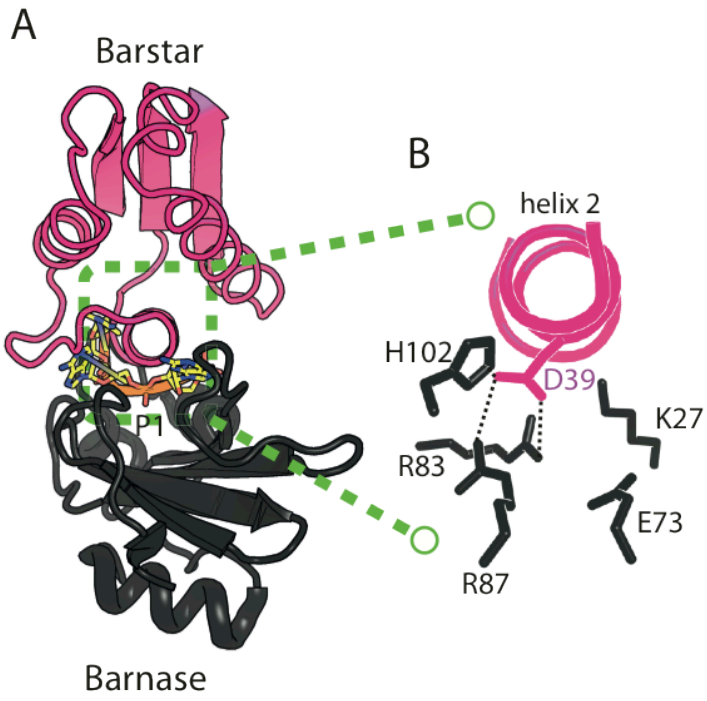

Figure 1.6: The barnase-barstar complex
(A) The structural superimposition of barnase (black) with its cognate RNA (yellow sticks) and the barnase- barstar (pink) complex shows that barstar binds in the active site of barnase. The position of the RNA cleavage is shown as P1. (B) Zoom-in in the active site of the barnase-barstar complex shown in (A). D39 forms salt bridges with the two Arg residues of the active site (pdbID: 1BRN)[103].

Signal tranduction is another key cellular process that invloves a variety of protein-protein interactions. However, in this case a very high affinity complex between the interacting partners would not be beneficial for the cell, since in most cases the proteins have to transiently associate and dissociate in order for the signal to be transferred across different cellular compartements. Normally, protein-protein interactions in signal transduction are mediated by the recognition of linear or discontinuous protein epitopes that bring two proteins together in order to achieve their function. For example, dimerisation of receptor tyrosyl kinases enables them to phosphorylate each other and initiate signal transduction[104,105]. However, a permanent dimer would activate apoptotic mechanisms, highlighting the necessity of an association of weaker affinity in this case. In this example the dimerisation is mediated by distinct protein domains known as SH3 domains (Src homology domain-3). SH3 domains have a characteristic 5-stranded β-barrel structure[106,107] and recognise proline rich areas with μM to nM affinities[108]. SH3 domains cluster conserved aromatic residues on one side that creates a hydrophobic surface, key for the macromolecular recognition[109]. This observation represents the origin of their affinity to polyproline type-II conformations (PPII) since these peptides also provide hydrophobic surfaces for binding (Figure 1.7). SH3 domains, apart from being crucial for signal transduction, provide a good example of dynamic binding. NMR studies have shown that these domains are able to bind different peptide ligands in different conformations[110] with the orientation of the ligand in the binding site depending not on the interface residues but on residues that precede it.

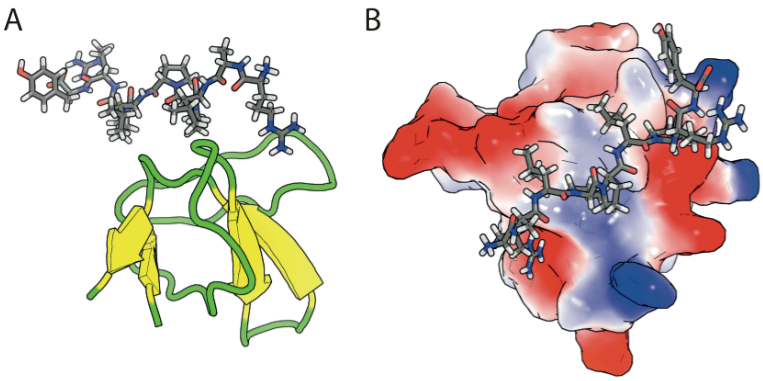

Figure 1.7: A SH3 domain bound to a PPII peptide.

(A) Cartoon representation of the SH3 domain (yellow) bound to a class I PPI I peptide (grey sticks). (B) Surface representation of the complex shown in (A). The electrostatic surface of the SH3 domain is shown (pdbID: 1RLQ)[110].

In other cases, such as the electron transfer system or the intergin mediated adhesion system, much weaker interactions (μM – mM affinities) are required between the proteins involved and often biomolecular recognition causes large conformational changes[111]. The importance of these weak intermolecular associations, even though underestimated in many cases, is nowadays increasingly appreciated and will be discussed in the next section.

1.5 Transient interactions: from encounter complexes to intrinsically disordered proteins

1.5.1 Encounter complexes: the importance of pre-orientation, insights from the phosphotranferase system

Protein-protein interactions are the outcome of a productive collision between two molecules that randomly translate in solution and the satisfaction of the orientational restraints arising from the rotation of the partners to form the stereo-specific complex. The diffusion rate limit for macromolecular collision, as calculated by the Einstein-Smoluchowski equation is 10^9–10^{10} $M^{-1}s^{-1}$, however the association rates for protein-protein interactions are expected to be at least 3 to 4 orders of magnitude slower due to the orientational restraints applied upon binding[112-114]. Surprisingly, very high rates of association are observed between protein molecules, that often approach the diffusion limit ($5x10^9$ $M^{-1}s^{-1}$ for the barnase-barstar association[115]) and therefore require the introduction of additional factors in the mechanism. Thus, it has been proposed that biomolecular recognition proceeds by the formation of non-specific, rapidly inter-converting species known as encounter complexes[70,116]. Fersht and co-workers, performed one of the first studies of encounter complexes by analysing a series of charge mutants and the salt dependence of the barnase-barstar interaction. They showed that the transition state of the formation of the barnase-barstar complex involves short-lived weak

electrostatic interactions, which seem to be generic for other interacting systems[71].

More recently, advanced NMR techniques have been used to study, in atomistic details, the encounter complexes formed by proteins involved in the bacterial phosphotranferase system[69,117]. The bacterial phosphotranferase system (PTS) transports sugars across the membrane[118] through a network of protein-protein interactions. In the first step of this reaction a central enzyme known as enzyme I (EI) extracts α phosphate from the phosphoenolpyruvate and transfers it to the first carrier protein, HPr (histidine protein) which is then transferred to the membrane (Figure 1.8)[119]. Clore and co-workers were able to reveal the existence of an ensemble of encounter complexes that precede the formation of the specific EIN-HPr interaction, which remarkably correlates with the electrostatic surface potential of EIN[69], highlighting the importance of electrostatic interactions in the initial stages of biomolecular recognition (more details on the experimental strategy used to tackle this system will be given in Section 1.5.4.2). These results show that the correct pre-orientation of the binding partners to form a productive encounter complex before the formation of the functional complex is generally the strategy for overcoming the slow diffusional rates for large biomolecules and thus these transient interactions are crucial for achieving fast association kinetics[120].

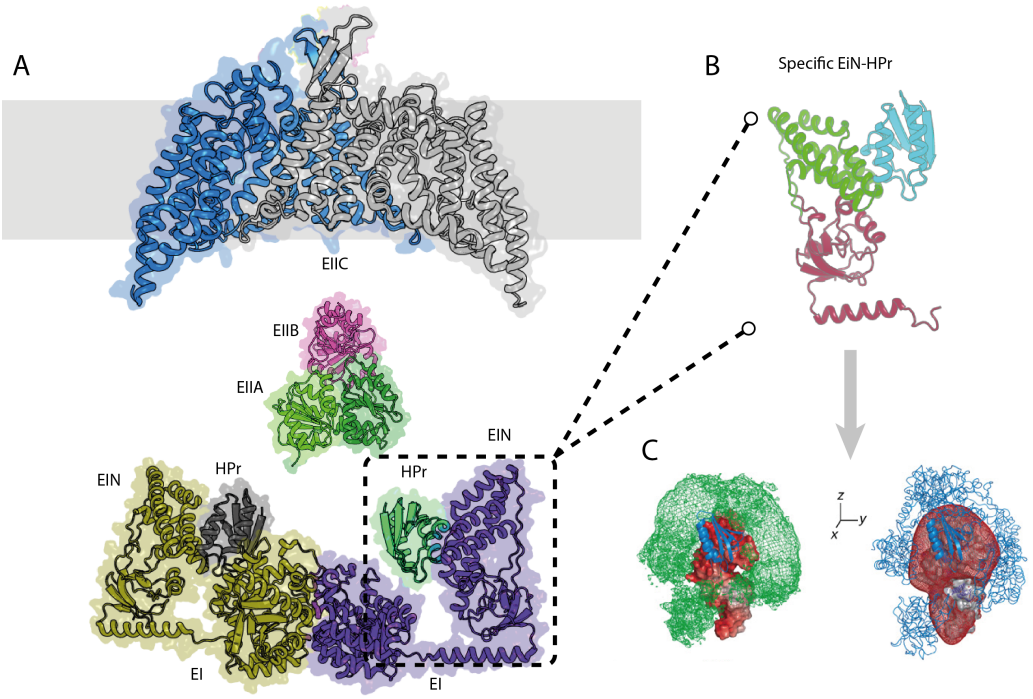

Figure 1.8: Protein-protein interactions in the PTS.
(A) Overview of proteins participating in PTS. Cartoon representation of the EI-HPr dimer of dimers (bottom)[121] responsible for the initial stages of the reaction. The phosphate is then transferred to the carrier proteins EIIA and EIIB[122] before reaching its final destination at the membrane transporter EIIC[123]. (B) Zoom-in of the stereo-specific complex of the N-terminal domain of EI (EIN – green) and HPr (blue)[124]. The C-terminal domain of EI is shown in red. (C) Encounter complexes in the EIN-HPr interaction. Right-the encounter complexes (N=20) are shown as a probability density map (green mesh) around the structure of the stereospecific complex (EIN-red surface, HPR-blue cartoon). Left-Correlation of the encounter complex distribution (blue cartoons with the electrostatic surface potential of EIN (red mesh). Panel adapted from[69].

1.5.2 The role of conformational disorder in recognition

Proteins are not static molecules. They exhibit motions that span a range of timescales (from ns to ms or slower)[125,126]. Protein dynamics have been connected with various aspects of protein function, such as enzymatic catalysis[127], regulation of protein synthesis[128] and many others[129]. Proteins that possess at least one region that does not have a fixed secondary or tertiary structure are known as intrinsically disorder proteins (IDPs) and lie at the extreme of the spectrum for molecular motions since their disordered regions can sample a very large space of heterogeneous conformations[130,131]. As monomeric proteins show different degrees of motion, macromolecular

assemblies can span a whole spectrum of dynamics[132,133]. IDPs that are involved in macromolecular binding are often mentioned as 'hub' proteins since they use their disordered regions to bind to many different partners, playing in this way a major role in protein-protein interaction networks[131,134]. They often represent good examples of the conformational selection theory[19], since binding stabilises one of the many conformations the IDP can sample in its unbound state[135]. Therefore, in many cases IDPs undergo coupled folding and binding at least at some part of the protein[136]. This phenomenon has been proposed to be the origin of the weak and transient interactions that IDPs are often participating in, since the entropic penalty for the folding of a disordered region results in higher K_d[134,135]. Therefore, IDPs are advantageous in systems where transient interactions are desirable or systems where the interactions have to be switched on or off quickly, such as cell signalling[137,138]. IDPs also have various other advantages in comparison with their folded counterparts. Upon binding they can adopt conformations that are sterically hindered for structured proteins, such as wrapping around their protein partner in order to maximise BSA[134,139]. From an evolutionary point of view, disordered regions are not likely to lose binding capacity or functionality upon mutation[134,140] and can act as mediators of shuffling protein domains[141]. As an outcome, these properties make them favourable for selection. Based on all these characteristics, IDPs can promote splicing, expose or hide binding interfaces, promote protein modularity, explaining their prevalence in the eukaryotic kingdom (more than 25% of the proteome)[142].

An example of an IDP involved in a biologically important complex is the interaction of SiC1 with its cyclin-dependent kinase (CDK) partner (Cdc4)[143]. Phosphorylation of the N-terminus of SiC1 creates binding sites for Cdc4, of which most are located in the first 90 residues. Intriguingly, even if Cdc4 contains only one binding site for SiC1, at least six phosphorylations on the latter are required for efficient binding, while individual binding sites have very weak affinities[144,145]. In this case, binding is mediated by all activated binding sites of SiC1 that interact with Cdc4 in a dynamic equilibrium, exchanging between their bound and unbound states[146]. Thus, SiC1 gains structure only locally (only on the bound epitope), while its other binding sites remain

disordered (Figure 1.9)[147]. Whereas the role of protein dynamics in the unbound state is well established, such an elevated degree of dynamicity in the complexed form is sometimes difficult to envisage. The mechanism by which protein motions are related to molecular recognition is a question that gained scientific attention only recently[148] and will be discussed in Section 1.5.3.

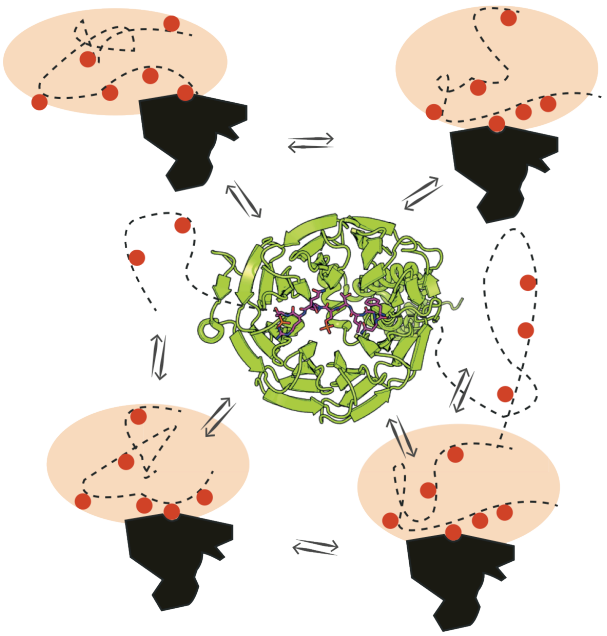

Figure 1.9: Dynamic complexes.
Schematic of the proposed mechanism of binding of SiC1 (black) to Cdc4 (dashed lines). Different binding epitopes of SiC1 (red circles) are in dynamic equilibrium between each other. Only the single bound epitope gains structure as shown in the crystal structure of the complex (middle), while the rest of the protein remains unfolded. Even though only one interface is present at each time, the remaining, unsatisfied, binding sites contribute to the energetics of binding by long range electrostatic interactions. Adapted from[134].

1.5.3 The dynamic nature of protein-protein interactions

The last example of a protein complex shown in Figure 1.9 depicts well the dynamic nature of protein-protein interactions. Even though the role of protein dynamics is well appreciated nowadays in a variety of aspects of protein functionality, their role in biomolecular recognition is still unclear.

Recent advances in structural biology have allowed the investigation of the link between structure and dynamics in the unbound versus the bound state of several proteins[135,149,150], and the quantification of conformational changes

upon binding. However, a first glimpse on this phenomenon was provided long ago using X-ray crystallography and comparing the structures of the apo or bound proteins[151]. While rigid body associations are often seen in biology (as in the case of the barnase-barstar complex, where the structures of the individual subunits have an RMSD of <1Å in the bound versus the unbound form), this is not always the case[152,153]. One of the first examples of a binding induced conformational change was the interaction between trypsinogen and the pancreatic trypsin inhibitor (PTI)[154]. Upon binding, the N-terminal region of trypsinogen is getting cleaved to form mature trypsin. Interestingly, the disordered loops of the precursor are ordered in the mature protein, showing that the molecule is undergoing a disorder-to-order transition upon interaction[155]. More pronounced conformational changes have been observed upon the GTP-coupled interaction of the G_{α} subunit with the $G_{\beta\gamma}$ heterodimer, including a complete remodelling of the protein surface, involving motions in the main chain and even alterations of the secondary structure[156,157]. However, it has to be noted that X-ray studies performed at cryogenic temperatures, introduce a serious bias towards reduced molecular motions[158] and thus crystallographic studies may underestimate the degree of flexibility in both the apo or complexed protein.

The question that emerges is whether there is a correlation between the dynamics of the free state and the conformational changes observed upon interaction. Recently, Julie-Forman Kay and colleagues proposed the relative solvent accessible surface area (A_{rel}) as a measure of protein flexibility upon binding[159]. This simple descriptor is the ratio of the observed ASA of a protein divided by the value expected for a fully folded protein, and thus describes protein flexibility in the free state. A correlation between A_{rel} and binding-induced conformational changes was observed, showing that flexible proteins tend to undergo significant structural changes upon complexation, while their stable counterparts are less likely to do so (Figure 1.10). Even though there is a good correlation between binding affinity and BSA for stable rigid-body associations, this link falls apart when binding induces conformational changes making the prediction of K_d based on structure challenging[160].

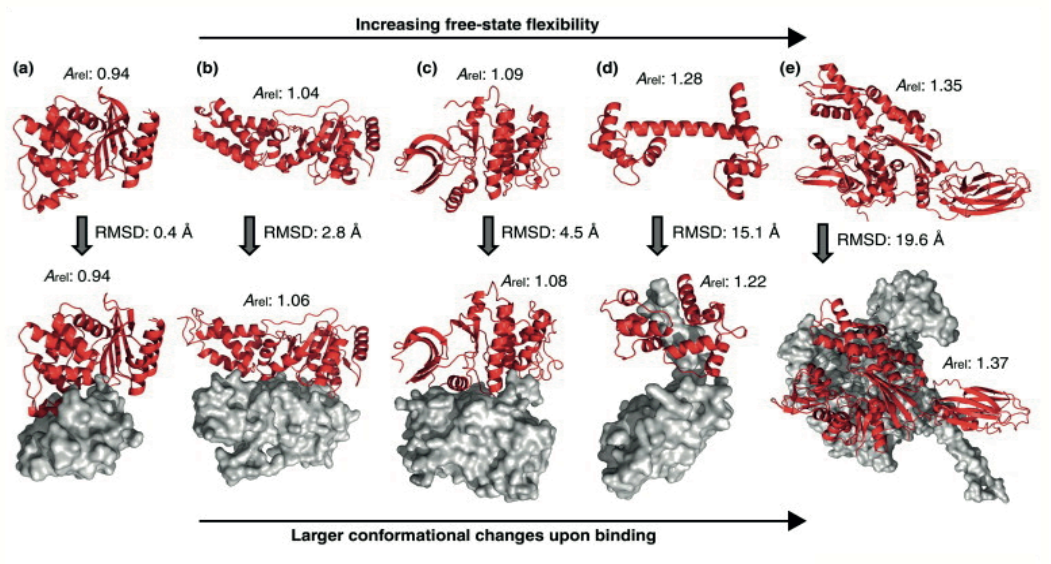

Figure 1.10: Relation between protein flexibility and conformational changes. Increased dynamics in the free state leads to pronounced conformational changes upon binding (left to right). (A) β-lactamase (pdbID:1M40) interacting with β-lactamase inhibitor (pdbID:1JTG). (B) Ffh (pdbID:1LS1) binding to FtsY (pdbID:1RJ9). (C) CDK-2 (pdbID:2R3I) interacting with cyclin A (pdbID:2CCH). (D) Calmodulin (pdbID: 2R3I) in complex with nitric oxide synthase (pdbID: 3HR4). (E) HSc70 (pdbID: 1YUW) interacting Sse1 (pdbID: 3C7N). Panel reproduced from[148].

Although a disorder to order transition upon binding is common especially if binding involves an IDP, the opposite phenomenon, (even though much more challenging to envisage and investigate), has also been observed for the interaction of p53 with the chaperone Hsp90[161]. Using NMR spectroscopy Park *et al.* were able to show that p53 even though stable is isolation, adopts a molten globule state upon interaction with Hsp90, as shown by increased exchange line broadening in the NMR lines of regions of the protein that are not surface accessible and therefore cannot be part of the interface[161].

These findings highlight the broad range of effects that protein-protein interactions can have on the subunits involved, ranging from gain of structure, conformational rearrangements or even destabilisation upon binding. Even in cases where the structure itself remains unchanged after recognition, protein dynamics/flexibility might be altered as a consequence of binding or in opposition; protein flexibility can potentially determine the fate of biomolecular recognition. The link between protein dynamics/flexibility and

binding was recently demonstrated for the RNA binding protein U2AF65, where changes in the population of a minor protein state directly affect binding affinity, in a classic conformational selection mechanism[19], showing how protein dynamics modulate protein-RNA association[162]. As these phenomena usually involve rapidly inter-converting conformations that exist in solution for a very short period of time or are remarkably dynamic, lacking defined structural properties (such as dynamic complexes or IDPs) their detection and characterisation using standard biophysical methods has proven problematic. However, recent advances in biophysical tools allowed their detection and even structural elucidation and they will be discussed in the next section.

1.5.4 Methods to study lowly populated, dynamic macromolecular associations and individual subunits

The link between protein dynamics and biomolecular recognition constitutes a fascinating scientific question. In the early stages of biophysics and structural biology, elucidation of the structural properties of the system in question was considered to be sufficient in order to understand the underlying mechanistic details of protein function. However, more recently, and following technological advances, the field has moved towards a direction where protein dynamics are considered equally important in the quest to understand how proteins exert their function.

Protein dynamics can be classified according to the timescale of the molecular motions involved. In general terms, motions in the fast timescale (ns - μs) involve deviations around an equilibrium position which are driven by the kinetic energy of the protein and are generally represented by small concerted movements of small structural elements inside the same energy well. Since such motions can sample numerous substates, studying fast protein dynamics can yield information about the entropy of the system[125,163]. Dynamics on a slower timescale (ms) almost always involve transition(s) from one equilibrium position to another, which is separated by distinct kinetic barriers. These motions include local unfolding, interdomain motions and conformational changes. Of course, proteins are not static when they jump

from one energy well to another, but they undergo less severe molecular motions in the fast timescale.

Experimental approaches to study dynamic protein states have always been a challenging problem for the scientific community. Concerning motions on the fast timescale, NMR relaxation methods have been traditionally the method of choice[22,164-166], whereas the recent development of time-resolved X-ray crystallography offers the potential to examine reactions that happen in the fs timescale in real time[167-169]. Motions on a slower timescale, because of their nature, allow experimental tools to capture intermediates much more easily. Techniques such as fluorescence, infrared and raman spectroscopy, circular dichroism and electron paramagnetic resonance can be valuable in providing kinetic information but they can only reveal a picture of the ensemble pool. The situation becomes even more challenging when the population of the inter-converting species is low (<5%), making them invisible for almost all classic biophysical methods. Lowly populated species in exchange with the native state are often important in excertting biological function as depicted in numerous examples including biomolecular recognition, enzymatic catalysis and protein folding[170]. However, the recent development of single molecule fluorescence techiques overcomes this drawback. With the use of single molecule Förster resonance energy tranfer (smFRET) which can act as a 'molecular ruler', structural information on rarely populated species can be obtained[171,172]. smFRET allows the determination of just a single distance, which in the presence of a high resolution structure can be translated into information about conformational changes or domain rearrangments[173-175]. However, smFRET and other low resolution techniques fail to provide information in the atomic level which is at least desirable, if not necessary, in order to understand the properties of the system under investigation. Recent advances in NMR spectroscopy have allowed the study of 'invisible' protein states, making NMR the only current tool to investigate sparsely populated protein conformations, in atomic detail. More specifically, relaxation dispersion NMR spectroscopy (RD) and paramagnetic relaxation enhancement (PRE) studies can detect molecules that are only 0.5-1% of the total population

in solution[80] and allow a full structural, kinetic and thermodynamic characterisation.

1.5.4.1 Carr-Purcell-Meiboorn-Gill (CPMG) relaxation dispersion: probing 'invisible' protein states

More than fifty years ago Carr-Purcell and Meiboorn-Gill described the so-called Carr-Purcell-Meiboorn-Gill (CPMG) pulse scheme[176,177], focusing on backbone dynamics of small ligands in the millisecond to microsecond timescale. In these initial studies exchange to alternative conformations was considered to be an undesired, or even annoying, phenomenon which prevented the extraction of "more relevant" information on protein dynamics. Fortunately, this was far from true. NMR techniques had been long used to study chemical exchange process in small molecules[178] but it took considerable effort and time to apply the same idea to more complicated biomolecular systems. The problem was overcome in the pioneering work of Loria and coworkers in 1999[179] and since then numerous applications have been evolved that reveal new insights into protein folding, protein-protein interactions and many other biologically relevant aspects[180-184].

The basic idea behind relaxation dispersion experiments is to quantify the effective line width of peaks in spectra as a function of the number of refocusing pulses in a CPMG pulse scheme. Chemical exchange in proteins frequently involves interconversion between a highly populated ground state A, and a lowly populated excited state B. If the exchange is slow on the NMR timescale (where the rate of exchange k_{ex} is smaller than the chemical shift difference between the species $\Delta\delta\omega$, $kex<\Delta\omega$) then two separate peaks should be visible in the NMR spectrum (assuming that the two spin states develop different chemical shifts). However, if the populations of the inter-converting species are severely skewed towards one of them (the main species), the direct detection of the minor peak is normally impossible[183]. In this instance, it is possible to detect the chemical shift of the minor state by observing changes in the line shape of the main peak. Parameters that can be extracted from such an experiment include k_{ex} from which the populations of the conformations in

equilibrium can be calculated but also the chemical shifts of the excited state, which provide structural insights that can lead to the structural characterisation of the lowly populated species that cannot be obtained by any other technique. Importantly, the development of methodologies that allow the accurate determination of the 3D structure based solely on chemical shifts[185], allows the characterisation of the 'invisible' species in atomic details.

For relaxation dispersion experiments it is crucial that the probes (spins) evolve at different frequencies (chemical shift) in states A and B (noted by ω_A, ω_B accordingly). If the above is not the case, then no useful information can be extracted by this experiment. If we consider a period of time, t, where the spin exists only in state A and a refocusing pulse is applied in the middle of this period then the frequency evolution of the spin over this time is zero and the spin is 'refocused'. On the other hand if during time t there is exchange between states A and B, the refocusing pulse is not able to fully refocus the spin (these occasions are noted grey in Figure 1.11A). In a scenario like this, at the end of the pulse scheme the spin would have evolved at a frequency which depends on the time it spends in each state and the number of exchange events. If the frequency of the refocusing pulses increases, then there is higher probability for complete refocusing after each pulse, resulting in sharper peaks (Figure 1.11B). By measuring the intensity of the peaks as a function of the frequency of the refocusing pulses a CPMG profile is obtained (Figure 1.11B).

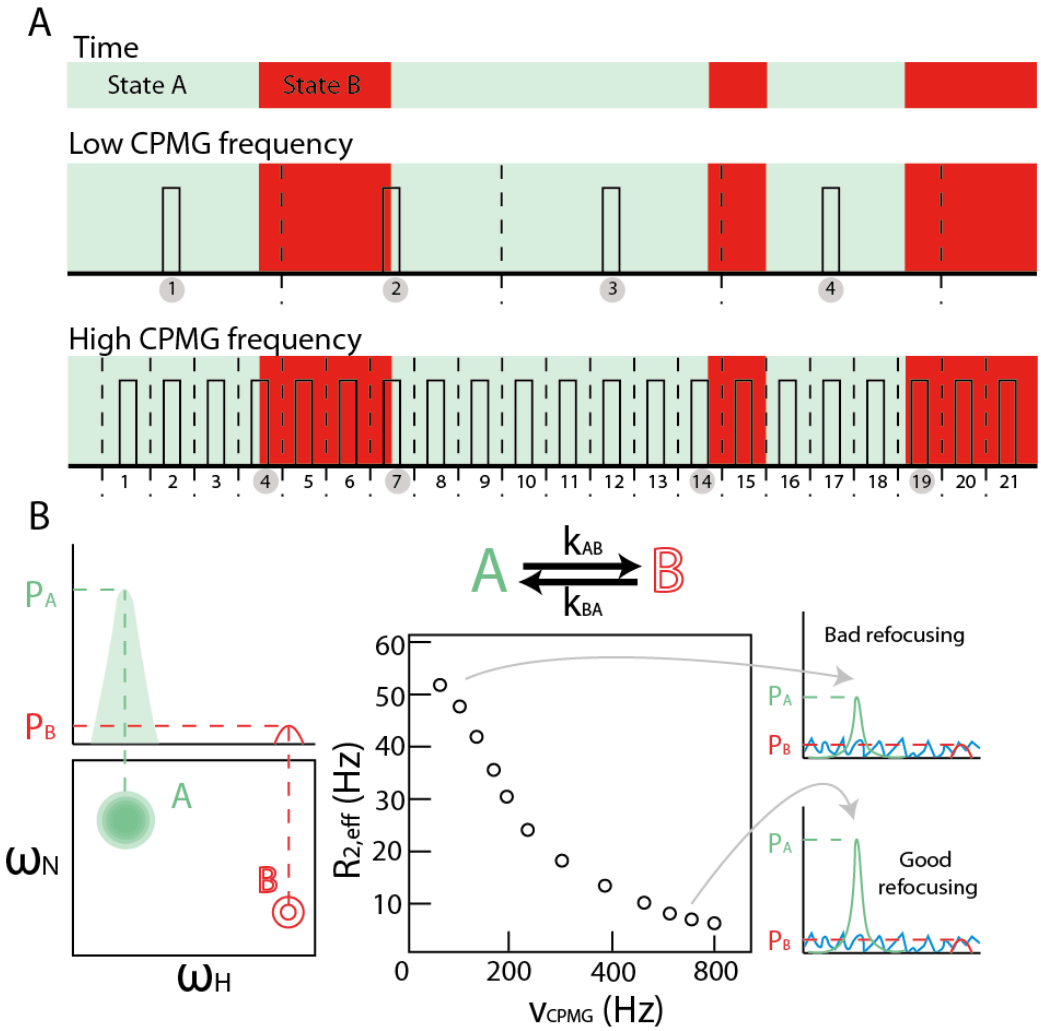

Figure 1.11: Relaxation dispersion NMR.
(A) A spin exchanges between states A (green) and B (red). 180° refocusing pulses (open boxes) are applied at different frequencies leading to a modulation of the relaxation rates. Pulses for which the refocusing is not complete, and the recovery of the signal is poor are noted with grey circles. Apapted from[170]. (B) Simulated 1D and 2D spectra of the slow exchanging system shown in (A) with two visible peaks corresponding to the major state A and the minor state B. In cases where the population of state B is low, direct detection of its chemical shift is not possible. The middle panel shows a simulated CPMG profile of relative peak intensity ($R_{2,eff}$) as a function of the frequency of the CPMG pulse train. When the refocusing pulses are applied with a low frequency, the refocusing is poor resulting in a signal of reduced intensity. However, increasing the pulse frequency improves refocusing (compare with A). Direct detection of the minor peak is prohibited as it is masked in the spectral noise (blue lines).

Even though relaxation dispersion is a relatively new technique it has been used to investigate protein folding[186,187], protein-ligand interaction[188] or protein aggregation[189,190], enzymatic catalysis[191] and protein degradation. Its main disadvantage is the relatively short window on the exchange regime that

it has access to (millisecond timescale). However this problem can be overcome by examining the relaxation in the rotating frame ($R_{1\rho}$)[165,192] or using more sophisticated pulse sequences[193]. It is often used to probe intramolecular conformational dynamics but since the exchange rate of the majority of protein-protein interactions falls within the detection limits of this technique (μs –ms) RD NMR has been used to study the kinetics and the conformational changes upon protein association[190,194]. In the present thesis, relaxation dispersion will be used in order to link protein structure and dynamics in the early stages of amyloid fibril formation.

1.5.4.2 Paramagnetic relaxation enhancement (PRE): characterising transient states and complexes

Much like the CPMG relaxation dispersion methodology, paramagnetic relaxation enhancement (PRE) has a long history. In fact the equations describing the PRE phenomenon were first described in Solomon's famous paper alongside the nuclear Overhauser effect (nOe) in 1955[195]. PRE arises from the magnetic dipolar interactions (in most cases) between a nucleus and an unpaired electron of the paramagnetic probe. While both nOe and PRE show the same r^{-6} distance dependence (between the neighboring nuclei-nOe, or the nucleus with the paramagnetic probe-PRE) owing to the large magnetic moment of the electron the PRE phenomenon extends to much larger distances, up to 35Å depending on the paramagnetic group used)[196-198]. The effect of the dipole-dipole interaction is an increase in the relaxation rates of the nuclear magnetisation, which can be measured using standard relaxation pulse schemes (Figure 1.11). Typically, relaxation rates (normally ^1H) are measured in the presence of a spin label and subtracted from their corresponding values when the spin label is inactive (reduced) resulting in the measurement of the PRE Γ_2 rate[197]:

$$\Gamma_2 = R_{2,para} - R_{2,dia} \qquad \text{equation (7)}$$

$$r = \left\{ \frac{K}{\Gamma_2} \left(4\tau_c + \frac{3\tau_c}{1+\omega^2\tau_c^2} \right) \right\}^{1/6} \qquad \text{equation (8)}$$

where $R_{2,para}$, $R_{2,dia}$ are the relaxation times in the paramagnetic and diamagnetic sample respectively, r is the distance between the spin label and the nucleus, τ_c is the correlation time of the electron–nucleus interaction (assumed to be the global correlation time of the protein for nitroxide spin labels), ω is the Larmor frequency of the spin (proton) and K is $1.23*10^{-32}$ cm^6s^{-2}. For a more detailed description of the PRE phenomenon see Section 2.1.2.

The subtraction in eq. 7 cancels out all relaxation processes that are common in the reduced and oxidised states, and leaves the paramagnetically-induced relaxation as the only difference[197]. Alternatively, the same measurement can be performed by comparing the peak intensities in a simple HSQC-type of experiment in a sample containing the oxidised (I_{ox}) versus the reduced (I_{red}) spin label (Figure 1.12A). The I_{ox}/I_{red} ratio can then be converted to PRE rates or directly into distances. Since the direct measurement of the PRE rates requires very high signal to noise ratio to allow robust fitting and is more time consuming, determination of the I_{ox}/I_{red} ratio is usually the method of choice, especially for proteins of higher MW (>40KDa)[162,197,199-202]. However, this can result in biased PRE rates because of other NMR phenomena contributing to the observed peak intensity in the presence of the paramagnet (e.g. increase in the T_1 relaxation), which are not taken into account. In this thesis all PRE experiments were carried out by direct determination of the Γ_2 rate in a full ^1H-T_2 relaxation experiment with 5 or 6 time-points (Figure 1.12B). Other phenomena, apart from the PRE, occurring in paramagnetic systems include pseudo contact shifts (PCSs)[203-205] and residual dipolar couplings (RDCs)[206] that were mainly used to investigate metal binding proteins, but those are beyond the scope of this thesis.

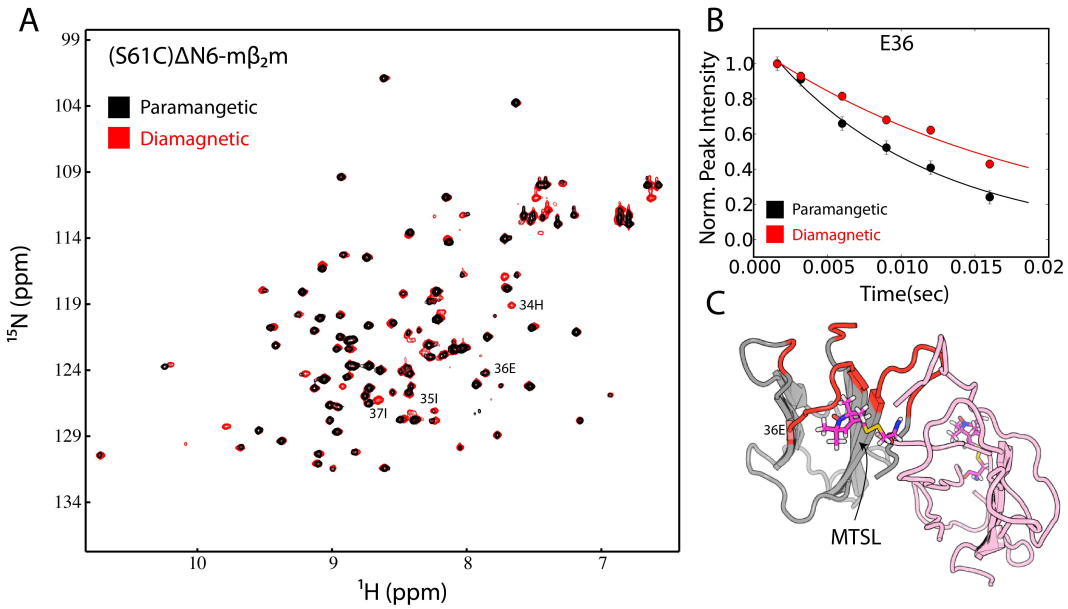

Figure 1.12: Principles of paramagnetic relaxation enhancement.
(A) Overlay of a ^{1}H-^{15}N HSQC spectrum with the spin label oxidized (paramagnetic sample – black) with the same spectrum after reduction of the spin label (diamagnetic sample – red). Missing peaks from the paramagnetic spectrum belong to amides that locate closely to the position of the spin label. (B) ^{1}H-R_2 relaxation rates (paramagnetic-black, diamagnetic-red) for the E36 amide resonance. Solid lines represent fits to single exponentials. (C) Location of the MTSL spin label in the complex (intermolecular PRE). Two protein subunits are shown in cartoon representation and the MTSL side chain is shown in sticks. Areas with increased PRE Γ_2 rates are coloured red. The example data used for this figure will be discussed in Chapter V.

The problem with nOe restraints in structure calculations relates to the severe resonance overlapping observed in NOESY-type of experiments, even in spectra of high dimensionality, which makes the unambiguous assignment of nOe crosspeaks difficult, if not impossible[207]. Different approaches have been implicated in order to deal with this issue[208], but due to the nature of the problem they tend to be erroneous, resulting in significant delays in data analysis and interpretation. PRE-based restraints, on the other hand, are very straightforward to assign because the nuclei involved are known based on previously performed through-bond assignment experiments. Based on eq. 9, PRE rates can be converted into distances and therefore used as nOe restraints in a classic NMR structure calculation approach. However, this methodology requires the introduction of upper and lower limits to the PRE distances resulting in a significant loss of information. Alternatively, Clore and co-workers presented recently a sophisticated theoretical frame to allow direct

back-calculation of Γ_2 rates from three dimensional (3D) structure, which is advantageous since it provides a tool for the full description of the PRE phenomenon in 3D space[209,210]. Theoretical and practical considerations about how this is achieved will be presented in more detail in Section 2.1.2.

Owing to all the aforementioned advantages of the PRE technique, it has been used to assist the structure calculation of large proteins[199], characterise the domain organization of multidomain systems[162,211,212], study protein-DNA interactions[213-215] or to interrogate intrinsically disordered proteins[216,217]. However, one of the most fascinating features of this technique is its ability to provide unique information about exchanging systems. In this case the use of the PRE-derived information depends highly on the timescale of exchange. If the motions lie in the slow exchange regime, the PRE data can offer a significant improvement in the quality of NMR structures in comparison with the classical nOe type of calculations (Figure 1.13). Such an approach is depicted in the case of membrane proteins[199] or DNA-protein complexes[209] (Figure 1.13), where the implementation of PRE and RDC NMR restraints, significantly improves the quality of the structures obtained.

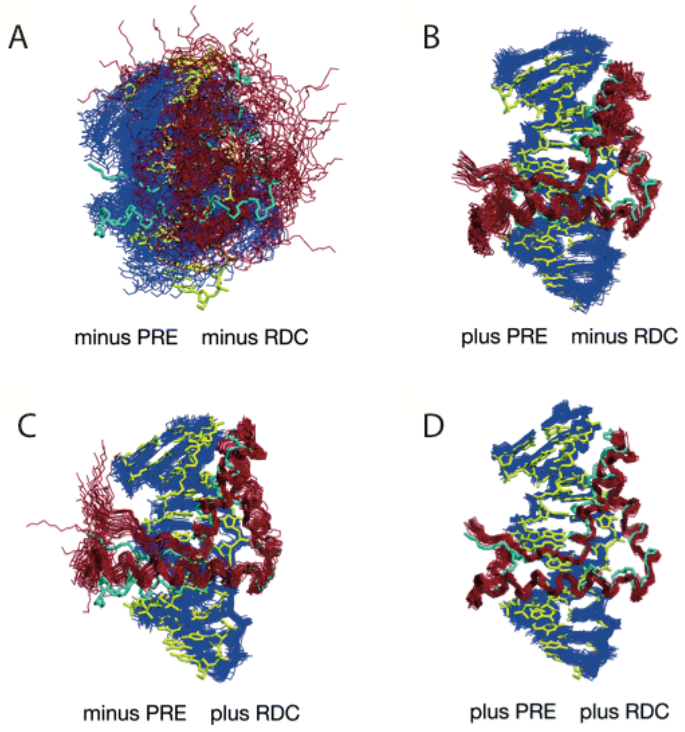

Figure 1.13: PREs in structure calculations.
Simulated annealing calculations for the SRY-DNA complex where only a single intermolecular nOe restraint is used. The 30 lowest energy structures with no

PRE and RDC (A), only PRE (B), only RDC (C) and both PRE and RDC (D) restraints are shown. The structure calculated with the full set of 168 intermolecular nOe restraints is shown as cyan ribbon for comparison. Adapted from[209].

On the other hand if the timescale of exchange is fast, application of the PRE methodology allows the characterisation of dynamic processes and the identification and structural elucidation of lowly populated intermediate species. More specifically, based on McConnell's equations[178], the apparent PRE rate (Γ_2) is highly dependent on the rate of the exchange event (k_{ex} $=k_{AB}+k_{BA}$) between states A and B. If the exchange takes place in the slow PRE regime, then the presence of the minor species B has no effect on the measured PRE Γ_2^{app} rate which in this case represents only the major species A. However, for larger k_{ex}, the Γ_2^{app} rate is highly influenced by the minor state B (see Figure 1.14). In conclusion, when $k_{ex}>>\Gamma_{2,B}-\Gamma_{2,A}$ (fast exchange regime), the apparent Γ_2 rate is the weighted populated average of the Γ_2 rates of the two species[210]:

$$\Gamma_2^{app} = p_A\Gamma_{2,A} + p_B\Gamma_{2,B} \qquad \text{equation (9)}$$

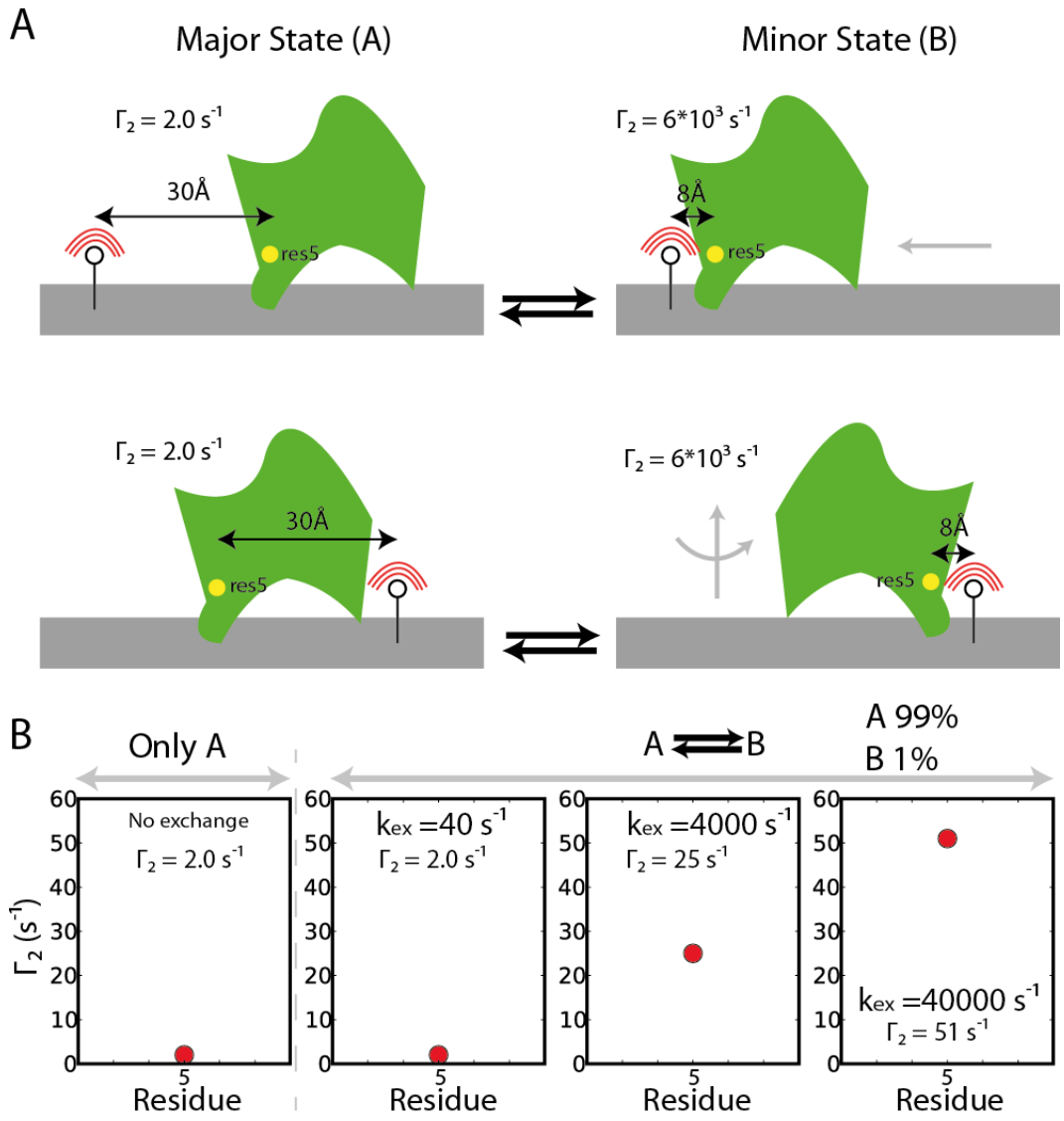

Figure 1.14: PREs in exhcanging systems.
(A) Shematic of an exchanging sytem where the major state A locates far from the spin label (exemplified by residue 5- cyan dot) giving rise to very small PRE effect, whereas the minor state B has a large Γ_2 rate. (B) Simulated PRE rates for residue 5 in (A) as a function of the exchange rate. As k_{ex} increases, the observed rate is dominated by the presence of the minor state B, while the exchange phenomenon does not have any effect when the exchange is slow. Adapted with permission from[210]. © 2009 American Chemical Society.

This aspect of the PRE phenomenon makes this technique ideal for studying transient reactions that have very short lifetimes, such as transient macromolecular interactions or conformational sampling in IDPs. Using this approach Clore et al. observed that the PRE rates for the EIN-HPR interaction cannot be explained only by the stereospecific complex, whose structure is available (Figure 1.8B)[122]. These data therefore, allowed the detection and structural determination (since PRE is a structural and not kinetic method as

the CPMG) of an ensemble of 20 transient encounter interactions between the proteins shown in Figure 1.8C[69].

1.6 Protein misfolding – Amyloid

1.6.1 Protein misfolding in disease

Several human diseases are connected with protein misfolding that affects normal cell function in a variety of ways. A classic example is cystic fibrosis. In this case, mutations in a transport protein result in the misfolding and therefore reduced secretion of the protein in amounts not sufficient for proper functionality[218,219]. In the same category fall numerous diseases that result in protein aggregation and the formation of insoluble material known as amyloid[220]. These diseases affect a variety of organs and tissues such as liver, kidney and brain and have gained an intense scientific interest over the years[221-223]. Amyloid deposits share a common fibrillar morphology in most cases, reporting on the same underlying structural architecture of their core (known as cross-β), which is closely related in all cases, even though the amino-acid sequences of the proteins involved are not related to each other[224-226]. Amyloids have been connected to numerous pathologic disorders including the Alzheimer's disease, Parkinson's disease and Creutzfeltd Jacobs disease[227-229]. Despite the numerous studies published up to date, the debate about the mechanism of cytotoxicity and the identification of the species on pathway to fibril formation that represent the cytotoxic entity is still active nowadays. However, the observation that fibril load is not correlated with the appearance of pathological symptoms[230-237], has focused attention on early stage oligomeric species or other aggregates before the formation of the end product of fibrils, as responsible for cell death[238-243]. Despite the numerous studies reporting on the toxicity of oligomers, the role of fibril themselves in the cytotoxic mechanism cannot be disregarded[244,245].

Recently, amyloid fibrils have been connected to various functional roles apart from cell death, such as storing hormones, transcription factors, protective shells and many others, suggesting that the amyloid fold is not toxic in all cases

and thus the unique mechanical properties of these molecules can be used to design nanomaterials[246-249].

1.6.2 What is amyloid? Common structural features

Back in the 19th century, in the early stages of amyloid research, the deposits found in diseased tissue were thought to consist of starch due to the blue colour they adopt after staining with iodine[250]. Later on, the proteinaceous nature of those deposits was revealed[223].

Amyloid deposits are structures of fibrillar morphology[225,251]. The fibrils are usually straight, unbranched filaments, ~100Å in diameter as shown in transmission electron micrographs (TEM), they bind amyloid specific dyes such as the fluorescent dye Thioflavin T (ThT) and Congo Red, which shows the characteristic red-green birefrigerence when imaged under polarised light. One of the first structural features of amyloid fibrils revealed in the 1960's by X-ray diffraction was their 'cross-β' architecture[252], which is shared by the vast majority of amyloid fibrils independently of the precursor protein and the conditions employed. It consists mainly of two reflections visualised by fibre diffraction experiments: one equatorial at ~10Å, which is thought to be due to the packing of β-sheets, perpendicular to the axis of the fibre, and one meridional at 4.7Å, which is due to the packing of adjacent β-strands along the axis of the fibre[21] (Figure 1.15). A variety of techniques have been applied over the years in the pursuit of structures of proteins in the amyloid form, ranging from limited proteolysis, scanning electron microscopy, atomic force microscopy, cryo-electron tomography, solid-state NMR, electron paramagnetic resonance spectroscopy (EPR) and X-ray crystallography[253,254]. Remarkably, even though all amyloid fibrils share a common cross-β fold in their hydrophobic core, their overall structure is polymorphic, a phenomenon that can be due to different packing arrangements of the β-strands (parallel, antiparallel, head –to-head, tail-to-tail), or due to the different arrangement of protofibrils inside the mature fibre[255,256].

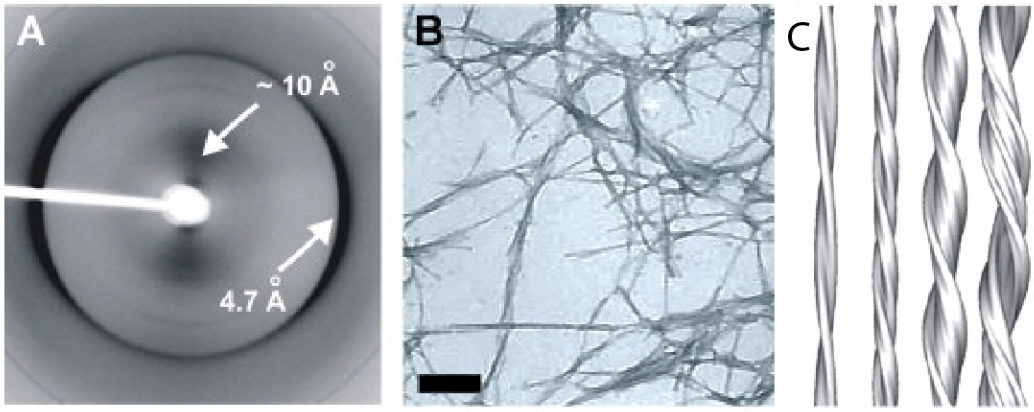

Figure 1.15: Features of amyloid fibrils.
(A) Characteristic cross-β fibre diffraction pattern of amyloid fibrils showing the typical meridional and equatorial reflection at 4.7 and 10Å respectively. (B) A typical long straight morphology is visualised by EM (bar denotes 100nm). (C) Low resolution 3D reconstruction of fibrils made by insulin, showing 2, 4 or 6 protofilaments, all with different morphologies. Adapted from[224].

1.6.3 Mechanisms of amyloid formation

In order to design therapeutic strategies that target the formation of amyloid and therefore prevent the appearance of pathology, it is crucial to understand the principles that govern fibril formation. The elucidation of a generic mechanism for fibril assembly has been difficult so far mainly because of the heterogeneity in the aggregation pathways different proteins follow. Even though the structure of amyloid is very similar independent of the protein precursor, the mechanistic details for its assembly may vary dramatically. Despite the numerous models that have been proposed over the years to explain mechanistic details (such as the conformational change hypotheses where a major conformational change leads to the formation of the cross-β fibril core[30], or the native-like aggregation, according to which fibril assembly proceeds via native-like intermediates[257]), it is widely appreciated that the general frame for amyloid formation is nucleation growth[258,259]. As with any nucleation-dependent reaction, amyloid assembly shows a lag phase, which can vary in length depending on the aggregating protein and the conditions employed, followed by a rapid elongation phase (Figure 1.16). During the lag phase, protein-protein interactions play a crucial role, with protein molecules associating in a transient manner to form nuclei that may have, or lack, the potential to elongate, thus creating a very heterogeneous pool of on- or off-pathway species. The detection and characterisation of these metastable

species even though experimentally challenging, is key to therapy, since they have been identified as membrane disrupting/cytotoxic entities[260-262]. When the nucleus has reached a critical size, the soluble to insoluble transition commences by attracting monomeric molecules to the surface of the pre-formed nucleus (elongation phase) in a process that is enthalpically driven (Figure 1.16). This nucleation-dependent amyloid growth can be followed experimentally by the amyloid specific dye thioflavin T (ThT) which binds predominantly to cross-β structures, resulting in a characteristic sigmoidal curve, representative of amyloid growth (Figure 1. 16).

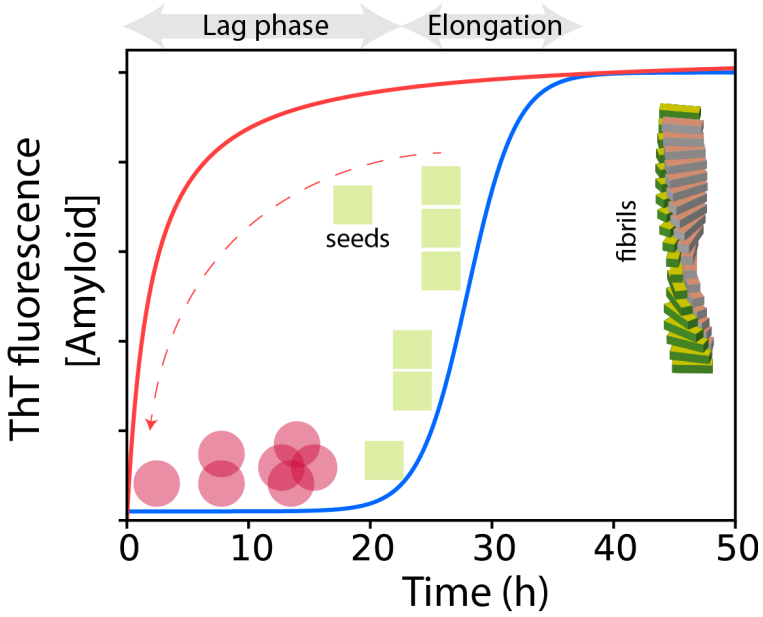

Figure 1.16: Nucleation growth and seeding.
Simulated curves for a nucleation-dependent growth mechanism with a lag time of 22h (blue line). Native-like oligomeric species (red spheres) accumulate during the lag time, before the appearance of cross-β species (green rectangles) that elongate to form amyloid fibrils. Seeding the reaction with performed nuclei (seeds) diminishes the lag phase and the reaction proceeds fast to completion (red line). Adapted from[224].

However, the experimentally observed microscopic rates of fibril formation cannot be explained adequately by the homogenous primary nucleation mechanism shown in Figure 1.16. Instead, secondary processes in which already formed nuclei/fibrils assist *de novo* formation of cross-β species, in either a monomer-independent (fibril fragmentation)[263,264] or monomer-dependent manner (secondary nucleation, where the surface of preformed nuclei catalyse the soluble-to-insoluble transition of monomeric protein)[265],

have been shown to dominate fibril assembly, highlighting possible new therapeutic avenues.

Regarding the phenomena that precede the formation of the critical nucleus various models have been proposed over the years. The observation that the protein precursor needs to at least partially unfold in order to adopt the cross-β structure led to the 'conformational change' hypothesis, which was strongly supported by the fact that under denaturing conditions most proteins have the ability to aggregate[266-268]. However, in the case of natively folded proteins, that aggregate under close to physiological conditions, the scenario must be more complex, since these proteins are unlikely to expose amyloidogenic segments under these conditions. In such systems protein dynamics and the ability to visit rarely populated, potentially more aggregation-prone species in equilibrium have been connected to increased amyloidogenicity[257]. A <1% increase in the population of non-native species for human lysozyme[268] and HypF-N[269,270] is able to tip the balance towards misfolding and aggregation showing that cooperative folding might have evolved as a protection mechanism against aggregation for thermodynamically unstable proteins.

1.6.4 Protein-protein interactions in the early steps of amyloid formation: glimpses on inhibition

In the pursuit of linking structural properties to amyloidogenicity, many studies have focused on analysing the properties of monomeric amyloid precursors. Unfolding energies, hydrophobicity, propensity to form inter/intra-molecular contacts, protein dynamics have all been linked to increased amyloid propensity[271-273]. The wealth of data available, allowed the generation of algorithms capable of predicting amyloidogenicity, based on either the physicochemical properties of the sequence involved or even based on structure (reviewed in[274]). On the other hand, scientific research has focused on elucidating the structure of the end product amyloid fibrils in order to establish their role in cytotoxicity and/or protein function. To that extent the structures of small amyloid forming peptide segments have been solved by X-ray crystallography[256,275,276], solid state NMR models have been proposed for

a variety of amyloid protofilaments, including $A\beta_{40}$[277-279] and the full C-terminal domain of HET-s[280], until the most recent characterisation of the first structure of an essentially complete amyloid fibril (and not only a protofilament) using a combination of solid-state NMR and electron microscopy[281]. However, the characterisation of the oligomeric species in between those two end points has proved difficult because of their transient nature and vast heterogeneity. These species are often termed 'amyloid oligomers' and their short lifetimes, polymorphic nature, dynamic properties and the variety in mechanisms of formation, make them inaccessible to most current biophysical tools. In addition, the distinction between off- and on-pathway interactions remains one of the most challenging aspects of their characterisation. Amyloid oligomers on the other hand, show increased ability to penetrate lipid bilayers and enhanced cytotoxic potential in comparison to monomeric or fibrilar forms of the same protein precursors, highlighting their importance in the aggregation pathway[241,282,283].

Recently, a distinction between on- and off-pathway assemblies in the early stages of amyloid formation was made possible using fluorescence techniques[284]. Cremades *et al.*[285] and Campioni *et al.*[286] in two independent studies showed that human α-synuclein and bacterial HypF aggregate through the formation of oligomeric species of different size and structure and as a result different cytotoxicity. In the later case, single molecule fluorescence techniques and the analysis of individual species in solution allowed the direct observation of the inter-conversion between the different oligomeric forms of α-synuclein. These findings highlight the dynamic nature of protein-protein interactions in the early stages of amyloid formation, giving rise to species of the same overall shape but different in structural details sufficient to make them toxic (or not toxic) to the cell. Both studies, even though they provide insights into the overall architecture, size and shape of the oligomeric assemblies fail to reveal their structure in atomistic detail. Eisenberg and co-workers reported the first high resolution structure of a supposedly toxic oligomer that appears to be on pathway to fibrils derived from a peptide segment of αB-crystalin[287]. Its structure revealed a novel barrel-like cylindrical fold, termed cylindrin, that is only marginally destabilised in

comparison to the classic steric zipper form that amyloid peptides often adopt in crystals (Figure 1.17). Interestingly, the cylidrin architecture appears to be more generic and not only related to αB-crystalin[288].

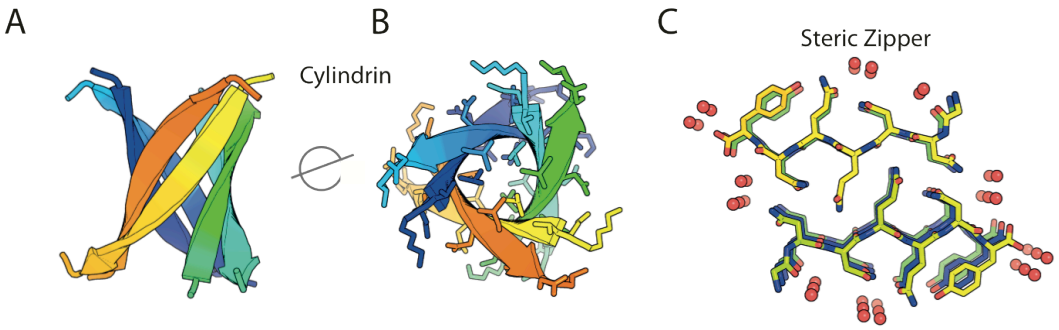

Figure 1.17: Crystal structure of a toxic oligomer.
Structure of the KVKVLGDVIEV (K11V) peptide of αB-crystalin. In its oliogmeric form, K11V assembles into a cylindrical β-barrel (cylindrin, pdbID:3SQN)[287] (A-B). The structure of the steric zipper of the GNNQQNY peptide from Sup35 in its amyloid form is shown in (C) for comparison. In comparison to the steric zipper, the cylindrin structure has a 0.75 score for shape complementarity in the interface in comparison to 0.80 for the steric zipper and a BSA per strand of 85Å^2 rather than 135Å^2 for the steric zipper, suggesting that it is somehow less stable[287].

A lot of research has focused on identification of potent inhibitors of amyloid formation in an effort to delay or diminish the onset of disease[289]. Even though the majority of these studies focus on the discovery and characterisation of small molecule inhibitors such as polyphenols or green tea extracts[290-297], the role of protein-protein interactions is increasingly appreciated. Biomolecular association of transthyretin subunits within the native tetramer stabilises the native fold and prevents release of the amyloidogenic monomers that aggregate through the formation of native-like intermediate species.[298,299]. Similarly, interface mutations in superoxide dismutase-1 (SOD-1) that stabilise the dimeric form of the protein, decrease the propensity to form amyloid[300], while interaction of lysozyme with a single chain antibody restores the global cooperativity of folding and reduces amyloidogenicity[301]. *In vivo*, nature has evolved an entire system to maintain protein solubility where chaperones that act as bodyguards of the proteome by facilitating folding or preventing misfolding play a central role. Protein-protein interactions are key components of the so-called 'proteostasis'[302,303] network as exemplified by Hsp70[304].

1.6.5 Beta$_2$-microglobulin

1.6.5.1 Structure – Normal function

Beta$_2$-microglobulin (β_2m) constitutes the light chain of the class I human Major Histocompatibility Complex I (MHC class I)[305,306], a crucial player of the immune system of humans responsible for presenting self and non-self antigens to T cells via the T cell receptor (TCR)[306]. MHC I retrieves polypeptides from the cytosol and displays them on the cell surface where they can be recognised by T cells[306]. It consists of 8 β-strands, rearranged in an antiparallel manner in a classic immunoglobulin-like fold with a Greek key topology (Figure 1.18). Importantly, β_2m has five proline residues in its sequence, with one of them, Pro 32 adopting the thermodynamically unfavourable cis conformation, indicative of conformational restrictions in the area surrounding it (BC loop).

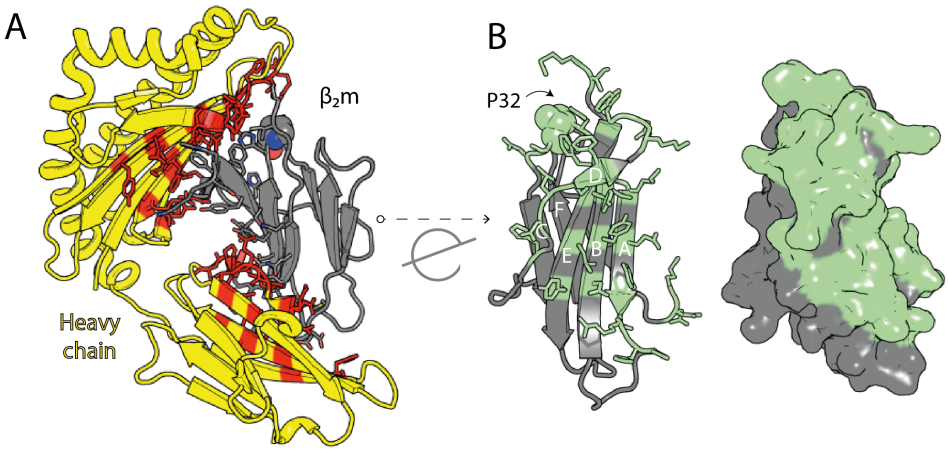

Figure 1.18: B_2m in its MHC-I bound and free form.
(A) Cartoon representation of the full human MHC-I complex (pdbID: 3MYJ)[307]. The heavy chains (α1, α2, α3) are shown in yellow and β_2m in grey. Interface residues were identified using a cut-off of 5Å^2 of BSA and are shown as red or grey sticks. P32 is highlighted in spheres. (B) Structure of monomeric β_2m shown as cartoon (left) or surface representation (right). Residues that participate in the interface with the heavy chain are shown in sticks and in green colour.

MHC-I is a key component of the immune system with a vital role in self/non-self recognition. Antigenic peptides originating from proteasomal degradation of self or non-self proteins[308,309] are transported to the endoplasmic reticulum, where MHC-I is assembled. After recognition, the full complex is transferred to

the membrane of all nucleated cells, in a process for which correct binding of β_2m is crucial[310,311]. Thus, the role of MHC-I molecules is to present self or non-self peptides at the cell surface where they can get recognised by the T-cell receptor. Complex resistance mechanisms link the activation of cytotoxic T-cells only after the recognition of a non-self antigen leading to cell death.

1.6.5.2 B_2m-related amyloidosis

As β_2m is non-covalently bound on the heavy chain of MHC-I it is constantly turned over in the body. In healthy individuals, free β_2m is removed by filtration in the kidneys leading to a plasma concentration of 0.16μM[312,313]. Due to the inability of dialysis treatment to clear β_2m from the blood, patients suffering from renal failure and undergoing dialysis treatment for several years observe an elevated level of β_2m in serum. The concentration of circulating β_2m can increase up to 20-fold (3.2μM), leading to the accumulation of the protein in the form of amyloid fibrils in osteartícular tissues in a pathological condition known as dialysis related amyloidoisis (DRA)[312,314] (Figure 1.19)

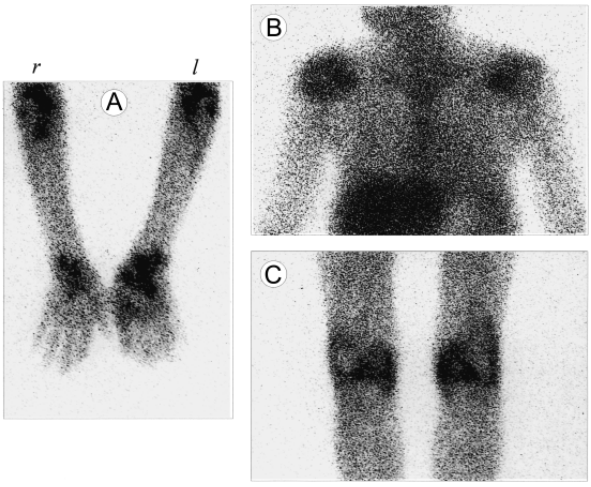

Figure 1.19: Accumulation of β_2m in joints.
A patient suffering from DRA imaged with indium (In-111)-labelled β_2m. Radioactive β_2m accumulates in joints (elbows, knees, wrists, hips, and ankles). Staining of the liver is due to uptake of the tracer and does not indicate amyloid fibril accumulation. Adapted from[315].

However, the increased concentration of β_2m in serum does not correlate with disease and hence, additional factors must be involved in the appearance of DRA[314,316]. Factors proposed involve post-translational modifications[317-321],

interaction with physiological factors such as serum amyloid protein, apolipoprotein E, fatty acids, collagen or the presence of ions such as Cu^{++} [322-330]. Moreover, the presence of β_2m amyloid fibrils in uraemic patients that do not undergo dialysis suggests that the self-association of wild-type β_2m (hβ_2m) molecules is insufficient for the onset of disease[331,332].

The accumulation of β_2m amyloid fibrils in joints is believed to be predominantly due to the high affinity of β_2m for collagen, which results in fibril deposition on the surface of the cartilage in all peripheral joints[314,315]. Although amyloid fibrils are detectable a few months after dialysis, the symptoms of DRA appear at least 5 years latter. They generally include carpal tunnel syndrome, bilateral chronic arthaglias and arthrophagy which can result in complete joint destruction caused by the recruitment of macrophages and monocytes in response to inflammation caused by fibrils[333,334].

Very recently, the presence of β_2m amyloid fibrils was associated with an autonomic neuropathy accompanied with gastrointestinal symptoms caused by a hereditary systemic amyloidosis[335]. The four members of the same family suffering from the disease, showed normal concentrations of β_2m in the plasma and no amyloid deposition in joints, in contrast to DRA. Deposition of fibrils occurred mainly in spleen and adrenal glands and fibrils consisted of the D76N β_2m variant, which showed also increased amyloidogenicity *in vitro* in opposition to hβ_2m. Intriguingly, the crystal structure of the D76N-β_2m variant showed minor differences to hβ_2m, apart from the formation of two new hydrogen bonds that stabilize the EF loop. The decreased stability of this variant is not sufficient to explain its increased amyloidogenicity (as many other β_2m mutants show decreased stability but no amyloidogenicity[336]), highlighting the lack of mechanistic details on β_2m aggregation.

1.6.5.3 Investigating β_2m aggregation *in vitro*

B$_2$m can assemble to amyloid fibrils that possess all the typical amyloid features from its unfolded state. Acid unfolded β_2m (pH 2.5) self-assembles to long straight amyloid fibrils that give a characteristic cross-β pattern, bind congo red and show a nucleation-dependent polymerisation[263,337].

Fragmentation of already made β_2m fibrils that are then used as seeds diminishes the observed lag time as shown in Figure 1.16, but also plays a crucial role in *de novo* fibril formation as revealed by detailed modelling of the observed kinetics of fibril formation at low pH[263]. As seen in Figure 1.18, binding of β_2m to the heavy chains of MHC-I is mediated by hydrophobic residues in the DE loop and the E strand (L54, F56, W60, F62, Y63, L65, Y67). The same regions of the protein are predicted to be the most amyloidogenic parts of the protein by aggregation prediction algorithms, suggesting an imbalance between regulation of protein function (binding to MHC-I) and aggregation propensity. Supportive of this hypothesis is the finding that the W60G mutation virtually abolishes the ability to bind to MHC-I, enhances protein stability and greatly reduces the amyloidogenic potential of this variant[338]. Importantly, slight changes in the conditions of the aggregation reaction (from pH 2.5 to pH 3.6 and increased ionic strength) result in a complete remodeling of the resulting fibrils[336,339]. B_2m fibrils at pH 3.6 are short in length (<500nm) and display a curved morphology as an outcome of an amyloid forming reaction with no significant lag time. Analytical ultracentrifugation (AUC) and electrospray ionization mass spectrometry (ESI-MS) experiments showed that polymerisation under these conditions is consistent with a monomer addition model in vast contrast with the nucleation–dependent reaction at pH 2.5[340]. Long-straight amyloid fibrils formed at pH 2.5 have been subjected to extensive analysis by solid state NMR (ssNMR)[341,342] to reveal a parallel in register arrangement of the β- strands in the fibril in agreement with electron paramagnetic relaxation (EPR) studies[343] and in opposition to the anti-parallel configuration found in the native state[344], suggesting that under physiological conditions a dramatic conformational change has to occur before amyloid assembly can proceed. 3D cryo-EM reconstruction of tomograms of hβ_2m fibrils formed at pH 2.5 revealed a dimer of dimers as the building block of these assemblies[345]. Interestingly, fibrils assembled at pH 2.5 were found able to interact with lipid bilayers and cause membrane damage mainly through their ends[346,347], suggesting that they are directly involved in cytotoxicity[346,347].

While amyloid formation proceeds readily from the unfolded or partially structured state (pH 2.5 and pH 3.6 respectively) this is very different from the *in vivo* conditions (pH ~7.0) where fibril assembly has to commence mainly from a natively folded precursor. As hβ_2m is incapable of fibril formation at neutral pH, there must exist another species that is populated at equilibrium, which shows an increased amyloid potential. Proteins such as transthyretin, immunoglobulin light chains and lysozyme are known to aggregate through a partially folded species that retains its native-like characteristics and thus complete unfolding of the protein is not necessary for amyloid formation to commence. The absence of a high kinetic barrier allows the protein to visit the aggregation-prone state through fluctuations of the native fold (protein dynamics) under physiological conditions in a process termed 'native-like aggregation'[257,271,348-351]. Following that observation, a long pursuit has started in order to identify folding intermediates on pathway to native hβ_2m that are potentially more aggregation prone than the native state. In these studies, two folding intermediates were identified with the first (I_1) being populated faster than milliseconds, characterised by typical burst phase kinetics in fluorescence experiments[352], and the second species (I_2) that forms milliseconds after I_1, that displays many native-like characteristics[352]. Folding of I_2 to the native state is hampered by increased energy barriers resulting in a very slow folding phase that takes place in milliseconds to seconds at 30^{o}C[353,354]. Detailed kinetic studies accompanied by site directed mutagenesis at Pro32 identified the cis/trans isomerisation of this residue as the limiting factor in the folding to the native state and therefore the trapped folding intermediate was termed I_T[353,354] (I_T and I_2 even though might represent different species but will be termed I_T throughout for simplicity). The rate constant for the transition of I_T to the native state (N) is 0.3-0.5h^{-1}, reminiscent of the rate of cis-trans Pro isomerisation[355], while double jump experiments and substitution of Pro with Val at position 32[353,354] diminish the slow folding phase, revealing the critical importance for Pro32 isomerisation in the folding of β_2m. Importantly, the concentration of I_T is directly related to the ability of the monomeric protein to extend already made β_2m fibrils seeds[356,357], showing not only that proline isomerisation is crucial for folding to the native state but also highlighting the

importance of the population of this folding intermediate in altering the balance between folding and aggregation.

While kinetic analysis shows that I_T is populated up to 3.7±1.4 % at pH 7.0[357], this value is beyond the detection limits of many biophysical techniques, making this species an invisible protein state in solution. However, the need to overcome a high kinetic barrier results in relatively prolonged lifetimes for I_T, which is populated up to 20% following rapid dilution from the unfolded state[354] and opens the way for its characterisation using ultra-fast structural methods. Indeed, the direct characterisation of this slow folding intermediate was made possible by using SOFAST NMR experiments[358], to reveal conformational changes around Pro32, namely the BC and FG loops[359,360]. However these elegant studies fail to characterize the structural properties of residues in the very close proximity of the epicenter Pro32, since resonances in this area are broadened beyond detection.

Of the many variants designed to probe the energy landscape of hβ_2m, the one involving deletion of the N-terminal 6 residues is of particular importance. This truncated variant, termed ΔN6, comprises ~30% of β_2m molecules in fibrils from patients suffering from DRA[317] (but not from the β_2m-associated systemic amyloidosis[335]) making this variant a physiologically relevant β_2m species. Recent studies using analytical size exclusion chromatography (ASEC) and NMR demonstrated the structural similarities between ΔN6 and I_T, providing an excellent tool for the investigation of this lowly populated state of hβ_2m[272,361,362]. Crucially, ΔN6 shows an increased amyloidogenicity at neutral pH by being able to self-assemble to amyloid-like fibrils without the addition of co-solvents.

The high resolution NMR structure of ΔN6 has been solved recently and revealed a native-like fold with ~1Å backbone RMSD to hβ_2m and, as expected, a trans peptidyl bond at P32[272]. However, the rotation of the main chain at position 32 causes a substantial rearrangement of the hydrophobic core involving residues in the DE and BC loops mainly, and also in distant sites of the protein (Figure 1.20A, B). To accommodate the rotation of the peptidyl bond at

P32, the side chains of F30, F62, although buried in the hydrophobic core in the native state, are now solvent exposed, while the cavity created due to their displacement is now filled by the side chains of P32, L64 and I35 (Figure 1.20A, B). These conformational changes result in global destabilisation of the protein by 3.8 kJ*mol^{-1}, a high propensity for oligomerisation as shown by ASEC and increased ps-ns dynamics in the DE and BC loops, phenomena that combined together designate ΔN6 as one of the most aggregation-prone β_2m variants studied to date under close to physiological conditions. The crystal structure of ΔN6 in its dimeric form and in complex with a nanobody that served as a crystalisation chaperone has also been solved, to reveal a domain swapped configuration, where the C-terminal G strands have exchanged between the monomers, although the intramolecular conformational changes shown in the NMR structure of monomeric ΔN6 are still prevalent[363] (Figure 1.20C).

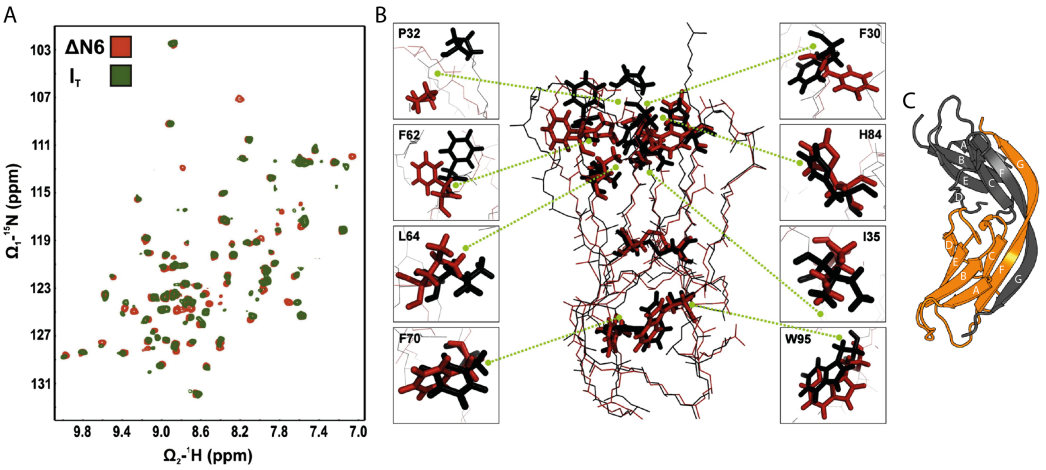

Figure 1.20: ΔN6 as a structural mimic of I_T.
(A) ^{1}H-^{15}N-HSQC spectra of ΔN6 (red) and I_T (green) obtained 3 min after refolding of hβ_2m from 25mM sodium phosphate, 8M urea, pH 7.5. (B) Repacking of the hydrophobic core. Side chains that undergo significant conformational changes in the structure of ΔN6 (red) in comparison to their position in hβ_2m (black) are shown in sticks. The protein backbone is shown as ribbon (pdbIDs: 2XKS, 2XKU). Panels adapted from[272]. (C) Structure of the domain swapped ΔN6 dimer. The monomers (orange and grey cartoon) have exchanged their G strands, with the FG loop acting as a hinge (pdbID: 2X89)[363].

The aggregation of ΔN6 shows a characteristic pH dependency, with no fibrils formed at pH 8.2 and enhanced amyloid potential at pH 6.2, a phenomenon which is also depicted in its pH-dependent T_2 dynamics[272]. This observation provides a link between protein dynamics and amyloid potential, suggesting that small differences in pH affect either the intramolecular motions that are

crucial for initiation of fibril formation or the intermolecular interactions between the ΔN6 monomers. The increased dynamics at low pH (6.2), conditions under which ΔN6 maintains a very native–like fold, could also indicate that proton uptake increases the population of an additional protonated intermediate subsequent to the formation of ΔN6 with increased propensity to aggregate (Figure 1.21).

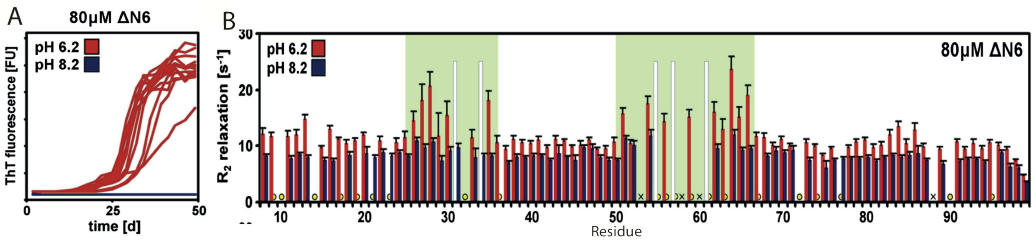

Figure 1.21: pH links dynamics and amyloidogenicity.
(A) Aggregation kinetics of 80μM ΔN6 at pH 6.2 (red) and 8.2 (blue), followed by ThT fluorescence. (B) NMR T_2-relaxation rates for 80μM ΔN6 at pH 8.2 (blue bars) and pH 6.2 (red bars). Areas showing increased dynamics (BC and DE loop) are highlighted in green boxes. Panels adapted from[272].

Apart from being able to self-assemble into fibrils at neutral pH, ΔN6 is remarkable in its ability to convert hβ$_2$m to an amyloid competent state, in a manner reminiscent of the way that prions transmit their infectivity[364]. Even when incubated in substoichiometric ratios (1:100 ΔN6:hβ$_2$m) monomeric ΔN6 enhances the amyloid potential of hβ$_2$m, a phenomenon which seems to be mediated by the partial unfolding of the A strand[272] (Figure 1.22).

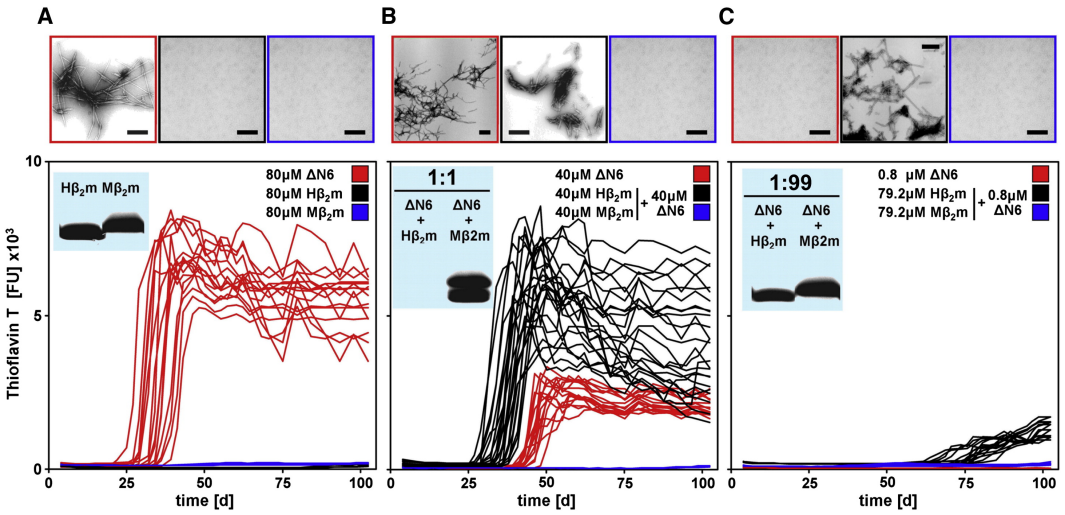

Figure 1.22: Promotion and inhibition of amyloid formation.
(A) Aggregation kinetics of 80μM ΔN6 (red), 80μM hβ$_2$m (black) and 80μM mβ$_2$m (blue). (B) Aggregation kinetics of 40μM ΔN6 (red) alone, incubated in a

1:1 molar ratio with hβ$_2$m (black) or with mβ$_2$m (blue). (C) Aggregation kinetics of 0.8μM ΔN6 (red) alone, or incubated with 79.2μM hβ$_2$m(black) or with 79.2μM mβ$_2$m(blue). Upper panels show electron micrographs of the end product of each reaction using the same colour coding. Insets represent SDS-PAGE analysis of the soluble fraction obtained after spinning down the fibril pellets. Adapted from[272].

While ΔN6 is one of the most fibrilogenic β$_2$m molecules, the murine version of β$_2$m (mβ$_2$m) shows a highly diminished aggregation propensity. Even under conditions that severely bias hβ$_2$m towards fibril formation (pH 2.0 and increased ionic strength-1.5M NaCl) this molecule is not able to self assemble into amyloid fibrils and only small spherical aggregates were visible in electron microscopy[365]. Preliminary experiments showed that when mixed at stoichiometric concentrations, biomolecular collision between ΔN6 and mβ$_2$m inhibits amyloid formation (Figure 1.22). Interestingly both mβ$_2$m and hβ$_2$m share a high sequence similarity with ~70% sequence identity and ~90% sequence homology. In structural terms, the similarity between those two molecules is striking, with less than 1.5Å RMSD difference (Figure 1.23).

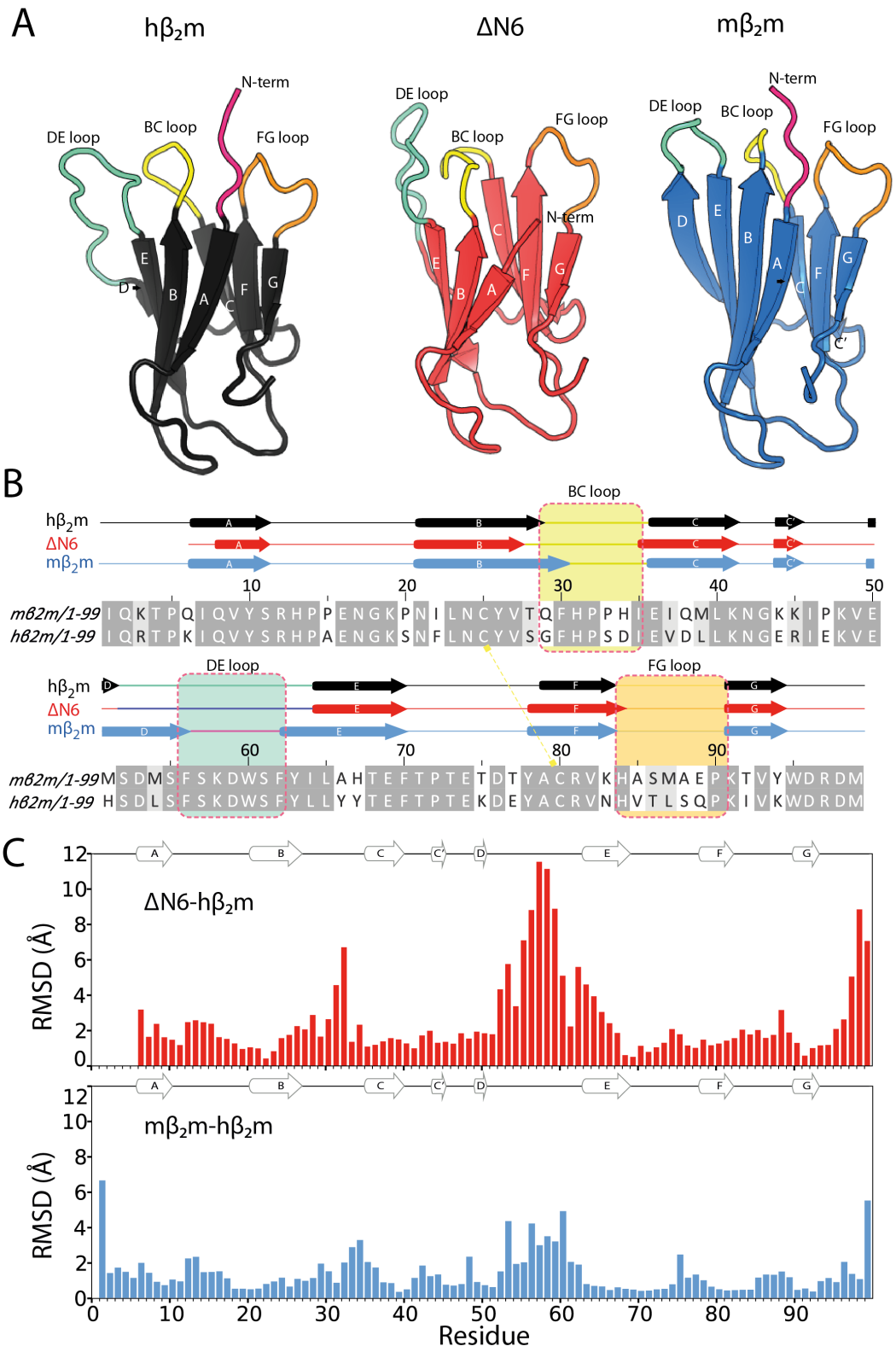

Figure 1.23: Structural analysis of β₂m variants.
(A) Structures of hβ₂m, ΔN6 and mβ₂m (left to right). The average backbone RMSD between hβ₂m and mβ₂m is 1.4Å, whereas for hβ₂m and ΔN6 it has a value of 1.1Å. (B) Sequence alignment of hβ₂m and mβ₂m. Regions in close proximity to P32 (BC loop, DE loop, FG loop) are highlighted in different colours on the structures in (A) and with dashed boxes in (B). (C) Per-residue

RMSD between hβ2m and ΔN6 (top) or hβ2m and mβ2m (bottom). RMSD values between the Cα atoms were calculated in XPLOR-NIH[366].

However, despite the remarkable overall similarity between these molecules, the mutations in the mβ2m sequence in the majority of cases involve the insertion or deletion of charged residues, a phenomenon that results in significant differences in the distribution of the electrostatic potential of mβ2m and hβ2m (Figure 1.24).

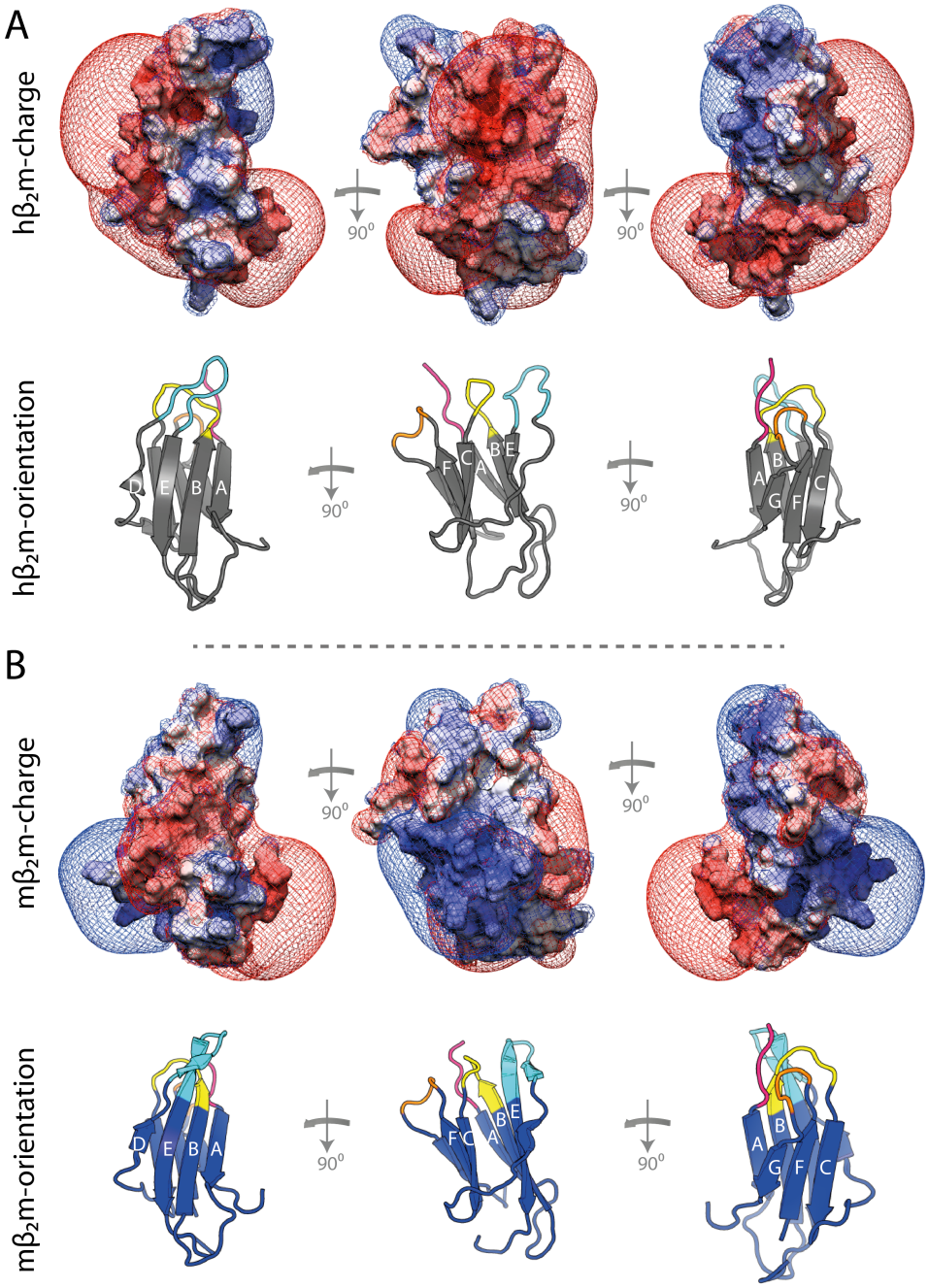

Figure 1.24: Charge distribution in hβ2m and mβ2m.
(A) The surface of hβ2m is coloured based on the electrostatic potential of the molecule calculated as ±2kT (blue positive, red negative, white no charge). The

electrostatic potential isosurface is also shown as mesh (absence of contour lines-no charge). A cartoon representation of hβ$_2$m is shown in the same pose underneath. The DE loop is coloured in cyan, the BC loop in yellow, the FG loop in orange and the N-terminus in pink. (B) As in (A) but for mβ$_2$m. Each column shows the same pose of hβ$_2$m and mβ$_2$m to allow comparison.

Even though hβ$_2$m, ΔN6 and mβ$_2$m all show remarkably different amyloidogenicity *in vitro,* they are predicted to have similar aggregation propensity as predicted by theoretical algorithms. Despite the wealth of studies on β$_2$m aggregation, *in vitro* and *in vivo,* we are still far from understanding the principles that govern its 'native-like' aggregation and even more predict the relationship between protein dynamics, biomolecular recognition and amyloid assembly.

1.7 Aims and Scope

B$_2$m in its different forms as presented in Figure 1.23, represents a remarkable opportunity to interrogate the origins of amyloidogenicity. The finding that excursions from the native fold lead to the formation of native-like, aggregation-prone intermediate species provides valuable clues on the mechanism of fibril formation by β$_2$m, but at the same time raises a plethora of important questions. Why is mβ$_2$m not prone to aggregation? Does it fold through an I_T state? Are folding intermediates crucial for the aggregation of any β$_2$m sequence or is this phenomenon only related to hβ$_2$m (interestingly mβ$_2$m also has a cis-peptidyl bond at position 32)? How do protein dynamics relate to the ability to populate such intermediates at physiological conditions? What are the steps that follow the formation of amyloidogenic species? Perhaps one of the most intriguing findings presented in Section 1.6.5.3 is that the biomolecular collision between proteins that are remarkably similar both in terms of sequence and in terms of structure, results in dramatically different impacts on amyloid formation. What is the mechanism of promotion (hβ$_2$m-ΔN6), inhibition (mβ$_2$m-ΔN6) and nucleation (ΔN6-ΔN6) of fibril assembly? Are the surfaces involved in each case different and what are the properties of the interfaces? What is the link between biomolecular recognition and protein dynamics in the early stages of fibril assembly? Finally, what is the effect of binding on the interacting partners?

The aim of this thesis is to answer the aforementioned questions and eventually discover the causative link behind '**Biomolecular Recognition and Protein Dynamics in Amyloid Aggregation'**. To that direction, I use an *in vitro* approach to investigate the clinically relevant system of β_2m, under close to physiological conditions, using (mainly) NMR in combination with a variety of other biophysical and computational techniques.

Chapter I: Introduction

The introductory chapter provides the general framework of the thesis by highlighting the importance of protein-protein interactions inside the cell, and presenting the key determinants and the scientific nomenclature of protein association. The mysteries and current knowledge about the world of amyloid are then discussed, to move on to a detailed update on the current knowledge about the amyloidogenic protein β_2m.

Chapter II: Theory, Materials and Methods

This chapter starts by discussing the theoretical background of the main NMR methods used, to carry on with a detailed presentation of all experimental protocols used in this thesis.

Chapter III: Investigating the link between structure/dynamics and amyloidogenicity; In this chapter NMR methods were used to study the structure and dynamics of mβ_2m and investigate the details of its folding pathway in solution. The results revealed clues on the reduced aggregation propensity of this molecule.

Chapter IV: Protein-protein interactions define the course of amyloid assembly

This chapter investigates the nature of the pH-dependent aggregation of ΔN6 and the effect of the mβ_2m-ΔN6 interaction on the course of fibril assembly using a variety of biophysical tools. The results unravelled mechanistic details about the efficient inhibition of amyloid formation by protein-protein interactions.

Chapter V: Transient macromolecular association during amyloid assembly in atomic details; This chapter explores the properties of the intermolecular interactions that lead to inhibition, promotion and nucleation of fibril assembly, using NMR methods suitable to detect 'invisible' protein states. The development of a sophisticated docking protocol allowed the visualisation of transient macromolecular binding and the mapping of the surfaces involved in each case in atomic detail. The analysis revealed striking differences in the effect of protein association on the interacting partners and as an outcome a novel model explaining inhibition, promotion and nucleation of fibril assembly is proposed.

Chapter VI: Summary & conclusions

The final chapter summarises the findings and puts the new insights obtained in this thesis about the aggregation of β_2m in the broad spectrum of amyloid. Furthermore, the link between biomolecular recognition and protein dynamics is discussed in a more general context using examples from the literature.

CHAPTER II

THEORY, MATERIALS & METHODS

2.1 Theory of NMR methods used

A brief introduction into the theory of the main NMR experiments used in this thesis will be given here. For detailed description of NMR phenomena refer to [367-369].

2.1.1 NMR relaxation

All nuclei possess a spin angular momentum, which in the case of NMR visible nuclei can take two values, +1/2 or -1/2, corresponding to states α and β. In the absence of an external magnetic field the two states are randomly populated and thus no net magnetization is generated. However, upon introduction of an external magnetic field B_0, the populations of the two states are redistributed following Boltzman's distribution, so as the energetically more favourable state α is slightly more populated, giving rise to a net magnetisation vector along the +z axis. Therefore, an equilibrium state is created which can be disturbed by the application of radio-frequency pulses. The process of re-establishing the equilibrium state is called relaxation, and is heavily related to global or local protein motions, yielding information about protein dynamics in a range of timescales relevant to protein function[164]. Notably, the magnetisation along the z axis is called longitudinal magnetisation, whereas a 90^o pulse rotates the magnetisation vector around the x axis, placing it on the x-y plane and creating the transverse magnetisation.

There are two main mechanisms of relaxation. First, as the molecule tumbles in solution, each individual spin experiences a fluctuating local magnetic field due to the motions of nearby nuclei, yielding time-dependent dipolar interactions which is the source of the dipolar relaxation mechanism. Additionaly, the fluctuations in the local magnetic field arising from the electronic enviroment, a phenomenon which is the origin of the chemical shift, also contribute to

relaxation in a mechanism known as chemical shift anisotropy (CSA)[368]. NMR relaxation occurs by three simultaneous processes. Recovery of the net magnetisation on the +z axis (T_1 – longitudinal relaxation), loss of coherence in the x-y plane (T_2 – transverse relaxation) or cross-relaxation to nearby nuclei. In a 1D-T_1 experiment a 180° pulse transfers the magnetisation to the –Z axis and the spins are left to relax for a period t. A 90° pulse is then applied, bringing the magnetisation back on the X-Y plane for detection (Figure 2.1).

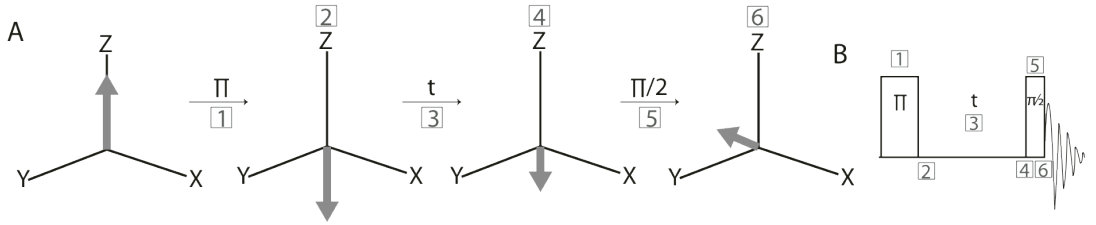

Figure 2.1: Inversion-recovery experiment for the measurement of T_1 relaxation. An inversion pulse (1) brings the net magnetisation to the –z axis (2). Relaxation occurs for a period t (3) and a 90° pulse rotates the remaining magnetisation vector around the y axis to end up in –x and the signal is then acquired. Signal intensity will decrease with increasing recovery periods (t) resulting in a mono-exponential curve.

To measure the transverse relaxation, a 90° pulse brings the equilibrium net magnetisation in the x-y plane and the spins are allowed to precess according to their chemical shift for time t. A refocusing 180° pulse inverts the magnetisation, which after another delay of time t, will end up either on the x, or –x axis. At this point coherence is re-established in the x-y plane and the length of the net magnetisation vector depends on the delay t. In a 2D experiment the intensity of each peak will decay exponentially with increasing relaxation delay t, yielding the R_2 relaxation rate ($R_2=1/T_2$) (Figure 2.2).

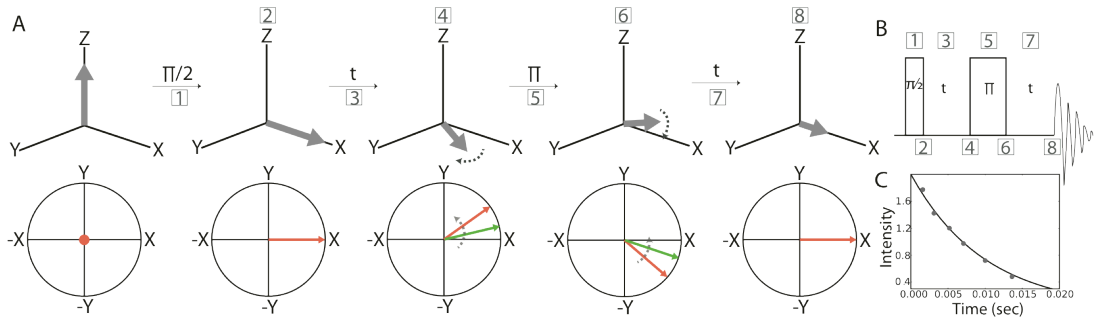

Figure 2.2: Measuring the relaxation in the transverse plane.
(A) A diagram of the net magnetization vector is shown on top and a top view of the x-y plane in the bottom. Only two spins are shown for simplicity. The pulse program used is shown in (B). A 90° pulse (1) brings the net

magnetization in the x-y plane (2) and precession is allowed for time t (3). During t spins precess with different frequencies based on their chemical shift and thus coherence is lost in the x-y plane. A 180° pulse is then applied (6) and the magnetisation is reverted. After exactly time t the spins will be refocused on the x axis, re-establishing coherence and the signal is recorded immediately. (C) A simulated T_2 relaxation curve of peak intensity versus relaxation time t is drawn using $R_2=200s^{-1}$.

The final relaxation mechanism occurs when two nuclei are close in space (<5Å). In this case they can exchange magnetization between each other because of dipolar interactions through space leading to the nOe. This cross-relaxation mechanism is proportional to the inverse of the sixth root of the distance between the two nuclei.

2.1.2 Paramagnetic NMR

In the presence of a paramagnet, nuclei show a different magnetic response owing to various molecular phenomena such as: dipole-dipole interactions, Curie-spin relaxation and pseudocontact shifts (reviewed in[210]). The paramagnetic effect can be detected by an increased relaxation rate of nearby nuclei (PRE), pseudocontact shifts (PCS) and residual dipolar couplings (RDC) (Figure 2.3). PRE can be measured in any paramagnetic system while PCS and RDC require an anisotropic g-tensor, meaning that it is necessary to assign the paramagnetic and diamagnetic states separately, a task that can be challenging [370,371]. Only the PRE phenomenon is considered in this thesis.

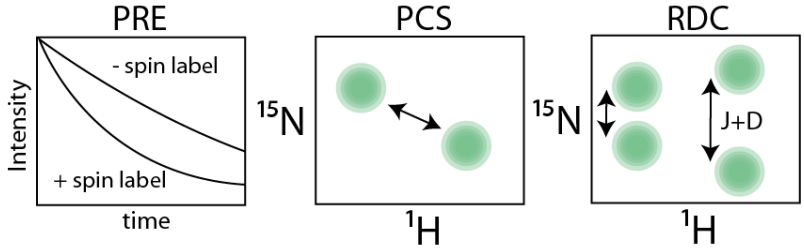

Figure 2.3: Different ways to observe paramagnetic NMR phenomena. Increased relaxation rates (PRE), pseudocontact shifts (PCS) and residual dipolar couplings (RDC). RDCs arise from the partial alignment of paramagnetically tagged molecules in the magnetic field, a phenomenon that rescues a portion of the through-space dipolar coupling, normally not observed because of the random orientations of magnetic vectors in an isotropic solution. Adapted from[210].

PRE experiments usually involve direct or indirect measurement of the relaxation times of H^N nuclei. ^1H nuclei have advantages in paramagnetic experiments with the main one being that the PRE for protons is much larger in comparison to carbon or nitrogen nuclei[209]. The effect of the PRE in the longitudinal and transverse relaxation rates is described by the Solomon – Bloembergen equations (SB)[195]:

$$\Gamma_1 = \frac{2}{5}\left\{\frac{\mu_0}{4\pi}\right\}^2 \gamma_1^2 g^2 \mu_B^2 S(S+1) J_{SB}(\omega_I) \qquad \text{equation (9)}$$

$$\Gamma_2 = \frac{1}{15}\left\{\frac{\mu_0}{4\pi}\right\}^2 \gamma_1^2 g^2 \mu_B^2 S(S+1)\left[4J_{SB}(0) + 3J_{SB}(\omega_I)\right] \qquad \text{equation (10)}$$

where g is the electron g-factor, γ_1 is the proton gyromagnetic ratio ω_I is the angular frequency of proton, μ_0 is the permeability of vacuum, S the electron spin quantum number and $J_{SB}(\omega)$ is the generalised spectral density function for the correlation function[1]:

$$J_{SB}(\omega) = r^{-6}\frac{\tau_c}{1+(\omega\tau_c)^2} \qquad \text{equation (11)}$$

where r is the distance between the paramagnet and the magnetic nucleus and τ_c the total correlation time, defined as $\tau_c = \tau_r^{-1} + \tau_s^{-1}$, where τ_r is the correlation time of the macromolecule and τ_s is the effective electron relaxation time. In a generalised form, equations 11, 12 can be expressed as a function of τ_c:

$$\Gamma_m = r^{-6} f_{SB,m}(\tau_c), \text{ m=1 or 2} \qquad \text{equation (12)}$$

[1] The correlation function is used to describe the time-dependence of the random molecular motions and is defined as the ensemble average of the product of the local magnetic fields at time t and the same fields after time τ:

$$C(t) = < B(t)_{loc} B(t+\tau)_{loc} >$$

Fourier-transform of the correlation function gives the spectral density function, which describes the probability of finding motion in a particular frequency.

Solomon – Bloembergen (SB) equations make two approximations. First, they assume that the electron relaxation time is unpaired to macromolecular correlation time, an assumption which is valid for nitroxide spin labels (such as MTSL) but not for metal ions such as Mn^{2+} [209,210]. Second, the dipole-dipole interactions involved are assumed to be rigid, an assumption that seems reasonable since the PRE vectors are long (up to 35Å) and thus less vulnerable to small fluctuations at either end of the vector. However, a paramagnetic tag with a flexible side-chain such as MTSL can sample a large conformational space, making this assumption not valid and introduces the need for an extension of the SB equations to include a formula that describes the internal motions of the PRE vector. Recently, Clore and coworkers introduced a novel framework that provides a model-free formalism for the incorporation of the internal motions of the PRE vector into the Solomon-Bloembergen equations[209,210], similar to the model-free approach used to interpret NMR relaxation data[372]. The correlation function for internal motions ($C_{l(t)}$) of the interaction vector can be written as:

$$C_I(t) = S^2 + (1 - S^2)\exp(-t/\tau_i) \qquad \text{equation (13)}$$

where S^2 is the order parameter and τ_i is the correlation time for the internal motions. Incorporation of eq. 12 into eq. 13 gives the model-free formalism of the PRE interaction:

$$\Gamma_m = S^2\left\langle r^{-6}\right\rangle f_{SB,m}(\tau_c) + (1 + S^2)\left\langle r^{-6}\right\rangle f_{SB,m}(\tau_t) \quad \text{equation (14)}$$

where τ_t is the total correlation time defined as $\tau_t = (\tau_r^{-1} + \tau_s^{-1} + \tau_i^{-1})$.

Eq. 14 offers two significant advantages over the conventional approach, namely, the ensemble dependence of the r^{-6} factor and the incorporation of the motional effects of the PRE vector in the form of τ_i and S^2. When dealing with flexible paramagnetic groups, that can sample a large conformational space, the model-free analysis, even though more computationally expensive, offers a much better representation of the PRE interaction. The computational strategy used to back-calculate the PRE from structure will be described in Section 2.3.11.

2.2 Materials

2.2.1 Chemicals

All chemicals were purchased by Sigma-Aldrich Co. (Poole, Dorset, UK) and were of analytical grade unless otherwise stated. *S*-(2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl-oxyl)methyl methanesulfonothioate (MTSL) was purchased by Toronto Chemicals (Toronto, Ontario, Canada), ^{15}N-labelled ammonium chloride and ^{13}C-labelled glucose for NMR studies were purchased from Cambridge Isotope Laboratories (Massachusetts, USA).

2.2.2 Bacterial strains used

The strains of *Escherichia coli* whose genotype is detailed below were employed during this project:

E.coli BL21 (DE3) pLysS- : E. coli B F–dcm ompT hsdS(rB– mB–) gal λ(DE3)[pLysS Camr]

E.coli DH5a: F- φ80lacZΔM15 Δ(lacZYA-argF)U169 deoR recA1 endA1 hsdR17(rk-, mk+) phoA supE44 thi-1 gyrA96 relA1 λ-.

2.2.3 Vectors

The pINK vector was used throughout the study[339]. The plasmid expressing mβ_2m was kindly provided by Prof. David Eisenberg. The human β_2m is inserted in a pET23a vector using the NdeI and HindIII restriction sites, under the control of a T7 promoter.

2.2.4 Oligonucleotides

The oligonucletides used in this study are shown in Table 2.1:

Construct	Primer sequence
S20C-ΔN6	Forward: 5' AGAGAATGGAAAGTGCAATTTCCTGAATTG 3’ Reverse: 5' CAATTCAGGAAATTGCACTTTCCATTCTCT 3'
S33C-ΔN6	Forward: 5' CTGGGTTTCATCCATGCGACATTGAAGTTG 3' Reverse: 5' CAACTTCAATGTCGCATGGATGAAACCCAG 3’
D38C-ΔN6	Forward: 5' CCGACATTGAAGTTTGCTTACTGAAGAATGG 3’ Reverse: 5' CCATTCTTCAGTAAGCAAACTTCAATGTCGG 3'

S61C-ΔN6	Forward: 5' GCAAGGACTGGTGTTTCTATCTCTTG 3' Reverse: 5' CAAGAGATAGAAACACCAGTCCTTGC 3'
T73C-ΔN6	Forward: 5' CACTGAATTCACCCCCTGCGAAAAAGATGAGTATG 3' Reverse: 5' CATACTCATCTTTTTCGCAGGGGGTGAATTCAGTG 3'
I92C-ΔN6	Forward: 5' GTCACAGCCCAAGTGCGTTAAGTGGGATC 3' Reverse: 5' GATCCCACTTAACGCACTTGGGCTGTGAC 3'
F56E-mβ_2m	Forward: 5' GTCAGATATGTCCGAGAGCAAGGACTGGTC 3' Reverse: 5' GACCAGTCCTTGCTCTCGGACATATCTGAC 3'
W60E-mβ_2m	Forward: 5' GTCCTTCAGCAAGGACGAGTCTTTCTATATCCTGG 3' Reverse: 5' CCAGGATATAGAAAGACTCGTCCTTGCTGAAGGAC 3'

Table 2.1: Primers. Primers were designed (according to the Stratagene protocol[373]) to incorporate single point mutations into the sequence of ΔN6 or mβ_2m. Purchased from MWG Operon.

2.2.5 Bacterial media

Luria Bretani (LB) medium was prepared by dissolving 25g of granulated LB (Melford) in 1L of deionised water before autoclaving at 121°C for 20min. The HDMI medium was used for preparing labelled proteins for NMR studies (^{15}N, ^{13}C or more complicated carbon labelling –see Table 2.2) .

	LB	HDMI
Bacto-tryptone	10g	
Yeast extract	5g	
NaCl	10g	
$Na_2HPO_4^-$		7.5g
$K_2HPO_4^-$		10g
$K_2SO_4^-$		9g
$KH_2PO_4^-$		10g
NH_4Cl*		1g

Table 2.2: Bacterial Media. Reagents for the bacterial media used in this study. All salts are made to 1L with deionised water and autoclaved in 2L flasks (when using the HDMI medium only 500mL of media are included in each flask to achieve better aeration of bacteria). *Can be replaced by heavy isotopes

Just before culturing, the HDMI medium was supplemented with the following filter sterilised supplements (added to each litre): 1M $MgCl_2$ (2mL), 1M $CaCl_2$ (100μL), 20% (w/v) D-glucose-$^{12}C_6$ (20mL or 10mL of D-glucose-$^{13}C_6$ or carbon labelling).

2.3 Methods

2.3.1 General analytical methods

2.3.1.1 Estimation of protein concentration by UV absorption at 280nm

Dilutions of the original sample were prepared where necessary, to give meaningful absorbance values A_{280}, ideally below 1. Protein concentration was determined using the Beer-Lambert's law using extinction coeffients of 20065$M^{-1}cm^{-1}$ for hβ_2m, ΔN6 and cysteine mutants of ΔN6, 21342$M^{-1}cm^{-1}$ for mβ_2m and 12950$M^{-1}cm^{-1}$ for F56E/W60E mβ_2m, determined by Gill and Von Hipper's law[374]:

$$E(Prot) = Number(Tyr)*Ext(Tyr) + Number(Trp)*Ext(Trp) + Number(Cystine)*Ext(Cystine)$$

where (for proteins in water measured at 280nm): Ext(Tyr) = 1490$M^{-1}cm^{-1}$, Ext(Trp) = 5500$M^{-1}cm^{-1}$, Ext(Cystine) = 125$M^{-1}cm^{-1}$.

2.3.1.2 Quantification of DNA

A diluted DNA sample (typically 1/200 in sterile water) was used to measure UV absorption at 260nm given that A_{260}=1.0 equates to 50μg/mL double stranded DNA, 40μg/mL RNA or single stranded DNA (all calculations were performed automatically by the appropriate application on a Ultrospec 1200 spectophotometer).

2.3.1.3 Preparation of agar plates

2.5g of granulated LB medium were dissolved in 100mL of deionised water and then 1.5g of bacto-agar (Melford) was added. The solution was autoclaved at 121^{o}C for 20 min and the appropriate antibiotics were added when the temperature dropped below ~50^{o}C. Approximately 25mL of the molten agar mixture was then poured into 100mm diameter petri dishes and was allowed to set at room temperature. Plates were stored at 4^{o}C for less than 15 days.

2.3.1.4 Agarose gel electrophoresis

Agarose gels were routinely used to monitor results of polymerase chain reaction or the success of plasmid purification. 3% (*w/v)* agarose gels were

prepared by dissolving 4.5g of agarose (Melford) into 150mL of TAE (40mM TrisHCl, 1mM EDTA) buffer. Gels were stained by the addition of 15µL of SIBR safe stain (Invitrogen). Samples are made by 1:5 dilution into the loading dye, loaded onto the gel which is then run at 100V using 1X TAE buffer as the running phase. DNA was then imaged by trans-illumination.

2.3.1.5 Sodium dodecyl-sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

Gel electrophoresis was routinely used to monitor over-expression and purification of recombinant proteins. A two layered-gel system consisting of a stacking gel (to load the samples) and a resolving gel (to separate the proteins) was used. Tris-tricine buffered SDS-PAGE gels were used throughout this study (see Table 2.3). Gels were stained in Instant Blue (Expedeon).

Solution	Resolving gel(mL)	Stacking gel(mL)
30% (*w/v*) Acrylamide: (0.8% (*w/v*) bis-acrylamide	7.5	0.83
3M Tris.HCl, 0.3% (w/v) SDS pH 8.45	5.0	1.55
H_2O	0.44	3.72
Glycerol	2.0	-
10% Ammonium persulphate	0.1	0.1
Tetramethylethylenediamine (TEMED)	0.01	0.01

Table 2.3: SDS-PAGE gels. Recipe for Tris-tricine buffered SDS-PAGE gels (volumes shown are sufficient to cast two 8cm X 10cm mini gels using a 1.5mm spacer).

2.3.2 Molecular Biology

2.3.2.1 Site-directed mutagenesis

In vitro site-directed mutagenesis was used to insert point mutations into the plasmid encoding ΔN6 or mβ_2m. Mutagenesis was performed using the QuikChange thermal cycling method (Stratagene) and the primers listed in Table 2.1. The presence of the correct mutation and the absence of unwanted additional mutations were examined by DNA sequencing, performed by GATC Biotech.

2.3.2.2 Digestion of methylated plasmid DNA

In order to destroy the parental methylated plasmid DNA, PCR products were mixed with 1µL of *Dpn*I endonuclease (New England Biolabs) and the appropriate amount of NEB 4 buffer (New England Biolabs). The reaction mixtures were incubated for 3h or overnight at 37°C.

2.3.2.3 Transformation of plasmid DNA

Competent cells were prepared by re-suspending the cell pellet of an overnight LB culture in 100mM $CaCl_2$. Cells were incubated for 10min harvested by centrifugation at 400rpm and re-suspended in 100mM $CaCl_2$ containing 30% (*v/v*) glycerol. A 50µL aliquot of previously prepared calcium-competent DH5a or BL21 (DE3) pLysS- *E. coli* cells was mixed with 2µL of plasmid DNA. Cells were incubated for 30min at 4°C, followed by a 45s heat shock at 42°C and a further incubation at 4°C for 5min. 100µL of LB medium was added to the transformation mixture, and the transformation mixture was incubated whilst shaking at 200rpm for 1h at 37°C before being plated on a pre-warmed LB-agar selective plate containing carbenicilin and/or chloramphenicol. The plate was incubated overnight at 37°C.

2.3.2.4 Isolation of plasmid DNA

10-20mL of liquid culture containing 100µg/mL carbenicilin, inoculated with cells containing the relevant plasmid were incubated at 37°C and 200rpm overnight in an orbital shaker. Cells were pelleted by centrifugation at 13,000*g* for 5min the next morning. Plasmid DNA was then isolated from the cells using a Plasmid Purification kit (Qiagen) following the manufacturer's instructions and stored at 20°C .

2.3.3 Protein purification

The pINK plasmid containing hβ_2m or ΔN6 gene was transformed in *E.coli* cells of the BL21 DE3 plysS- strain. Starter cultures were generated by inoculating 100mL of LB medium with cells containing the relevant gene and 50µg/mL carbenicilin and 50µg/mL chloramphenicol and incubate overnight at 37°C, 200rpm. 2L flasks containing 1L of LB medium or the appropriate medium for ^{15}N, ^{13}C or more complicated labeling were inoculated with 10mL of starter

culture. Cells were incubated at 37°C, 200rpm until they reach OD_{600} of ~0.6 and then the expression of β_2m was induced by the addition of Isopropyl β-D-1-thiogalactopyranoside (IPTG-final concentration of 1mM). Expression was allowed to continue overnight at 37°C and cells were harvested next morning using a Heroaus continual action centrifuge performing at 15000rpm. The cell pellet containing β_2m as inclusion bodies, was chemically lysed by the addition of 50-100mL of lysis buffer (100µg/mL lysozyme, 50µg/mL DNAse I, 50µg/mL phenylmethanesulfonyl fluoride (PMSF), 10mM TrisHCl pH 8.0). Further cell disruption was performed using a constant cell disrupter system (Constantsystems) at a high pressure of 20.0kpsi. Inclusion bodies were separated using centrifugation (15,000*g* using a Sorvall SS34 rotor) in a Beckman centrifuge for 40min, 4°C and the inclusion body pellet was washed with 10mM TrisHCl pH 8.0 buffer four times. Finally, β_2m was solubilised overnight in 10-20mM TrisHCl pH 8.0 containing 8M urea (MP biomedicals) and refolded by dialysis (3000MW cutoff) against 2-5L of the same buffer but lacking urea. The refolded protein was centrifuged 30min, 15000rpm (Sorvall SS34 rotor) to pellet insoluble material and the supernatant was loaded on a Q-Sepharose (GE Healcare) self-packed (200-300mL) column already equilibrated with 2 column volumes of 20mM TrisHCl pH 8.0 for anion exchange purification. Bound protein was eluted with a gradient of 0-400mM NaCl (in the same buffer) over 800mL and was freeze-dried after dialysis in dH_2O or concentrated using 3000MW cutoff centricons (Avanti LTD). Freeze-dried protein was re-suspended in 10mM sodium phosphate buffer pH 7.0, filtered through 0.2µM filters (Fisher Scientific) and gel-filtered using a HiLoad Superdex-75 Prep column (Amersham Biosciences), calibrated with a standard gel filtration calibration kit (GE Healthcare). The monomer peak was collected, concentrated, aliquoted and stored at -80°C or freeze-dryed. The purification protocol had to be optimized for the purification of mβ_2m or CYS-mutants of ΔN6. The modified protocols will be discussed in Section 5.2.1. A yield of ~70mg/L of LB culture was obtained typically and protein purity and identity were checked by SDS-PAGE and ESI-MS analysis respectively (the latter performed by Dr. James Ault, University of Leeds).

2.3.4 Fibril growth

Samples containing 0.6-60μM protein, 10mM sodium phosphate pH 6.2-8.2, 83.3-86.6 mM NaCl, 0.02% *(w/v)* sodium azide, 10μM ThT, were incubated at 37°C in 96 well plates (Corning Incorporated, Costar), using a plate reader (FLUOstar OPTIMA for a period of 7-28 days. *De novo* fibril formation was performed by (orbital) shaking at 600 rpm. Fibrils were pelleted by centrifugation at 14100g in a bench-top centrifuge for 15min and were visualised using transmission electron microscopy (TEM), Section 2.3.6 or atomic force microscopy (AFM), Section 2.3.7. ThT data were normalized between 0 and 1 (set as the average of the last 100 datapoints) and fitted to the generalized sigmoidal function (eq. 15) using in house scripts written in Python (www.python.org).

$$F(t) = A + \frac{K - A}{(1 + Qe^{-B(t-M)})^{1/v}} \qquad \text{equation (15)}$$

where A is the pre-transition baseline, K the plateau of the fluorescence, B the growth rate and M the time of maximal growth. Q and ν are parameters affecting the transition from and to the exponential growth phase respectively.

2.3.5 Electron Microscopy

TEM was routinely used to confirm the presence of fibrils and the morphology of fibrillar samples. Fibrils prepared as above were diluted 0-10 fold in 18MΩ H_2O and then placed on a glow-discharged colloidion-coated copper grids (provided by M. Fuller, University of Leeds) and incubated for 30s. Excess sample was removed by blotting with filter paper and fibrils were then stained with 4%(*w/v*) uranyl acetate for 30s. Excess uranyl acetate was removed by further blotting and images were collected on a CM10 electron microscope at 80keV.

2.3.6 Atomic force microscopy

Long straight hβ_2m fibrils were deposited onto freshly cleaved mica surfaces (Agar Scientific) after dilution (with filtered 18MΩ water) to 0.4μM monomer

equivalent concentration to ensure uniform surface coverage and dispersion of the sample. After dilution 20μL of the sample were immediately deposited onto the centre of the mica. Each sample was incubated for 5min before washing with 1mL of sterile water and drying under a gentle stream of N_2 gas. AFM imaging was performed using a Dimension 3100 Scanning Probe Microscope (Veeco Instruments) and PPP-NCLR silicon cantilever probes (Nanosensors, Neuchatel, Switzerland) with a force constant of 48N/m. Height and phase images were acquired and processed using the supplied software NanoScope 6.13r1 to remove sample tilt and scanner bow before analysis.

2.3.7 Analytical size exclusion chromatography

Samples for analytical size-exclusion chromatography were made by incubating monomers or a mixture of monomers (ΔN6, hβ_2m, mβ_2m) in 20mM sodium phosphate buffer pH 6.2, 7.2, or 8.2 containing 0.02% *(w/v)* sodium azide in different ratios at 25°C or 37°C for 2-3 days. Protein samples were centrifuged for 10min 13000g in a benchtop centrifuge and equilibrated at 4°C for ~10min before loading on an analytical Superdex-75 10/300 column. The running buffer used was 100mM TrisHCl pH 8.0 and all experiments were carried out using an AKTA explorer system at 4°C, unless otherwise stated.

2.3.8 Analytical ultracentrifugation (AUC)

Proteins were dissolved in 10mM sodium phosphate pH 6.2-8.2, 83.3-86.6mM NaCl to a final concentration of 60-120μM and then dialysed in the same buffer overnight. Sedimentation velocity experiments were performed by loading 410 μL of sample in aluminium centrepieces equipped with sapphire windows and inserted in a AN60 Ti four-cell rotor. Absorbance data were measured at a speed of 48000rpm using a wavelength adjusted based on protein concentration to give an absorbance reading <1Au. Data were analysed by fitting a continuous distribution model for the sedimentation coefficient in SEDFIT[375].

2.3.9 Labelling with MTSL

5mM DTT was added to 2-3mg of protein in 20mM sodium phosphate buffer pH 7.2 and the mixture was incubated at room temperature for 30min. DTT was removed by gel filtration using a NAP-10 column (GE Healthcare) and 10X molar excess over the total free thiol concentration of MTSL (Toronto chemicals, stored in acetonitrile) was added immediately. The sample was incubated at room temperature for 4-5h and the excess of MTSL was removed by buffer exchange in 10mM sodium phosphate buffer pH 6.2 using a PD-10 (GE Healthcare) desalting column. Labelling was confirmed by ESI-MS analysis performed by Dr.James Ault (University of Leeds). The protein was aliquoted and stored at -80^{o}C.

2.3.10 NMR spectroscopy

2.3.10.1 Backbone assignments

Assignments of the backbone atoms of mβ_2m were performed using samples of 500μM or 750μM uniformly labelled (^{15}N, ^{13}C) protein in 10mM sodium phosphate buffer pH 6.2, 83.3mM NaCl, 0.02% (*w/v*) NaN_3, 10% (*v/v*) D_2O. 3-dimensional NMR experiments were recorded at 25^{o}C using Varian Inova spectrometers (Agilent) operating at proton frequencies of 500MHz (HNCA, HNCO, CBCA(CO)NH, HN(CA)CO) and 750MHz (HNCACB), equipped with a room temperature or cryogenic probe, respectively. Experiments were carried out using 32 scans per increment in the indirect dimension, 2048 complex points and spectral windows adjusted to highest and lowest chemical shifts observed in each case.

In 3D-NMR experiments the magnetisation is transferred through bonds via the J-coupling to nearby nuclei. Because the J coupling can extend as far as 4 bonds, some of these experiments will give rise to peaks corresponding to the current spin system or residue (i) or the proceeding residue i-1. For instance in a HNCA experiment (Figure 2.4A), the magnetisation is transferred to ^{15}N starting from ^{1}H and then to Cα using the N-Cα coupling. Since the amide nitrogen is coupled to the Cα of the i residue (one bond) and to the Cα of the i-1 residue (two

bonds), two peaks with different intensities (with the one belonging to the i residue being the most prevalent) are visible in the final spectrum (Figure 2.4C). The chemical shift is evolved in ^{1}H, ^{15}N, ^{13}C dimensions, resulting in a 3D experiment, with the ^{13}C dimension being a projection of the ^{1}H-^{15}N plane seen in a classic HSQC spectrum. More complicated magnetisation pathways result in significant loss of sensitivity, a phenomenon which makes 3D experiments problematic especially in the case of larger proteins. The CBCA(CO)NH experiment for example (Figure 2.4B), starts from ^{13}C–attached protons (either Cα or Cβ) and the first pulse train transfers the magnetisation to ^{13}C (either Cα or Cβ) and the chemical shift is evolved simultaneously. From ^{13}C the magnetisation is then transferred to ^{13}CO and then to ^{15}N to end up at $^{1}H_N$. In this case each ^{15}N plane will include two peaks corresponding to the Cα and the Cβ of the i-1 residue (Figure 2.4C). Using a combination of the HNCA and CBCA(CO)NH spectra identification of the Cα-1 resonance can be achieved (Figure 2.4). All spectra were processed in nmrPiPE[376] using linear prediction in the nitrogen dimension and assignments were performed in CCPNMR analysis software[377].

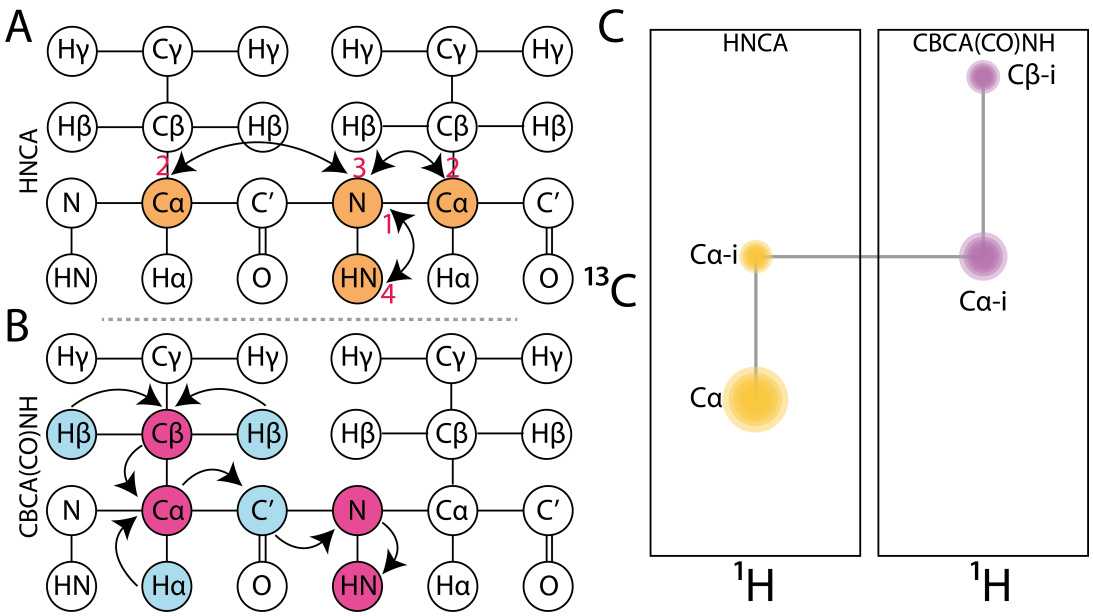

Figure 2.4: 3D-NMR spectra for assigning backbone resonances.
Magnetisation pathways for the HNCA (A) and CBCA(CO)NH (B) experiments (see Section 2.3.10.1 for more details). (C) Simulated HNCA (left) and CBCA(CO)NH (right) spectra.

Taking advantage of the observation that chemical shifts are sensitive reporters of secondary structure in proteins, HN, CA, CB, HA, CO, N chemical shifts were

used to predict the φ, ψ angles (and therefore the secondary structure elements) of mβ$_2$m using the TALOS+ software[378]. Briefly, TALOS+ uses a database of proteins with known structure and chemical shifts to find the best 10 matches against the input chemical shifts. The average φ, ψ torsion angles are then reported as the predicted values for the protein in question. Additionally, using a similar approach prediction of the order parameters (S^2)- a measure of protein motions in the ns-μs timescale can be made[378,379].

2.3.10.2 H/D exchange NMR

In a H/D exchange experiment, there are numerous parameters that affect the rate of exchange such as temperature, pH and amino acid sequence. Hydrogen atoms in native proteins are protected from exchange with deuterons by the formation of hydrogen bonds, implying that an 'opening' event has to occur prior to H/D exchange. Scheme 2 summarises the H/D exchange reaction in a native protein[380,381]:

$$NH_{closed} \underset{k_{cl}}{\overset{k_{op}}{\leftrightarrows}} NH_{open} \overset{k_{ch}}{\leftrightarrows} ND \qquad \text{scheme (2)}$$

where k_{op} is the rate of protein opening (unfolding), k_{cl} is the rate of protein refolding and k_{ch} is the intrinsic rate of H/D from the open state. As it is evident from scheme 2 the observed rate of exchange (k_{obs}) upon exposure of a protonated protein in a deuterated buffer is given by:

$$k_{obs} = \frac{k_{op}k_{ch}}{k_{op} + k_{cl} + k_{ch}} \qquad \text{equation (16)}$$

Assuming that for a natively folded protein the rate of opening is much slower than the rate of closing, eq. 16 can be simplified as follows:

$$k_{obs} = \frac{k_{op}k_{ch}}{k_{cl} + k_{ch}} \qquad \text{equation (17)}$$

Based on eq. 17 different exchange mechanisms emerge depending on the relationship between k_{cl} and k_{ch}. When the rate of closing is much faster than the intrinsic rate of H/D exchange ($k_{cl} >> k_{ch}$) eq. 17 translates into:

$$k_{obs} = \frac{k_{op}k_{ch}}{k_{cl}} = K_{op}k_{ch} \qquad \text{equation (18)}$$

This mechanism of H/D exchange is known as EX2 exchange and usually is the outcome of small, non-cooperative fluctuations in native fold. Importantly the determination of the equilibrium position K_{op} allows the calculation of the free energy of unfolding. The opposite scenario where $k_{cl} >> k_{ch}$ allows the removal of k_{cl} from eq. 17, resulting in EX1 exchange:

$$k_{obs} = k_{op} \qquad \text{equation (19)}$$

Thus, EX1 exchange reports on the kinetics of the unfolding process, which is the outcome of large scale motions or due to the protein visiting the globally unfolded state.

Samples for H/D exchange experiments were prepared in 10mM sodium phosphate buffer pH 6.2 and then freeze dried. Freeze-dried protein was dissolved in 10mM sodium phosphate buffer, apparent pH 6.2 prepared in 100% (*v/v*) D_20 buffer containing 83.3mM NaCl, placed into the NMR tube after manual mixing and the loss of intensity of amide proton was monitored by SOFAST ^{1}H-^{15}N HSQC spectra[382] at 25°C (mβ_2m) or 37°C (hβ_2m). The dead time of the experiment was 3-10min while the experiment time varied between 10-15min. The loss of signal was fitted to single exponentials (two parameters) using in house software written in Python (www.python.org). Error calculation was performed based on the noise of the experiment.

2.3.10.3 T_1, T_2 and ^{1}H-^{15}N nOe relaxation

For ^{15}N relaxation experiments, a concentrated stock of protein (ΔN6 or mβ_2m) was buffer exchanged to 10mM sodium phosphate pH 8.0 or 10mM sodium phosphate pH 6.2, using 3000MW cutoff centricon centrifugal filters (Amicon), snap frozen in liquid N_2 and stored in -80°C. On the day of the experiment,

protein was diluted with the same buffer but containing 10% (*v/v*) D_2O, 0.02% (*w/v*) NaN_3, 86.6mM or 83.3mM NaCl to a final protein concentration of 750 or 80μM and placed in the NMR tube. All relaxation experiments were performed at 25°C, using a 600MHz Varian Inova spectrometer equipped with a room temperature probe. The temperature of the instrument was calibrated beforehand, using a sample of 99.5% (*v/v*) methanol to eliminate small fluctuations in the probe temperature that could be crucial in a relaxation measurement. T_1 experiments were performed using a series of 10 experiments with 0.16, 0.32, 0.48, 0.64, 0.80, 0.96, 1.12, 1.44 s relaxation delays, while 18.8, 37.7, 56. 6, 75.5, 94.4, 113.2, 132,1, 151.0 ms were the delays used in the T_2 experiments. Error estimation was performed using duplicates and a ^{1}H-^{15}N HSQC spectrum was recorded before and after acquisition of any relaxation experiment to ensure sample quality. All pulse sequences used contained a heat compensation element to avoid radiation-induced heating of the sample during the experiment and a long inter-scan delay of 2.2μs was used to ensure full re-equilibration before each scan. For ^{1}H-^{15}N nOe cross-relaxation experiments saturation of amide protons was achieved with a 120° pulse for 3.5 s prior to the experiment.

2.3.10.4 Relaxation dispersion CPMG NMR experiments

All CPMG experiments were performed using 0.6mM samples in 10mM phosphate buffer including 83.3-86mM NaCl at pH 6.2-8.2. Backbone ^{15}N transverse relaxation dispersion experiments were carried out as described in[179] utilising a Varian Inova spectrometer performing at 500MHz using a fixed relaxation delay of 48ms and interpulse delays that varied from 0.375 to 12ms. Spectra were acquired in an interleaved manner with 90 increments and 128 scans per incremental delay to ensure high signal to noise ratio. Resulting spectra were processed using NMRPipe[376] and peak intensities were measured in NMRview[383]. $R_{2,effective}$ values were calculated using eq. 20 and were fitted to the Bloch-McConnell[178] equations using the software NESSY[384].

$$R_{2,eff} = \frac{I_x / I_0}{\tau_{cpmg}} \qquad \text{equation (20)}$$

where I_x is the peak intensity in each experiment, I_0 is the peak intensity in the reference spectrum (with no CPMG train applied) and τ_{CPMG} is the fixed relaxation delay of 48ms.

2.3.10.5 PRE experiments

60μM MTSL-labelled ΔN6 was mixed with 60μM ^{15}N-labelled hβ_2m, mβ_2m, or ΔN6 in 10mM sodium phosphate buffer pH 6.2, 10% (*v/v*) D_2O, 0.02% (*w/v*) NaN_3. H_N-PRE data were measured as the difference of the proton transverse relaxation (R_2) rates between oxidised and reduced (by addition of 3mM ascorbic acid) MTSL-ΔN6. Data were recorded at 25°C using a ^{1}H-^{15}N correlation based pulse sequence with 6 or 5 time-points (0.0016-0.016ms), utilising a Varian-Inova 750MHz spectrometer equipped with a cryogenic probe. Fitting of exponential functions to the intensity data was carried out using in house software with 1000 Monte Carlo steps for error estimation. The error was estimated from the noise of the experiment.

2.3.10.6 Real-time refolding

The increased energy required for proline isomerisation provides a suitable timeframe for monitoring the refolding process by NMR in real-time. In the folding mechanism of hβ_2m proposed by Jahn et al. folding of the I_T to the native state takes places in minutes allowing the direct observation of the inter-conversion process. Protein samples were made in 10mM sodium phosphate buffer pH6.2 and freeze-dried. Unfolding was initiated by dissolving the freeze-dried protein in 60μL of the same buffer containing 8M urea at 37°C for 1h, and the protein was then refolded by rapid 10 fold dilution in 25mM sodium phosphate pH 6.2, 10% (*v/v*) D_2O, 0.02% (*w/v*) NaN_3. The refolding from I_T to the native state was monitored by a series of SOFAST ^{1}H-^{15}N HSQC spectra[382,385] at 25°C, with 80 increments in the indirect dimension, 2 scans per increment and 512 complex points, resulting in a total acquisition time of 45 s. Refolding data were analysed by grouping the observed resonances into two categories: peaks of the pure I_T state that do not show significant overlap with peaks of the native state and peaks belonging to the native state. Resonances that showed significant peak overlap between the two protein states were excluded from the analysis. The appearance of the native and the loss of the I_T

signal were fitted to a saturation recovery or exponential decay function respectively for each peak individually and after averaging of the intensities in each group.

2.3.10.7 NMR chemical shift perturbations for affinity calculations

To monitor the interaction between different proteins, NMR titration experiments were carried out using ^{1}H-^{15}N HSQC at 25°C. To calculate the affinity of the ΔN6-mβ$_2$m interaction a 80μM ^{15}N-mβ$_2$m sample was titrated with 0, 40, 80, 160, 320 μM ^{14}N-ΔN6. Similarly, a 80μM ^{15}N-hβ$_2$m sample was titrated with 0, 40, 80, 160, 320, 480 μM ^{14}N-ΔN6. The total chemical shift perturbation (CSP) was calculated using the formula:

$$\Delta\delta\omega = \sqrt{(5 * \delta^{1}H)^{2} + (\delta^{15}N)^{2}} \qquad \text{equation (21)}$$

where $\delta^{1}H$ and $\delta^{15}N$ are the observed changes in the proton and nitrogen chemical shifts respectively. Residues showing a CSP larger than two standard deviation from the average were considered significant and CSP data were fitted to a simple binding hyperbola to extract K_d values:

$$\Delta\delta\omega = \Delta\delta\omega_{\max} \frac{[L_T] + [U_T] + K_d - \sqrt{([L_T] + [U_T] + K_d)^2 - 4[L_T][U_T]}}{2[L_T]}$$

equation (22)

where $[U_T]$, $[L_T]$ is the concentration of the ^{14}N-labelled and ^{15}N-labelled protein used respectively and $\Delta\delta\omega_{max}$ the largest CSP observed for the residue in question. K_ds were extracted in a Monte Carlo analysis with 1000 steps performed using in house scripts in Python (www.python.org). The error was estimated as the standard deviation of residues that do not show significant CSP.

2.3.11 Simulated annealing calculations

Following Section 2.1.2 that set out the theoretical framework for the interpretation of PRE, a computational protocol was set up in order to accurately back-calculate PRE rates from 3D structures. XPLOR-NIH[366]

contains a potential energy term for the PRE (PREpot) which describes the minimisation of the observed (Γ_2^{obs}) versus the calculated (Γ_2^{cal}) rate, weighted by a weight term w[209]:

$$E_{PRE} = k_{PRE} \sum_i w_i \left\{ \Gamma_2^{obs}(i) - \Gamma_2^{cal}(i) \right\}^2 \qquad \text{equation (23)}$$

where k_{PRE} is the force constant. The usual approach when weighting experimental restraints is to use the error of the measurement to weight the contribution of each individual restraint to the total energy of the potential term. However the PRE is a specific case in that a residue showing a high PRE rate will also have a higher absolute error associated with it, in comparison with another residue with no significant PRE. This is not difficult to imagine since the higher PRE rate will result in faster loss of the signal into the noise, making fitting erroneous. Thus, weighting the observed PRE rates according to their associated error will result in severely underweighting large Γ_2^{obs} rates, which contain all the useful information. Therefore, and as proposed by Iwahara et al.[209], the weight term w in order to overcome this issue takes the form of:

$$w_i = \frac{1}{\delta} \frac{\Gamma_2^{obs}(i)}{\Gamma_2^{obs,\max}(i)} \qquad \text{equation (24)}$$

where $\Gamma_2^{obs,max}$ is the highest PRE value in the dataset and δ_i the error of the measurement.

To account for the large conformational space sampled by the paramagnetic group, the flexible side chain of MTSL was represented by multiple conformers N=5 and the energy of the PRE potential term was ensemble averaged over all of them simultaneously. Since the multiple conformations of the MTSL moiety are not physically relevant, atomic overlap between these atoms was allowed[209,211,386].

The computational strategy employed included two PRE potential terms (arising from S61C-ΔN6 and S33C-ΔN6) and classic geometry restraints to

restrict deviation from bond lengths, angles and dihedrals. Resonances which were completely absent from the oxidized spectrum due to the residue being close to the spin label were incorporated in the protocol as nOe-type of restraints with an upper bound of 11.5Å and a lower bound of 9Å[197]. Additionally, chemical shift perturbations observed upon binding were incorporated as sparse, highly ambiguous intermolecular distance restraints as described in[387]. For residues showing significant chemical shift perturbations from protein A all hydrogen, nitrogen and oxygen atoms were set to interact with all hydrogen, nitrogen and oxygen atoms of residues showing altered chemical shifts from protein B. As chemical shifts can be influenced by numerous factors upon protein-protein interaction, the treatment of the derived data undertaken here, results in a loose potential term that is unlikely to bias the structure calculation. Finally, the protocol also included a weak radius of gyration restraint (R_{gyr}) calculated as $2.2N^{0.38}$, where N is the number of atoms in the complex. R_{gyr} is required in order to prevent bias towards more extended structures and tends to underestimate the true value of the radius of gyration[388].

The aforementioned potential terms were used in a rigid-body energy minimisaton/ simulated annealing in torsion angle space protocol to minimize the difference between the observed and calculated Γ_2 rates, starting from random orientations. The first step in the structure calculation consisted of 5000 steps of energy minimisation against only the sparse chemical shift restraints, followed by simulated annealing dynamics with all the potential terms active where the temperature is slowly decreased (3000-25K) over 4fs. During the hot phase (T=3000K) the PRE and nOe terms were underweighted to allow the proteins to sample a large conformational space and they were geometrically increased during the cooling phase. Proteins were treated as rigid bodies until the initiation of the cooling phase, where side chains were allowed to float (semi-rigid body calculation). The final step included torsion angle minimisation using all potential terms. The agreement between the experimental and the back-calculated PRE data is described by the PRE Q factor defined as:

$$\left[\sum i(\Gamma_{2,i}^{obs} - \Gamma_{2,i}^{cal})^2 / \sum i(\Gamma_{2,i}^{obs})^2\right]^{1/2} \qquad \text{equation (25)}$$

Ensemble calculations where the interacting species are represented as multiple states were carried out as before but in this case the Γ_2^{cal} is calculated as the average value between the conformers. The length of the dynamics was adjusted to the ensemble size (e.g for double the number of conformers dynamics need to be run for twice as long) to ensure convergence. The population of each conformer was set by specifying its weight in the calculation. For ensemble calculations the average Q factor was calculated either by averaging the individuals Q factors of the best scoring ensembles (Qe) or by calculating the Q factor of an average predicted dataset (the Q factor of the ensemble of ensembles – Qee).

The calculation of the mβ_2m homodimer described in Section 3.2.3 was performed using the same protocol but without the use of the PRE restraints. Only sparse chemical shifts and geometry restraints were used in the simulated annealing run.

CHAPTER III

INVESTIGATING THE LINK BETWEEN STRUCTURE/DYNAMICS AND AMYLOIDOGENICITY

3.1 Introduction

Over the last 15 years, various molecular phenomena have been proposed as the triggering event for amyloid aggregation including conformational changes[271,272], global or partial unfolding[350,356], or the formation of the critical nucleus. Since the aforementioned processes are normally interlinked, it is often difficult to de-convolute the origin of enhanced or hampered aggregation propensity of a particular protein. Considering how these events affect the probability of a protein precursor to form amyloid fibrils is at the origin of our understanding of the mechanism of how proteins aggregate.

As discussed in Section 1.6.5.3, the amyloidogenicity of hβ_2m is directly related to the concentration of its folding intermediate I_T in solution[357]. However, I_T itself shows a remarkable pH-dependent amyloid potential, with increased amyloidogenicity at pH 6.2 in comparison to pH 8.2. This phenomenon is well depicted on ΔN6 which is shown to closely mimic I_T's properties[272]. Protein dynamics in the ms timescale were found to correlate very well with the ability to form amyloid fibrils for ΔN6 (increased dynamics – increased amyloidogenicity at pH 6.2) providing a link between structure and amyloidogenicity[272]. In this Chapter the origin of mβ_2m's reduced amyloidogenicity at physiological conditions is investigated. Towards that direction, a complete characterisation of the structure and dynamics of mβ_2m in solution at neutral pH (6.2) was carried out. After successful expression and purification of mβ_2m and its isotopically enriched (^{15}N, ^{13}C) forms, assignment of its backbone atoms was performed using triple resonance NMR techniques. A full analysis of pH and/or concentration-dependent protein dynamics over a range of timescales in solution and investigation of the folding pathway of mβ_2m revealed clues on the origin of increased/decreased amyloidogenicity for the different β_2m variants. Comparison with similar studies performed on ΔN6

and hβ_2m that served as positive and negative controls respectively allowed the link between structure and aggregation propensity to be made. Notably, the results showed that thermodynamic stability is not related to aggregation propensity, an observation that is consistent with the results found for some, but not all other systems[389], showing that amyloidogenicity is linked to the need to overcome kinetic barriers on pathway to fibril formation.

3.2 Results

3.2.1 Optimisation of the purification protocol for mβ_2m

In order to perform NMR studies on mβ_2m the purification protocol had to be optimised. Based on the purification protocol described in Section XX (Materials and Methods) for hβ_2m, after transformation of the plasmid containing the murine β_2m gene, the protein was expressed overnight in *E.coli* cells of the BL21 DE3 plysS- strain, after induction with IPTG. The cell pellet containing mβ_2m as inclusion bodies, was chemically lysed by the addition of 50-100mL of lysis buffer (see Section 2.3.3) and further cell disruption was performed using a constant cell disrupter. Mβ_2m was solubilised from inclusion bodies overnight in 20mM TrisHCl pH 8.5 containing 8M urea and refolded in the same buffer without urea. Initial purification was performed by anion exchange on a Q-Sepharose column which was equilibrated with 2-5 column volumes of 20mM TrisHCl pH 8.5. The bound protein was eluted in a gradient of 0-500mM NaCl (in the same buffer) over 500mL and was freeze-dried or concentrated. Freeze-dried protein was re-supsended in 10mM sodium phosphate pH 8.2 and further purification was performed by gel filtration on a HiLoad Superdex-75 Prep column. The peak corresponding to monomeric protein was collected and was either dialysed against water and freeze-dried or concentrated, aliquoted and stored at -80^{o}C.

Using this modified protocol, purification of mβ_2m typically yielded ~50mg of pure protein/per litre of LB culture (and approximately half of that yield for minimal media), showing that at least 2 samples of high purity labelled protein

for NMR studies could be obtained from just 1L of labelled media. A typical purification is shown in Figure 3.1.

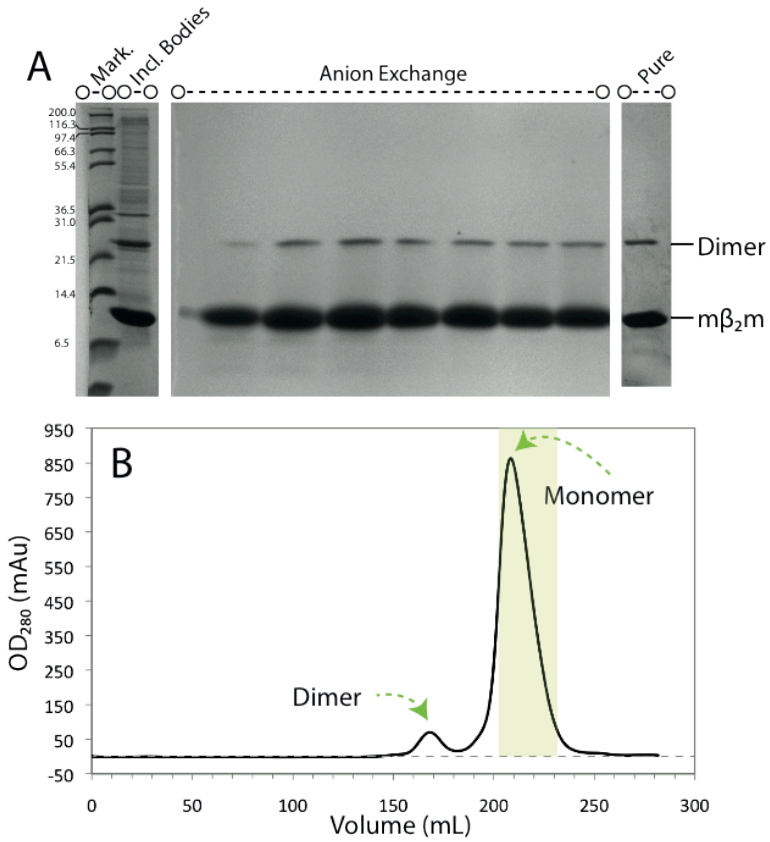

Figure 3.1: Purification of mβ$_2$m.
(A) SDS-PAGE analysis of samples taken at different purification steps (shown on top). The pure protein is shown in the last lane. (B) Gel filtration chromatogram. The fraction collected for storage is highlighted in a yellow box.

3.2.2 Mβ$_2$m and hβ$_2$m share similar structural properties in solution

As seen in Section 1.6.5.3 mβ$_2$m, hβ$_2$m and ΔN6 all show different degrees of amyloidogenicity with both hβ$_2$m and ΔN6 being able to form fibrils from the unfolded state (pH 2.0), while the latter is also capable of amyloid-like aggregation at pH 6.2. In contrast, mβ$_2$m does not aggregate into amyloid under any tested conditions despite having a very similar amino-acid sequence[365]. However, the small differences in the amino acid sequence could potentially translate into large deviations in the physico-chemical properties of mβ$_2$m compared with hβ$_2$m, thus explaining its reduced amyloidogenicity. Much theoretical work has been focused in predicting the ability of a protein to form amyloid fibrils based on the properties of its sequence, with numerous prediction algorithms being the outcome of this research[274]. The algorithm

Zyggregator used here, takes into account parameters such as β-strand and α-helix propensity, hydrophobicity, charge and the presence of gatekeeper residues to compute a per residue as well as an overall score of amyloid propensity, termed Zyggregator score[390]. Results for all 3 proteins shown in Figure 3.2, reveal very similar patterns of amyloid potential for all proteins, suggesting that the difference in the ability to form fibrils cannot be explained by this algorithm. Therefore, the structural properties of mβ$_2$m were investigated, seeking to answer what is the origin of its reduced amyloid potential.

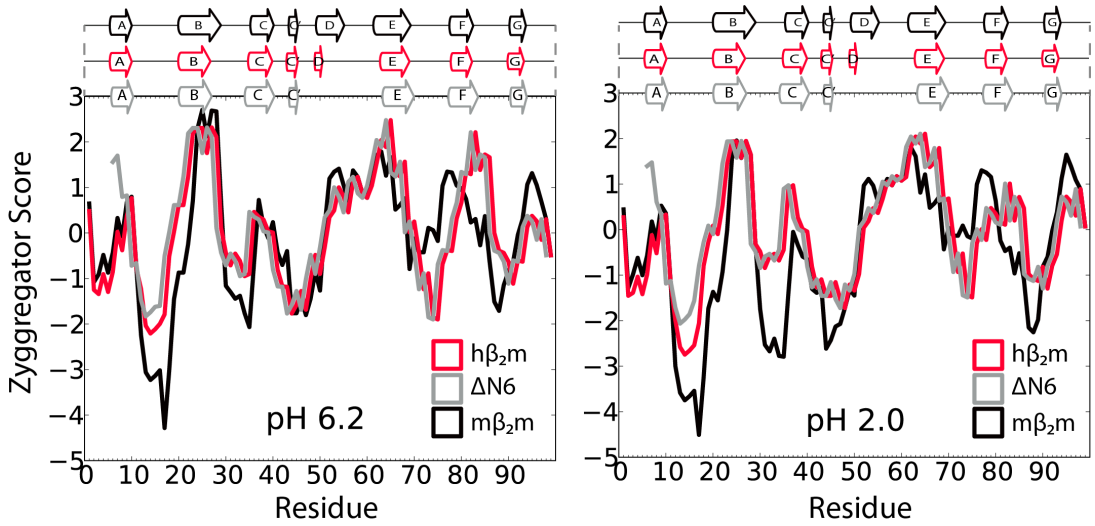

Figure 3.2: Predicted amyloidogenicity for hβ$_2$m, mβ$_2$m and ΔN6.
Per-residue plots for the zyggregator score[390] for hβ$_2$m (red), ΔN6 (grey) and mβ$_2$m (black), at pH 6.2 (left) and pH 2.0 (right). The secondary structure elements of each protein are shown on top of each panel using the same colour code.

Despite the fact that there are numerous high-resolution X-ray structures of mβ$_2$m, all of them are in complex with the murine MHC-I heavy chains and in a crystal form. However, binding to the MHC-I heavy chain is known to cause conformational changes in the D-strand and the DE loop of the β$_2$m monomer and more importantly, aggregation occurs only after its dissociation from the heavy chain of MHC-I *in vivo*[312,313]. Indeed, previous studies comparing the hydrogen exchange (H/D) properties of free and MHC-bound β$_2$m showed that dissociation of β$_2$m causes a dramatic destabilisation of the native fold and presumably the protein is then able to visit one or more amyloidogenic states on pathway to fibril assembly[391]. Therefore, to study monomeric mβ$_2$m in solution, uniformly labelled (^{15}N-, ^{13}C- labelled) protein was expressed and

purified to homogeneity (see section 3.2.1) and a set of standard triple resonance NMR experiments (HNCA, HNCO, CBCACONH, HNCACO, HNCACB) were performed in order to assign the backbone resonances of mβ$_2$m under close to physiological conditions (10mM sodium phosphate pH 6.2, 83.3mM NaCl) (Figure 3.3), 25^{o}C. Examples slices of HNCA, CBCACONH, HNCACB spectra and the assignment procedure is shown in Figure 3.3. The resulting assigned ^{1}H-^{15}N HSQC spectrum of mβ$_2$m is shown in Figure 3.4. A table of the assigned chemical shifts can be found in Appendix I.

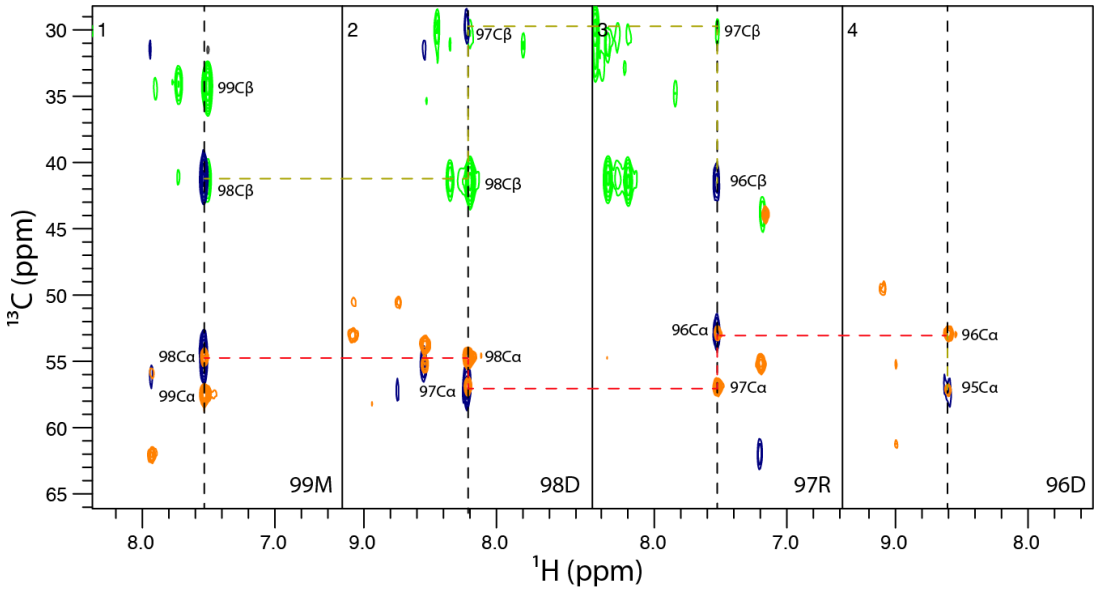

Figure 3.3: Assigning mβ$_2$m.
Examples slices of the 500MHz HNCA(orange), CBCACONH(blue), and 750MHz HNCACB(green) spectra for residues 99M, 98D, 97R and 96D. The assignment walk for Cα and Cβ is highlighted in red and yellow dashed lines respectively.

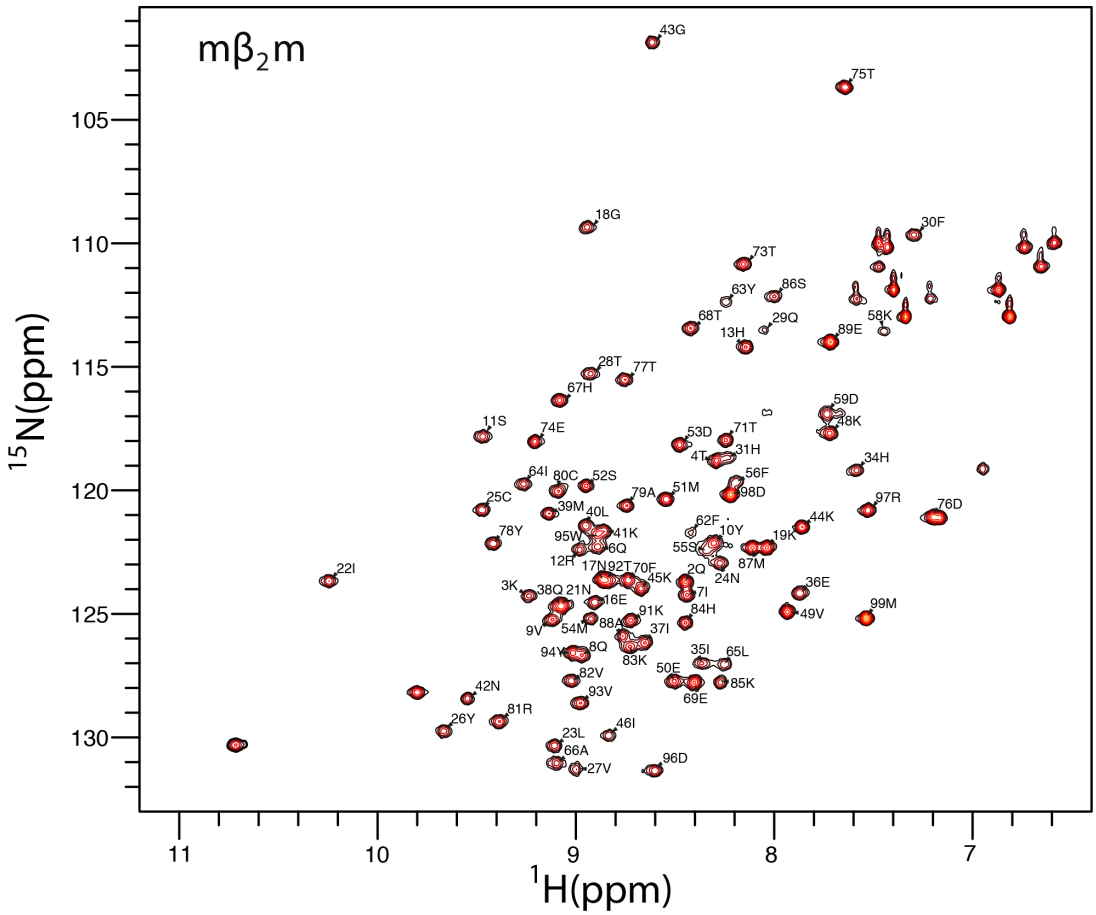

Figure 3.4: The assigned 500MHz ^{1}H-^{15}N HSQC spectrum of 500µM mβ$_2$m. Unassigned peaks correspond to side-chain resonances. Missing assignments for residues 57, 60, 61 suggest that those resonances are either too broad to be detected or overlap with other resonances.

^{1}H-^{15}N amide backbone assignments were obtained for 86 out of 90 non-proline residues and ~85% of all backbone atoms. The very high quality ^{1}H-^{15}N HSQC spectrum, shown in Figure 3.4, reveals high signal dispersion and sharp lines, indicative of a well folded protein that undergoes little, if no, chemical exchange under the conditions exploited. Comparison of the chemical shifts between mβ$_2$m and hβ$_2$m reveals significant chemical shift differences that are spread throughout the molecule and are not localised in specific regions of the protein (Figure 3.5).

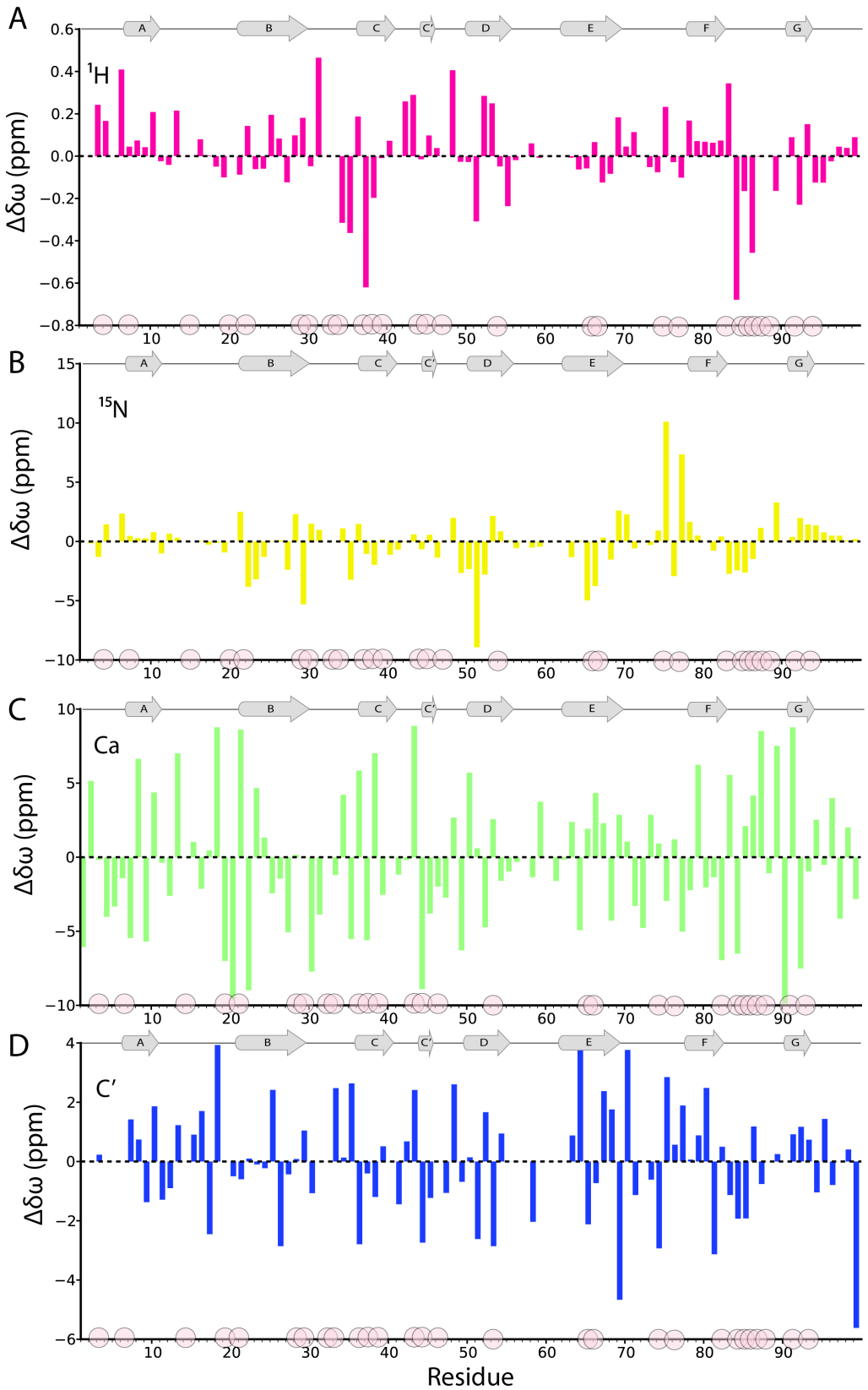

Figure 3.5: Chemical shift analysis of mβ$_2$m and hβ$_2$m.
Chemical shift differences of ^{1}H (A), ^{15}N (B), ^{13}C$_a$ (C), ^{13}C' (D) between mβ$_2$m and hβ$_2$m at 10mM sodium phosphate pH 6.2, 25^{o}C. The crystallographically determined secondary structure elements are shown as grey ribbons on top of the panels. Pink spheres denote the positions of single point mutations between mβ$_2$m and hβ$_2$m.

Taking into account the high sequence identity between the two molecules (Figure 1.23), the profound chemical shift differences observed in Figure 3.5 were unexpected. The hypothesis that emerges from these data is that the hydrogen bonding network that stabilises the native fold is significantly different in the human and murine version of β_2m and this might be reflected in the potential to form amyloid. To determine whether this is the case, the obtained chemical shifts were incorporated in TALOS+[378] in order to predict the secondary structure content of mβ_2m (see Section 2.3.10.1). Results are shown in Figure 3.6.

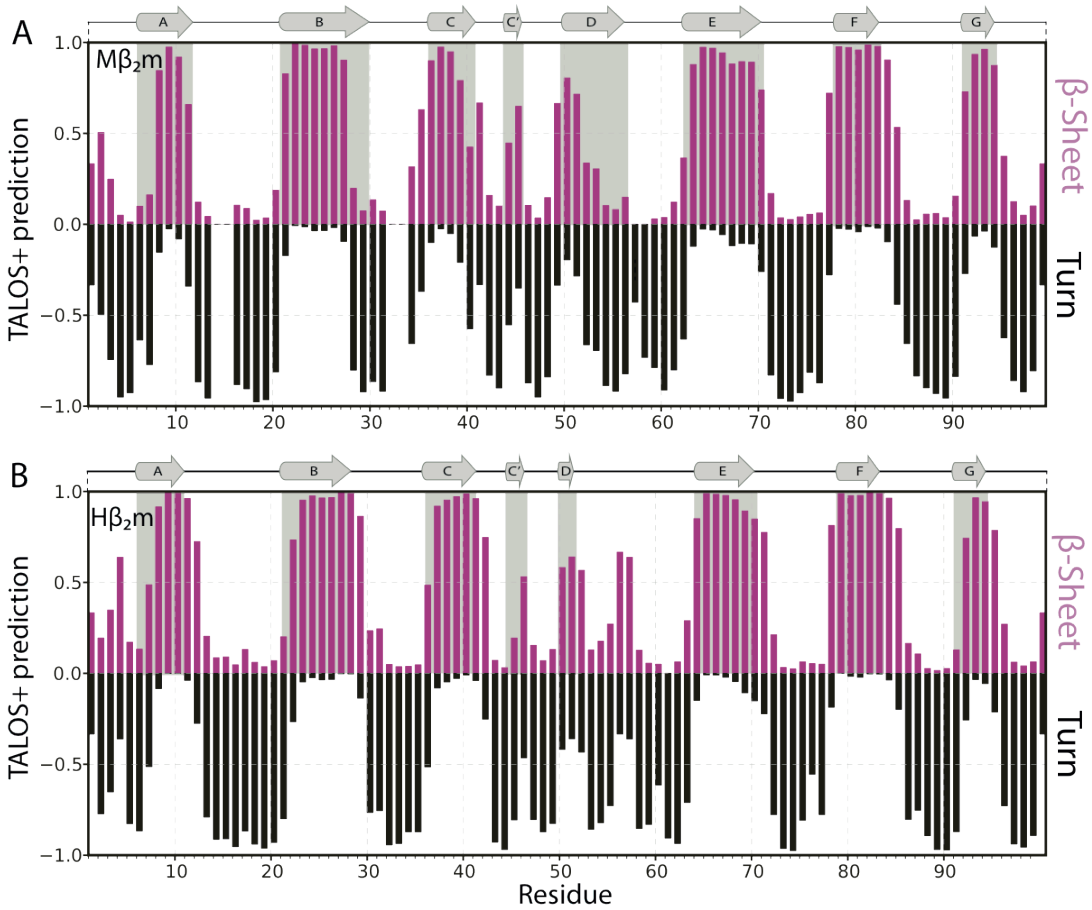

Figure 3.6: Secondary structure content for mβ_2m and hβ_2m.
TALOS+ prediction of the secondary structure elements in mβ_2m (A) and hβ_2m (B). Predictions for a residue lying in a β-strand are shown in purple , while predictions for β-turn are shown in black. Predictions for α-helix are insignificant and are not shown for simplicity. Chemical shifts for hβ_2m were extracted from the 17165 BMRB entry[272]. The secondary structure elements of the MHC-I bound mβ_2m and monomeric hβ_2m are shown on top of the panels as grey ribbons.

This analysis revealed a similar organisation of the secondary structure elements in both hβ_2m and its murine counterpart (Figure 3.6), suggesting that

both proteins adopt very similar structures in solution (and not only in crystals), with mβ2m presumably having a more disordered D strand. Crucially, based on the chemical shift data the D strand region in monomeric mβ2m seems to be more tangled in comparison to the crystal structure of the MHC-bound complex, suggesting multiple conformations in this area of the molecule, a phenomenon that has been previously connected to increased amyloidogenicity for hβ2m[392]. D-strand dynamics is a known feature of β2m. This part of the protein takes part in the recognition process by the heavy chain of MHC-I (Figure 1.18) and therefore a disordered D strand is observed for MHC-I bound hβ2m (but not mβ2m) molecules. On the other hand, in monomeric hβ2m, a breakage in the middle of the D strand region that involves a rotation of residue D53, results in a discontinuous strand comprising of residues 50-51 and 55-56[393,394]. Interestingly, a rare conformer that shows a completely straight D strand and therefore is hypothesized to facilitate the association of β2m monomers via their edge strands, was trapped by X-ray crystallography (Figure 3.7A)[344].

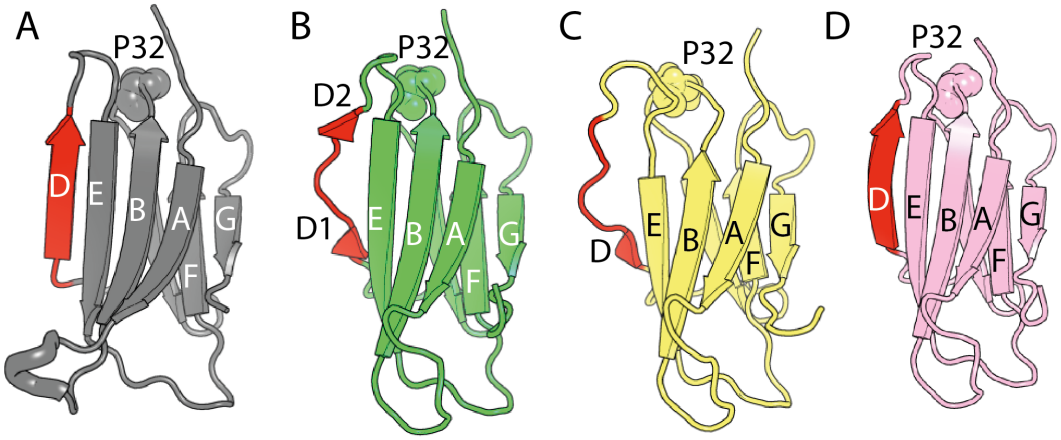

Figure 3.7: D strand dynamics.
(A) X-ray Structure of monomeric hβ2m (grey cartoon) showing a straight D strand (pdbID: 1LDS)[344]. (B) MHC-bound hβ2m (green cartoon), solved by X-ray crystallography with a discontinuous D-strand (pdbID:3MYJ)[307] and (C) monomeric NMR structure of hβ2m (yellow cartoon) showing a disordered D strand (pdbID:2XKS)[272]. The D strand region is coloured red and P32 is shown in spheres. (D) MHC-bound mβ2m (pink cartoon, pdbID: 1LK2_B),

ΔN6, the most aggregation-prone hβ2m variant currently known at neutral pH, exhibits pH-dependent aggregation, with no fibrils formed at pH 8.2 in contrast with pH 6.2, where the protein shows enhanced amyloidogenicity (Figure 1.21). This behaviour is reflected not only in the μs-dynamics of the protein (see

Section 1.6.5.3) but also in its chemical shifts, that change their positions in a pH-dependent manner, suggestive of pH-dependent conformational changes and/or protein association (Figure 3.8A). To assess whether mβ_2m also displays any pH-dependent conformational changes, its $^{1}H/^{15}N$ chemical shifts were monitored at low (6.2) and high (8.0) pH by ^{1}H-^{15}N HSQC spectra (Figure 3.8).

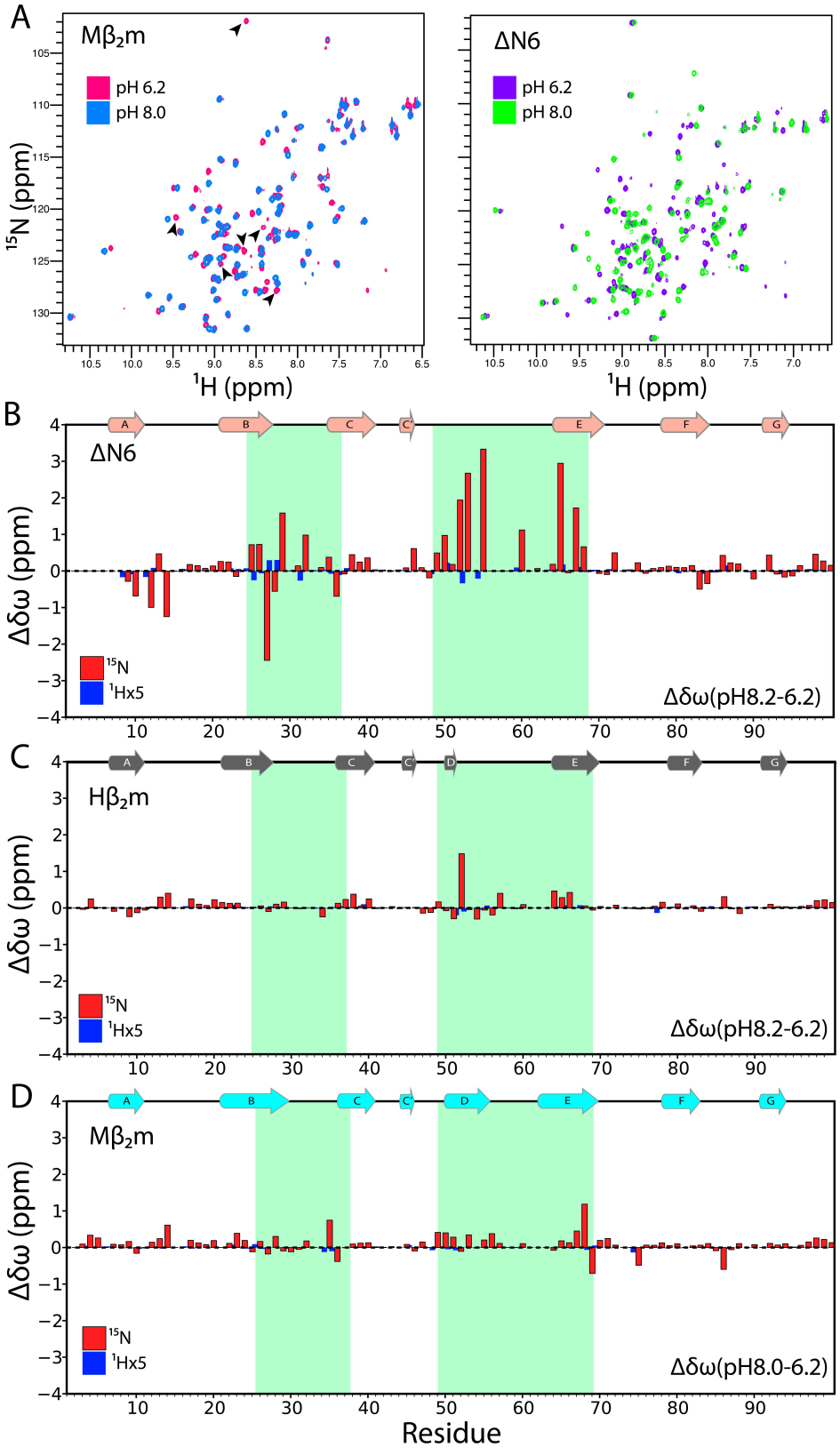

Figure 3.8: pH- dependent chemical shifts of ΔN6, hβ2m or mβ2m.
(A) ^{1}H-^{15}N HSQC spectra of 80µM mβ2m at pH 6.2 (pink) or at pH 8.0 (blue) or ΔN6 at pH 6.2 (purple) or pH 8.2 (green). Resonances missing from the spectrum of mβ2m (G43, K45, M54, K85) at pH 8.0 presumably due to deprotonation are marked with black arrows. (B) Diagram comparing chemical shifts differences in ^{15}N (red bars) or ^{1}H(x5) (blue bars) between 80µM of ΔN6

at pH 8.2 and 80μM of ΔN6 at pH 6.2 (10mM sodium phosphate, 25°C). (C) As in (B) but for hβ$_2$m at pH 8.2 and 6.2. (D) As in (B) but for mβ$_2$m at pH 8.0 and 6.2. The secondary structure elements of each protein are shown above each plot as a cartoon representation. Areas with significant chemical shift perturbations for ΔN6 are highlighted in green boxes.

As shown in Figure 3.8, the pronounced chemical shift differences upon pH titration observed for ΔN6 are absent for mβ$_2$m, which does not respond to alterations of the pH, in a similar manner to hβ$_2$m. This finding further highlights the importance of the conformational changes at low pH (6.2) as the origin of the increased amyloidogenicity of ΔN6. Interestingly, some resonances of mβ$_2$m are no longer detected at pH 8.0, presumably due to deprotonation and/or rapid exchange with the solvent, a phenomenon that is not observed for hβ$_2$m or ΔN6 under identical conditions. These findings indicate that mβ$_2$m is less protected from exchange with the solvent compared with ΔN6 and hβ$_2$m, a phenomenon that was investigated further in Section 3.2.4.

Apart from being very sensitive structural probes, chemical shifts can be used to predict the dynamic properties of the vectors involved. Based on the observation that random coil chemical shifts (RCI) are usually representative of highly flexible regions, NMR order parameters (S^2)[165] –usually obtained by model-free analysis of relaxation experiments[372]- can be estimated. S^2 values provide an indication of the amplitude of motions a particular vector experiences in solution. CS-predicted S^2 values are shown in Figure 3.9.

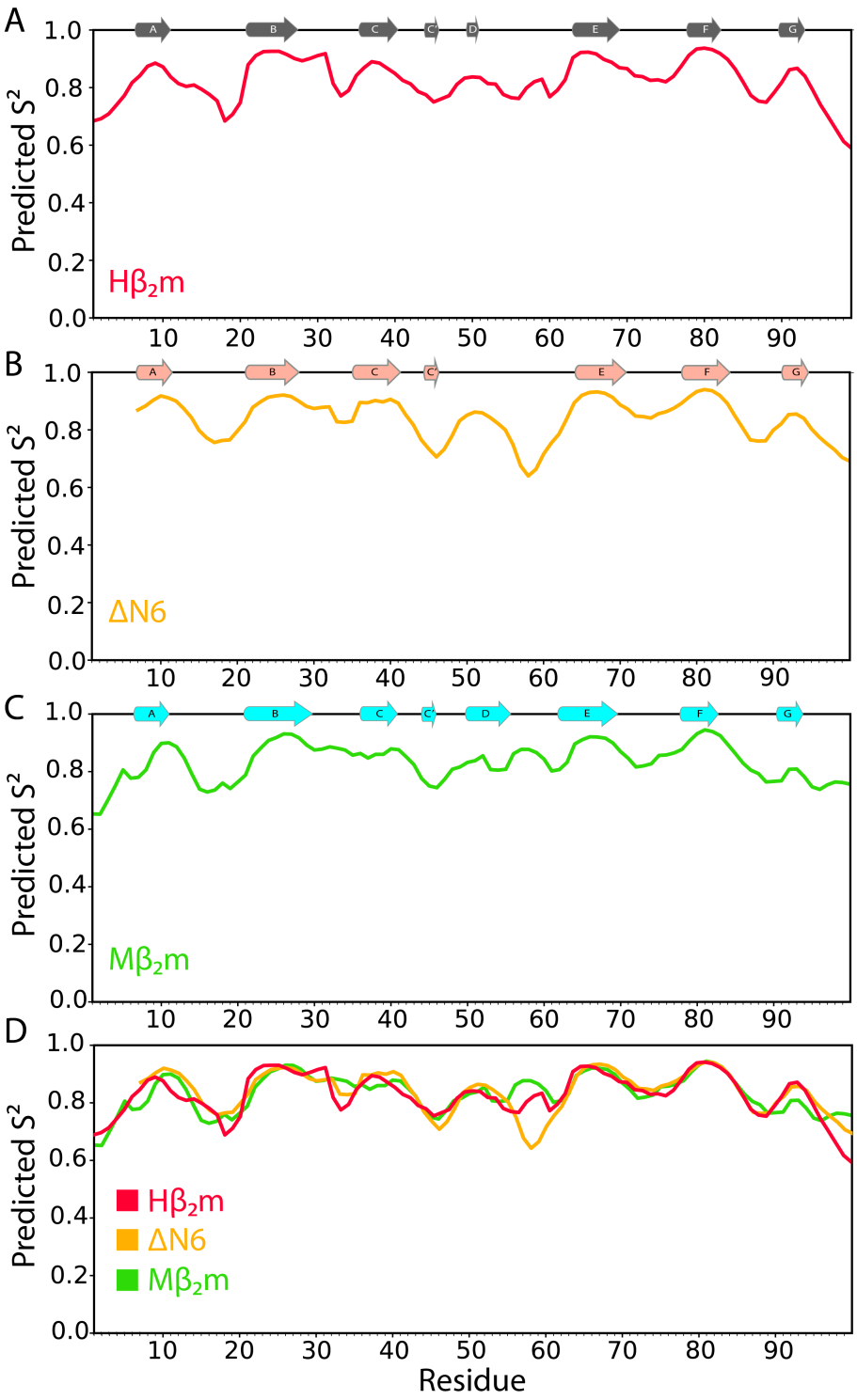

Figure 3.9: Predicted S^2 values of mβ₂m, hβ₂m and ΔN6.
TALOS+[378] predicted S^2 values for hβ₂m (A), ΔN6 (B) and hβ₂m (C) based on their chemical shift index. A value of 1.0 indicates a completely rigid vector, whereas a value of 0.0 indicates a vector that is freely moving in solution. An overlay of all the data is shown in (D).

Assessing protein dynamics based on chemical shifts, at least in the case of β₂m molecules, seems to be problematic, since as seen in Figure 3.9 the pattern of the predicted order parameter is similar for all three proteins. ΔN6 is predicted

to be more flexible in the DE loop, however this could be due to the fact that resonances of residues in that region could not be assigned, presumably because of exchange line broadening.

Data presented so far reveal the high level of similarity that $m\beta_2m$ and $h\beta_2m$ share in solution and provide some clues on their dynamical properties. However, since it has been shown before that fast protein dynamics can dictate the amyloidogenicity of ΔN6 and consequently $h\beta_2m$[272], the dynamics of $m\beta_2m$ in solution were next investigated in detail.

3.2.3 Investigation of fast protein dynamics reveals dimerisation hot-spots

In a previous study, the intramolecular motions of ΔN6 were investigated using NMR to reveal that ΔN6 is in exchange with one or more lowly-populated species that has/have increased conformational dynamics (as shown by increased R_2 rates, Figure 1.21). The population of these species in equilibrium with the native state of ΔN6 is increased at low pH (pH 6.2, Figure 1.21) an observation that does not occur for $h\beta_2m$ and hence it was hypothesized that these lowly-populated species are responsible for the increased amyloid potential of ΔN6 under the conditions employed[272]. To assess whether $m\beta_2m$ is in exchange with lowly-populated, potentially more amyloidogenic species a complete characterisation of the internal motions of the protein was carried out using ^{15}N-relaxation methods at pH 6.2 and 8.0, 25^{o}C. As expected, $m\beta_2m$ does not show any alteration in its T_1 or HETnOe dynamics as was previously observed for $h\beta_2m$ and ΔN6[272] (Figure 3.10).

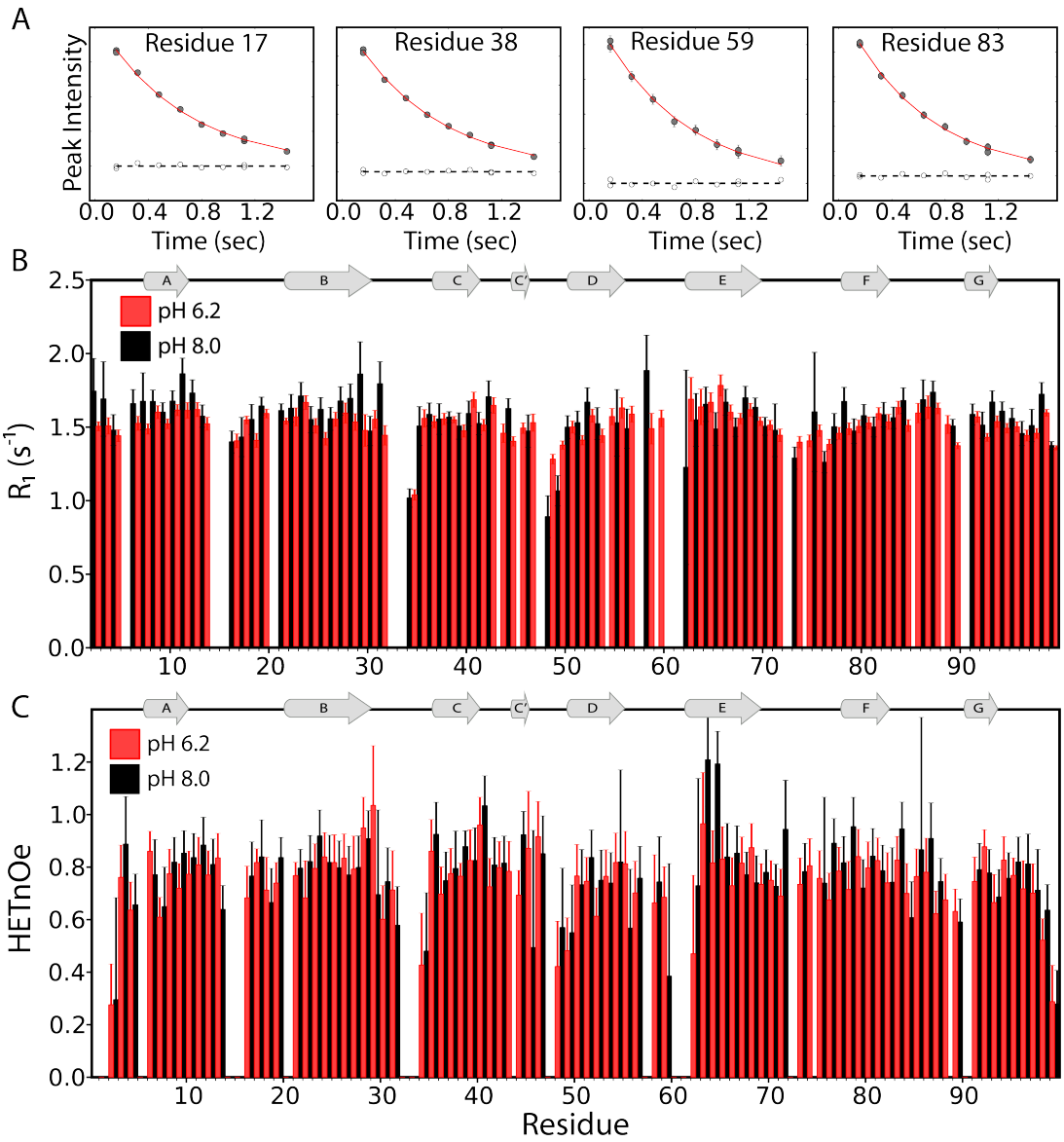

Figure 3.10: Picosecond-nanosecond protein dynamics of mβ$_2$m.
(A) Example raw data are shown for the T_1 relaxation of 80μM mβ$_2$m, at pH 6.2, 25°C. The peak intensity is plotted as a function of relaxation time for residues 17, 38, 59, 83. The red lines represent fits to single exponentials and open circles denote the residuals of the fit. (B) Longitudinal T_1 relaxation (R_1=1/T_1) for 80μM mβ$_2$m at pH 6.2 (red bars) compared with 80μM mβ$_2$m at pH 8.0 (black bars). (B) {^{1}H}-^{15}N heteronuclear nOe rates for 80μM mβ$_2$m at pH 6.2 (red bars) and 80μM mβ$_2$m at pH 8.0 (black bars). The buffer used in all experiments was 10mM sodium phosphate pH 6.2/8.0, 83.3mM/86.6mM NaCl. Secondary structure elements are shown as cartoons on top of the panels.

Characteristic of a well ordered, structured domain these data are in accord with previous findings for hβ$_2$m and ΔN6. The effect of pH on mβ$_2$m dynamics on the ps-ns timescale was also investigated by ^{15}N T_2-relaxation measurements. Whereas an increase in the dynamics of ΔN6 is observed as the

protein becomes protonated at pH 6.2 (Figure 1.21), this phenomenon is not observed for mβ2m (Figure 3.11).

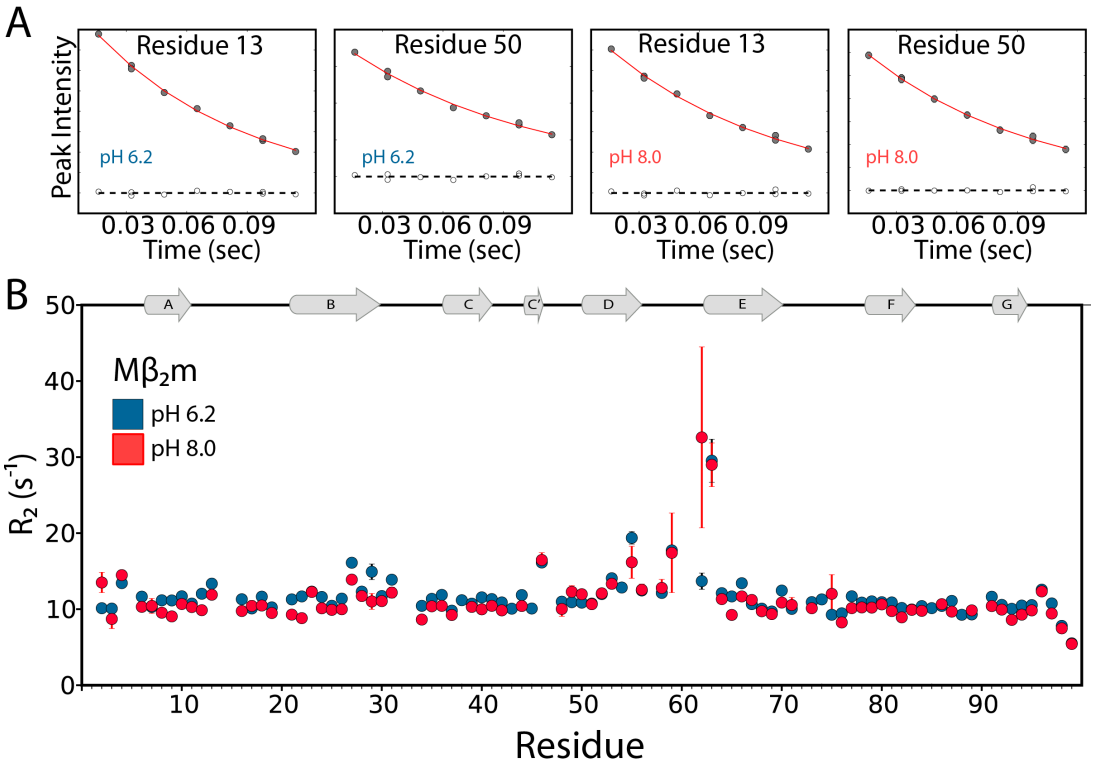

Figure 3.11: pH-dependent dynamics of mβ2m measured by T_2 relaxation. (A) Example raw data are shown for the T_2 relaxation of 80μM mβ2m, at pH 6.2 or pH 8.0, 25°C. The peak intensity is plotted as a function of relaxation time for residues 13 and 50 at pH 6.2 (left) or for the same residues at pH 8.0 (right). The red lines represent fits to single exponentials and open circles denote the residuals of the fit. (B) R_2-relaxation rates ($R_2=1/T_2$) for 80μM mβ2m at pH 6.2 (blue) and pH 8.0 (red). The buffer used in all experiments was 10mM sodium phosphate pH 6.2/8.0, 83.3mM/86.6mM NaCl. Secondary structure elements are shown as grey cartoons on top of each panel.

The fact that the R_2 rates for ΔN6 at pH 6.2 are higher for all resonances than those at pH 8.2, suggests an increased tumbling time at low pH (Figure 1.21). In support of this scenario, the R_2 dynamics of ΔN6 show the same response upon increasing protein concentration (80-500μM) suggestive of a pH-dependent oligomerisation event[272]. In contrast, mβ2m remains vastly monomeric even at pH 6.2, with no dependence of the R_2 on pH, over the range studied. The impact of protein concentration on the dynamics of mβ2m was also investigated here at pH 6.2. The increased average R_2 and decreased R_1 rates over all residues seen in Figure 3.12A, B could be the outcome of protein association or increased viscosity at high protein concentration which prevents free tumbling of the molecules in solution. However, the product of R_1*R_2 can

be made independently of the viscosity of the solution[395]. In this analysis the enhanced R_1*R_2 at high protein concentration shows clearly that mβ_2m undergoes weak dimerisation as the protein concentration is increased, with residues 26-30 showing the most pronounced effect, possibly because this region is taking part in the dimerisation interface (Figure 3.12C). Importantly, in a recent study the use of NMR and SAXS suggested that hβ_2m dimerises at high protein concentrations, with the apical part of the protein around Pro 32 being involved in the dimer interface. A similar tendency was also observed for the more aggregation-prone I_T state[396].

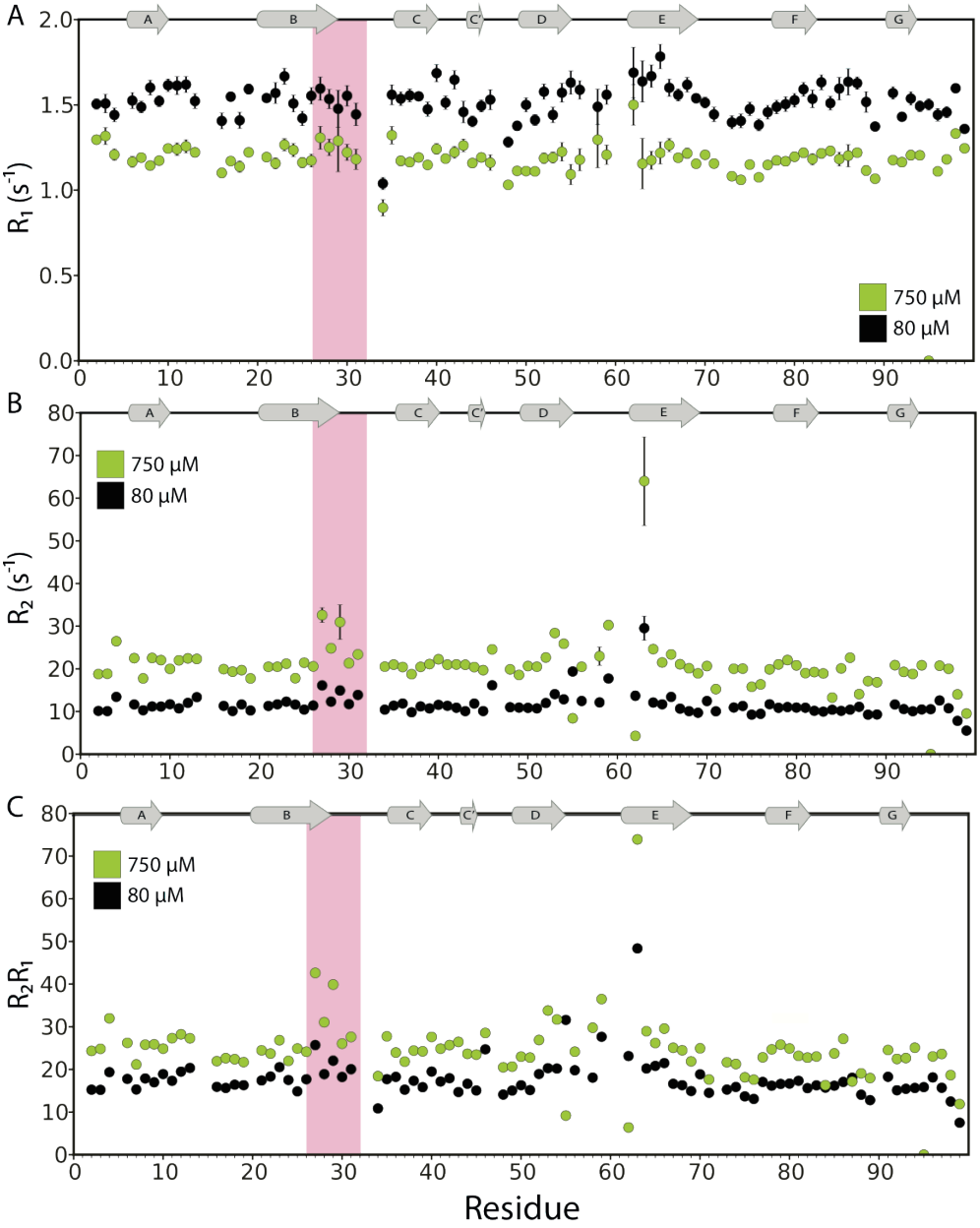

Figure 3.12: Concentration-dependent ps-ns dynamics.
(A) Longitudinal T_1 dynamics ($R_1=1/T_1$) for 80µM mβ2m at pH 6.2 (black) compared with 750µM mβ2m at pH 6.2 (yellow) in 10mM sodium phosphate buffer. (B) Transverse T_2 dynamics ($R_2=1/T_2$) for 80µM mβ2m at pH 6.2 (black) compared with 750µM mβ2m at pH 6.2 (yellow). (C) Comparison of the R_1*R_2 product for 80µM mβ2m at pH 6.2 (black) and 750 mβ2m at pH 6.2 (yellow). Secondary structure elements are shown as grey cartoons on top of each panel and areas (residues 26-30) which show a differential-site specific increase in dynamics upon increased protein concentration (in addition to the global enhancement) are highlighted in pink boxes. For residues for which the error bars are not visible, those are smaller than the marker size.

Based on these data, an estimation of the correlation time (τ_c) can be made using the relaxation times T_1 and T_2 in the formula:

$$\tau_c \approx \frac{1}{4\pi\nu_N}\sqrt{6\frac{T_1}{T_2}-7} \qquad \text{equation (26)}$$

where ν_N is the resonance frequency of the detected nucleus (^{15}N) and T_1, T_2 are the longitudinal and transverse relaxation times respectively.

At 80μM a calculated τ_c of 6.22ns for mβ$_2$m, is fully consistent with a monomeric protein of 11kDa in size, while at 750μM the correlation time τ_c increases to 9.90 ns, suggesting a monomer-higher order species (presumably dimers) equilibrium. However, assuming spherical geometries for both monomers and since the value of τ_c is not that expected for a fully formed stable dimer (12.44ns) the dimerisation must be transient and of weak affinity even at this very high protein concentration (750μM). Thus, the results presented show that the increased motions in the BC and DE loop of mβ$_2$m, in close proximity to P32 that have been connected with the increased aggregation potential of ΔN6 are absent from mβ$_2$m, explaining its decreased amyloidogenicity at physiological conditions. However, mβ$_2$m dimerises weakly with residues 26-30 likely to be involved in the interface.

In support of this scenario significant chemical shift differences were observed when the protein concentration was increased 6-fold. Residues that show the largest chemical shift differences (larger than 3 times the standard deviation) form a contiguous surface including the BC, CD and FG loops (Figure 3.13A, B). Residues in the A strand that show smaller chemical shift perturbations are not part of this surface and thus were not considered part of the interface. Concentration-dependent chemical shift changes were converted to sparse distance restraints and were incorporated into a docking protocol using XPLOR-NIH[366]. The resulting structural models of the mβ$_2$m homodimer show an antiparallel rearrangement of the monomers, where the major intermolecular contacts are formed by residues in the BC and CD loops (Figure 3.13C). However, it has to be underlined that due to the limitations on the use of chemical shift perturbations as distance restraints in docking calculations the structures shown are only structural models and might deviate from the

real structure of the dimer. Furthermore, the relevance of the mβ$_2$m homodimer to its aggregation properties remains to be elucidated.

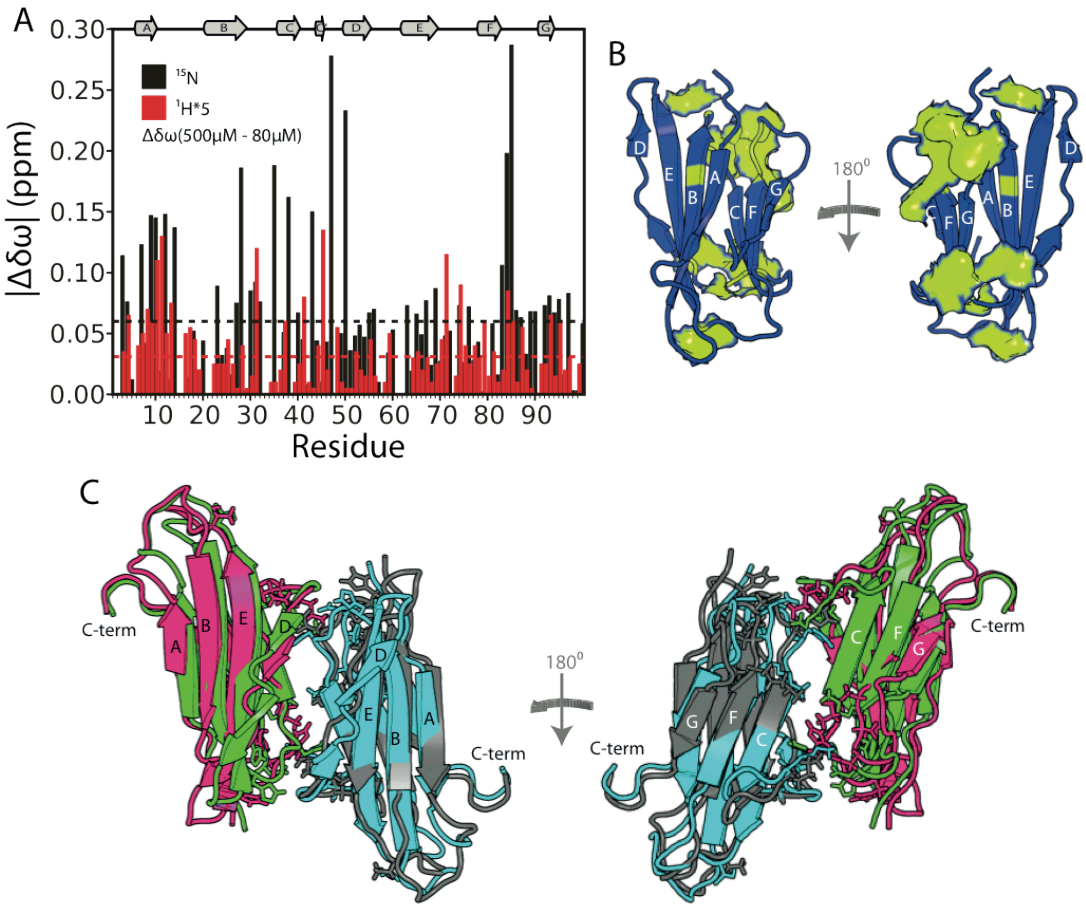

Figure 3.13: Modeling the mβ$_2$m homodimer.
(A) Chemical shift perturbations (CSP) (^{15}N-black and 5*^{1}H-red) observed between a sample comprising of 80μM and 750μM mβ$_2$m in 10mM sodium phosphate buffer pH 6.2. Secondary structure elements are shown on top as cartoons. (B) The structure of monomeric mβ$_2$m (1LK2_B) showing residues that showed a CSP larger than 3 standard deviations as yellow surface representation. (C) The lowest energy structure of the homodimer obtained by using the CSPs shown in A and B as distance restraints in a XPLOR-NIH docking protocol, is shown in pink and grey cartoons. The structure obtained by specifying the residues highlighted in B as 'active' residues in HADDOCK[397] is shown in green and cyan cartoons for comparison. Residues with high CSP (mapped on the monomeric structure in B) are shown as sticks.

3.2.4 Thermodynamic and kinetic stability are not related to amyloidogenicity

To gain insights into the kinetic and/or thermodynamic energy barriers that stabilise the native fold, the hydrogen-deuterium exchange dynamics (H/D exchange) of mβ$_2$m and hβ$_2$m were investigated. In these experiments, a

protein concentration of 80 μM was used, conditions under which the proteins are predominantly monomeric. Lyophilised protein was dissolved in 100% D_2O at pH 6.2 and its H/D exchange properties were investigated by following the loss of signal intensity in a series of ^{1}H-^{15}N SOFAST-HMQC experiments at 25°C. Since both proteins have been shown before to exchange in a complex EX1 and EX2 mechanism[391], thermodynamic analysis based on the measured H/D exchange rate (k_{ex}) is prohibited. Interestingly, the results obtained showed that $m\beta_2m$ exchanges 1-2 orders of magnitude faster than $h\beta_2m$ as depicted in Figure 3.14. Surprisingly perhaps, the least protected (most dynamic) protein is not able to form fibrils at pH 6.2.

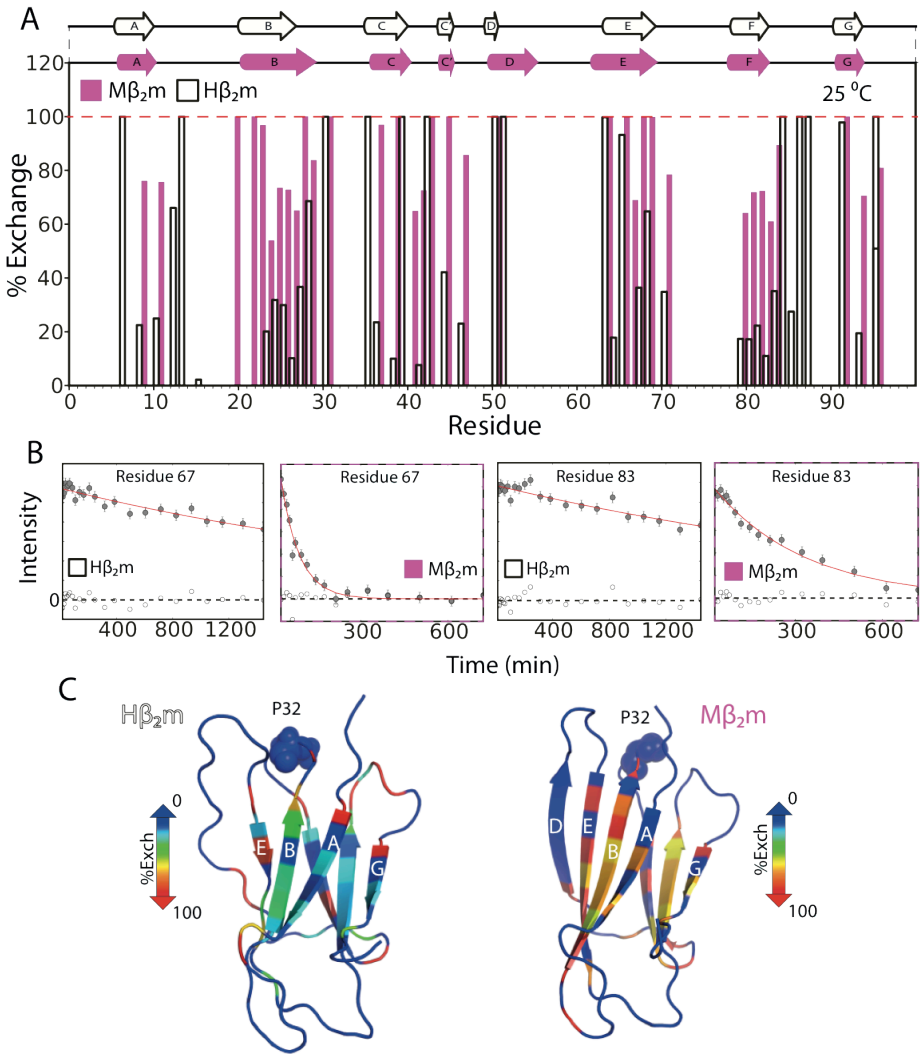

Figure 3.14: H/D exchange dynamics for $m\beta_2m$ and $h\beta_2m$.
(A) Diagram of % H/D exchange for 80μM $m\beta_2m$ (pink) and 80μM $h\beta_2m$ (open bars) at pH 6.2, 25°C. % Exchange is described as the ratio of the intensity at the end of the experiment (1450 min for $h\beta_2m$ and 725 min for $m\beta_2m$) and the intensity at ~0 min after H/D exchange was initiated. Missing bars denote residues that exchange too fast and therefore, their resonances are absent from

the spectrum. Black/white and purple ribbons on top of the panel indicate the secondary structure elements of hβ$_2$m and mβ$_2$m respectively. (B) Raw data showing the difference in the H/D exchange rates of mβ$_2$m and hβ$_2$m. Representative curves are shown for residues 67 (left panels) and 83 (right panels) for 80μM mβ$_2$m and 80μM hβ$_2$m. Fitted exponentials are shown as red solid lines and the residuals of the fit as open dots at the bottom. (C) The structure of hβ$_2$m (left) or mβ$_2$m (right) coloured according to the % H/D exchange. P32 is shown as spheres.

In addition to having the lower kinetic stability, mβ$_2$m is more unstable thermodynamically in comparison to hβ$_2$m and ΔN6 at pH 6.2 and 25^{o}C, with a ΔG^0 of -10.7±0.5 kJ/mol while the unfolding free energies for hβ$_2$m and ΔN6 are -22.5±0.9 kJ/mol and -11.3±0.5 kJ/mol respectively (Clare Pashley personal communication).

3.2.5 Investigating the folding pathway of mβ$_2$m by real-time NMR

As mentioned above, the amyloidogenicity of hβ$_2$m is encrypted in its folding intermediate I_T, which is only marginally populated at physiological conditions (~<4%). The population of this aggregation-prone species at neutral pH is presumably well below the critical concentration at which nucleation of amyloid formation commences, explaining why hβ$_2$m is not aggregation-prone under these conditions. Enhancing the population of I_T by the addition of co-solvents induces the ability of hβ$_2$m to aggregate at neutral pH revealing a linear relationship between the concentration of the intermediate species and fibril yield[398]. However NMR data presented in section 3.2.3 showed that mβ$_2$m represents a well folded, not dynamic protein entity, showing little if any exchange to additional species. These data naturally raise the question of whether or not mβ$_2$m folds through an I_T state and if it does, to what extent the murine I_T state possesses the same amyloidogenic properties as its human counterpart.

To answer these questions, real-time folding experiments were performed on mβ$_2$m using NMR. The protein was denatured for 1h at pH 6.2 by the addition of 8M urea and then refolded rapidly by 10-fold dilution in 25mM sodium phosphate pH 6.2 at 25^{o}C. Folding was then monitored by ^{1}H-^{15}N SOFAST-

HMQC experiments[382] that were acquired over time to gain insights on the structural features of any intermediate species that accumulate as well as on the folding kinetics. The high energy barrier for proline isomerisation results in a very slow folding phase (k_{app}= 0.3 – 0.5 h^{-1}) providing a timeframe that can easily be investigated by SOFAST methods that allow the recording of a spectrum as rapidly as 2min after the reaction was initiated[385]. Here, the experiment time was adjusted to 45 s and folding was monitored over a period of 100min for mβ_2m.

The spectrum of mβ_2m recorded ~4:00 min after the refolding was initiated, reveals the presence of 80 cross-peaks most of which overlay with resonances of the native mβ_2m. Interestingly, resonances in the N-terminal region (8-9), the BC loop (31-36) and the DE loop and the N-terminal region of the E strand (52-53, 56-61, 63-66) all of which lie in close proximity to P32, are absent from the spectrum presumably because of exchange line broadening (Figure 3.15). Similar observations have been made for hβ_2m, where the proline cis-trans isomerisation at position 32 leads to conformational changes in the apical region of the protein as depicted in the structure of ΔN6 (Figure 1.20)[359]. Additionally, enhanced chemical shift differences (Figure 3.15B) next to the areas of extreme line broadening (N-terminus, C-terminal region of the B and E strand, N-terminal region of the D strand and the F strand and the FG loop) further support the scenario of chemical exchange to additional conformations with these areas as the epicentre. Thus, these data suggest that the folding pathway of mβ_2m involves a slowly folding intermediate that seems to have similar structural properties to those of the I_T state of hβ_2m, and ΔN6.

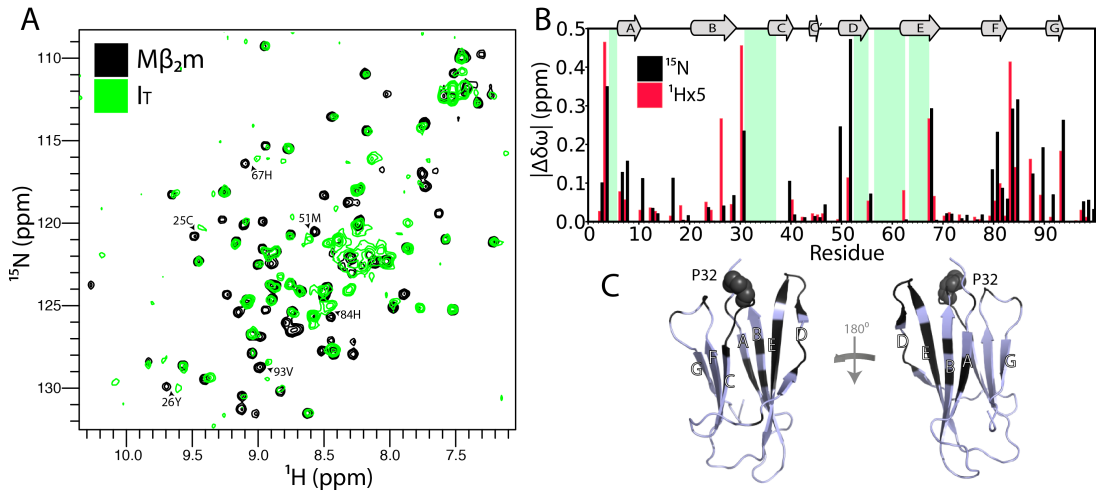

Figure 3.15: The I_T state of mβ$_2$m.
(A) ^{1}H-^{15}N SOFAST-HMQC spectra of 200μM native mβ$_2$m (black) and 200μM mβ$_2$m-I_T (green). The latter was obtained ~4min after the refolding was initiated at pH 6.2. (B) ^{1}H (red bars) and ^{15}N (black bars) chemical shift differences between mβ$_2$m-I_T and mβ$_2$m. Regions of increased line broadening are highlighted in green boxes. The secondary structure content of mβ$_2$m is shown as grey ribbons on top of the panel. (C) The residues whose resonances are broadened beyond detection in mβ$_2$m-I_T (black) are shown on the structure of mβ$_2$m (grey cartoon). P32 is shown as sphere.

However, the kinetics of the refolding reaction differ significantly between the two proteins. As shown in Figure 3.16, the folding of hβ$_2$m is completed within ~20min, under the conditions employed, while the spectrum of hβ$_2$m obtained only 3 min after folding was initiated contains resonances consistent with a ~30% population of the protein being already in the native state (Figure 3.16 top panels). On the other hand, after 20 min, mβ$_2$m still populates ~20% of its I_T state (with the native signature completely absent in the first spectrum recorded after ~4min), showing that the rate limiting proline isomerisation step is significantly slower in this case (Figure 3.16 middle and bottom panels). Figure 3.17 compares the rate of folding for the same residues in the context of hβ$_2$m or mβ$_2$m.

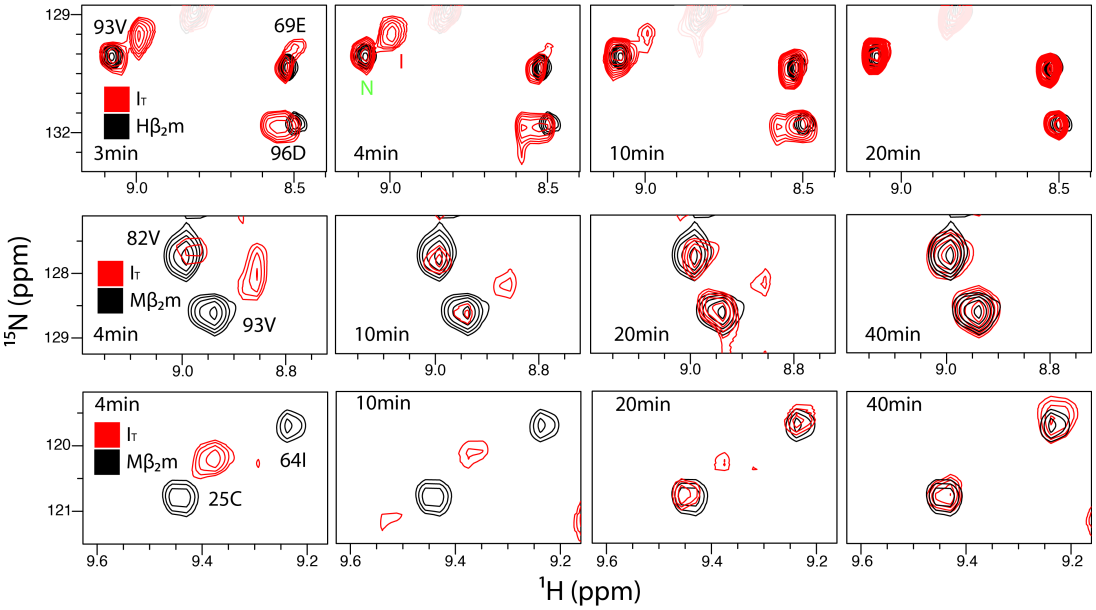

Figure 3.16: ^{1}H-^{15}N SOFAST HMQC spectra of hβ$_2$m and mβ$_2$m and their I$_T$ states. Zoom-in the regions of ^{1}H-^{15}N SOFAST-HMQC spectra recorded during the folding of hβ$_2$m (top) or mβ$_2$m (middle, bottom) in 10mM sodium phosphate buffer pH 6.2, 25^{o}C, 0.8M final urea concentration. Resonances for 93V, 69E and 96E are shown for hβ$_2$m and 93V, 25C and 64I for mβ$_2$m in a series of spectra recorded at 3, 4, 10, 20 min (hβ$_2$m) and 4, 10, 20, 40 min (mβ$_2$m) after folding was initiated.

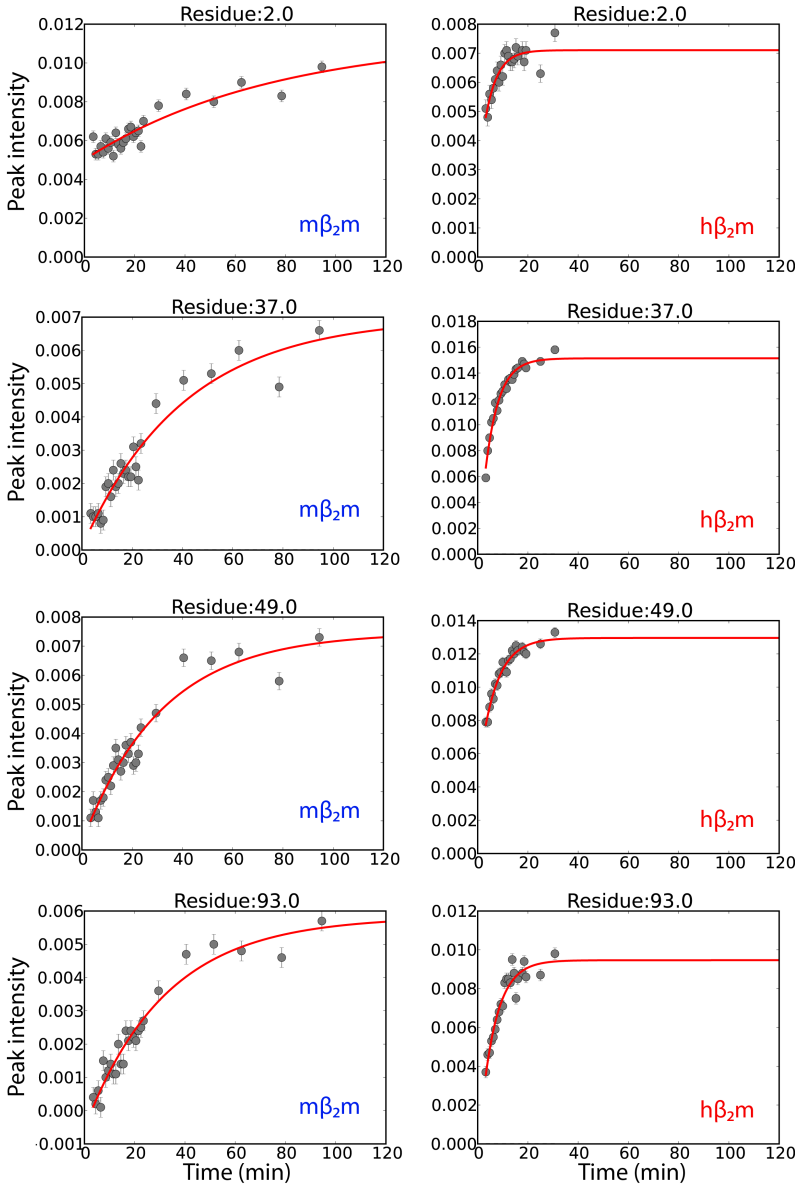

Figure 3.17: Folding kinetics of proline-isomerisation by real-time NMR. The appearance of the native signal for residues 2, 37, 49, 93 of mβ2m (left) or the same residues in hβ2m is plotted as a function of time. Red lines represent mono-exponential fits to the data. The missing initial amplitude for residues in hβ2m shows that ~20-40% of the reaction has already taken place in the dead time of the experiment, a phenomenon not observed for mβ2m.

A detailed analysis of real-time NMR data was then performed in order to extract kinetic parameters for the folding of mβ2m on a per-residue level (Figure 3.18). Resonances in the mβ2m spectrum were divided into two categories: those belonging to the native state and those belonging to the pure I_T state, (N-coloured green and I-coloured red in Figure 3.18). Overlapping resonances between the I and N states were excluded from the analysis.

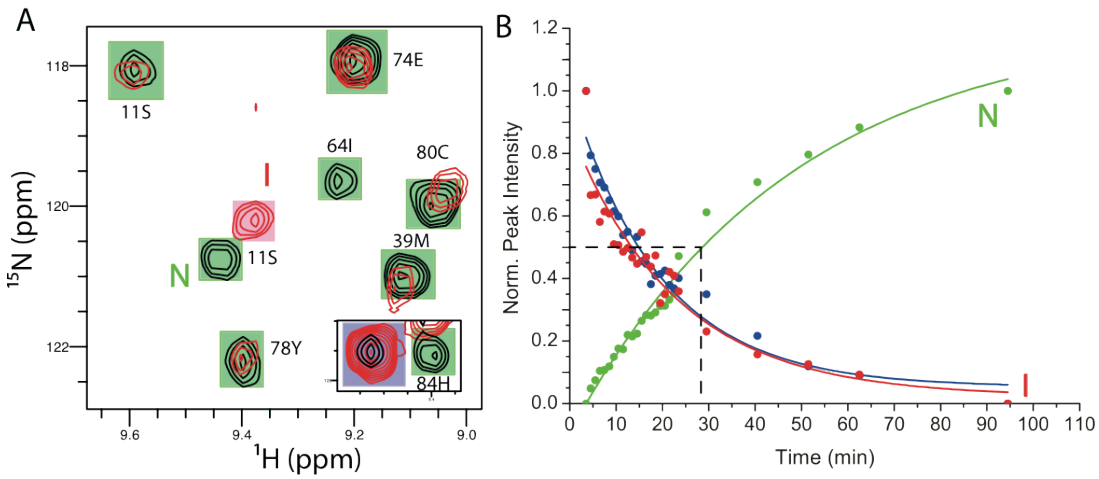

Figure 3.18: Folding rates and kinetic analysis of proline isomerisation. (A) Region of the ^{1}H-^{15}N SOFAST-HMQC spectrum of native mβ$_2$m (black) and I_T (3min in the folding reaction) (red). Native (N) and I_T (I) peaks are highlighted in green and red boxes respectively. Resonances that are present in the native spectrum but their intensity is decreasing as the folding reaction proceeds are shown in blue box in the inset. (B) Fitting of the appearance of the native (green) of disappearance of I_T signal (red). The sum of corresponding peak intensities is plotted as a function of time. The rate of disappearance of the peak intensity belonging to the third class of resonances (blue in A) is also shown in blue. Solid lines represent mono-exponential fits and dashed lines the 50% midpoint.

A folding rate (k_f^N) of 0.000353s^{-1} was obtained for mβ$_2$m at pH 6.2 and 25^{o}C, which is at least 5 fold slower than the equivalent rate for hβ$_2$m under similar conditions (pH 6.6, 24^{o}C) reported previously (k_f^N = 0.002 s^{-1})[360]. Interestingly, a 3rd class of resonances which are present in the native spectrum of mβ$_2$m with low peak intensities and unassigned (8 peaks in total) but whose intensities decrease as folding proceeds, was identified. The rate of signal loss for this class of resonances (blue in Figure 3.18B) is remarkably similar to the one observed for the I_T resonances suggesting that even under native-like conditions, mβ$_2$m undergoes chemical exchange to an intermediate state. These peaks most likely correspond to lysine side chain atoms and thus are not detectable in the amide relaxation experiments shown in Figures 3.11, 3.12. Notably, the k_f^I rate was found to be 0.00083 s^{-1}, two fold higher than k_f^N showing that the folding of mβ$_2$m is more complex than two state, and there is (are) additional NMR invisible state(s) on pathway to the native fold compared with hβ$_2$m.

A comparison of the structural properties of the I_T states of $h\beta_2m$ and $m\beta_2m$ based on their chemical shifts reveals a remarkable similarity between the two intermediate states. Resonances in the N-terminus, the BC and DE loops of both proteins suffer from severe line broadening, pointing to chemical exchange involving the regions around Pro32 (Figure 3.19). Since $m\beta_2m$ folds through an I_T state which seems to be similar to the aggregation-prone I_T of $h\beta_2m$, additional steps are potentially required for this intermediate to unravel it encrypted amyloidogenicity.

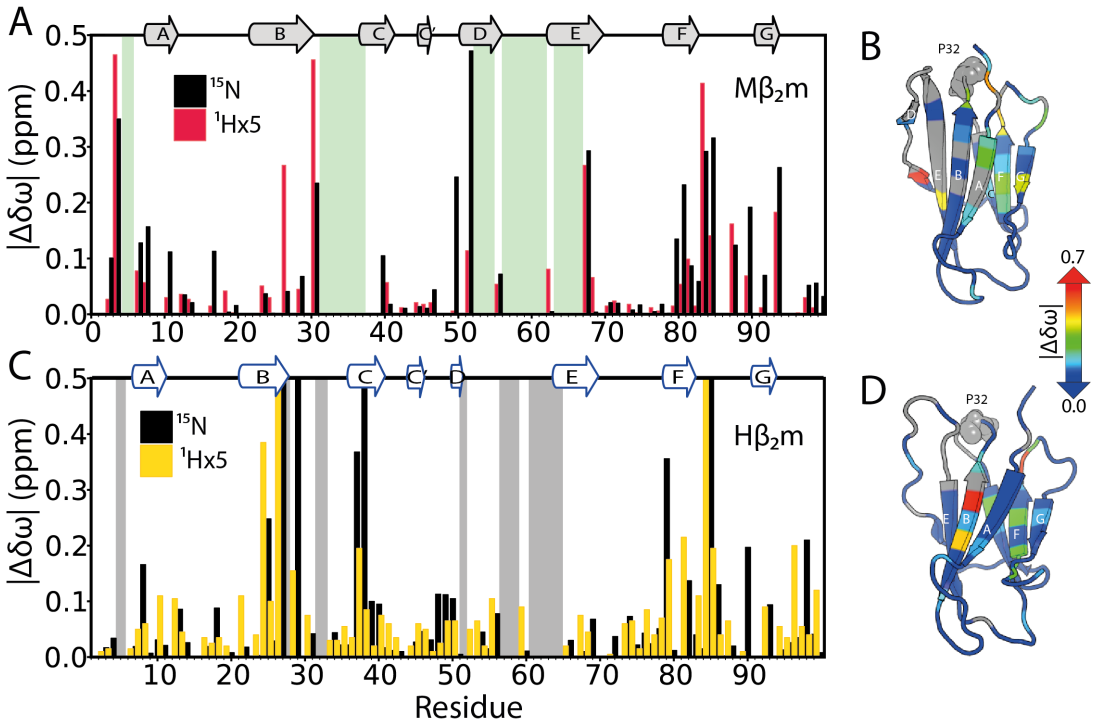

Figure 3.19: Similar structural properties for the human and murine I_T states. (A) ^{15}N (black) and $^{1}HX5$ (red) chemical shift differences between $m\beta_2m$ and its I_T state (spectrum obtained ~4min after refolding was initiated). Regions of increased line broadening are shown in green boxes. Redrawn from Figure 3.15. The chemical shift differences observed are mapped on the structure of $m\beta_2m$ in (B). Residues whose resonances are broadened beyond detection are shown in grey. Proline 32 is shown in spheres. (C) As in A, but for $h\beta_2m$ and the corresponding I_T state (spectrum obtained ~3min after refolding). (D) As in B but for $h\beta_2m$.

3.3 Discussion

The detailed NMR analysis of $m\beta_2m$ performed here revealed very similar properties to its human counterpart both in terms of structure (Figure 3.6) and dynamics (Figure 3.11, 3.12). As perhaps expected, both proteins show little response to pH by monitoring both their chemical shifts and ps-ns dynamics, in

marked contrast with ΔN6. The enhanced chemical exchange, which is linked to the high amyloidogenicity of ΔN6 at low pH (6.2), is absent for mβ$_2$m explaining its reduced amyloid potential under physiological conditions but do not distinguish it from hβ$_2$m, which shows the same behaviour. Additionally, mβ$_2$m was found to display the lowest thermodynamic stability as shown by the H/D exchange data supported by equilibrium unfolding experiments, in comparison to hβ$_2$m and ΔN6 and the lowest amyloidogenicity at the same time. This observation is in line with recent studies showing that thermodynamic stability is not connected with amyloid formation, but instead the protein has to overcome kinetic barriers that are most likely due to local unfolding/conformational changes on pathway to fibrils[389].

However, the structural characterisation of one of the least aggregation-prone β$_2$m variants (mβ$_2$m) performed here provides an opportunity to determine why the amyloidogenic properties of ΔN6 and/or hβ$_2$m, hβ$_2$m differ so substantially. In addition, it provides the possibility to determine features that control or modulate the amyloid-forming properties of these very similar proteins. Since chemical shifts represent one of the most sensitive structural probes, comparison of the chemical shifts of the three proteins allows a correlation between structure/dynamics and amyloidogenicity to be made. An extra control in this analysis is the non-amyloidogenic form of ΔN6 at pH 8.2 (Figure 3.20).

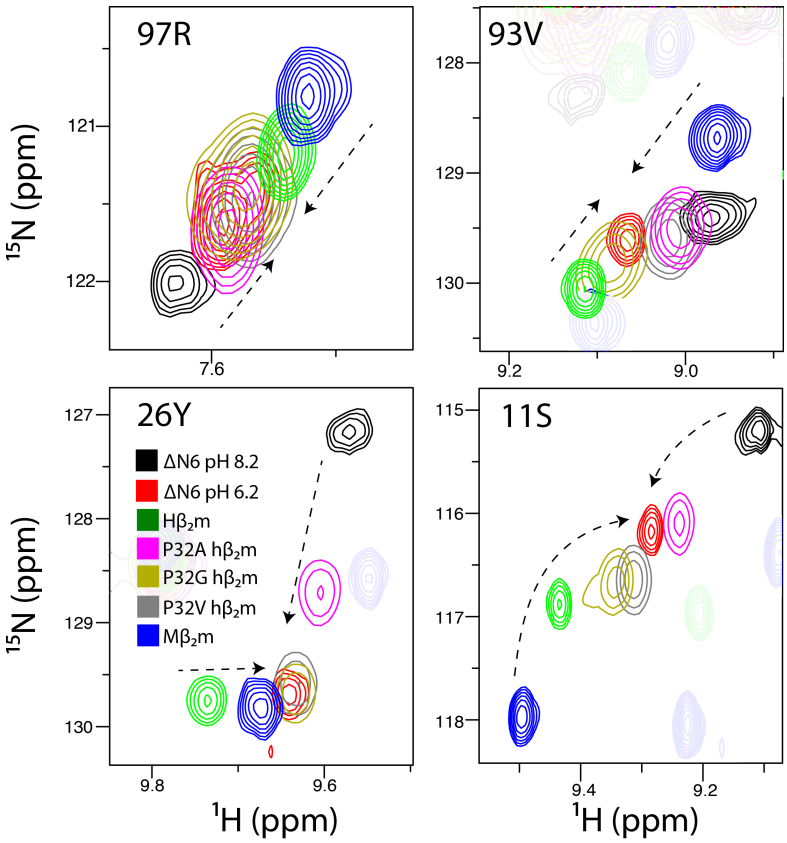

Figure 3.20: Zoom-in in regions of the ^{1}H-^{15}N HSQC spectrum of ΔN6, mβ$_2$m, hβ$_2$m and its variants. Chemical shift analysis of representative residues for 80μM mβ$_2$m pH 6.2 (blue), hβ$_2$m pH 6.2 (green), ΔN6 at pH 8.2(black) and ΔN6 in 10mM sodium phosphate pH 6.2, 25 ^{o}C (red). Resonances of 97R, 93V, 26Y and 11S are shown. The spectra of 500μM P32A hβ$_2$m (magenta), 500μM P32G hβ$_2$m (yellow), 500μM P32V hβ$_2$m (grey) at pH 7.5, 25^{o}C are also shown (recorded by Timo Eichner). Transitions to the most amyloidogenic state (ΔN6 at pH 6.2) are shown as dashed lines.

This analysis shows that the chemical shifts of ΔN6 at low pH (6.2) move towards their native - hβ$_2$m positions. As seen in Figure 3.20, the resonances of the highly amyloidogenic form of ΔN6 (at pH 6.2) lie (in the majority of cases) between those of the two non-amyloid-prone states (hβ$_2$m, mβ$_2$m or ΔN6 at pH 8.2). Interestingly, what this would suggest is that there could be two distinct non-amyloidogenic states for β$_2$m and transition to the more aggregation-competent species can occur from both directions. The cis-trans isomerisation at position 32 that leads to the reorganisation of the side chains in the hydrophobic core of the protein gives rise to pronounced chemical shift differences between hβ$_2$m and its I_T state (or ΔN6) as shown in Figure 1.20. Presumably, since these conformational changes are highly pH sensitive, at lower pH (6.2) the structure relaxes back to another intermediate (also

indicated by increased R_2 rates) which is more amyloidogenic in comparison with the species populated at pH 8.2, but contains some native-like signature in its structure, indicated by the similarity of the chemical shifts of ΔN6 at pH 6.2 with those of hβ_2m. In this scenario it is the combination of isomerisation at proline 32 and additional conformational changes downstream that are required to create a species that has increased amyloidogenicity. Accordingly, hβ_2m has an amyloidogenic signature already in its native structure/sequence that is released upon pro-isomerisation, while ΔN6 has gone through this rate limiting step, yet has to still change its conformation to form amyloid. In this context, mβ_2m is a mixture of both of these non-amyloid-prone states and as a consequence its chemical shifts lie closer to those of hβ_2m in some cases or closer to ΔN6 at pH 8.2 for others. Importantly, chemical shifts of other hβ_2m variants (including P32A, P32G, P32V) fall in the same lines as shown in Figure 3.20. In particular, according to previous studies, P32G was found able to elongate but not nucleate fibril assembly at neutral pH, with the elongation rate being directly correlated with the population of the I_T state for this mutant[357]. In contrast, P32V lacks the ability to nucleate and elongate amyloid fibrils[353], while P32A is proposed to be highly amyloidogenic when incubated with Cu^{++}[399].

This analysis, if correct could be useful in order to predict the amyloidogenicity of a given mutant by just looking at its chemical shifts. Of course and since the chemical shift differences between the different proteins are not linear, such a transition is most likely more complex than 2-state and involves more than one distinct molecular processes, that are inter-linked to each other.

Here, mβ_2m was found to fold through an I_T state similar in structure to that of hβ_2m, and that this species folds to the native state at a significantly slower rate (5-fold) than its human counterpart. This observation could suggest that the equilibrium position between the native and the I_T state is shifted towards the native state, resulting in very low populations of I_T for mβ_2m at physiological conditions. This supposition is supported by the absence of residues displaying ns-ms dynamics of the murine protein at neutral pH, measured by the T_2 relaxation rates of backbone amides. However, measurement of the unfolding

rates is required to confirm this view. Furthermore, and if the folding pathway of $m\beta_2m$ is similar to the one proposed for $h\beta_2m$, the protein should be able to access the amyloidogenic I_T much easier from the unfolded state. However, $m\beta_2m$ is not able to form fibrils even when it is acid unfolded, showing that further conformational changes on top of the structural changes caused by the proline isomerisation are needed for the protein to reach an amyloid-competent species.

The characterisation of the structure, dynamics and folding pathway of $m\beta_2m$ performed here, revealed that it lacks the dynamic properties of ΔN6 that make the latter an aggregation-prone species, even though it folds through an intermediate which possess similar structural properties as ΔN6. However, one of the key properties of human I_T, and consequently ΔN6 monomers, is their ability to interact with each other and oligomerise, suggesting that their increased amyloidogenicity is encrypted in their biomolecular interactions and the effect they have on the aggregation pathway. Chapters IV and V will address this hypothesis using a variety of biophysical tools.

CHAPTER IV

PROTEIN-PROTEIN INTERACTIONS DEFINE THE COURSE OF AMYLOID FORMATION

4.1 Introduction

The value of protein-protein interactions in defining protein function in general, and more specifically in protein aggregation, was emphasized in Chapter I. Protein-protein interactions are at the origin of most mechanistic models for amyloid assembly[258,263,265], but at the same time their nature has prevented detailed characterisation to date. In order to delay or diminish the onset of disease and prevent the appearance of the pathological symptoms, different strategies have been applied over the years[289]. These include: 1) stabilization of the non-amyloidogenic native state, a strategy that was successful in the case of human transthyretin and lead to the design of the first anti-amyloid drug in clinical trials[400], 2) strong binding to non-aggregation prone proteins (such as antibodies) that have the ability to disaggregate oligomers or protofibrils, as shown in the case of $A\beta_{40}$[401], 3) the design of modified versions of aggregating peptides that can bind to the elongating cross-β structure but incorporate an entity (such as a non-natural amino acid) that blocks further polymerisation (β-blockers)[402] or 4) the discovery of small molecules that interfere with assembly and/or cytotoxicity[292,294,296]. Even nature itself seems to have adopted a similar strategy by modulating protein-protein interactions in order to inhibit prion propagation in different fungal strains[364,403,404].

B_2m as a model system offers a remarkable opportunity to study how protein-protein interactions affect the course of amyloid assembly. Previous work has shown that biomolecular interaction between ΔN6 and $h\beta_2m$, even in substoichiometric ratios, converts the latter in an aggregation-prone state while the ΔN6-$m\beta_2m$ interaction inhibits amyloid assembly (Figure 1.22)[272]. In this Chapter, the structural properties of ΔN6 that lead to its increased amyloidogenicity are investigated. NMR studies provided a link between conformational dynamics and amyloid potential and also highlighted the

importance of investigating the interaction between ΔN6 monomers in the early stages of amyloid assembly. Furthermore, the efficiency of inhibition of ΔN6 amyloid assembly by mβ$_2$m is addressed here. An investigation of the mechanism of inhibition was carried out by applying a variety of biophysical techniques. Detailed analysis allowed the elucidation of mechanistic details on inhibition of amyloid formation by mβ$_2$m, which is able to redistribute the equilibrium between the oligomeric species established in the early stages of aggregation.

4.2 Results

4.2.1 Investigating the link between conformational dynamics and aggregation propensity

Even though ΔN6 constitutes one of the most amyloidogenic hβ$_2$m variants known to date, it takes more than 25 days for it to self-assemble *in vitro* into amyloid fibrils at neutral pH (Figure 1.21). In order to facilitate kinetic and structural characterisation of the first events in the nucleation mechanism, the protocol used to follow the aggregation of ΔN6, was recently updated by increasing the level of agitation, to allow for much shorter incubation periods[405]. Using this modified protocol and the same buffer used in all the NMR studies, ΔN6 was found able to form amyloid fibrils in ≤30h from the initiation of the reaction (Figure 4.1). All reactions were carried out at pH 6.2, conditions under which ΔN6 shows increased amyloidogenicity.

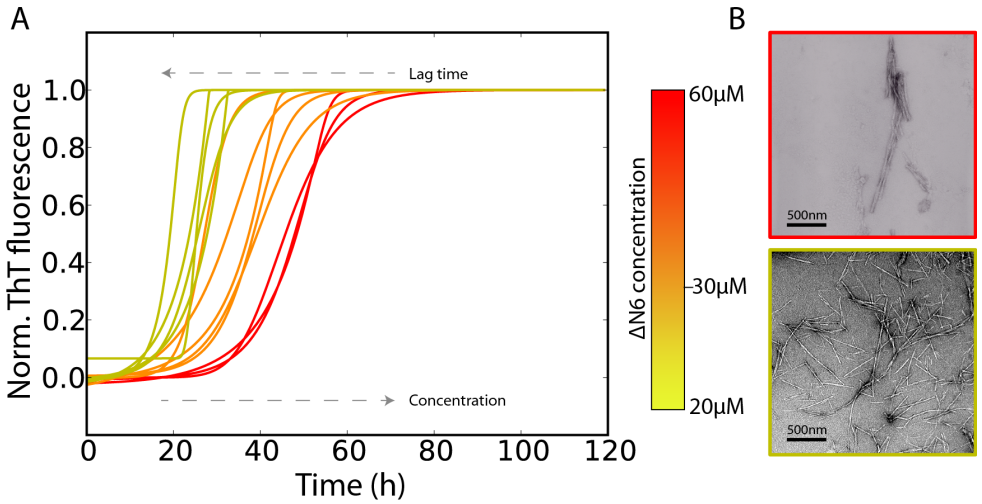

Figure 4.1: Aggregation kinetics of ΔN6 at pH 6.2.

(A) The aggregation of ΔN6 in 10mM sodium phosphate pH 6.2, 83.3mM NaCl, 37°C, 600rpm was followed by ThT fluorescence at different protein concentrations (colourbar). The curves shown represent fits of the generalised form of the sigmoid function (Section 2.3.4) to the raw data. Note that the lag time is reduced to ~30h rather than 25 days as shown in Figure 1.21. (B) Electron micrographs of the end point of the reaction (after (120h) of 60μM (top) and 20μM ΔN6.

The data shown in Figure 4.1 revealed a reverse concentration-dependence for the aggregation rate of ΔN6 under the conditions employed. Normally, for most aggregating proteins, increasing protein concentration results in a decreased lag time, since the possibility of productive protein-protein interactions is increased. However, this phenomenon is reversed in the case of ΔN6, suggesting the presence of an ensemble of interactions some of which are on pathway, leading to the formation of the critical nucleus and some being off-pathway, effectively reducing the concentration of monomers available, and thus delaying the formation of amyloid. A similar phenomenon has been observed for the aggregation of immunoglobulin light chains[406].

Importantly, the pH-dependence of the reaction is still preserved using the new protocol, with no fibrils formed at pH 8.2 but rapid fibril formation at pH 6.2 (Figure 4.2A-B). The enhanced line broadening observed in NMR spectra of ΔN6 at pH 6.2, in contrast with sharp linewidths at pH 8.2, is suggestive of chemical exchange phenomena (Figure 4.2C-D). The observation that the linewidth of ΔN6 resonances at pH 6.2 is also concentration-dependent (Figure 4.3) reveals an increased oligomerisation propensity for the protein under these conditions. Importantly, this phenomenon is absent in hβ_2m and mβ_2m at low pH (6.2), suggesting an origin for the increased amyloidogenicity of ΔN6 at pH 6.2 (Figure 4.2E-H).

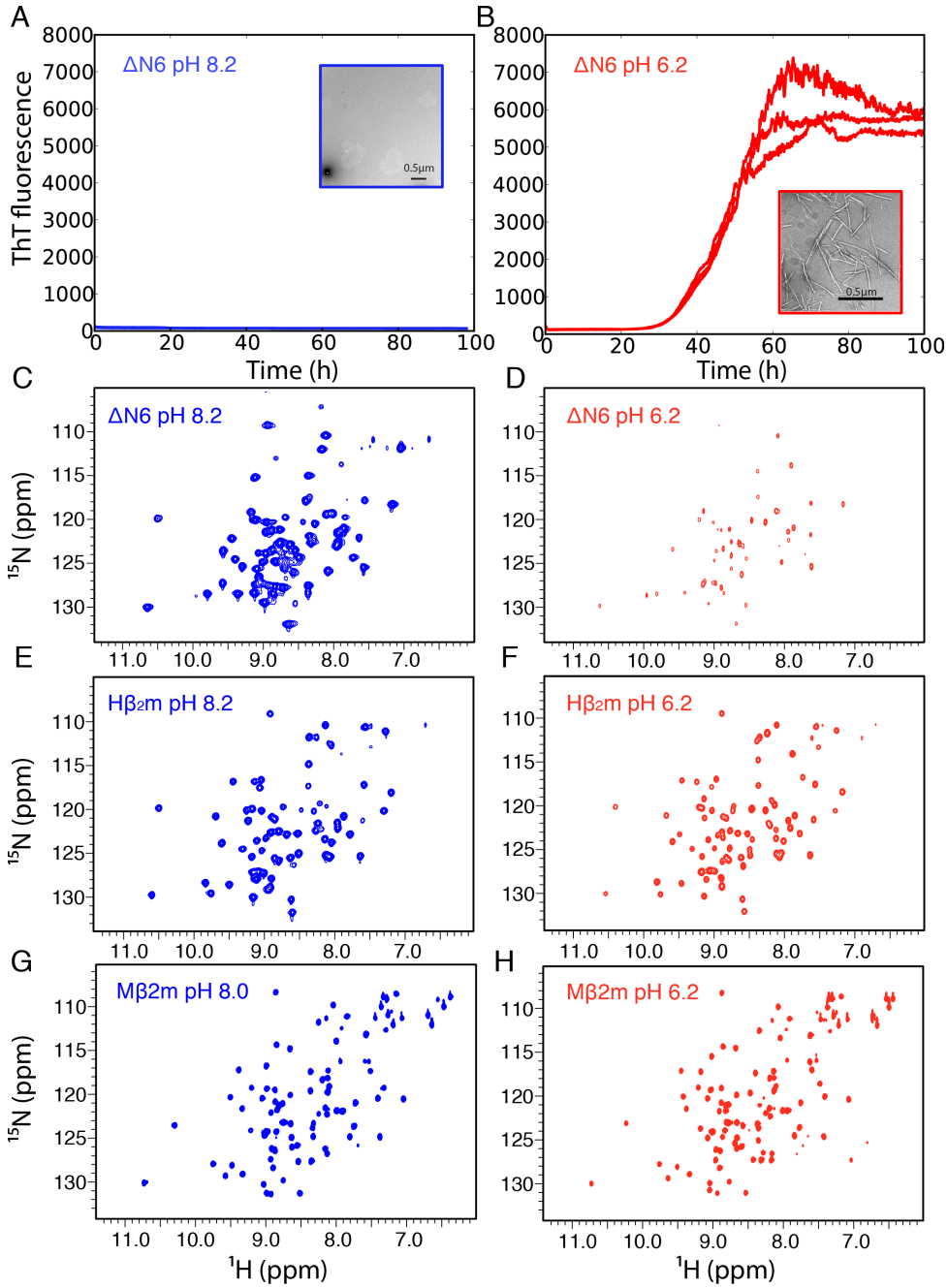

Figure 4.2: pH-dependent aggregation and dynamics.
Aggregation kinetics of 60μM ΔN6 in 10mM sodium phosphate buffer pH 8.2, 86.6mM NaCl (A) or pH 6.2, 83.3mM NaCl (B), 37°C. Electron micrographs of the end point are shown as insets. 500MHz ^{1}H-^{15}N HSQC spectra of 600μM ΔN6 in 10mM sodium phosphate at pH 8.2, 86.6mM NaCl (C) or pH 6.2, 83.3mM NaCl (D). The increased line broadening at pH 6.2 is also concentration-dependent (Figure 4.3), strongly pointing to the presence of high order species at low pH. The 500MHz ^{1}H-^{15}N HSQC spectra of 600μM hβ$_2$m in the same buffer and at pH 8.2 or 6.2 are shown in E and F respectively. G and H show the 750MHz ^{1}H-^{15}N HSQC spectra of 600μM mβ$_2$m in the same buffer and at pH 8.0 or 6.2 respectively. In C,D, E,F and G, H the contour levels were set to the same value to allow direct comparison. Lowering the protein concentration and increasing the number of scans results in spectra of higher quality for ΔN6 at pH 6.2.

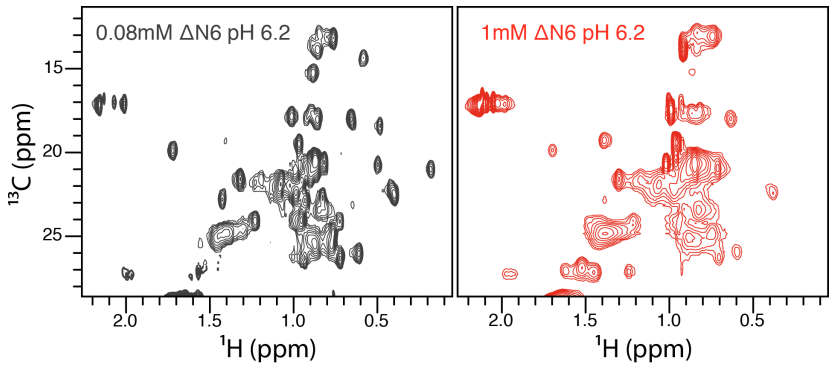

Figure 4.3: Concentration-dependent line broadening for ΔN6.
Zoom-in in the methyl region of the ^{1}H-^{13}C HSQC spectrum of 80μM ΔN6 in 10mM sodium phosphate pH 6.2, 83.3mM NaCl (grey) or 1mM ΔN6 under the same conditions (red).

Conventional NMR spectroscopy suffers from the fast relaxation times of protons resulting, in many cases, in pronounced line broadening in the direct dimension (^{1}H). Therefore phenomena such as exchange with solvent, conformational changes and/or partial unfolding, which are often physiologically relevant, can severely reduce the quality of NMR spectra. Recent technological advances have allowed the direct detection of carbon nuclei and thus all the problems related to the fast relaxation times of protons can be overcome[371,407-409]. Here, proton-less NMR[410] was used in an effort to gain more structural information on the molecular events that are responsible for the enhancement of the amyloidogenicity of ΔN6 at pH 6.2. ^{13}Cα-^{13}CO correlation spectra acquired at pH 8.2 (Figure 4.4A) show sharp lines and good dispersion, while lowering the pH to 7.2 results in enhanced line broadening (Figure 4.4B). This phenomenon was even more pronounced at pH 6.2, where the acquisition of carbon-detected NMR spectra was infeasible in a reasonable timeframe (48h, data not shown). Furthermore, pH-dependent chemical shift differences located mainly in the DE and BC loops and the N-terminal region were observed, in line with the data shown in Figure 3.8 (Figure 4.4C,D).

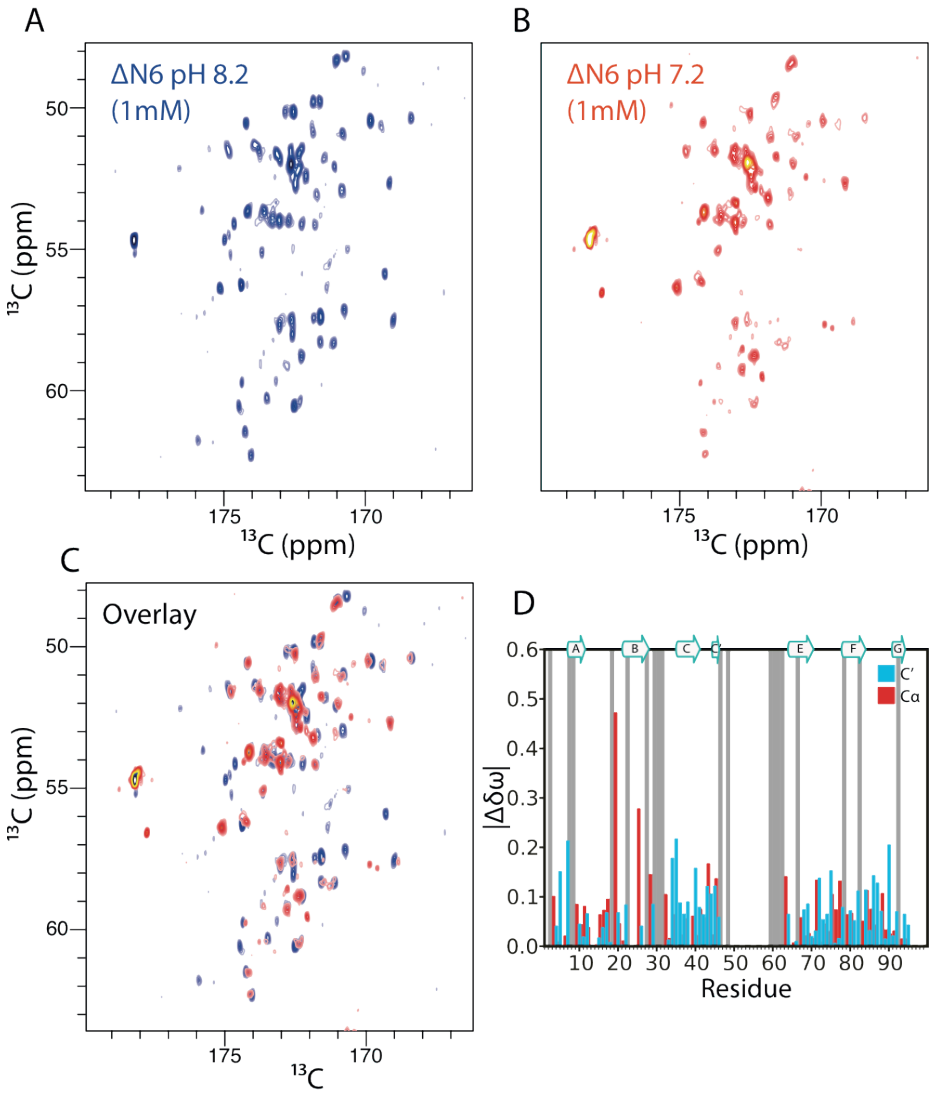

Figure 4.4: Direct carbon detection NMR.
900MHz ^{13}Cα-^{13}C' spectra of 1mM ΔN6 in 10mM sodium phosphate buffer pH 8.2 (A) or pH 7.2. An overlay of the spectra acquired is shown in C. Cα (red) and C' (cyan) chemical shift differences (absolute values) at pH 8.2 versus pH 7.2 are shown in D. Grey bars indicate residues that either show large chemical shift differences and thus assignments at pH 7.2 are ambiguous, or are broadened beyond detection. The secondary elements of ΔN6 are shown on top of the panel. All spectra were acquired in CERM, Florence, Italy.

These data showed that the chemical exchange that takes place at lower pH is so enhanced that even the use of direct carbon detection NMR is problematic. At the high protein concentrations (1mM) required for carbon detection, exchange to high molecular weight species is likely to be prevalent, making their direct characterisation challenging for NMR.

Further investigation of the exchange events that are connected with amyloid aggregation was performed by relaxation dispersion NMR (CPMG, Section 1.5.4.1). In these experiments exchange from the native state to alternative protein states was monitored under different conditions (varying pH) in an effort to kinetically and structurally characterise the 'invisible' protein states that, as indicated from the T_2 relaxation data shown in Figure 1.21, seem to be responsible for the increased amyloidogenicity of ΔN6 at pH 6.2. At basic pH (8.2), where ΔN6 is not prone to aggregation, most residues whose resonances were found to exchange to alternative states locate in the ABED β-sheet (Figure 4.5A, C, E). On the other hand, at pH 7.2 residues that show significant dispersion are spread throughout the structure of ΔN6, while the exchange contribution to the R_2 rate (R_{ex}) seems to be reduced, whilst R_2 itself (plateau) increases on average (Figure 4.5B, D, F). Under conditions in which ΔN6 is most amyloidogenic (pH 6.2) the severe line broadening shown in Figure 4.2D did not allow accurate characterisation of the underlying exchange processes. Increased oligomerisation at lower pH (7.2 or 6.2) could possibly explain the experimental data.

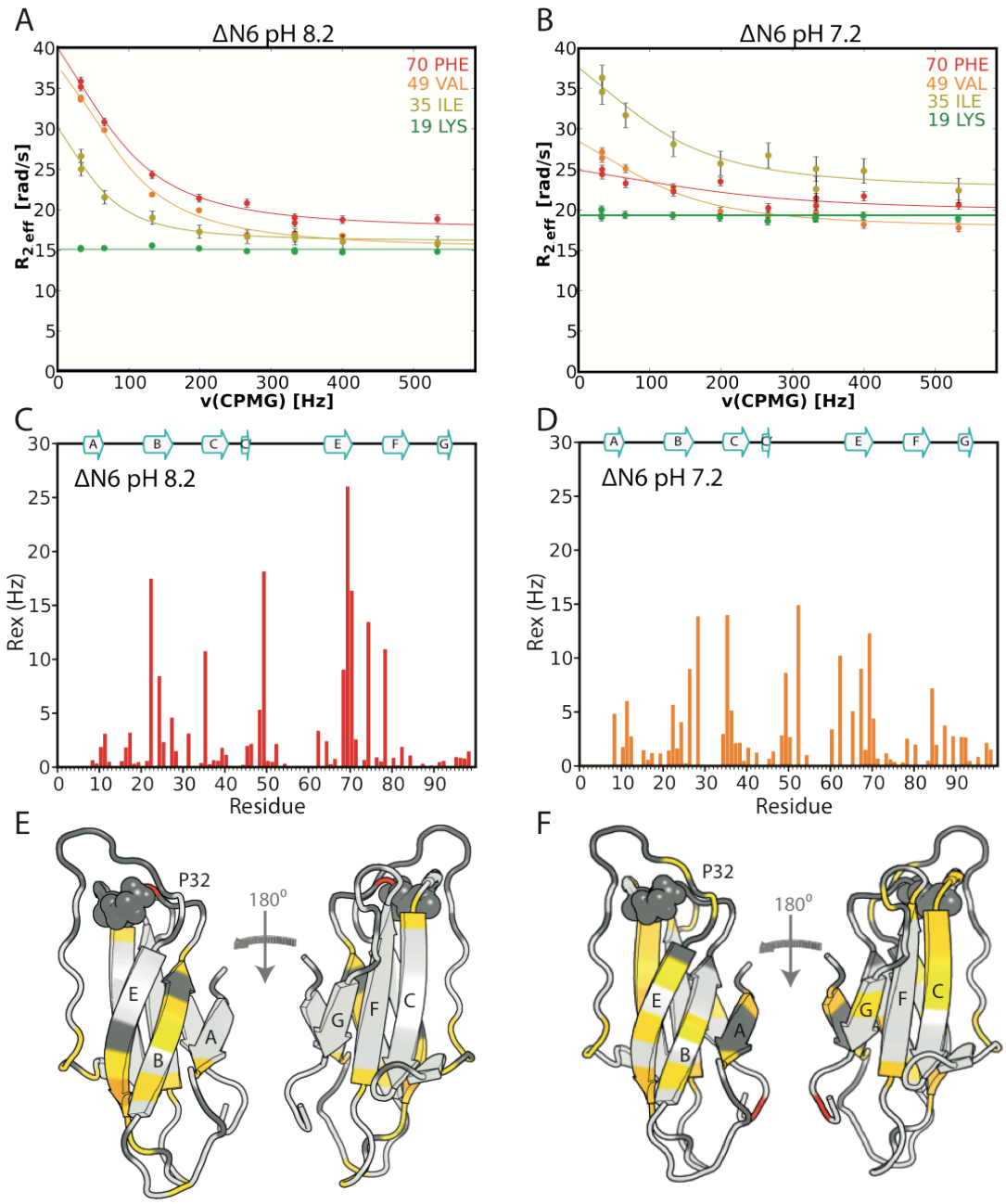

Figure 4.5: Relaxation dispersion NMR for ΔN6 at pH 8.2 or 7.2.
CPMG profiles for residues 19K, 35I, 49V, 70F at pH 8.2 (A) and 7.2 (B). Solid lines represent individual fits to the McConell equations[178]. A per residue histogram of R_{ex} defined as $R_{ex} = R_{2,eff}(vCPMG=0) - R_{2,eff}(vCPMG=533)$ at pH 8.2 is shown in C while the same plot at pH 7.2 is shown in D. The secondary elements of ΔN6 are shown on top of the panels. Residues that are fitted to a 2-state exchange model at pH 8.2 and pH 7.2 are mapped on the structure of ΔN6 in E and F respectively. Residues that are broadened beyond detection, or whose resonances overlap or have missing assignments are shown in dark grey. Proline 32 is shown in spheres. The protein concentration and the buffer used in all experiments was 600μM and 10mM sodium phosphate respectively.

Importantly the non-amylodoigenic hβ2m experiences reduced chemical exchange under all conditions explored when compared with ΔN6.

Interestingly the equilibrium between protein sub-states of hβ$_2$m is not sensitive to pH, in opposition to the results obtained for ΔN6 (Figure 4.6).

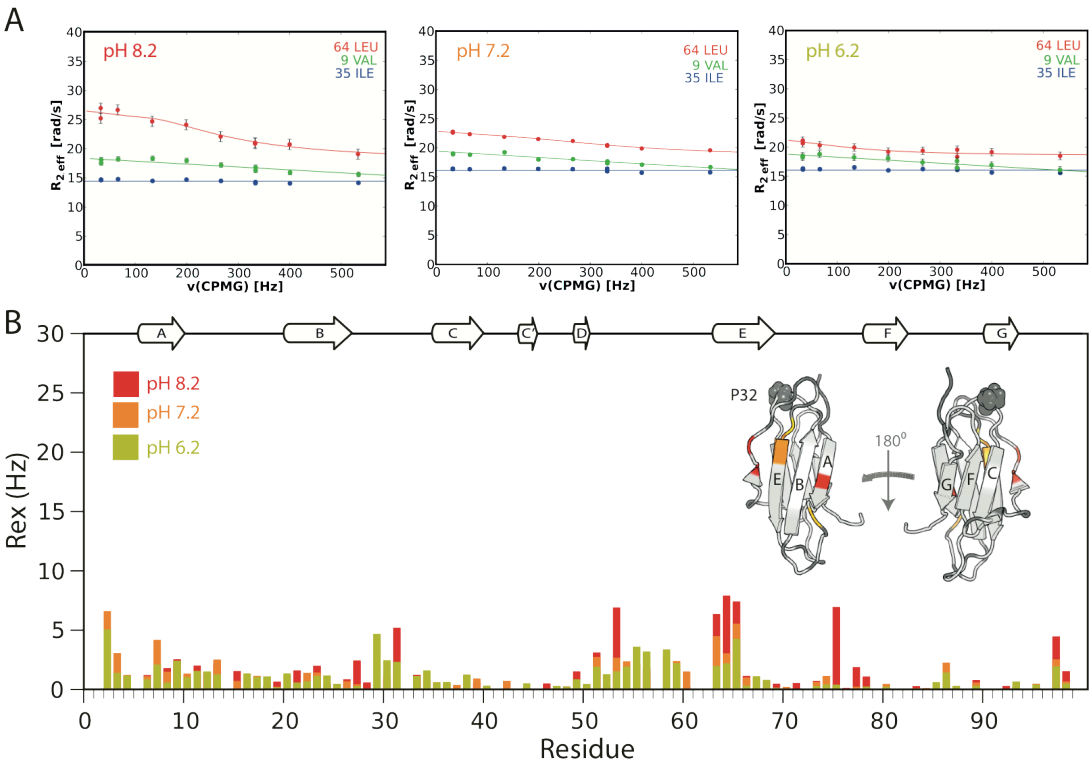

Figure 4.6: Relaxation dispersion NMR for hβ$_2$m at pH 8.2, 7.2 or 6.2. (A) CPMG profiles for residues 9V, 35I, 64L at pH 8.2 (left), 7.2 (middle) and 6.2 (right) of hβ$_2$m. Solid lines represent individual fits to the McConell equations[178]. (B) Histograms of R_{ex} defined as $R_{ex} = R_{2,eff}(vCPMG=0) - R_{2,eff}(vCPMG=533)$ at pH 8.2 (red), pH 7.2 (orange) or pH 6.2 (yellow). The secondary structural elements of hβ$_2$m are shown on top of the panels. The structure of hβ$_2$m coloured based on residues that undergo 2-state exchange (pH 8.2) according to the fitting program is shown as inset. Residues that are broadened beyond detection, or whose resonances overlap or have missing assignments are shown in dark grey. Proline 32 is shown in spheres. The protein concentration used in all experiments was 600μM and the buffer was 10mM sodium phosphate.

CPMG experiments allow the characterisation of the exchange event both in structural and kinetic terms, however in this case the mechanism involved is likely to be more complex than 2-state, complicating the data analysis. At high protein concentrations, the dispersion profiles observed may arise not only from intramolecular dynamics but also from the biomolecular collision between monomers or, most likely, due to a combination of both processes. Therefore further quantitative analysis of the CPMG data was not performed. However, even in a qualitative way these results link the increased conformational dynamics of ΔN6 at pH 6.2 and 7.2 to its increased

amyloidogenicity and highlight the importance of studying the molecular events responsible for this phenomenon using techniques sensitive enough to work with lower protein concentrations to simplify the interpretation of the data. The increased overall R_2 rates (defined as the plateau value of $R_{2,eff}$) for ΔN6 at pH 7.2 (Figure 4.5B) and also the observation that the residues showing significant dispersion at this pH are spread throughout the molecule (Figure 4.5F) point to a pH-dependent oligomerisation event (in line with data shown in Figures 4.2, 4.3, 4.4) that possibly affects the conformational properties of monomeric ΔN6, under the same conditions $h\beta_2m$ shows reduced dynamics and/or propensity to oliogmerise.

4.2.2 Substoichiometric inhibition and promotion of amyloid assembly

4.2.2.1 How efficient an inhibitor of aggregation is $m\beta_2m$?

The data presented in Figure 1.22 showed that $m\beta_2m$ acts as an inhibitor of the aggregation reaction of ΔN6 at neutral pH. However, since the affinity and the stoichiometry of the interaction between the two proteins remained unknown, the efficiency of this inhibitory effect could not be addressed. Whether $m\beta_2m$ can inhibit the aggregation of ΔN6 when the two proteins are mixed with a less than 1:1 molar ratio (excess of ΔN6), reminiscent of the finding that only one molecule of ΔN6 per 100 molecules of $h\beta_2m$ is sufficient to promote amyloid assembly, was therefore investigated. Proteins were mixed in different combinations and in substoichiometric ratios (less ΔN6 than $h\beta_2m$ or less $m\beta_2m$ than ΔN6) and the aggregation reaction was monitored by ThT fluorescence and negative stain EM (Figure 4.7).

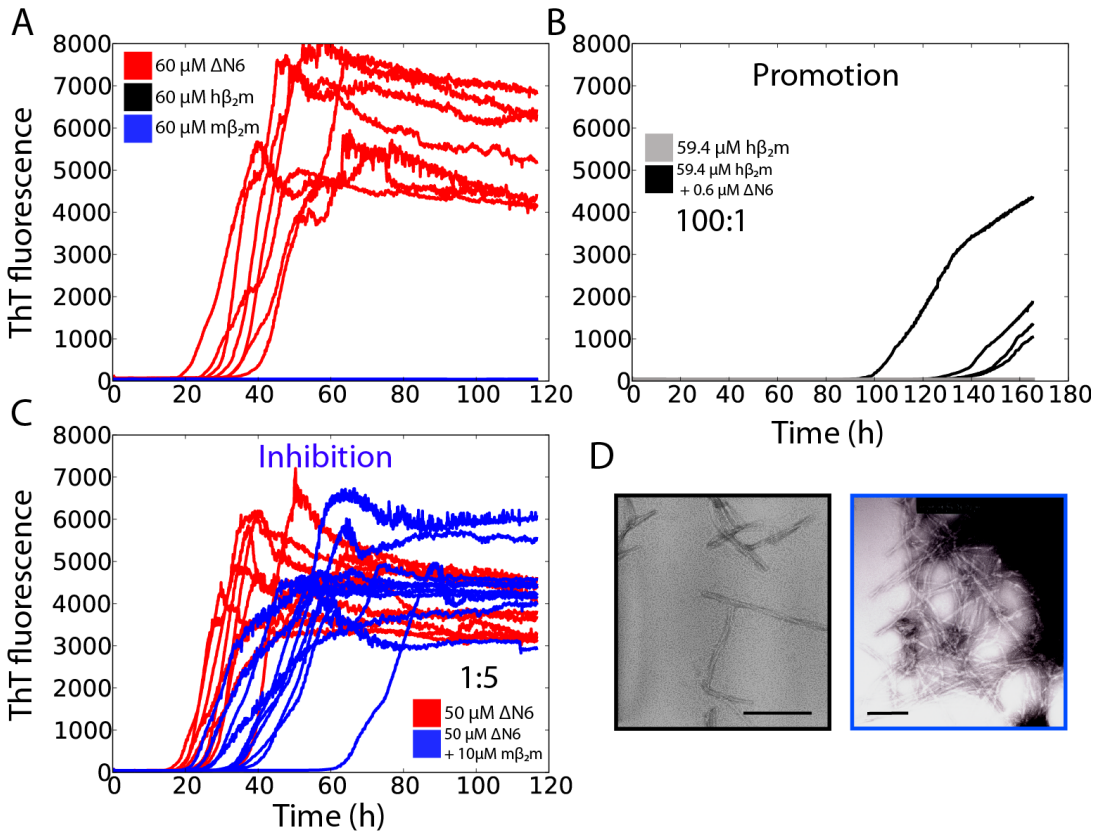

Figure 4.7: Substoichiometric promotion and inhibition of amyloid assembly. (A) Kinetics of the aggregation of 60μM ΔN6 (red traces), 60μM hβ2m (black traces), or 60μM mβ2m (blue traces) followed by ThT fluorescence at 37°C, 600rpm. (B) When hβ2m and ΔN6 are mixed in a 100:1 (59.4μM hβ2m:0.6μM ΔN6) ratio amyloid formation readily proceeds at pH 6.2 (black traces), while hβ2m alone shows no ThT signal (grey). (C) Addition of 10μM of mβ2m into 50μM ΔN6 causes a significant retardation of amyloid formation (blue traces) in comparison to 50μM ΔN6 alone (red traces in C) or 60μM ΔN6 (red traces in A). (D) Electron micrographs of 59.4μM hβ2m:0.6μM ΔN6 after 160h (black traces in B) –left, or 50μM ΔN6:10μM mβ2m after 120h (blue traces in C)-right. The buffer used for all experiments was 10mM sodium phosphate pH 6.2, 83.3mM NaCl, 0.02% (w/v) NaN_3. Bar represents 200nm.

Interestingly, small amounts of mβ2m were sufficient to significantly prolong the lag time of aggregation of ΔN6 showing that mβ2m is a potent inhibitor of amyloid assembly. Following that observation, an analysis of the effect of mβ2m on the aggregation of ΔN6 was performed over a wide range of concentrations, using ThT fluorescence and electron microscopy, in order to investigate the inhibitory ΔN6-mβ2m interaction in more detail. In these experiments the total protein concentration was kept constant at 60μM in order to allow direct comparison between the kinetics of the same amount of ΔN6, mixed with itself or mβ2m or nothing. For instance mixtures of 30μM ΔN6

with 30μM mβ$_2$m are expected to show an increased lag time in comparison to 30μM ΔN6 alone. However, data in Figure 4.1 show that 60μM ΔN6 aggregates with a longer lag time than 30μM ΔN6 alone and thus real inhibition would occur only if 30μM ΔN6 + 30μM mβ$_2$m show a longer lag time than 30μM ΔN6+30μM ΔN6. The full dataset obtained is shown in Figure 4.8 and a summary of the data in Figure 4.9.

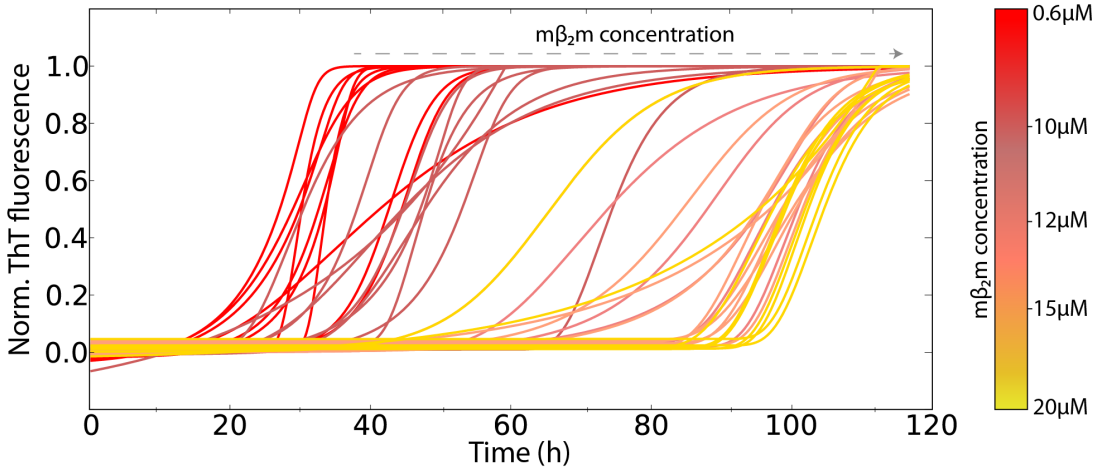

Figure 4.8: Increasing concentrations of mβ$_2$m increase the lag time of ΔN6 fibril assembly. Increasing amounts of mβ$_2$m (colourbar) were mixed with the equivalent amount of ΔN6 to give a total protein concentration of 60μM and the progress of amyloid formation was monitored by ThT fluorescence in 10mM sodium phosphate buffer pH 6.2, 83.3mM NaCl. Raw ThT data were normalised to a final signal of 1.0 and fitted to a sigmoidal function (see Section 2.3.4). Only the fitted curves are shown here for clarity. Mixtures with a molar ratio of 1:1 or excess of mβ$_2$m:ΔN6 did not show any significant increase in ThT fluorescence and are not shown (see Figure 4.9).

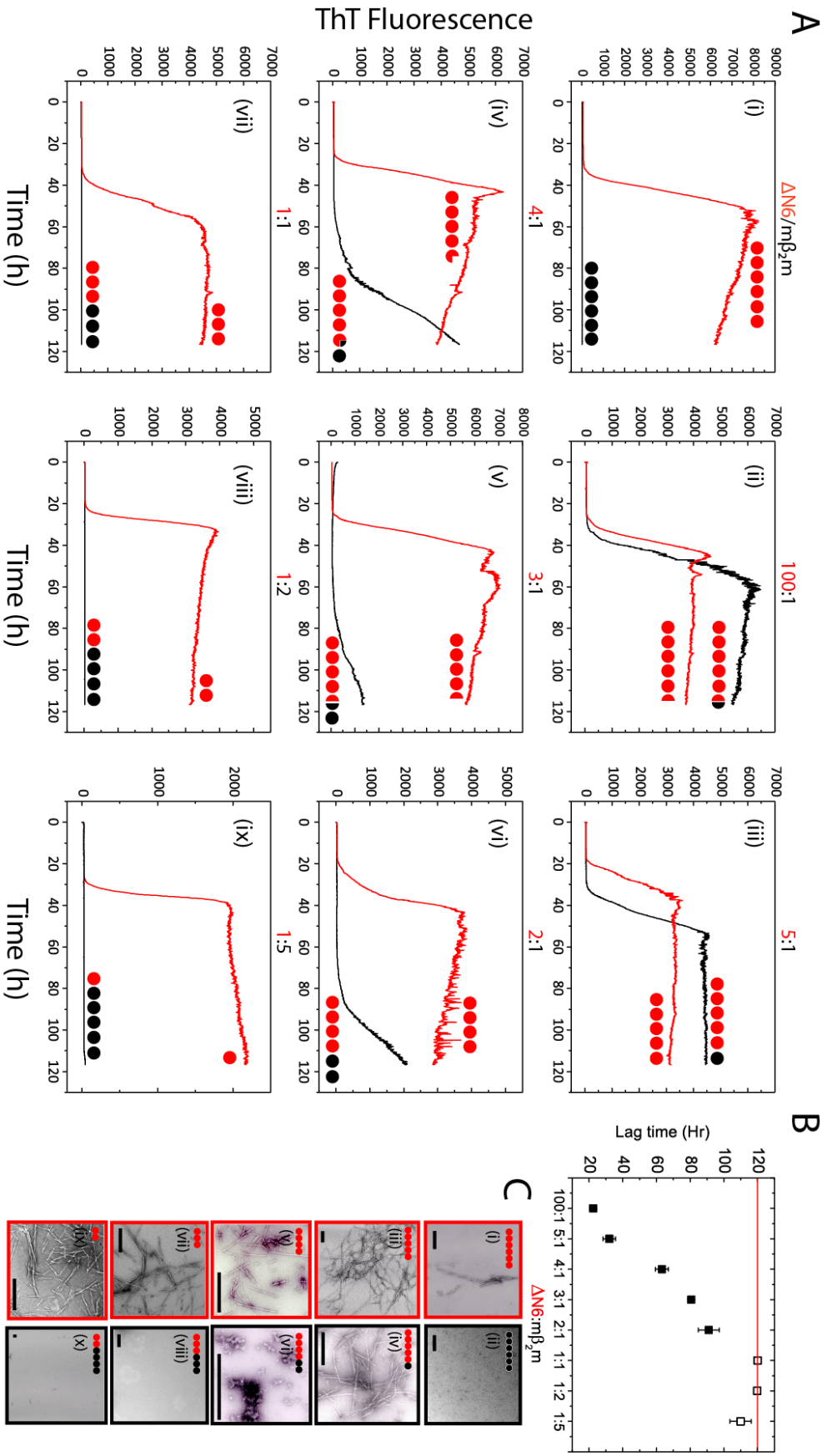

Figure 4.9: Inhibition of ΔN6 fibril formation by protein-protein interactions. (A) Aggregation kinetics of ΔN6 alone (red traces) or ΔN6 mixed with mβ2m in different molar ratios (ΔN6:mβ2m) measured using ThT

fluorescence. The trace with the median lag time of at least 10 replicates is shown. Black spheres and red spheres each represent 10μM of $m\beta_2m$ or 10μM of ΔN6, respectively. Each panel shows the aggregation kinetics of the protein mixture in black and the kinetics of the equivalent concentration of ΔN6 (alone) in red. The total protein concentration for all of the protein mixtures is 60μM (the reader should relate all black traces to the control of 60μM ΔN6 shown in the first panel). (B) Diagram of the lag time of fibril formation for different molar ratios of ΔN6 and $m\beta_2m$. Open symbols represent experiments where the lag time could not be estimated because the protein mixtures did not show an increase in ThT fluorescence by the end of the experiment (120h) (red line). Error represents SEM. (C) Negative stain electron micrographs of the end point of the reaction (after 120h) for the equivalent traces in A, highlighted by the red and black spheres. Electron micrographs of ΔN6 alone are highlighted in red boxes, while mixtures of ΔN6-$m\beta_2m$ are highlighted in black boxes. The buffer used for all experiments was 10mM sodium phosphate pH 6.2, 83.3mM NaCl. Scale bar denotes 500nm.

The data in Figures 4.8 and 4.9 show that when ΔN6 and $m\beta_2m$ are mixed in substoichiometric ratios, increasing amounts of $m\beta_2m$ cause a prolonged lag time of fibril assembly (Figure 4.9, panels ii-iv), while equimolar concentrations of ΔN6 and $m\beta_2m$ or an excess of the latter cause complete inhibition of amyloid formation over the 120h time-course followed here (Figure 4.9, panels vii-ix). Therefore, $m\beta_2m$ is a potent inhibitor of ΔN6 aggregation given its ability to stop the aggregation reaction even in low doses, a fact that is often not observed even for small molecule inhibitors of amyloid assembly[294]. Identifying the surfaces involved in the biomolecular recognition of ΔN6 by $m\beta_2m$ is thus of crucial importance and could pave the way to the discovery of small molecule inhibitors of aggregation. Data presented in Chapter V are trying to answer some of these questions.

4.2.2.2 Does $m\beta_2m$ redirect the aggregation pathway of ΔN6?

Since $m\beta_2m$ has such a drastic effect on the amyloid formation by ΔN6, we wondered whether the biomolecular interaction between the two proteins in the early stages of aggregation is able to alter the pathway of aggregation or even remodel the species associated with it. Data presented in Figure 4.9C (panel iv) provide some first clues on this question. In the middle of the elongation phase (after 120h), in the presence of $m\beta_2m$ (ΔN6:$m\beta_2m$ 2:1), proteinaceous oliogmeric species with a diameter of 50nm were observed, that

were not detected when ΔN6 or mβ$_2$m were incubated alone (Figure 4.9C-panels i,ii). Views of the oligomeric species detected at different magnifications are shown in Figure 4.10.

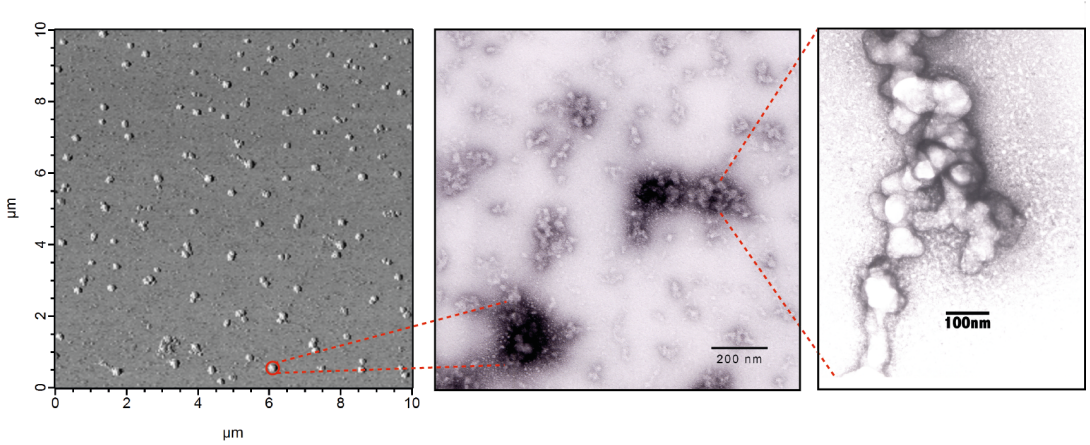

Figure 4.10: Mβ$_2$m-induced oligomers.
Images of oligomers detected in a sample comprising of 40µM ΔN6 and 20µM mβ$_2$m after 120h of incubation at 37^{o}C, using AFM (phase mode-left), or negative stain EM (middle and right) at different magnifications. The corresponding ThT trace is shown in black in Figure 4.9A, panel vi.

However, since the amyloid forming reaction has not proceeded to completion after 120h under the conditions employed (still in the elongation phase-Figure 4.9A, panel vi), the oligomeric species shown in Figure 4.10 might represent on-pathway species formed by ΔN6 that convert into amyloid-like fibrils later on in the aggregation reaction. To investigate this possibility a time-course experiment was performed, in which the aggregation of ΔN6 was monitored by ThT fluorescence and negative stain EM at different time points during assembly (Figure 4.11).

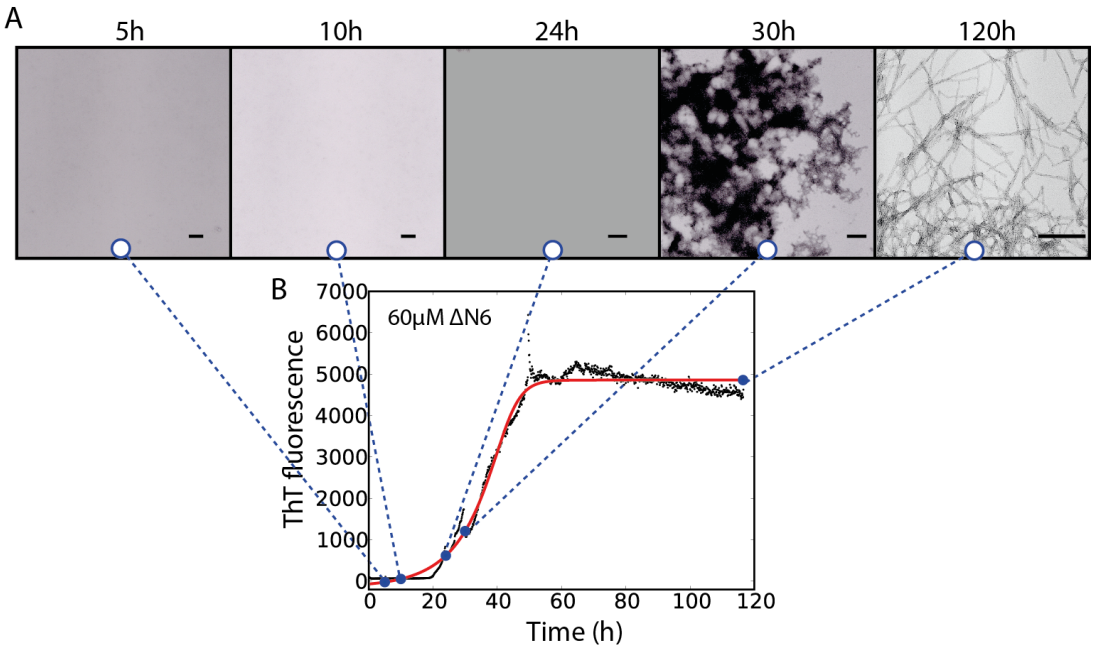

Figure 4.11: Time-course of ΔN6 fibril assembly.
Negative stain electron micrographs taken at different time-points (indicated on top of the panels) during the aggregation of 60μM ΔN6 in 10mM sodium phosphate pH 6.2, 83.3mM NaCl, 37°C, 600rpm (A). The corresponding ThT kinetics of the same well used to make the EM samples is shown in B. The raw data are shown as black dots, the fit to the data as red line and the time-points used for EM as blue dots. Bar represents 200nm.

In these experiments ΔN6 was found to aggregate through the formation of large dense aggregates (30h in Figure 4.11), which convert to amyloid fibrils during or after the elongation phase of fibril assembly. Importantly, the same phenomenon has been observed before for ΔN6 using different incubation conditions (50mM sodium acetate pH 5.0, quiescently)[363]. Interestingly, this aggregated form of ΔN6 does not have the spherical appearance of the species generated when mβ$_2$m is present (Figure 4.10) and therefore it is unlikely that they share the same underlying molecular signature. An intriguing possibility that emerges is that the spherical oligomers represent off-pathway kinetically trapped species and that the ΔN6:mβ$_2$m interaction is able to redirect or create alternative branches in the aggregation pathway of ΔN6, reducing the effective concentration of species readily available to form aggregation-prone entities. If the oligomers observed are off-pathway, they should have abolished the ability to seed fresh monomeric ΔN6. In order to test this hypothesis, the higher molecular weight species were separated by centrifugation at 14000rpm, for 5 min and were used as 10% *(v/v)* seeds for elongating fresh

ΔN6 monomer. The results in Figure 4.12 revealed that the oligomers observed after 120h of incubation of 40μM ΔN6 and 20μM mβ$_2$m diminish the lag time of fibril formation, which now proceeds readily at pH 6.2 (red in Figure 4.12). Importantly the oligomers have the same seeding capacity as fibril seeds made by ΔN6 alone (grey in Figure 4.12). Notably, seeds made by the incubation of 20μM ΔN6 and 40μM mβ$_2$m for 120h do not seed the reaction, which proceeds with the same lag time as 20μM ΔN6 alone (cyan in Figure 4.12). This finding is consistent with the lack of increased ThT fluorescence (Figure 4.9A, panel viii) and the absence of any high molecular weight species by EM (Figure 4.9B, panel x) for this mixture of proteins. Even though this experiment is not calibrated so the same amount of seeds are added in each case, the results suggest that the kinetically trapped oligomeric species, even though they seem to be part of an alternative branch of the aggregation pathway, are of cross-β nature (as shown also by their ability to bind the fibril specific antibody WO1- data not shown) and can potentially convert into fibrils.

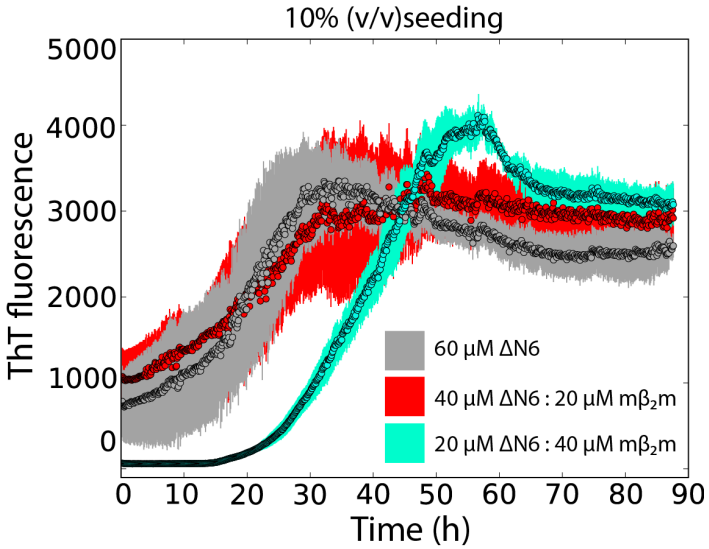

Figure 4.12: Seeding capacity of species made by the incubation of ΔN6-mβ$_2$m in different protein ratios. Mixtures of ΔN6 and mβ$_2$m (in 2:1 or 1:2 molar ratios) or ΔN6 alone, were incubated for 120h in 10mM sodium phosphate pH 6.2, 37^{o}C, 600rpm (kinetics are shown in Figure 4.9, panel vi, viii, i, respectively). Samples were centrifuged (14000rpm, 5min) using a bench-top centrifuge, the pellet was re-suspended in fresh buffer and was used as 10% *(v/v)* seeds when mixed with 60μM of ΔN6 monomer. The ability of species generated by ΔN6 alone (grey), ΔN6:mβ$_2$m 2:1 (red), ΔN6:mβ$_2$m 1:2 (cyan) to seed elongation was monitored by ThT fluorescence. The average of 5 replicates is shown as dots and the error bars represent one standard deviation of the mean.

The kinetically trapped oliogmeric species that can seed fibril formation are a late product of the ΔN6-mβ$_2$m interaction and are observed after 120h of incubation, in the middle of the elongation phase under the conditions employed (Figure 4.13). However, understanding how these species form during the lag time of fibril formation is of crucial importance in understanding the mechanism of inhibition of ΔN6 assembly and constitutes the topic of Section 4.2.4.

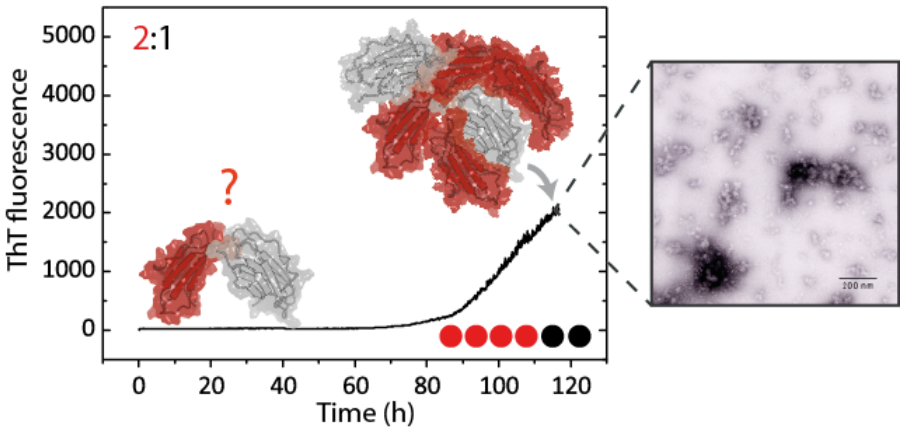

Figure 4.13: The importance of the early steps in amyloid inhibition.
The panel shows the aggregation kinetics of 40μM ΔN6 mixed with 20μM mβ$_2$m (reproduced from Figure 4.9A, panel iv). At the endpoint of the reaction spherical aggregates are observed (shown in the electron micrograph), however the details of the intermolecular interaction between ΔN6 and mβ$_2$m during the early stages of this reaction that lead to a significant retardation of the lag time of fibril assembly remain to be investigated.

4.2.4 The mechanism of inhibition of ΔN6 assembly by mβ$_2$m

The data presented so far show that the biomolecular association of ΔN6 with mβ$_2$m inhibits amyloid assembly even when the proteins are mixed in substoichiometric ratios. The dependence of inhibition on mβ$_2$m concentration and the appearance of novel species during assembly of these protein mixtures point to a kinetic phenomenon, where off-pathway non-amyloidogenic species are preferentially stabilised. However, at this stage the precise molecular details of the elongation reaction and how it is inhibited by mβ$_2$m remain elusive. In order to increase our understanding of the intermolecular interactions in the early steps of amyloid aggregation, an analysis using AUC was performed. Proteins were mixed, dialysed into the AUC buffer and the

molecular weight distribution of different protein combinations under fibril forming conditions was examined by sedimentation velocity AUC (Figure 4.14).

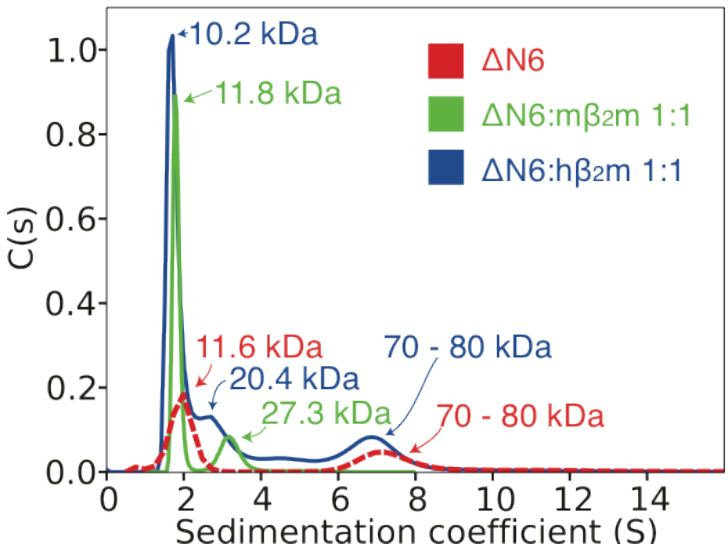

Figure 4.14: Mβ$_2$m alters the distribution of oligomeric species.
Sedimentation velocity AUC of ΔN6 alone (red), ΔN6 mixed with mβ$_2$m (green) and ΔN6 mixed with hβ$_2$m (blue) in 1:1 molar ratio in 10 mM sodium phosphate buffer pH 6.2, 83.3mM NaCl, 25°C. The estimated molecular weight of each peak is noted on top using the same colour code as the traces. ($MW_{\Delta N6}$: 11.13KDa, $MW_{h\beta2m}$: 11.81KDa, $MW_{m\beta2m}$: 11.77KDa).

These experiments revealed that under the conditions employed, ΔN6 is in equilibrium with higher order species (7-8mers). Interestingly, these species are not formed when ΔN6 is mixed with mβ$_2$m and instead a monomer-dimer equilibrium is established. The presence of high molecular weight species for ΔN6 at pH 6.2 is in line with the data shown in Figure 4.2D and 4.3, revealing that at least some degree of the line broadening observed at pH 6.2 originates from the presence of oligomers although other sources of increased linewidth such as fast conformational dynamics cannot be excluded. In the presence of mβ$_2$m the equilibrium between different protein states is remodeled and the monomer-oligomer equilibrium observed for ΔN6 alone is substituted by a monomer-dimer equilibrium. However, since the molecular weight observed by AUC for the dimeric species (27.3kDa) is higher than its theoretical value (~23kDa) exchange with trimer or higher order oligomers cannot be ruled out. Importantly, the interaction between ΔN6 and hβ$_2$m does not show this behavior. Instead, the high molecular weight oligomers are preserved.

As shown in Figure 4.9, panel iii, ΔN6 is able to self-assemble into amyloid fibrils of typical morphology in the presence of mβ$_2$m suggesting that the increased energy barrier provided by the ΔN6-mβ$_2$m interaction can eventually be overcome or in other words that mβ$_2$m kinetically traps species off pathway to ΔN6 fibril assembly. If the hypothesis is correct, the extend of inhibition should be directly proportional to the concentration of the off-pathway species (ΔN6-mβ$_2$m) and therefore to the concentration of mβ$_2$m added. To determine if this is the case, mixtures of ΔN6 and mβ$_2$m were incubated for a total of 2 weeks and the progress of fibril assembly was monitored by ThT fluorescence and negative stain EM (Figure 4.15). The results showed that all protein mixtures converted into amyloid-like fibrils over this extended time period. Consistent with these findings, a vast excess of mβ$_2$m (1:20 ΔN6:mβ$_2$m) prolongs the lag time of fibril assembly to more than 400h (Figure 4.15D).

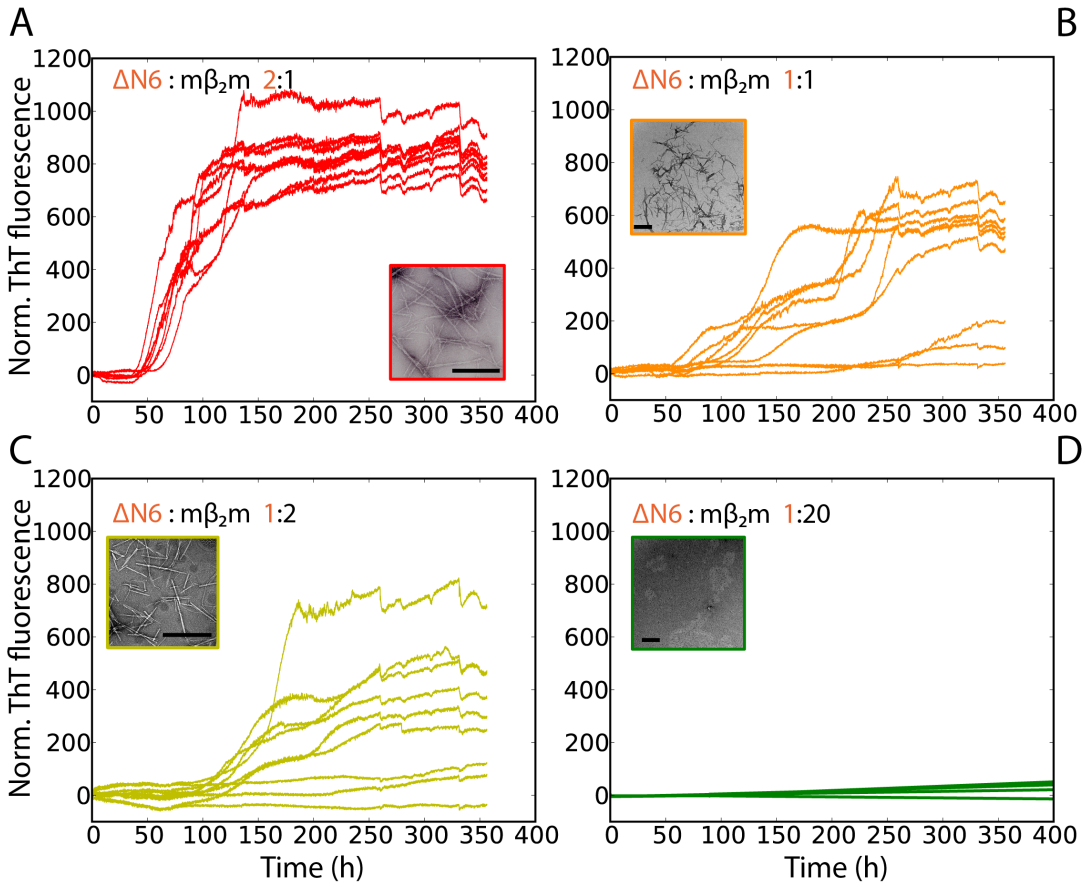

Figure 4.15: Kinetic inhibition of ΔN6 amyloid formation by mβ$_2$m.
Aggregation kinetics of ΔN6 mixed with mβ$_2$m in different molar ratios, 2:1 (40μM ΔN6 + 20μM mβ$_2$m) (A), 1:1 (30μM ΔN6 + 30μM mβ$_2$m) (B), 1:2 (20μM ΔN6 + 40μM mβ$_2$m) (C) and 1:20 (10μM ΔN6 + 200μM mβ$_2$m) (D), incubated for a total of 400h in 10mM sodium phosphate pH 6.2 83.3mM Nacl, 37°C, 600rpm. Electron micrographs of the end product are shown as insets. Bar represents 500 nm.

The fibrils were then pelleted by centrifugation and the supernatant was analysed by SDS-PAGE and analytical gel filtration. Even after this prolonged incubation period (2 weeks), mβ$_2$m is able to keep some ΔN6 in solution, while in its absence, all ΔN6 monomers are consumed into the fibrils (Figure 4.16). Importantly, the concentration of soluble ΔN6 at the end of the reaction seems to be independent of the amount of mβ$_2$m added, an observation that further supports a kinetic mechanism of inhibition since mβ$_2$m does not alter the critical concentration (C_c)[2] of fibril formation. In the mixtures of ΔN6 and mβ$_2$m, oligomeric species were observed by gel filtration, which most probably correspond to, or are related to the assemblies shown in Figure 4.10 (Figure 4.16).

[2] The critical concentration (C_c) corresponds to the concentration of protein above which the concentration of the soluble fraction does not increase at the end of the reaction and is equal to:

$$C_c = k_{on} / k_{off}$$

where k_{off} is the monomer dissociation constant from the polymer and k_{on} is the association constant for monomer addition.

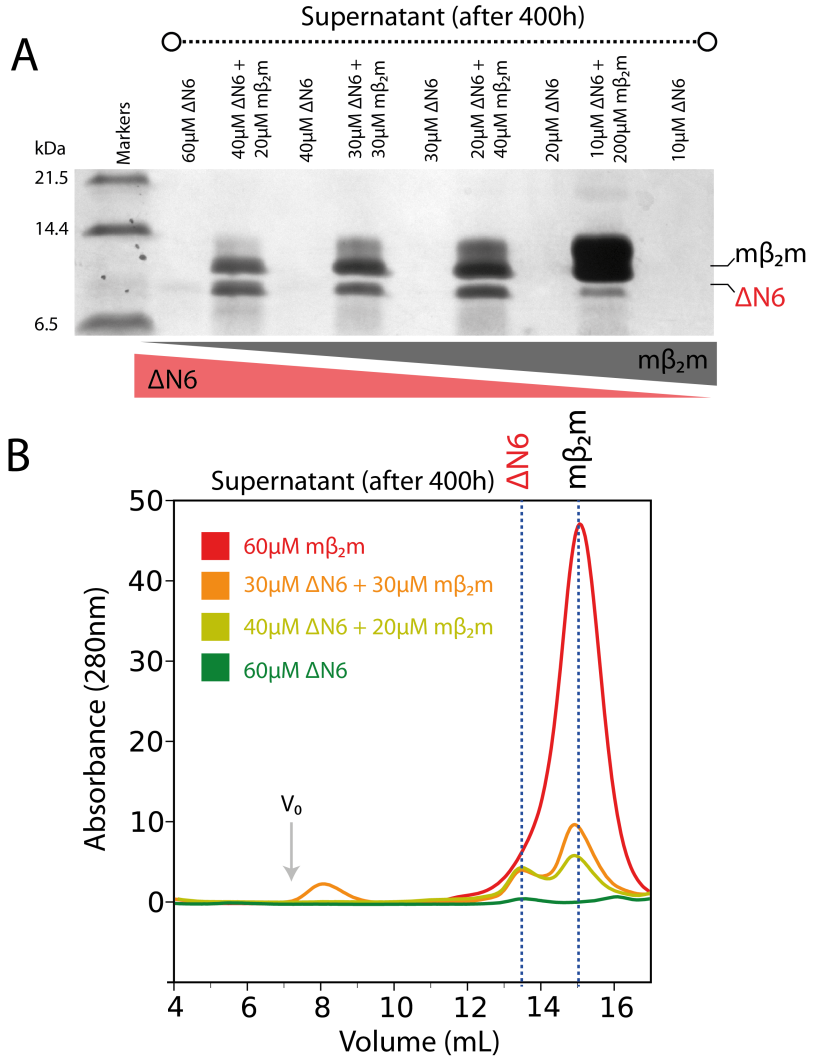

Figure 4.16: Analysis of the soluble fraction of different samples of ΔN6/mβ2m after 400h of incubation. After 400h of fibril formation (kinetics shown in Figure 4.15) the aggregates were pelleted by centrifugation and the soluble fraction was analysed by SDS-PAGE (A) and analytical gel filtration (B). Lanes in (A) represent: markers, 60μM ΔN6, 40μM ΔN6 + 20μM mβ2m, 40μM ΔN6, 30μM ΔN6 + 30μM mβ2m, 30μM ΔN6, 20μM ΔN6 + 40μM mβ2m, 20μM ΔN6, 10μM ΔN6 + 200μM mβ2m, 10μM ΔN6. (B) Analytical gel filtration of 60μM ΔN6 (green), 40μM ΔN6 + 20μM mβ2m (yellow), 30μM ΔN6 + 30μM mβ2m (orange) or 60μM mβ2m (red). Dotted lines indicate the expected elution volumes of monomeric ΔN6 and mβ2m. Gel filtration was performed in 100mM Tris HCl pH 8.0 using a Superdex 75 10/300 column (GE Healthcare). The void volume is marked with an arrow.

4.2.5 Is mβ2m incorporated into the ΔN6 fibrils?

The data in Figure 4.16 show that in the mixtures of ΔN6 and mβ2m the amount of mβ2m left in solution at the end of the reaction is always less than that expected based on the concentration of mβ2m added (e.g. the orange trace in

Figure 4.16 does not have half of the intensity of the red one). This observation suggests that some mβ$_2$m is associated with the ΔN6 fibril pellet either by mβ$_2$m being incorporated into the structure of the fibril or mβ$_2$m may associate with the ΔN6 fibril surface. One way to discriminate between those two scenarios is to examine the fibril yield. A linear relationship is observed between the intensity of the ThT fluorescence at the endpoint and the concentration of ΔN6 used in the fibril growth reaction (Figure 4.17A). When ΔN6 is mixed with mβ$_2$m in different ratios but keeping the total protein concentration constant and if mβ$_2$m were able to be incorporated into the fibrils of ΔN6, the plateau of the fluorescence should stay constant (horizontal dotted line in Figure 4.17B). However, a decrease in ThT fluorescence is observed as the amount of ΔN6 added in the reaction is decreased, indicating that mβ$_2$m is not integrated into the fibrils that form (Figure 4.17B). The slope of ThT fluorescence versus ΔN6 concentration is steeper for the ΔN6:mβ$_2$m mixtures than for ΔN6 alone (compare Figure 4.17A, B), suggesting that the ΔN6 fibrils made in the presence of mβ$_2$m have different structural properties than the fibrils assembled when ΔN6 is incubated alone, depicted in the ability to bind ThT. This well accepted concept of fibril polymorphism is a known phenomenon among β$_2$m fibrils assembled under different conditions[339,342] and was recently investigated for mixed ΔN6 and hβ$_2$m fibrils[405].

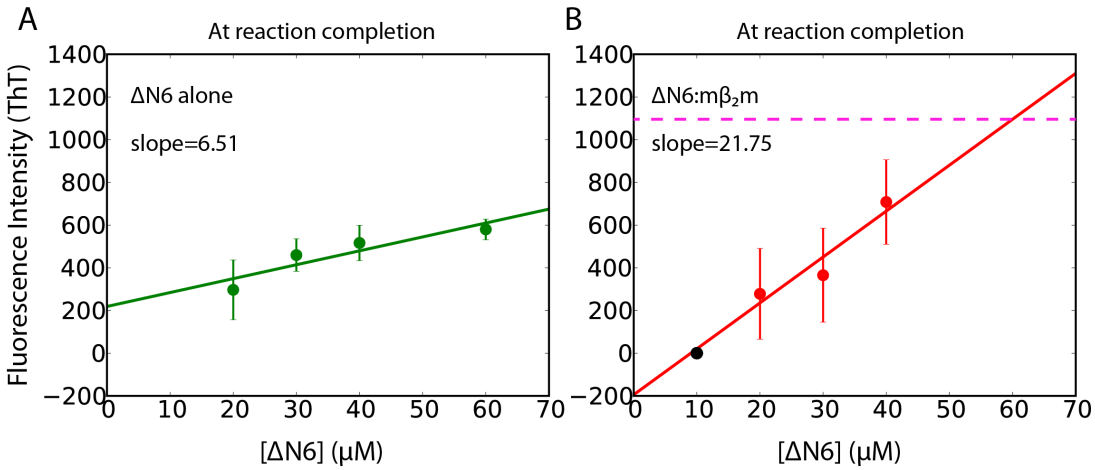

Figure 4.17: Calibrating the relationship between ThT fluorescence intensity at the endpoint and fibril yield. A linear relationship between ΔN6 concentration and ThT intensity is observed for ΔN6 alone (A) and the ΔN6:mβ$_2$m mixtures (B). Straight lines represent linear fits to the data. The dashed line in B shows the theoretical value calculated by assuming that mβ$_2$m is able to incorporate into the fibrils. (The endpoint is defined as the average of the last 100 datapoints of a fibril growth curve).

To provide definitive proof about the composition of the fibrils made by the incubation of ΔN6 and mβ$_2$m (after 400h), the pelleted fibrils were depolymerised by either incubation at 100°C for 10min or by incubation in 1,1,1,3,3,3-hexafluoro-2-isopropanol (HFIP) for ~12h and then analysed by SDS-PAGE and mass-spectrometry. In parallel, mβ$_2$m monomer was added to pre-assembled ΔN6 fibrils and an identical analysis was carried out. These experiments revealed the presence of a small amount of mβ$_2$m in the insoluble fraction when the proteins were mixed before fibril assembly (Figure 4.18A, B). Notably similar amounts of mβ$_2$m were found in the fibrillar fraction when the protein was added post fibril assembly (Figure 4.18C, D). These date therefore indicate that mβ$_2$m does not become incorporated into the ΔN6 fibrils and it is rather associated with the fibrils after they are formed.

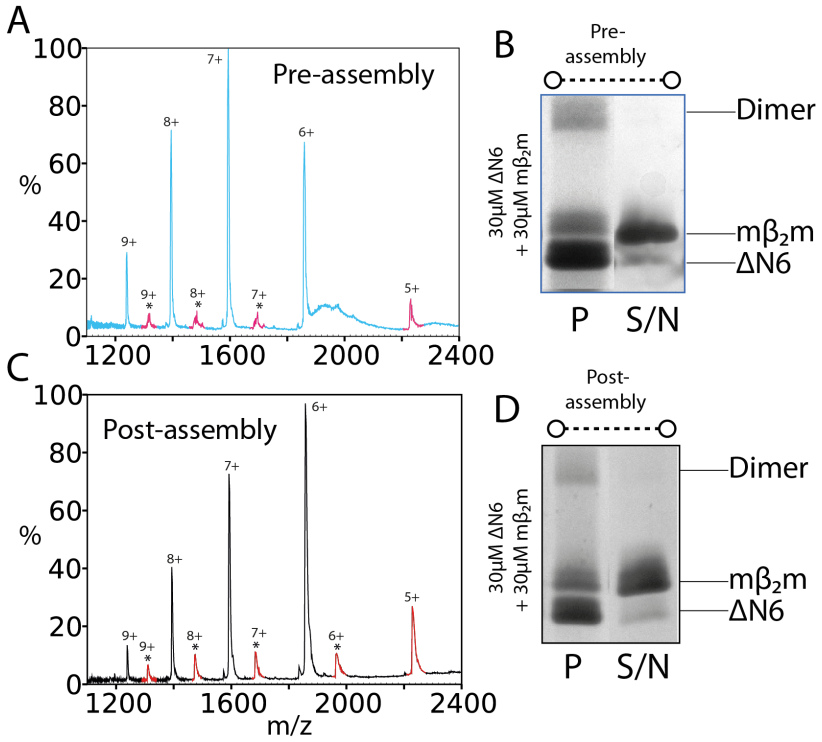

Figure 4.18: Analysis of the insoluble fraction of samples of mβ$_2$m co-incubated with ΔN6. (A) ESI-mass spectrum of the pellet formed from 30μM ΔN6 + 30μM mβ$_2$m where the proteins were mixed during fibril assembly. Fibrils formed after 400h of incubation (Figure 4.15) were pelleted by centrifugation, depolymerised in 100% *(v/v)* HFIP and subjected to analysis by ESI-MS. Peaks corresponding to mβ$_2$m are shown in red and are marked with an asterisk. (B) SDS-PAGE of the samples described in (A). After centrifugation the pellet (P) and the supernatant were analysed by SDS-PAGE. (C) As in A but when 30μM mβ$_2$m monomer were added to 30μM ΔN6 fibrils post assembly.

(D) As in (B) but for post assembly addition of mβ$_2$m to pre-formed ΔN6 fibrils. Mass spectrometry was performed by Dr. James Ault (University of Leeds).

In summary, the data presented here show that mβ$_2$m is able to delay fibril assembly by ΔN6, but the reaction will eventually proceed if the proteins are incubated for prolonged periods of time. Consequently, mβ$_2$m constitutes a polymerisation-incompetent species that cannot be incorporated into the fibrils but rather remains associated with them.

4.3 Discussion

4.3.1 Secondary processes dictate fibril nucleation

Even though the primary mechanism underlying fibril assembly is nucleated growth, an increasing amount of recent evidence suggests that simple nucleated polymerisation is not sufficient to explain the kinetics of fibril formation[263-265]. It has been shown recently using various systems, including acid unfolded hβ$_2$m, that primary aggregation pathways that consider only the concentration of initially monomeric protein precursors are poor descriptors of the kinetic process of assembly[263]. Instead, secondary processes that are dependent on the concentration of already formed fibrils have to be included in models describing aggregation[263,264]. In the case of hβ$_2$m at low pH (2.5), fibril fragmentation as the outcome of agitation was identified as the main secondary process that dominates the progress of amyloid aggregation under these conditions[263]. This is not difficult to imagine as agitation causes fibril breakage and the creation of additional fibril ends that act as seeds to induce secondary polymerisation. The data shown in Figure 4.1 suggest that a similar scenario might be taking place at neutral pH for ΔN6. Increased agitation (600rpm) results in a substantial decrease in the lag time of fibril formation that reduces the timescale of fibril formation from days to hours. However, under conditions of increased agitation the reverse concentration dependence of the lag time suggests that not all the surfaces generated are equally competent for aggregation and instead some off-pathway species are created.

4.3.2 Amyloid polymorphism

Even though amyloid fibrils formed from different protein precursors that do not share any significant similarities, either at the level of the amino acid sequence or in terms of structure, all share a cross-β architecture in the fibril core[224], the fibrils may differ significantly in their structural properties. This phenomenon, known as fibril polymorphism is well-characterised even in fibrils made from the same monomer precursor and may arise from differences in the packing of the side chains in the hydrophobic core, or different arrangements of the protofilaments inside the fibril[256,281,405,411,412]. In the case of $h\beta_2m$, small alterations in the pH of the fibril formation (from pH 2.0 to pH 3.6) result in drastic remodelling of the course of amyloid assembly and as a consequence fibrils adopt a worm-like structure instead of the typical long straight morphology seen at pH 2.0[336,339,413].

Recently the role of co-polymerisation and its effect on fibril morphology and structural properties was investigated using mixtures of different protein precursors. ΔN6 and $h\beta_2m$ were mixed in stoichiometric ratios under conditions in which the latter is not amyloidogenic (pH 6.2). In these studies $h\beta_2m$ was found to be incorporated into the fibrils resulting a 1:1 mixed fibril (Figure 4.19)[405].

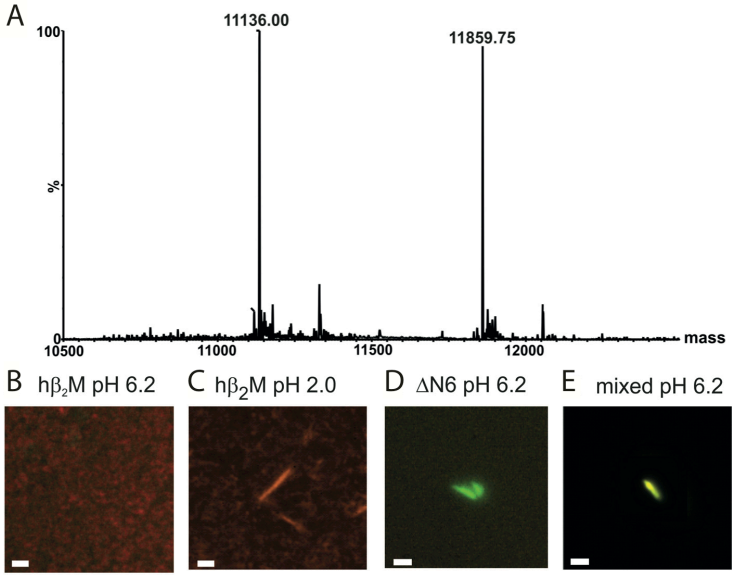

Figure 4.19: Copolymerisation of ΔN6 and $h\beta_2m$.
(A) ESI-mass spectrum of the depolymerisation products of fibrils made by a 1:1 molar ratio mixture of ΔN6 (11,136 Da) and $h\beta_2m$ (11,859.75 Da). (B)

Fluorescence microscopy images of hβ$_2$m labelled with 5(6)-carboxytetramethyl- rhodamine succinimidyl ester (TAMRA, Invitrogen) at pH 6.2 where no fibrils are observed, while TAMRA-labelled hβ$_2$m assembles into fibrils at pH 2.0 (C). (D) Fibrils of fluorescein-5-isothiocyanate (FTIC, molecular probes)-labelled ΔN6 at pH 6.2, (E) fibrils formed by an 1:1 mixture of FTIC-labelled ΔN6 and TAMRA-labelled hβ$_2$m at pH 6.2. The co-localisation of 'red' hβ$_2$m and 'green' ΔN6 results in fibrils coloured yellow. Scalebar = 5μM. Adapted from[405].

Here similar experiments were carried out on the heteropolymerisation of mβ$_2$m and ΔN6 to show that mβ$_2$m is not able to adopt the amyloid fold even in the presence of ΔN6 seeds. Thus, the behaviour of the two closely related molecules hβ$_2$m and mβ$_2$m is surprisingly different, even though they share many common structural features (Chapter III) and are both incapable of self-assembly into amyloid fibrils at neutral pH. This finding highlights the importance of ΔN6 as a potential *in vivo* 'catalyst' of aggregation as small amounts can reveal the hidden amyloid potential of hβ$_2$m leading to disease. On the other hand, and in line with data presented in Chapter III, its murine counterpart has abolished the ability to form fibrils even under conditions that highly favour aggregation such as the presence of fibrillar seeds.

4.3.3 Protein-protein interactions and the formation of species barrier.

Even though the majority of amyloid fibrils share similar structural features, independently of the organism of origin or of the protein involved, or the sequence of the protein precursor, nature has evolved mechanisms of facilitating assembly, or protecting against the accumulation of misfolded proteins. One such example can be found in prions; proteins that possess at least one β-rich conformation that is infectious by being able to transmit its structural and pathological properties onto normally folded prion monomers[414,415]. Interestingly, when misfolded prion molecules are transfected to different species they can lose their high infectivity, establishing a so-called "species barrier"[404,416]. This highlights the impact of different intermolecular interactions on the amyloid potential of the same protein molecule from different organisms[403,417]. However, the exact molecular mechanism by which different protein-protein interactions modulate the enhancement or inhibition of amyloid assembly remains unknown.

ΔN6 has been shown to posses prion-like properties in its ability to convert hβ$_2$m in an aggregation-prone conformation[272] reminiscent of the transmission of prion infectivity which force their natively folded counterparts to misfold and hence to adopt an assembly –competent conformation. However, the prion-like characteristics of ΔN6 are not only limited to its ability to endow prion-like propagation on hβ$_2$m, but also in its capability to form species barriers. The data presented here show that the interaction between the murine (mβ$_2$m) and ΔN6 variants of β$_2$m, which share a 70% sequence identiy and 90% sequence homology results in inhibition of the aggregation of ΔN6 by destroying pre-amyloid oligomers and shifting the equilibrium towards monomeric and dimeric species. Inhibition, however, is under kinetic control. Eventually, therefore, ΔN6 is able to self-assemble into amyloid fibrils over long timescales (Figure 4.20). Similarly, prion strains, although each able to aggregate and transmit their aggregation potential onto natively folded prion monomers from the same species, cannot stimulate the aggregation of prion strains from other species, both *in vivo* and *in vitro*, establishing in this way a species barrier[364,418,419]. In filamentous fungi, the mechanism of self-non-self recognition is exerted by the HET-S/HET-s prion strains. The N-terminal HeLo domain of HET-S, even though 97% identical in sequence to HET-s, can inhibit prion formation of its own prion-forming domain (PFD), but can also inhibit the propagation of the prion form of HET-s when added in trans[420]. Similarities to mβ$_2$m are obvious. A model for prion inhibition by HET-S has been proposed, where HET-S, although able to adopt the amyloid β-solenoid fold, is incompetent for further polymerisation[420]. However the exact moleular details of this mechanism remain unknown.

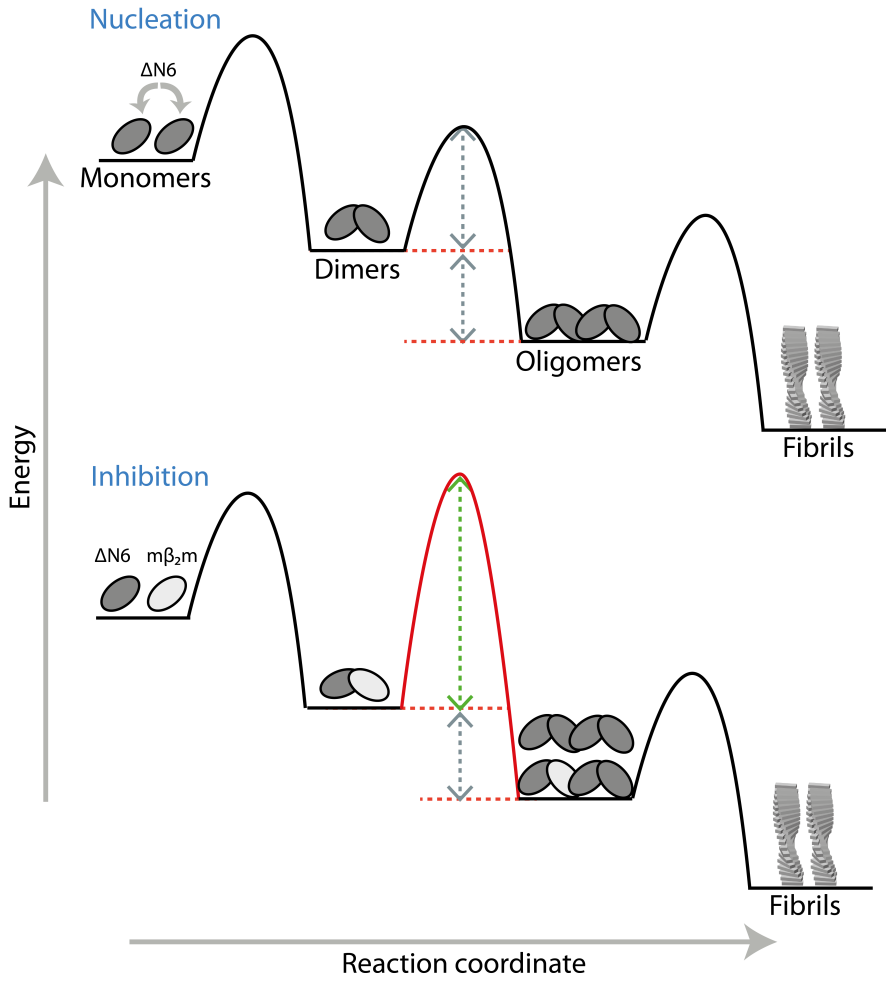

Figure 4.20: The mechanism of inhibition of ΔN6 fibril assembly by mβ$_2$m. Schematic showing a simplified version of fibril formation by ΔN6 alone (top) or when mβ$_2$m is present (bottom). During the nucleation of fibril assembly, ΔN6 monomers interact to form dimeric species and subsequently oligomers that convert to fibrils at the end of the reaction. A downhill polymerisation mechanism (where each species has lower energy than its precursor) is assumed for simplicity. When mβ$_2$m is added, the energy of the transition state from dimers to pre-amyloid oligomers is significantly increased resulting in kinetically trapped heterodimers and thus the establishment of a monomer-heterodimer equilibrium. However, and since the interaction does not alter the free energy of the oligomeric species which remain thermodynamically favoured, the reaction will eventually proceed to fibril assembly in a longer timescale. ΔN6 and mβ$_2$m are shown as dark grey and light grey ellipses, respectively.

Importantly, species barriers are not only found in prions. Rat islet amyloid polypeptide (IAPP) is reported not to form amyloid fibrils under any conditions tested and can aslo act as an inhibitor of the aggregation of its human counterpart[421], in a behaviour reminiscent of the mβ$_2$m-ΔN6 interaction. However, the interaction between rat and human IAPP in a 1:1 ratio causes

only a two-fold increase in the lag time of fibril formation[421], thus at least a 10-fold molar excess of rat IAPP is needed to suppress amyloid assembly. On the other hand, mβ$_2$m is a much more potent inhibitor of assembly in its capability to prolong the lag time of fibril formation ≥6-fold when added stoichiometrically. Another example of interactions between heterologous proteins during assembly, even though rare, has been reported before for IAPP and Aβ40 (25% sequence identity, 50% homology), providing a link between Alzheimer's disease and diabetes[422]. Chemical modification of IAPP by the addition of two methyl groups turns off self- (IAPP-IAPP), and hetero-polymerisation (IAPP-Aβ$_{40}$)[423-425], delaying the appearance of cytotoxic oligomeric assemblies in a similar way that mβ$_2$m affects the aggregation ΔN6. A link between transmission barriers and disruption of amyloid propagation thus appears to be a general phenomenon in amyloid assembly.

Mβ$_2$m presents a promising framework to design potential therapeutic stategies in order to treat β$_2$m-related amyloidoses. However, being itself a protein molecule, mβ$_2$m has limitted potential to act a a drug. Nevertheless, characterisation of the main principles that govern the biomolecular association of mβ$_2$m with ΔN6 that leads to inhibition of fibril assembly might lead to the discovery of new small molecules that mimic the inhibitory effect, opening the way to new therapeutic strategies. The investigation of the ΔN6-mβ$_2$m interaction and its effects, in atomistic detail is discussed in Chapter V.

CHAPTER V

TRANSIENT MACROMOLECULAR ASSOCIATION DURING AMYLOID ASSEMBLY IN ATOMIC DETAIL

5.1 Introduction

The studies on the amyloid-promoting hβ$_2$m-ΔN6 interaction and the inhibitory mβ$_2$m-ΔN6 interaction presented in Chapter IV suggest that the inhibition of amyloid formation is taking place by the accumulation of off-pathway kinetically trapped ΔN6-mβ$_2$m heterodimers which significantly delay the onset of amyloid formation. What are the structural and dynamic properties of these dimeric species? Do they differ from their ΔN6-hβ$_2$m or ΔN6-ΔN6 counterparts and how do the different interactions affect the properties of the proteins involved? Answering these questions constitutes the main goal of this Chapter. Since the nature of the interactions in the early stages of amyloid formation in general, and in the case of β$_2$m specifically, is transient and because these (dimeric) species are not stable enough to be interrogated by X-ray crystallography, NMR was used to investigate their properties. As described in Section 1.5.4.2 PRE techniques are ideally suited to study weak intermolecular interactions[212,386] such as those between ΔN6-hβ$_2$m, ΔN6-mβ$_2$m or ΔN6-ΔN6. The rationale behind these experiments is described in Figure 5.1.

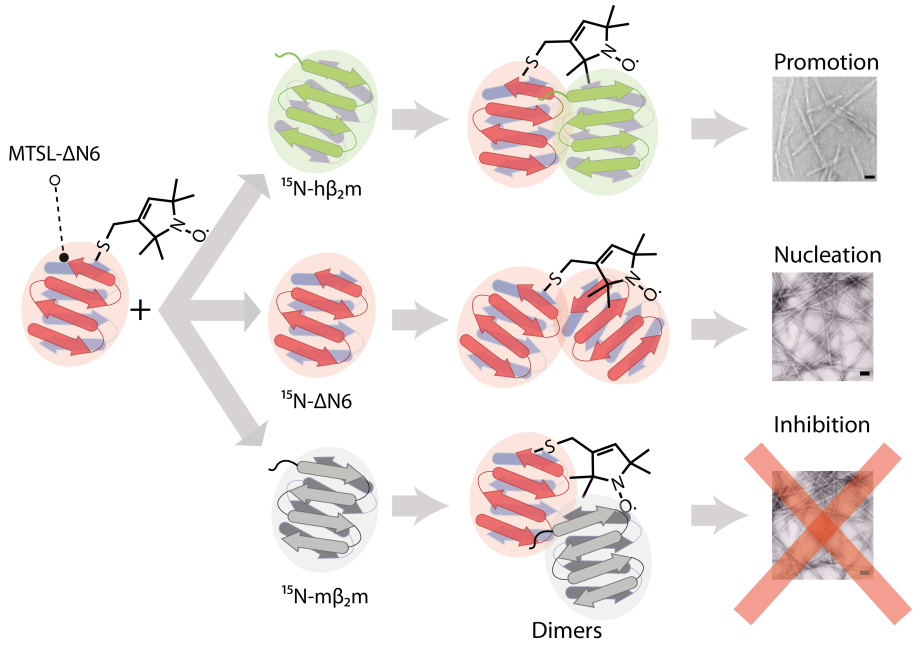

Figure 5.1: Experimental design for the intermolecular PRE experiments.

A paramagnetically (MTSL)-labelled (^{14}N, ^{12}C) ΔN6 monomer (red) is mixed with a ^{15}N-labelled monomer (green-hβ_2m, grey-mβ_2m or red-ΔN6) for NMR studies. A range of (encounter) complexes are formed, of which some might survive in solution only for a limited amount of time, and thus are not detectable by the majority of biophysical techniques. Using PRE techniques and different protein combinations the aim of this Chapter is to characterise the biomolecular interactions that lead to the promotion (ΔN6-hβ_2m), nucleation (ΔN6-ΔN6) or inhibition (ΔN6-mβ_2m) of fibril assembly. The small molecule shown is MTSL. (Bar denotes 100nm).

Using the experimental design described in Figure 5.1, mapping of the interaction surfaces between ΔN6-mβ_2m, ΔN6-hβ_2m and ΔN6-ΔN6 was performed to reveal that the inhibitory (ΔN6-mβ_2m) and promoting (ΔN6-hβ_2m) interactions involve similar surfaces. However, further investigation showed that the structural details, affinities and their outcome on the conformational properties of the involved monomers differs markedly. Interestingly, the self-association of ΔN6 is more complex and potentially involves multiple surfaces.

5.2 Results

5.2.1 Design, expression, purification and MTSL-labelling of CYS-ΔN6 mutants.

To allow the characterisation of the interacting surfaces between different protein molecules (as shown in Figure 5.1) cysteine residues had to be introduced in solvent exposed sites on the surface of ΔN6, which could then be paramagnetically labelled with MTSL. The criteria for using residues to be mutated to Cys included: solvent accessibility, sequence conservation (preferably serine or threonine residues), and residues should lie at the termini of β-strands in order to avoid disruption of the native fold. Based on these criteria, residues S20, S33, D38, S61, T73 and I92 were chosen for site-directed mutagenesis to Cys (Figure 5.2A). All mutants expressed well, potentially yielding ~30mg of protein per L of culture (Figure 5.2B).

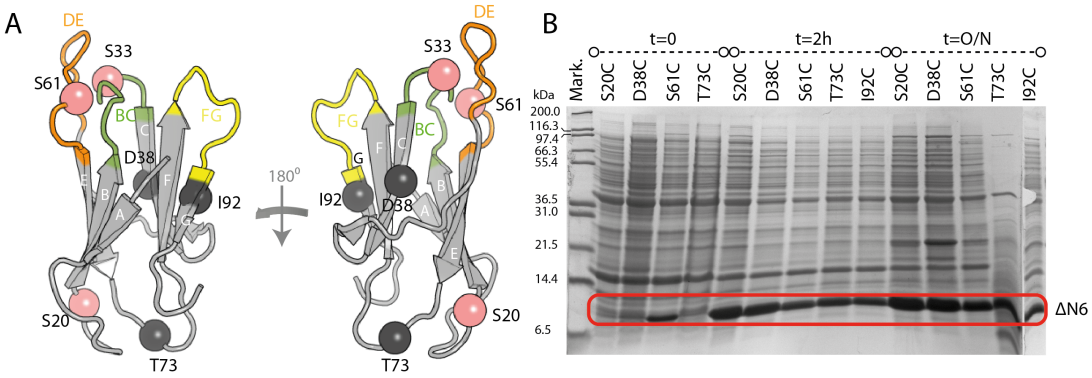

Figure 5.2: Design of Cys-ΔN6 mutants and their over-expression.
The residues (S20, S33, D38, S61, T73, I92) that were mutated individually to Cys are shown in spheres on the structure of ΔN6 in (A). From those only the ones coloured in pink (S20, S33, S61) were used in subsequent PRE studies. The BC, DE and FG loops are coloured in green, orange and yellow respectively. (B) Examples of over-expression of some of the Cys-ΔN6 mutants. Time after induction in noted on top.

To allow the purification of Cys-ΔN6 mutants the purification protocol described in Section 2.3.3 had to be modified. ΔN6 contains two native cysteine residues (Cys25 and Cys80) that form a disulfide bond in the native fold. During the refolding step of the purification, the formation of the native disulfide bond is not problematic since it is thermodynamically favourable and thus the refolding can take place in a TrisHCl buffer. However, the presence of an additional sulfur in the Cys-ΔN6 mutants raises the necessity of a buffer of the correct redox potential to refold these proteins. Failure to do so, results in protein precipitation or the formation of misfolded aggregates that bind to the Q-Sepharose anion exchange column with high affinity (they elute only after washing the resin 1M NaOH). To identify the conditions that allow complete refolding of the mutants a small scale refolding screen was performed for each protein[343]. Briefly, inclusion bodies were prepared as described for hβ_2m (Section 2.3.3) and 50μL aliquots of 5mg/mL protein were solubilised in 8M urea for 2h with rocking. Refolding was initiated by 5-fold dilution in seven different refolding buffers (Table 5.1) containing 25mM TrisHCl pH 8.0 and/or 2mM reduced glutathione, 0.2mM oxidized glutathione, 20mM NaCl, 0.64/0.4M L-arginine. The sample was further diluted (10-fold) in 25mM TrisHCl pH 8.0 and anion-exchange chromatography was performed using 5mL HiTrap columns (GE healthcare) in the same buffer and a gradient of 0-1M NaCl over 10 column volumes. Refolded protein elutes ~15mL after the salt gradient was

initiated (~200mM NaCl), while aggregated forms of ΔN6 have a higher affinity for the Q-Sepharose resin and elute after ~30mL (~600mM NaCl). The refolding screen was performed for all Cys mutants and the results for S61C-ΔN6 and S33C-ΔN6 are shown in Figure 5.3 as examples. While S61C-ΔN6 refolds only in the presence of L-arginine (0.4mM), S33C-ΔN6 refolds in all conditions tried (even in just TisHCl) showing that the conditions of refolding are case-sensitive (Figure 5.3A, B).

After the small scale screen for refolding conditions, large scale purification was carried out by 10-fold dilution of the unfolded protein in the selected buffer. Alternatively, the protein was diluted in the refolding buffer to give a final protein concentration of ~7mg/mL and the remaining urea was removed by dialysis into the same buffer. Further dialysis was performed against 20mM TrisHCl pH 8.0. Before purification using anion exchange chromatography, 5mM DTT was added and anion exchange was performed in 20mM TrisHCl, pH 8.0 using a gradient of 0-1M NaCl over 5 column volumes. The protein was dialysed against water, freeze-dried and re-suspended in 10mM sodium phosphate pH 7.0, 5mM DTT before gel filtration (Superdex 75, 10/300 column, GE Healthcare), which was carried out in the same buffer, but lacking DTT. Pure protein (Figure 5.3C) was labelled with MTSL using the protocol described in Section 2.3.9 and 100% labelling was confirmed by mass-spectrometry performed by Dr. James Ault (University of Leeds) (Figure 5.3D).

#	Additive	S20C	S28C	S33C	D38C	S61C	T73C	I92C
1.	-							
2.	2mM GSH, 0.2mM GSSH,							
3.	2mM GSH, 0.4mM GSSH							
4.	2mM GSH, 0.2mM GSSH, 0.4M Arg							
5.	0.4mM Arg							
6.	20mM NaCl							

Table 5.1: Screen for refolding buffers for Cys mutants of ΔN6.
Successful refolding conditions are highlighted in grey for each of the mutants. Buffers #3, #6 did not give rise to any peak corresponding to ΔN6 (GSH: oxidized glutathione, GSSH: reduced glutathione, Arg: L-arginine). The buffer used was 20mM TrisHCl, pH 8.0.

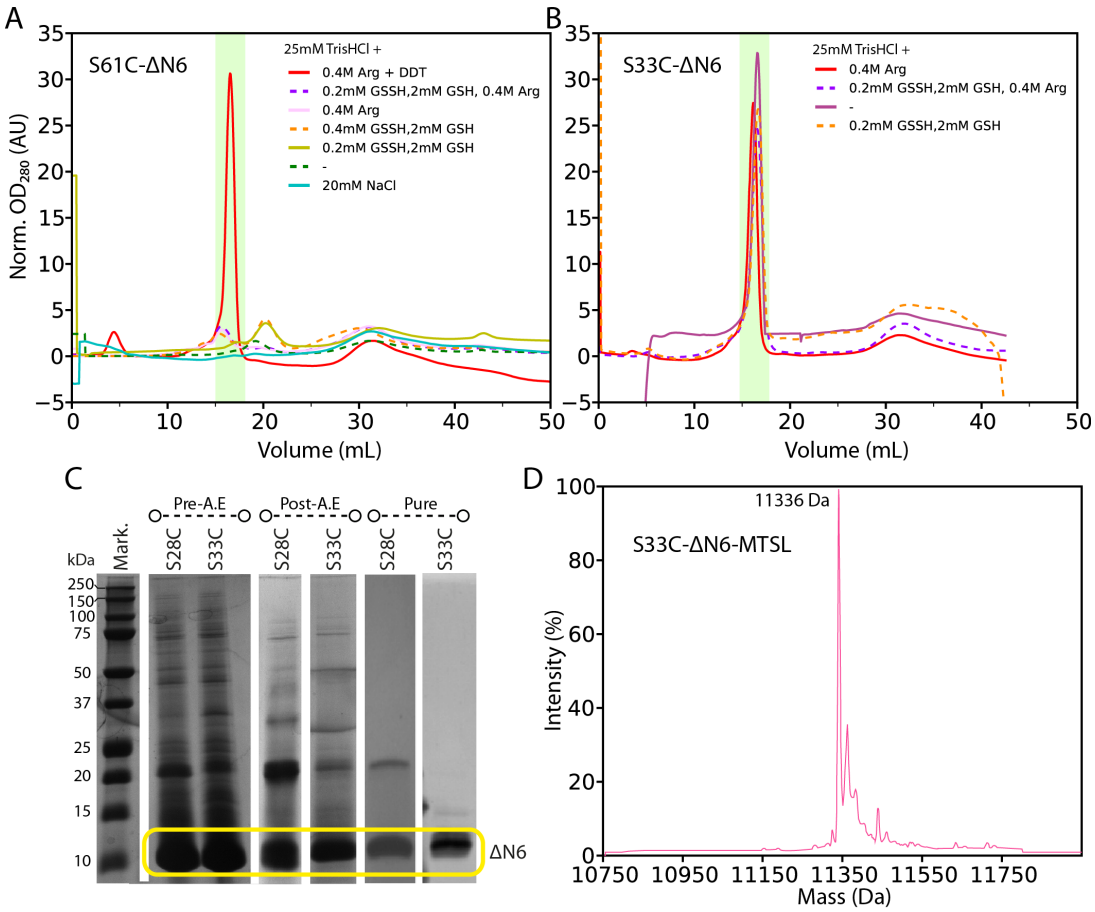

Figure 5.3: Optimising the refolding/labelling of S61C-ΔN6 and S33C-ΔN6. Example of anion-exchange chromatograms obtained during the refolding screen for S61C-ΔN6 (A) and S33C-ΔN6 (B). The peak corresponding to monomeric ΔN6 is highlighted in a green box, additional peaks correspond to contaminants or aggregated forms of ΔN6. (C) SDS-PAGE of S28C-ΔN6 and S33C-ΔN6 before (first two lanes) and after anion exchange (A.E) (middle two lanes) in buffer #5 shown in Table 5.1 and the pure products after gel filtration (last two lanes). The band corresponding to ΔN6 is highlighted in a yellow box. (D) ESI-Mass spectrum of MTSL-labelled S33C-ΔN6. Unlabelled S33C-ΔN6 has a mass of 11152.2Da and MTSL adds 185.4Da to give a mass for the fully labelled protein of 11137.6Da.

Using this methodology S61C-ΔN6, S33C-ΔN6 and S20C-ΔN6 were purified with yields of 2mg, 1mg, 1mg /L of culture.

5.2.2 Identifying the interaction surfaces

The spin-labelled ΔN6 variants generated were used in a series of PRE experiments as described in Figure 5.1 in order to investigate the surfaces involved in inhibition (ΔN6-mβ_2m), promotion (ΔN6-hβ_2m) or nucleation

(ΔN6-ΔN6) of fibril assembly. 60μM MTSL-labelled ΔN6 were mixed with 60μM ^{15}N-mβ_2m, 60μM ^{15}N-hβ_2m, or 60μM ^{15}N-ΔN6 and a PRE experiment was carried out as presented in Figure 1.12.

PRE data collected for the interaction between ^{14}N-ΔN6(S61C-MTSL) and ^{15}N-mβ_2m are shown in Figure 5.4B, panel i. Results show high PRE values (Γ_2>80 s^{-1}) for residues in the BC and DE loops and lower Γ_2 (<80 s^{-1}) for residues in the N-terminal 10 residues and the FG loop (Figure 5B, panel i). These regions cluster on the apical side of the molecule, surrounding P32 (Figure 5.4C). A similar PRE pattern was obtained using ΔN6 labelled at position 33 (Figure 5.4B, panel iii). The results therefore suggest, that the apical region of the molecule is involved in the interaction with ΔN6 that creates a kinetically trapped, off-pathway heterodimer that inhibits amyloid formation. Consistent with this, when the spin label is moved to position 20 (on ^{14}N-ΔN6), the Γ_2 rates in the BC and DE loops are reduced vastly (<25 s^{-1}) (Figure 5.4B, panel ii), suggesting that S20 is distant from the site of interaction. Therefore a head-to-head configuration is likely for the ΔN6-mβ_2m heterodimer, involving the BC and DE loops from both monomers creating the inhibitory complex.

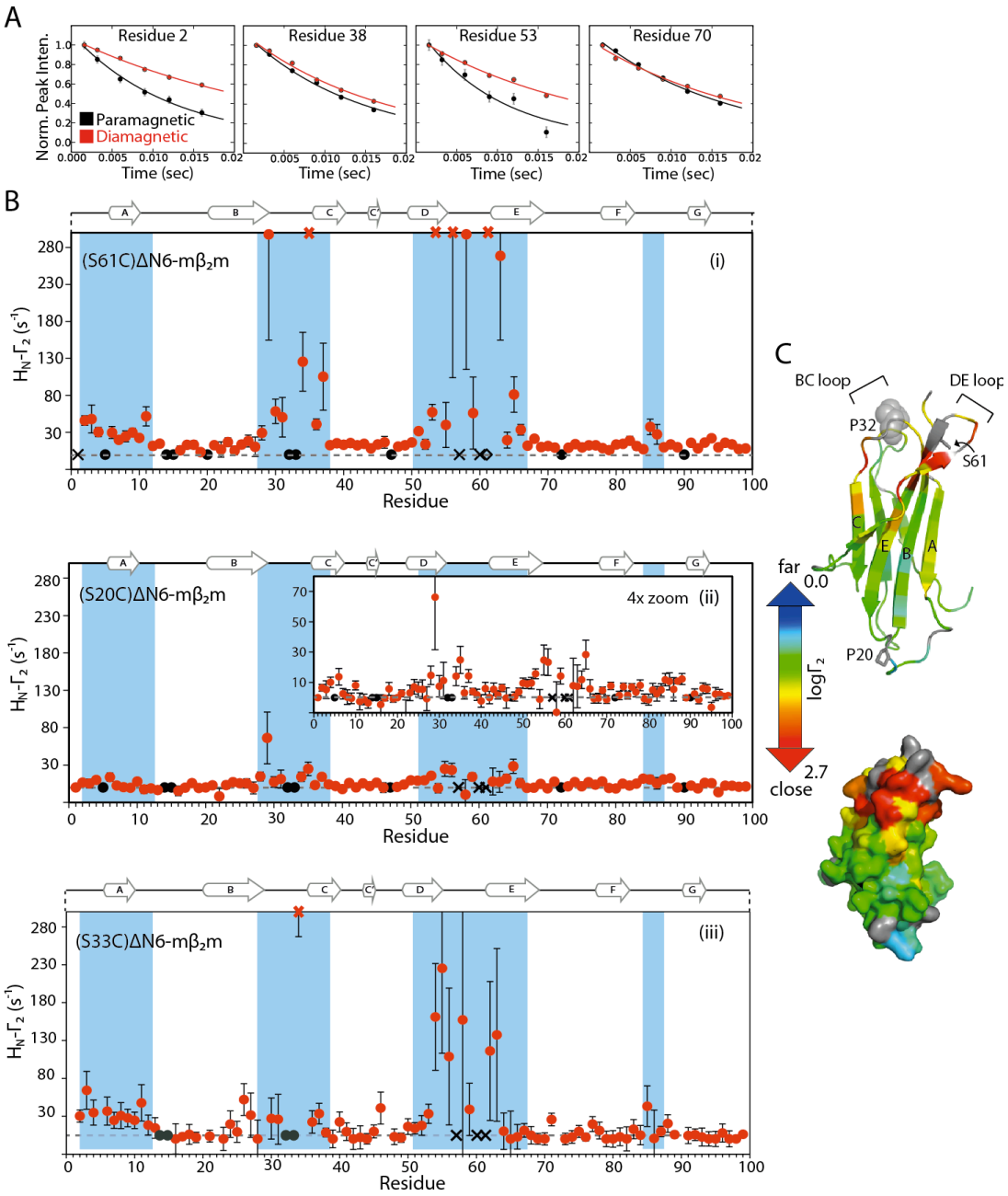

Figure 5.4: Interaction interfaces in amyloid inhibition (ΔN6-mβ$_2$m). (A) Example raw ^{1}H relaxation data for the (S61C)ΔN6-mβ$_2$m interaction are shown for residues 2, 38, 53, 70. Peak intensity is plotted as a function of relaxation time when MTSL is oxidised (paramagnetic-black) and after reduction with ascorbic acid (diamagnetic-red). Solid lines represent fits to single exponentials. (B) Intermolecular PRE profiles for the ΔN6-mβ$_2$m (60μM each) interaction when the spin label is attached on S61 (panel i), S20 (panel ii), or on S33 (panel iii) on ^{14}N-ΔN6. A zoom-in of the diagram in panel ii is shown as an inset. Red crosses indicate residues for which the Γ_2 rate is either too large to appear on this scale or that are broadened beyond detection when the spin label is oxidised and therefore the Γ_2 rate cannot be measured. Black dots represent proline residues or residues that are broadened beyond detection and black crosses denote residues for which the assignments are missing. The secondary structure elements of mβ$_2$m are shown on top of the graphs as ribbons. (C) Structure of mβ$_2$m coloured by the logarithm of the S61-Γ_2 rates.

The small PRE signal observed when ΔN6 is labelled at position 20 can be attributed to MTSL driven binding since incubation of free MTSL with ^{15}N-mβ$_2$m results in a very similar profile (Figure 5.5). In this case the PRE amplitude is higher than the one observed for the interaction of MTSL-S20C-ΔN6 with mβ$_2$m presumably due to the increased entropic penalty of a protein-protein interaction in comparison to binding of a small molecule (MTSL). Therefore data arising from the C20 site were not included in further analysis (see Section 5.2.4.1). The possibility that PRE rates from the C33 and C61 sites are also due to MTSL-mediated binding is excluded, because the amplitude of the effect is much higher for the protein-protein interaction and additional regions with increased Γ_2 rates appear in the N-terminus and the FG loop which are not seen for MTSL alone.

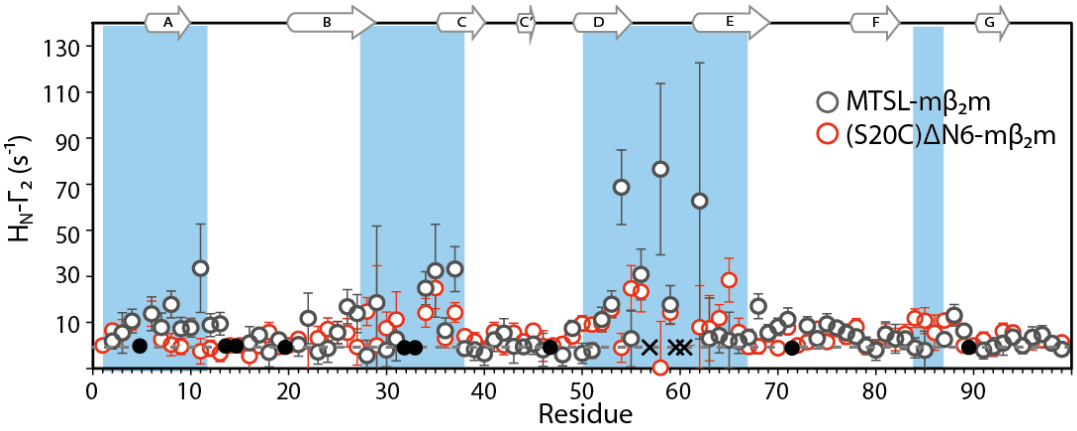

Figure 5.5: Background interactions mediated through MTSL.
PRE profiles for the interaction of 60μM free MTSL with 60μM ^{15}N-mβ$_2$m (open black symbols) and 60μM MTSL-(S20C)ΔN6 with 60μM ^{15}N-mβ$_2$m (open red symbols). Free MTSL binds in the DE and BC loops of mβ$_2$m and is also responsible for the very weak PRE effect seen with (S20C)ΔN6 and mβ$_2$m. A Similar phenomenon has been observed before for EDTA-based spin labels[386].

Having identified the surfaces involved in the inhibition of fibril assembly (ΔN6-mβ$_2$m) the interaction between ΔN6 and hβ$_2$m that leads to an enhancement of the amyloidogenicity of the latter was investigated using the same methodology, buffers and protein concentrations as described for the ΔN6-mβ$_2$m interaction. The results of these experiments are shown in Figure 5.6.

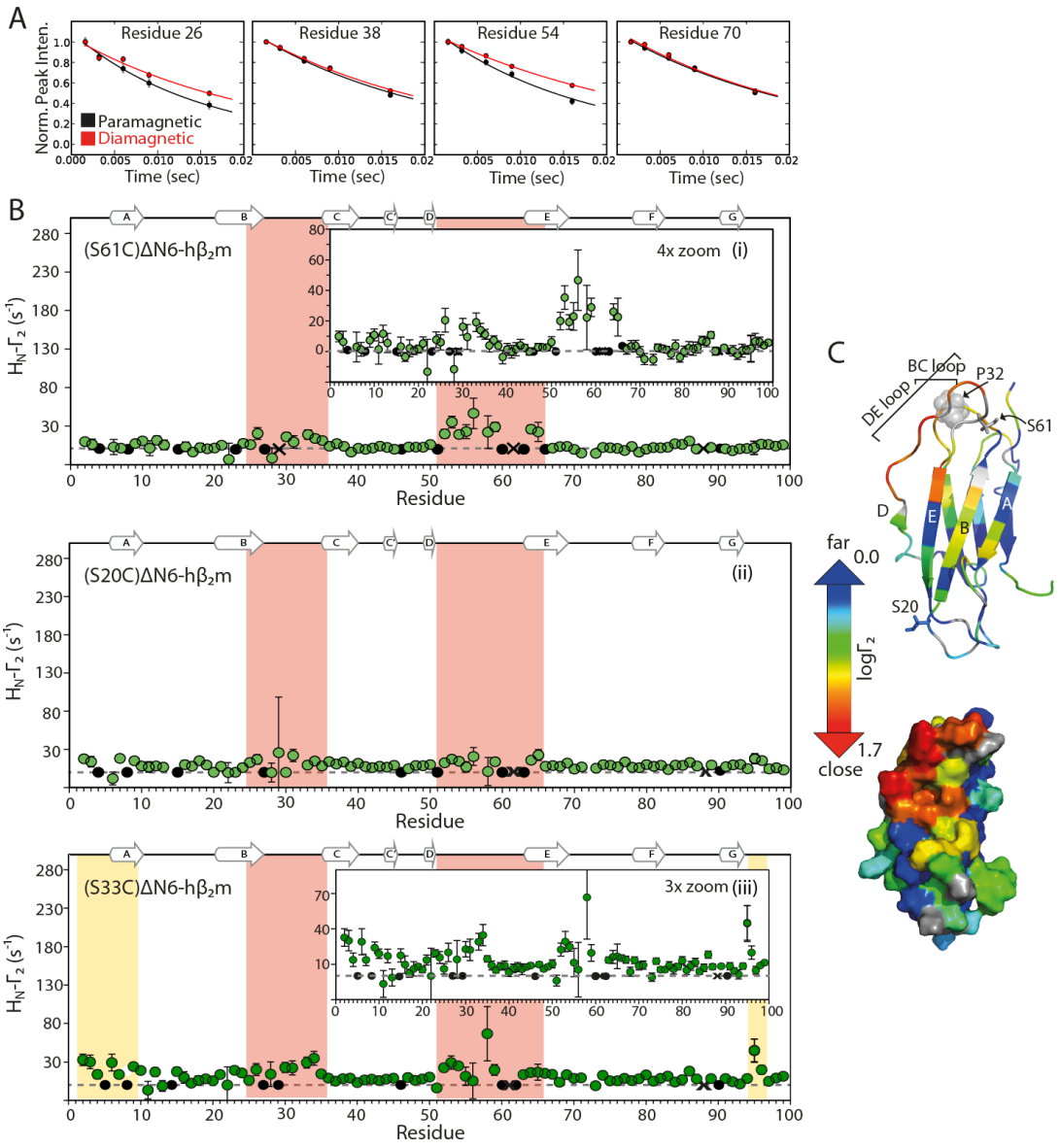

Figure 5.6: Interaction interfaces in promotion of amyloid assembly (ΔN6-hβ2m). (A) Example raw ^{1}H relaxation data for the (S61C)ΔN6-hβ2m interaction are shown for residues 26, 38, 54, 70. Peak intensity is plotted as a function of relaxation time when MTSL is oxidised (paramagnetic-black) and after reduction with ascorbic acid (diamagnetic-red). Solid lines represent fits to single exponentials. (B) Intermolecular PRE profiles for the ΔN6-hβ2m (60μM each) interaction when the spin label is attached on S61 (panel i) or S20 (panel ii) or S33 (panel iii) on ^{14}N-ΔN6. A zoom-in of the first and last diagram is shown as an inset. The BC and DE loop are highlighted in pink boxes. The two additional regions of increased PRE rates detectable only when the MTSL is attached at position 33 (N-terminus and residue 95, 96) are shown in yellow boxes. All data were collected in 10mM sodium phosphate pH 6.2. Black dots represent proline residues or residues that are broadened beyond detection and black crosses denote residue for which the assignments are missing. The secondary structure elements of hβ2m are shown on top of the graphs as ribbons. (C) Structure of hβ2m coloured by the logarithm of the S61-Γ2 rates.

PREs to ^{15}N-hβ_2m were measured using ^{14}N-ΔN6 spin labelled at positions 61, 20 or 33. The H_N-Γ_2 PRE profiles obtained (Figure 5.6, panels i, ii, iii) are similar in pattern to those obtained for the ΔN6-mβ_2m interaction for both S61C-MTSL-ΔN6 and S20C-MTSL-ΔN6, with the largest PREs observed for residues 55-65 in the DE loop and 26-34 in the BC loop, when the spin label is attached at position 61 (Figure 5.6C). Additional residues in the N- (residues 2-10) and C- termini (residues 95, 96 showing increased PRE rates are observed when the spin label is attached at position 33, which were not seen when the MTSL was attached at positions 61 or 20. However, the magnitude of the Γ_2 rates when S61C-MTSL-ΔN6 is mixed with hβ_2m is significantly reduced in comparison to the Γ_2 rates obtained when the same spin-labelled ΔN6 variant is mixed with mβ_2m (compare Figure 5.4B (panel ii) and 5.6B (panel ii). These results suggest that the K_d of the hβ_2m-ΔN6 association is greater than its mβ_2m-ΔN6 counterpart (and hence the population of the bound species is lower and/or the exchange rate of the mβ_2m-ΔN6 dimerisation is faster[210] (Figure 1.14). The different scenarios that can give rise to such a phenomenon were further investigated in section 5.2.3.

Finally, using the same methodology, the interaction between ΔN6 monomers that results in the nucleation of fibril formation was investigated. ^{14}N-labelled ΔN6 with the spin label attached at position 61 or 20 was mixed with ^{15}N-ΔN6 in equimolar concentrations (60μM each) and the Γ_2 rates were measured under identical conditions to those used for the ΔN6-mβ_2m or the ΔN6-hβ_2m interaction. Interestingly, the PRE profiles for the self-association of ΔN6 obtained using S61C-MTSL-ΔN6 are very similar to those observed for the inhibition or promotion of fibril formation, with the residues in the BC and DE loop experiencing high Γ_2 rates as observed for the inhibitory (ΔN6-mβ_2m) or promoting (ΔN6-hβ_2m) interactions (Figure 5.7). However, the interaction sites differ in detail with residues 8-9 and 27-28 in the A and B strands experiencing higher Γ_2 rates, compared with their counterparts in mβ_2m or hβ_2m (compare Figure 5.4B, 5.6B, 5.7A). Most significantly and very different from the other two scenarios, some of the Γ_2 rates obtained for the ΔN6-ΔN6 interaction appear to be independent of the position of the spin label, a phenomenon consistent with non-specific binding or binding with multiple

modes[38-40]. The presence of oligomers in the ΔN6 sample observed using analytical ultra-centrifugation and the increased line broadening discussed in Section 4.2.4 (Figure 4.14) are consistent with the latter explanation.

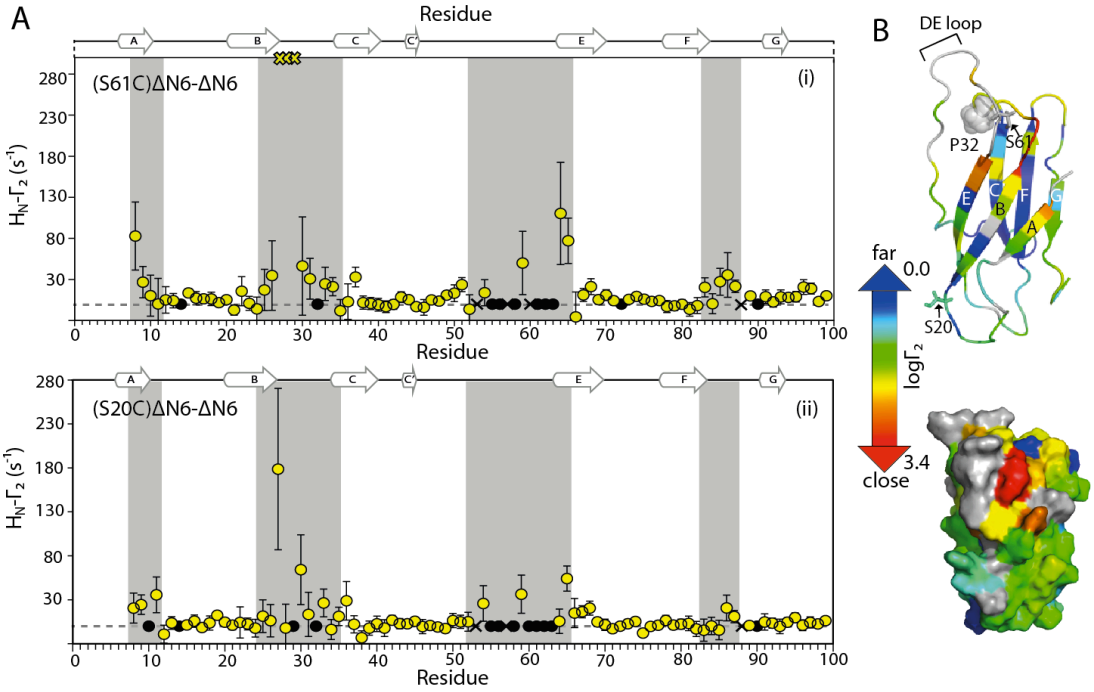

Figure 5.7: Interaction interfaces in nucleation of amyloid assembly (ΔN6-ΔN6). (A) Intermolecular PRE profiles for the ΔN6-ΔN6 (60μM each) interaction when the spin label is attached on position 61 (panel i) or 20 (panel ii) on ^{14}N-ΔN6 at pH 6.2. Significant PREs arising from the C20 site are observed which are unlikely to be caused from background binding of MTSL (compare with Figure 5.5). Yellow crosses indicate residues for which the Γ_2 rate is either too large to appear on this scale or that are broadened beyond detection when the spin label is oxidised and therefore the Γ_2 rate cannot be measured. Black dots represent proline residues or residues that are broadened beyond detection in a normal HSQC spectrum and black crosses denote residues for which the assignments are missing. The secondary structure elements of ΔN6 are shown on top of the graphs as ribbons. (B) Structure of ΔN6 coloured by the logarithm of the S61-Γ_2 rates.

As shown in Figure 4.2 ΔN6 exhibits pH-dependent aggregation and dynamics and therefore the PRE experiment for the ΔN6-ΔN6 interaction was performed at pH 8.2, conditions under which ΔN6 is not amyloidogenic and shows reduced conformational dynamics (see Figure 1.21 and 4.2). The results of these experiments shown in Figure 5.8 revealed that no significant PRE is detectable at this high pH using labelled protein at positions 61 and 20, in marked contrast to the data shown in Figure 5.7 (obtained at pH 6.2). These results confirm that the interactions observed at pH 6.2 are relevant to amyloid aggregation and are

not simply elastic random collisions between the protein molecules. The results therefore suggest that ΔN6 undergoes pH-dependent intermolecular interactions that are possibly at the origin of its increased or decreased ability to form fibrils.

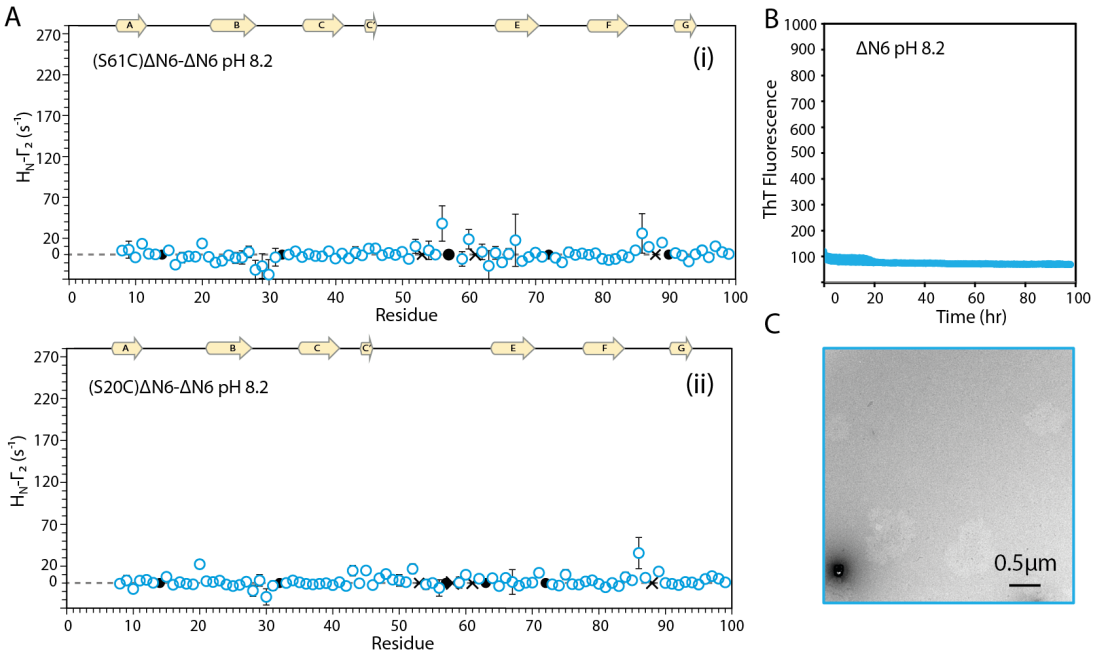

Figure5.8: pH-dependent interactions for ΔN6.
(A) PRE profiles for the ΔN6-ΔN6 interaction at pH 8.2 when the spin label is attached at position 61 (panel i) or 20 (panel ii). Black dots represent proline residues or residues that are broadened beyond detection and black crosses denote residues for which the assignments are missing. The secondary structure elements of ΔN6 are shown on top of the graphs as ribbons. (B) Aggregation kinetics of 60μM ΔN6 in 10mM sodium phosphate pH 8.2, 600rpm followed by ThT fluorescence. (C) Electron micrograph of the end point of the reaction (after 100h) shown in B.

Importantly, the presence of MTSL does not alter the properties of the biomolecular recognition in solution as shown by comparing the chemical shifts of ^{15}N-mβ$_2$m or ^{15}N-hβ$_2$m in the presence of MTSL-labelled ^{14}N-(S61C)ΔN6 or ^{14}N-ΔN6 (Figure 5.9).

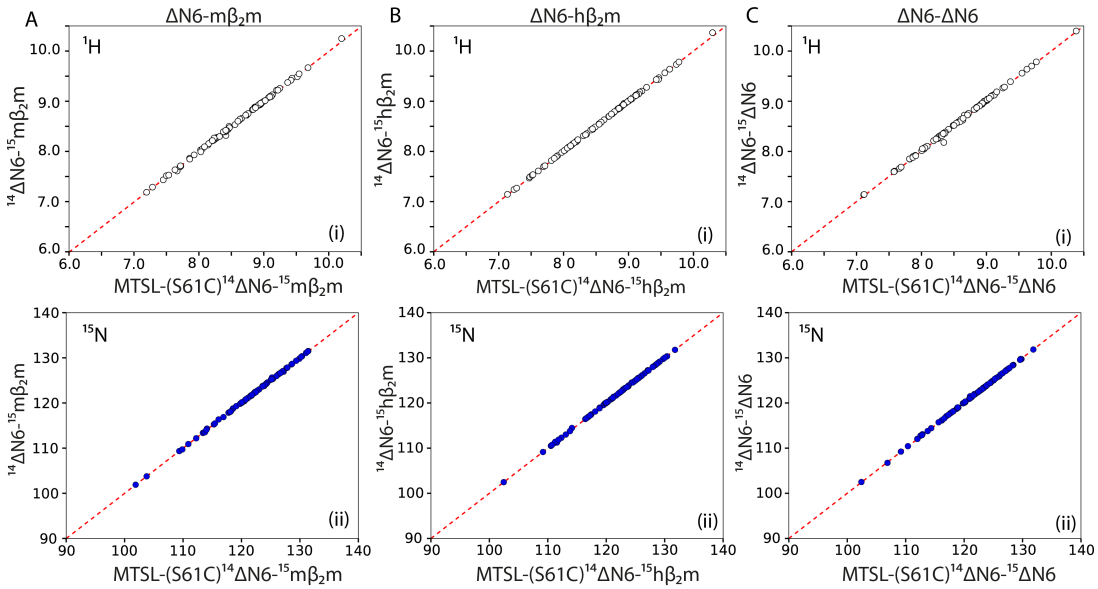

Figure 5.9: MTSL does not alter the interface.
(A) ^{1}H (panel i) and ^{15}N (panel ii) chemical shift correlation plots for the chemical shifts of ^{15}N-mβ$_2$m when mixed with MTSL-labelled ^{14}N-(S61C)ΔN6 (x axis) or with ^{14}N-ΔN6 (y axis). (B) As in A but for ^{15}N-hβ$_2$m. (C) As in A but for ^{15}N-ΔN6. Red dashed lines correspond to linear fits to the data.

5.2.3 Measuring the affinity of the ΔN6-mβ$_2$m and the ΔN6-hβ$_2$m interactions

Following from the PRE data presented in Section 5.2.2, further investigation was carried out using unmodified versions of the proteins in order to characterise the nature of the binding site and determine the K_d of the different interactions. When ^{15}N-mβ$_2$m was mixed with ^{14}N-ΔN6 changes in the ^{1}H-^{15}N HSQC spectrum were obvious even when the proteins were mixed in substoichiometric ratios. These included a global decrease in peak intensity, which was more pronounced for peaks in the BC and DE loops (Figure 5.10). This observation is consistent with the PRE data presented in Figure 5.4, showing that the same residues experience increased PRE rates, suggesting that these parts of the molecule are in the dimer interface. The global loss of intensity can be attributed to the generation of species with increased tumbling time, also consistent with dimer formation as also shown by AUC in Figure 4.14. On the other hand when ^{15}N-hβ$_2$m was mixed with ^{14}N-ΔN6 changes in the ^{1}H-^{15}N HSQC spectrum were observed only in samples of equimolar ratios or excess of ^{14}N-ΔN6. The differential line broadening for peaks in the BC and DE loops is detectable in this case as well, but the global loss of the intensities is

not as profound as observed for the mβ2m-ΔN6 interaction. These data suggest a reduced affinity of ΔN6 for hβ2m in comparison to mβ2m (Figure 5.10).

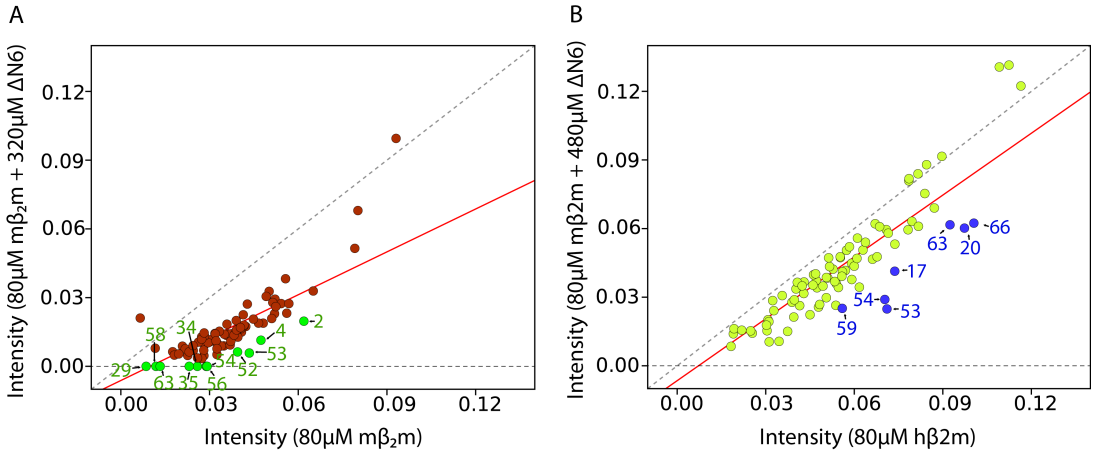

Figure 5.10: Line shape analysis of the mβ2m-ΔN6 or hβ2m-ΔN6 interactions. (A) Correlation of the peak intensities of a sample comprising 80μM ^{15}N-mβ2m (x axis) versus the peak intensities of sample of 80μM ^{15}N-mβ2m mixed with 320μM ^{14}N-ΔN6. The red solid line represents a linear fit to the data and the black dashed line has a slope of 1 and is drawn to guide the eye. Residues showing increased line broadening are coloured green. (B) As in A but for 80μM ^{15}N-hβ2m versus 80μM ^{15}N-hβ2m mixed with 480μM ^{14}N-ΔN6. Residues showing enhanced line broadening are coloured blue. Note the different slope of the red lines in A and B, indicative of an increased global loss of intensity for the mβ2m-ΔN6 interaction.

The interaction of ΔN6 with both hβ2m and mβ2m was found to be in the fast to intermediate exchange regime, thus giving rise to small but significant chemical shift changes upon protein association (Figure 5.11).

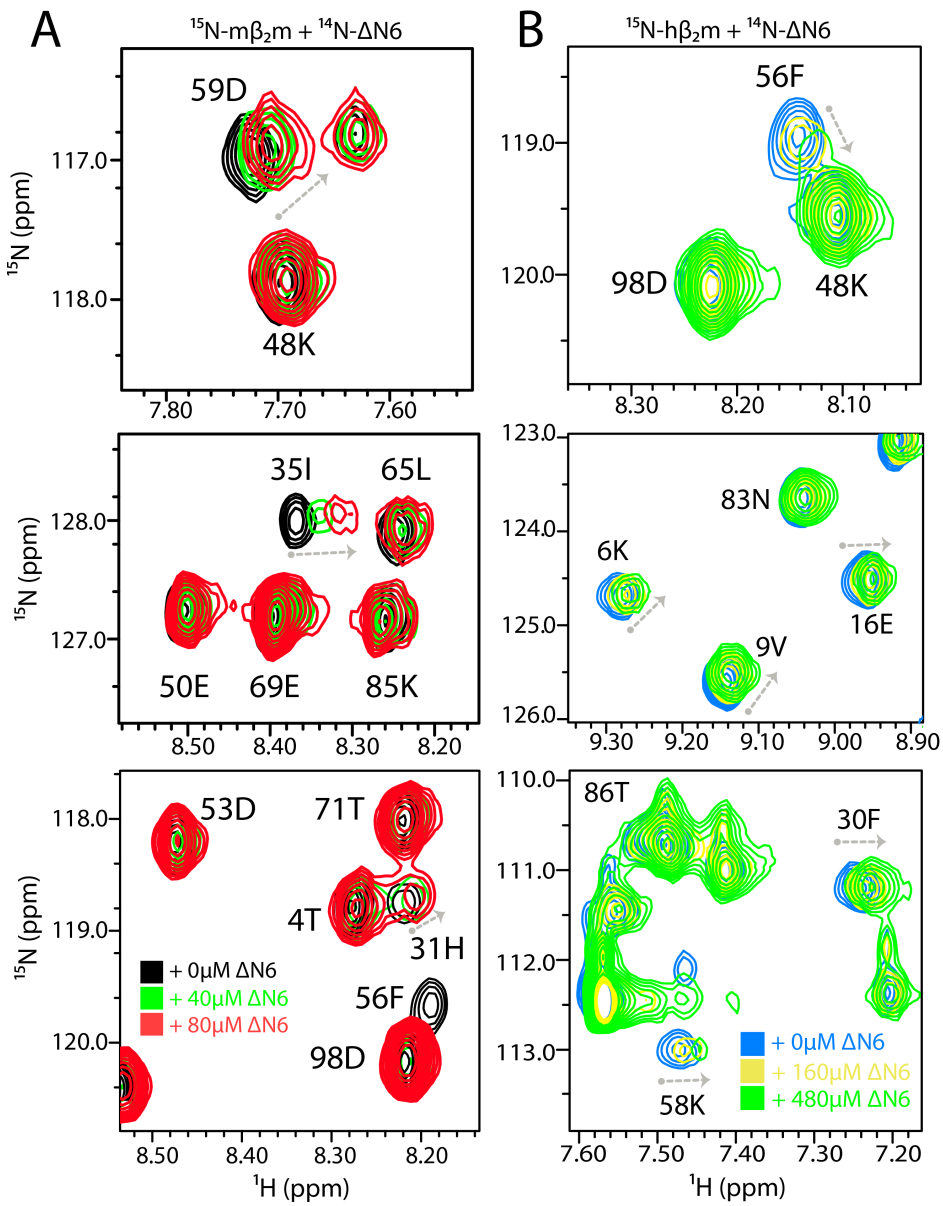

Figure 5.11: Chemical shift perturbations upon binding.
(A) Zoom-in in regions of ^{1}H-^{15}N HSQC spectra of 80μM of ^{15}N-mβ$_2$m alone (black) or 80μM of ^{15}N-mβ$_2$m mixed with 40μM ^{14}N-ΔN6 (green) or with 80μM ^{14}N-ΔN6 (red). Residues that show significant chemical shift differences are noted with grey arrows. Residue 56F is already broadened beyond detection when the proteins are mixed in substoichiometric ratios. (B) Zoom-in in regions of ^{1}H-^{15}N HSQC spectra of 80μM of ^{15}N-hβ$_2$m alone (blue) or 80μM of ^{15}N-hβ$_2$m mixed with 160μM ^{14}N-ΔN6 (yellow) or with 480μM ^{14}N-ΔN6 (green). Residues that show significant chemical shift differences are noted with grey arrows.

A full titration of ^{14}N-ΔN6 into either ^{15}N-mβ$_2$m or ^{15}N-hβ$_2$m was performed as described in Section 2.3.10.7 and the total chemical shift perturbation (CSP) data calculated using the formula in eq. 21 (Section 2.3.10.7) were globally fitted to binding curves to extract binding affinities. This analysis revealed an

apparent K_d for the inhibitory mβ2m-ΔN6 interaction of ~68μM, while the affinity of the amyloid promoting hβ2m-ΔN6 interaction is ~7 fold decreased, resulting in a K_d of ~500μM (Figure 5.12).

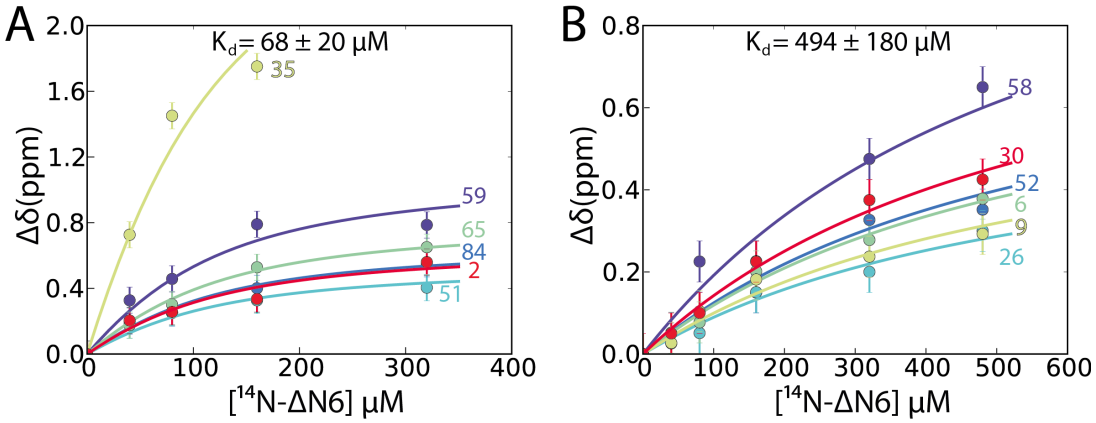

Figure 5.12: Altered binding affinities for the mβ2m-ΔN6 and hβ2m-ΔN6 complexes. (A) Binding curves for residues of mβ2m that show chemical shift differences upon addition of ΔN6. The total CSP is calculated using the formula in eq. 21, Section 2.3.10.7). Data for residues 2, 35, 51, 59, 84, 85 and the indole proton of W60 were globally fitted using 5 concentrations of ^{14}N-ΔN6 (0, 40, 80, 160, 320μM). (B) Binding curves for the titration of ^{14}N-ΔN6 into ^{15}N-hβ2m. Residues 6, 9, 26, 30, 52, 56, 58 and 97 were used for the fitting against 6 points (0, 40, 80, 160, 320, 480μM of ^{14}N-ΔN6).

The chemical shift differences upon addition of ^{14}N-labelled ΔN6 to ^{15}N-labelled mβ2m located in the BC and DE loops of the latter, in agreement with the PRE data shown in Figure 5.4. This finding provides further evidence that the chemical modification of ΔN6, by the addition of MTSL does not alter the properties of the mβ2m-ΔN6 interaction (Figure 5.13A). On the other hand, the residues of hβ2m that show significant chemical shift differences upon interaction with ΔN6 are more spread out on the structure of hβ2m, while residues in the A strand of hβ2m (but not of mβ2m) show increased chemical shift differences upon binding (this phenomenon reveals differences in the two inferfaces and will be discussed in Section 5.2.6) (Figure 5.13B). In this analysis the protein concentrations used were adjusted so the same % bound of ^{15}N-labelled protein (~42%) exists in solution in both cases, in order to allow direct comparison of the data.

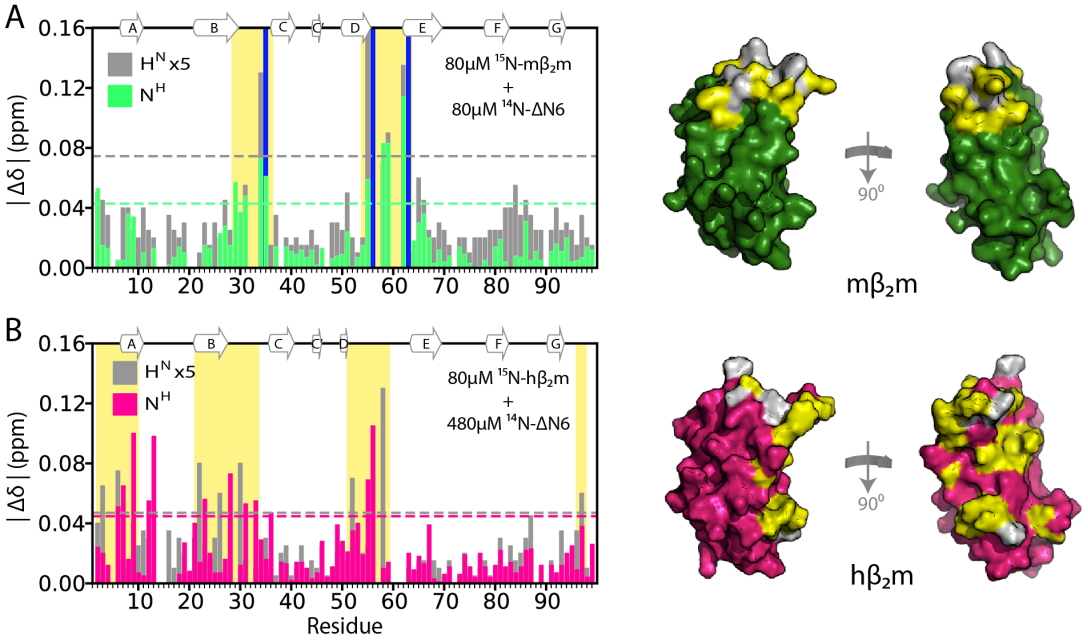

Figure 5.13: Histograms of chemical shift differences upon binding.
(A) ^{1}H (grey) and ^{15}N (green) chemical shift differences for resonances of ^{15}N-labelled mβ_2m upon interaction with ^{14}N-labelled ΔN6 in a 1:1 molar ratio (80 μM each- ~30% mβ_2m bound). All chemical shift differences locate to the top half of the molecule (BC and DE loops) shown as a surface representation (orange). Red bars denote residues that show significant CSP upon addition of 40μM ΔN6, but are broadened beyond detection due to exchange line broadening when the concentration of ΔN6 is increased to 80μM. (B) ^{1}H (grey) and ^{15}N (pink) chemical shift differences when ^{15}N-labelled hβ_2m and ^{14}N-labelled ΔN6 are mixed in a 1:6 molar ratio (80μM : 480μM - ~30% hβ_2m bound). Residues that show high PRE rates are highlighted in yellow boxes and shown in yellow on the structure of hβ_2m. Dotted lines represent 2 standard deviations of the mean over the entire dataset.

The calculation of the binding affinities allows the estimation of the % bound ^{15}N-labelled protein in the PRE experiments presented in Figures 5.4, 5.6. For the mβ_2m-ΔN6 interaction and using 60μM of both proteins, the bound population of mβ_2m is calculated to ~36%, while in the hβ_2m-ΔN6 complex the % bound hβ_2m is ~9.8%. Extrapolating the (S61C)ΔN6-hβ_2m PRE data to a population of bound hβ_2m of 36% produces a profile that is similar in amplitude to the data obtained for the (S61C)ΔN6-mβ_2m interaction (Figure 5.14), showing that the difference in amplitude of the PRE rates predominantly is due to differences in the population of the bound complex and not to altered exchange rates between the subunits of the two complexes (Figure 5.14).

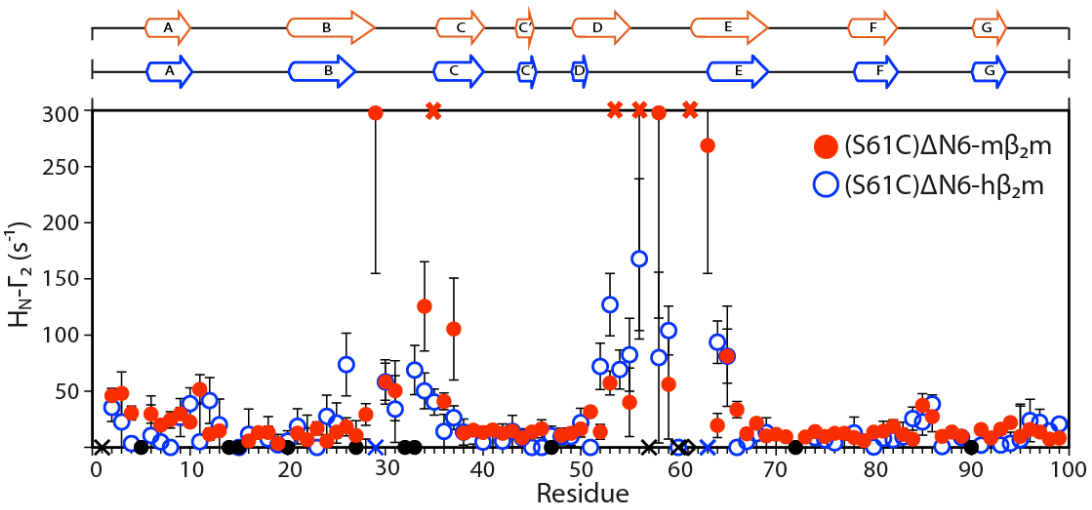

Figure 5.14: The PRE amplitude depicts differences in the population of bound species. The PRE profile for the (S61C)ΔN6-mβ$_2$m interaction is shown in red dots reporting on a population of bound ^{15}N-mβ$_2$m of 25%. The PRE data for the (S61C)ΔN6-hβ$_2$m interaction shown in Figure 5.6B (panel i) originating from 6.2% of bound ^{15}N-hβ$_2$m , are here extrapolated to a population of bound ^{15}N-hβ$_2$m of 25% and are shown as blue open dots.

To investigate the relationship between the population of the bound complex with the amplitude of the PRE data further, intermolecular PRE experiments were carried out using three different concentrations of MTSL-labelled(S61C)-ΔN6, which was mixed with ^{15}N-mβ$_2$m in molar ratios 1:1, 1:3 and 1:15 (ΔN6:mβ$_2$m). The exact concentrations used were as follows, 60μM ΔN6 : 60μM mβ$_2$m (~36% bound mβ$_2$m) , 50μM ΔN6 : 150μM mβ$_2$m (~21% bound mβ$_2$m) and 10μM ΔN6 : 150μM mβ$_2$m (~4.5% bound mβ$_2$m) The full dataset obtained is shown in Figure 5.15A. As expected the observed PRE rates reduce in amplitude as the population of the bound ^{15}N-mβ$_2$m reduces from ~36% to ~21% and ~4.5%. The PRE data for residues 1-12 in the N terminus and the A strand were summed for each MTSL-ΔN6 concentration, and are plotted against the %bound ^{15}N-mβ$_2$m in Figure 5.15B (PREs only from the region were chosen since they display the smallest error in the measurement). The resulting curve can be fitted to a binding isotherm with a K_d of 68μM, showing that the amplitude of the PRE rates follows the thermodynamics of binding.

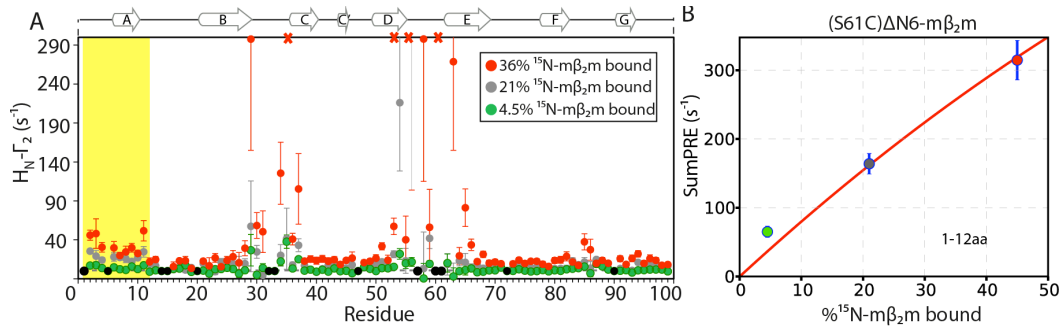

Figure 5.15: Concentration-dependent PREs for the ΔN6-mβ$_2$m interaction. (A) Intermolecular PRE profiles obtained using ^{15}N-labelled mβ$_2$m and MTSL-(S61C)ΔN6 : 60μM ^{15}N-mβ$_2$m(red), or 50μM MTSL-(S61C)ΔN6 : 150μM ^{15}N-mβ$_2$m(grey) or 10μM MTSL-(S61C)ΔN6 : 150μM ^{15}N-mβ$_2$m(green). Data for the N-terminus and the A strand are highlighted in a yellow box. The secondary structure elements of mβ$_2$m are shown on top of the graph as cartoons. (B) PRE data for residues 1-12, arising upon titration of ^{15}N-mβ$_2$m with MTSL-(S61C)ΔN6. Data for the specified residues at each titration point were summed and are plotted using the same colour as in A. The solid red line represents a binding curve using the calculated K_d of 68μM for the ΔN6-mβ$_2$m association.

5.2.4 Quantitative structural investigation of protein-protein interactions in the early stages of amyloid formation

The PRE data presented in Figures 5.4, 5.6 were analysed in a quantitative way by performing simulated annealing calculations in torsion angle space in order to minimise the agreement between the experimentally observed and back-calculated Γ_2 rates. To generate a dataset suitable for this analysis, PREs arising from position 61 and 33 for the ΔN6-hβ$_2$m interaction were measured in two independent experiments and the average PRE value for each residue was used for fitting. For the ΔN6-mβ$_2$m interaction, resonances in the BC and DE loop disappear in the oxidized spectrum when the spin label is attached at positions 33 and 61. To obtain an estimate for the Γ_2 rate for these residues, the PRE experiments (using spin-labelled ΔN6 at positions 33 and 61) were repeated with varied protein concentrations to: 1) improve the signal-to-noise ratio and 2) reduce the concentration of the bound complex so that a more accurate Γ_2 rate can be measured (titration data for (S61C)ΔN6-mβ$_2$m are shown in Figure 5.15). PREs were then extrapolated to their values for a 1:1 complex using the measured K_d of 68μM and the resulting dataset was used for quantitative analysis of the structural properties of the resulting complex.

The docking protocol developed to visualise the way the proteins interact in solution is described in Section 2.2.11. Briefly it consists of three steps: 1) starting from random positions the proteins 'come together' as rigid bodies, by performing an energy minimisation of the chemical shifts observed in the interface, 2) a simulated annealing run is performed using the PRE and the chemical shift data as restraints with the side-chains left free to rotate, and 3) the final step consists of an energy minimisation in torsion angle space. Section 5.2.4.1 will present the data for the inhibitory (ΔN6-mβ_2m) and 'catalytic' (ΔN6-hβ_2m) interactions while Section 5.2.4.2 will discuss the ΔN6-ΔN6 interaction.

5.2.4.1 Inhibition of amyloid formation by suppression of conformational states

Using a single conformer representation of the inter-converting species (while the MTSL side chain is represented by a 5-membered ensemble), reasonable agreement between the observed and the calculated PRE data was achieved for both the inhibitory and the amyloid-promoting interactions. For the ΔN6-mβ_2m association a Q factor (eq. 25, Section 2.3.11) of 0.52 was obtained for the best scoring structure, which decomposes to 0.41 and 0.65 for the data arising from position 33 and 61 respectively (Figure 5.16A-C). The structure displaying the lowest Q factor shows a head-to-head association with the DE loops of the two monomers participating in the interface (Figure 5.16D). Reasonable convergence, with the 10 lowest energy structures sharing a backbone RMSD of 4Å between their mβ_2m subunits suggest that the structure shown in Figure 5.16D represents the main heterodimer species in solution. The data obtained when the MTSL was attached at position 20 were not used during fitting as they were considered to be due to non-specific MTSL binding (see Figure 5.5) and thus their calculated values should be (close to) zero.

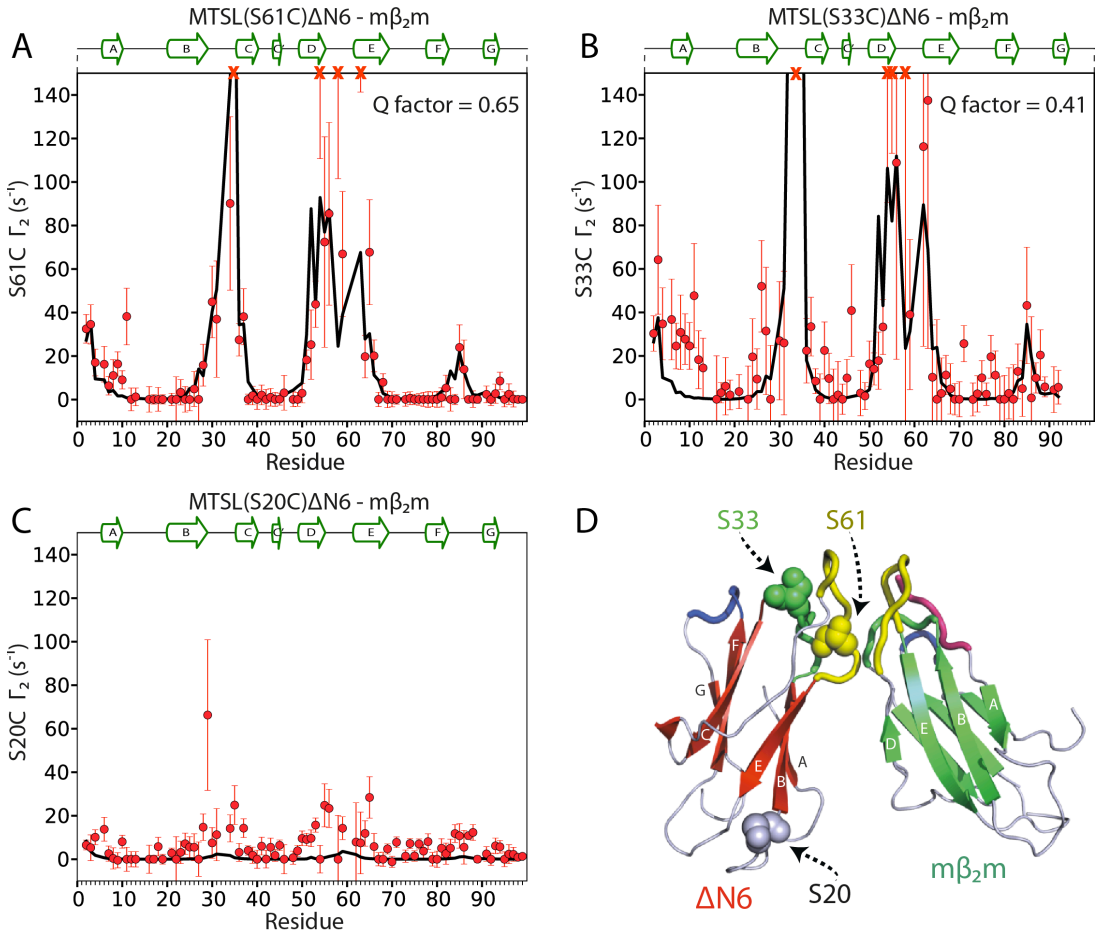

Figure 5.16: Single conformer representation (N=1) for the ΔN6-mβ2m interaction. PRE data fitting for the ΔN6-mβ2m interaction when the spin label is attached at position 61 (A), 33 (B) or 20 (C). Black lines represent the fit to the data and red dots the experimentally measured Γ_2 rates. Data arising from position 20 were not used in the fitting (see Figure 5.5). Red crosses denote residues that show large PRE rates that do not fit in the scale used here. (D) The lowest Q factor dimer structure of the ΔN6-mβ2m heterodimer. ΔN6 and mβ2m are shown in a cartoon representation (red and green colour respectively). The DE loop of mβ2m is highlighted in blue, the BC loop in green and the FG loop in pink. The positions at which the spin label is attached are shown as spheres on ΔN6.

Close analysis of these data, however, reveals significant outliers. Specifically, residues in the DE loop of mβ2m for both the C33 and C61 sites and in the N-terminal region when MTSL is attached at position 33 do not follow the predicted PRE pattern (Figure 5.16A, B).

In the case of the interaction of ΔN6 with hβ2m a single conformer representation (N=1) yields an overall Q factor of 0.65 (0.73 for the data from position 33 and 0.41 for the position 61 respectively) (Figure 5.17A, B). Again,

the lowest Q factor structure shows a head-to-head dimer that shares a similar overall configuration with the ΔN6-mβ$_2$m (Figure 5.17D). However, terrible convergence (the backbone RMSD between the hβ$_2$m subunits of the 10 lowest energy structures is calculated at 20Å) suggests that there is not only structure in solution. The majority of the intermolecular contacts are again mediated through the DE lops of both monomers as shown in Figure 5.17D. Outliers locate in the N-terminal 10 residues for both 33 and 61 sites and the DE loop or residues 95, 96 in the C-terminus when the spin label is attached at position 33. Again, data arising from position 20 were not used during fitting and their calculated values should be (close to) zero (Figure 5.17D). The PRE data for the inhibitory ΔN6-mβ$_2$m and the promoting ΔN6-hβ$_2$m interaction were calculated using the same population of bound species to allow direct comparison (36% in both cases).

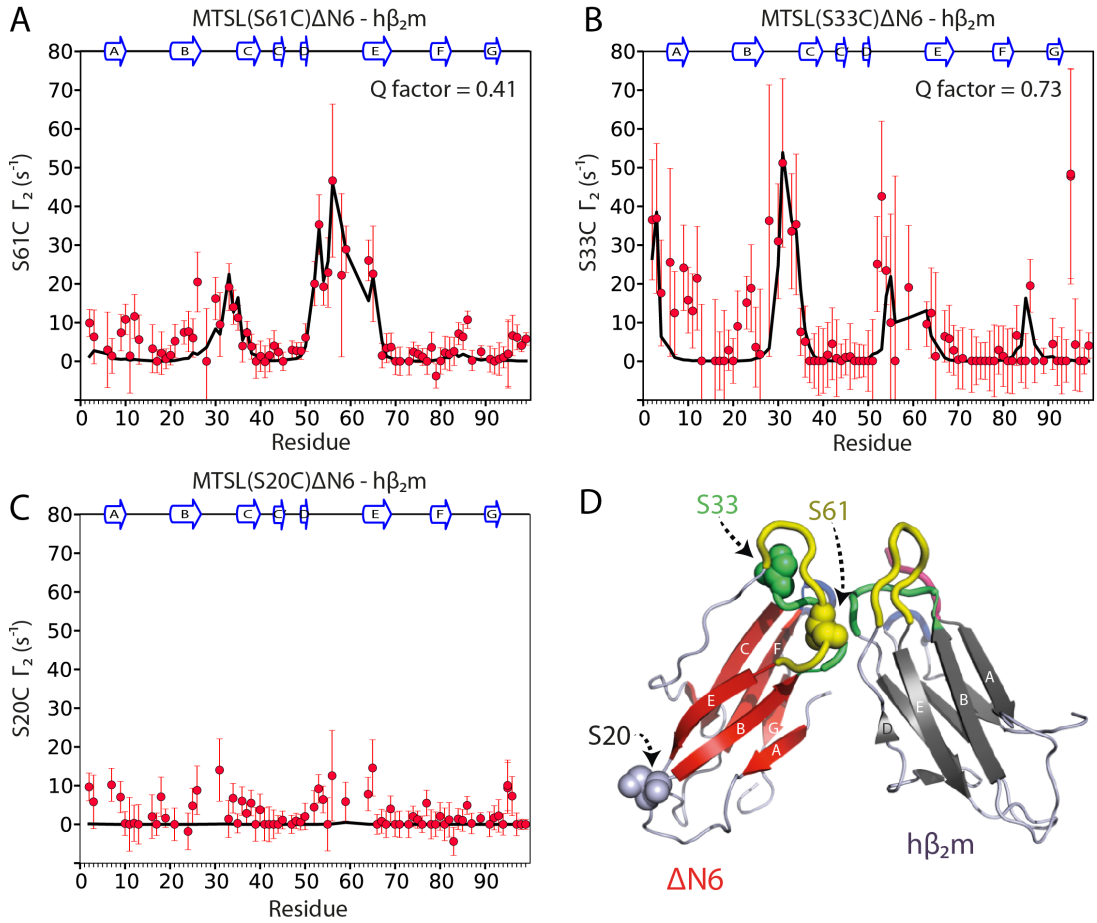

Figure 5.17: Single conformer representation (N=1) for the ΔN6-hβ$_2$m interaction. PRE data fitting for the ΔN6-hβ$_2$m interaction when the spin label is attached at position 61 (A), 33 (B) or 20 (C). Black lines represent the fit to the data and red dots denote he experimentally measured PRE rates. Data arising from position 20 were not used in the fitting. (D) The lowest Q factor dimer structure of the ΔN6-hβ$_2$m heterodimer. ΔN6 and hβ$_2$m are shown in a

cartoon representation (red and green colour respecitvely). The DE loop of $h\beta_2m$ is highlighted in blue, the BC loop in green and the FG loop in pink. The positions at which the spin label is attached are shown as spheres on ΔN6.

The results shown in Figures 5.16, 5.17 show that a single dimer conformation is sufficient to describe the PRE data adequately well. However, to fully describe the protein association in both cases, an ensemble of species is needed. This is not surprising since the system under investigation is complex and involves aggregation-prone molecules. The theoretical framework described in Section 2.1.2 allows the ensemble averaging between multiple conformations of the interacting species, if those are in fast exchange between each other (see Figure 1.14, Section 1.5.4.2 and Section 2.1.2). Thus, averaging between two (N=2), three (N=3) or six (N=6, only for the ΔN6-$h\beta_2m$ interaction) alternating states was carried out as described in reference [209]. For the ΔN6-$m\beta_2m$ interaction a rapid decrease in the Q factor is observed when N is increased to 2 (Q=0.39) with no further significant decrease when N=3 (Q=0.38) as monitored by both Qe and Qee (see Section 2.3.11) which almost overlay. Notably a Q factor of 0.32 would be obtained if the quality of the fit was comparable with the error of the measurement (Figures 5.18A and 5.19A panels i-iii). In the case of the ΔN6-$h\beta_2m$ association, a similar behaviour is observed for Qee, which plateaus after N=2, but not for Qe which is significantly higher in all cases. This observation reveals that the entire ensemble (100 structures) is required to describe the PRE data for the ΔN6-$h\beta_2m$ association and can be explained by the stochastic orientation of the $h\beta_2m$ subunit inside the ensemble (Figures 5.18B and 5.19B). On the other hand an individual ΔN6-$m\beta_2m$ ensemble member can describe the data more readily (as Qee and Qe overlay). Thus, an ensemble size of 2 was chosen for further analysis (PRE fits are shown in Figure 5.20). Further analysis to identify whether the improvement in the fit is statistically significant or can be improved by consideration of even larger ensembles of conformers will be needed.

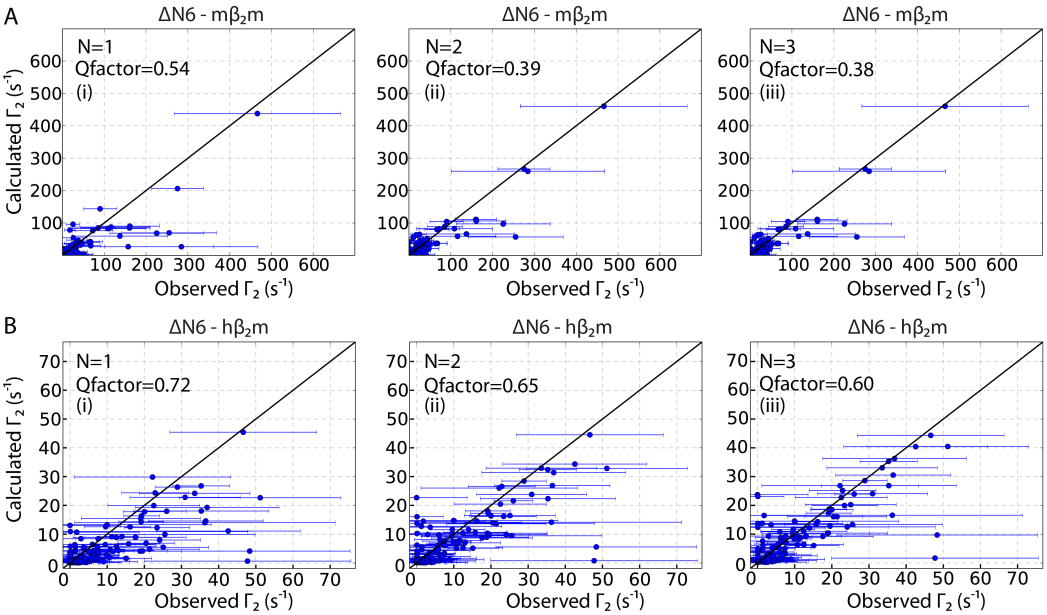

Figure 5.18: Multiple conformer representation (N=2 or N=3) of β2m interactions. Correlation plots between the observed and the calculated Γ_2 rates for the ΔN6-mβ2m (A) and the ΔN6-hβ2m (B) interactions. The single conformer representation (N=1) is shown in panels i, two conformers (N=2) in panels ii and three conformers are shown in panels iii. The calculated Γ_2 rates shown represent the average values (per residue), back-calculated from 50 independent calculations. All ensemble members were equally weighted during the calculations.

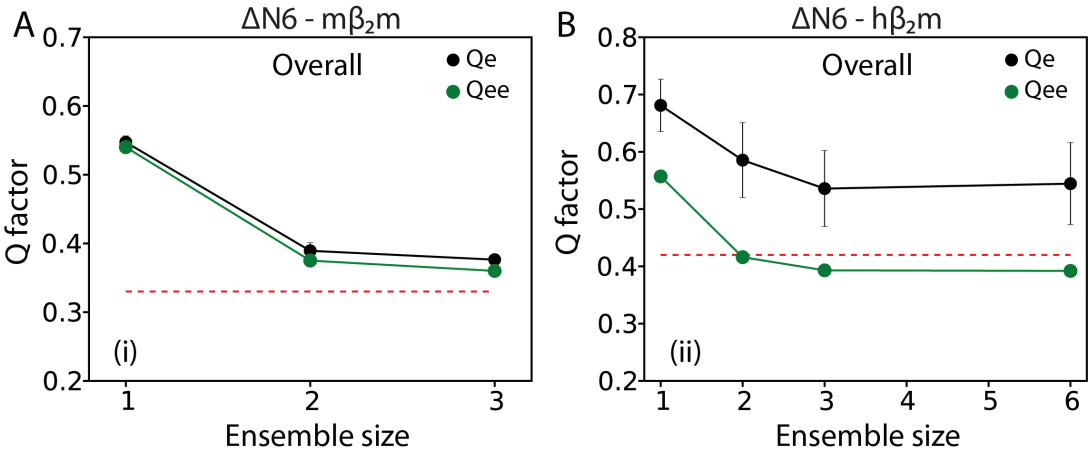

Figure 5.19: Dependence of the Q factor on the ensemble size.
Plots of the calculated Q factor versus the size of the ensemble of interacting species for the ΔN6-mβ2m (A) and the ΔN6-hβ2m (B) interactions. The overall Q factor is shown. The Q factors shown represent the average values (per residue) back-calculated from 50 independent calculations (Qe) or the Q factor of the average dataset (Qee) and the error bars denote one standard deviation of the mean. Dashed red lines represent the best Q factor possible obtained for each dataset, if the quality of the fit is comparable to the error of the experiment and is calculated as described in[386]. Briefly, errors for each PRE rate were multiplied by a random number between -1 and 1 and this product was then added to the measured PRE rate. A Q factor was then calculated

between the experimentally measured data and this 'artificial' dataset. This process was repeated 100 times and the average Q factor is reported here (red lines).

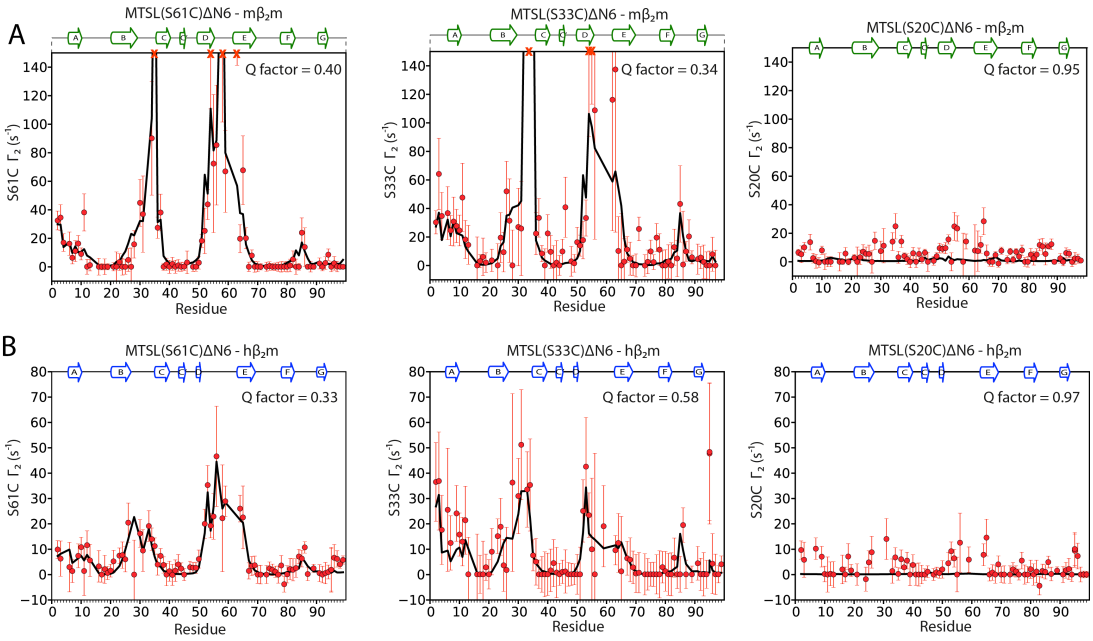

Figure 5.20: Agreement between experimental and back-calculated PREs. PRE data fitting for the ΔN6-mβ2m interaction (A) or for the ΔN6-mβ2m interaction, when the spin label is attached at position 61 (left), 33 (middle) or 20 (right) using an ensemble size of two (N=2). Black lines represent the fit to the data and red dots denote he experimentally measured PRE rates. Data arising from position 20 were not used in the fitting.

The ensemble calculations presented so far assume an equal population for each of the ensemble members. However, this might not be the case. The data shown in Figures 5.17 and 5.18, where the ensemble size is 1, show that the majority of the PRE data can be explained using a single conformation and thus the population of the additional state(s) must be low. To test this hypothesis, a series of calculations was performed by adjusting the population of the major and minor species using an ensemble size of 2. As shown in Figure 5.21 a higher population of the minor species is needed to optimise the fit to the ΔN6-hβ2m data (the Q factor flattens at a population of ~20% for the minor species) compared with the data for the ΔN6-mβ2m interaction (the Q factor plateuaes at ~10%). Cross-validation with statistical tools (such as jack-knife calculations) will be needed to avoid over-fitting of the data.

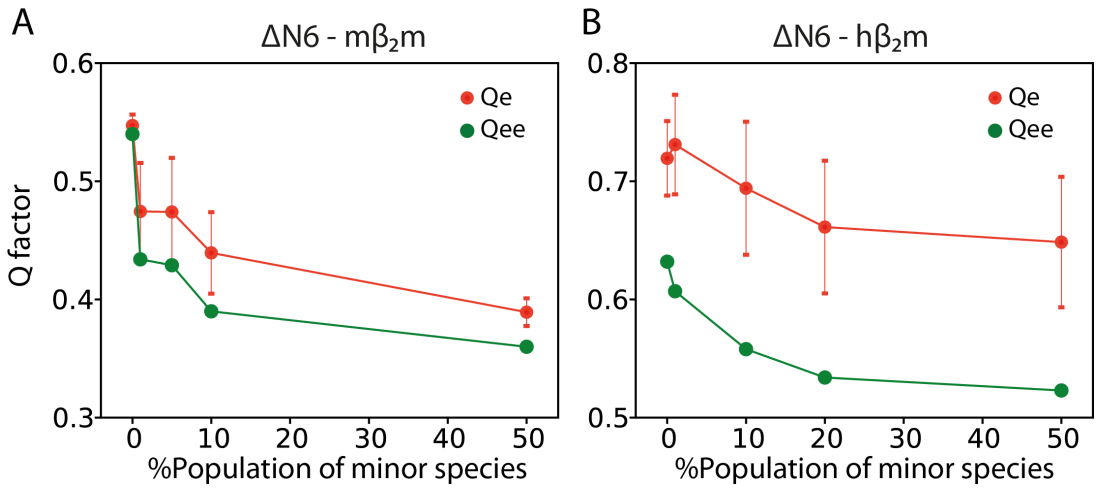

Figure 5.21: Population of alternative protein states.
The dependence of the Q factor on the population of the minor species is shown using an ensemble size of 2 for the ΔN6-mβ$_2$m (A) and the ΔN6-hβ$_2$m (B) interactions. The Q factors shown represent the average values back-calculated from 50 independent calculations (Qe) or the Q factor of the average dataset (Qee) and the error bars denote one standard deviation of the mean.

The ensemble calculations, in contrast with the traditional NMR structure calculations, do not yield a unique structure. Instead they provide a range of different possible solutions. In the case of a dimer calculation as discussed here, the calculations yield a distribution of possible orientations between the interacting species. The results can be represented most readily in structural terms using an atomic probability density map as described in references [69,426]. The distribution of mβ$_2$m around ΔN6 using an ensemble size of 2 is shown in Figure 5.21A and the complementary picture of ΔN6 molecules around mβ$_2$m is shown in Figure 5.21B. A summary of the data is shown in Figure 5.22C as a contact map.

As seen in Figure 5.22, mβ$_2$m molecules cluster around the DE loop of ΔN6, which is the only region of the molecule that makes significant contacts with mβ$_2$m. On the other hand ΔN6 shows a characteristic bimodal distribution around the DE loop of mβ$_2$m, with one cluster of molecules facing the β-sheet comprised of the A, B, E and D strands, while the second cluster of ΔN6 molecules locates opposite the edge strands D and C.

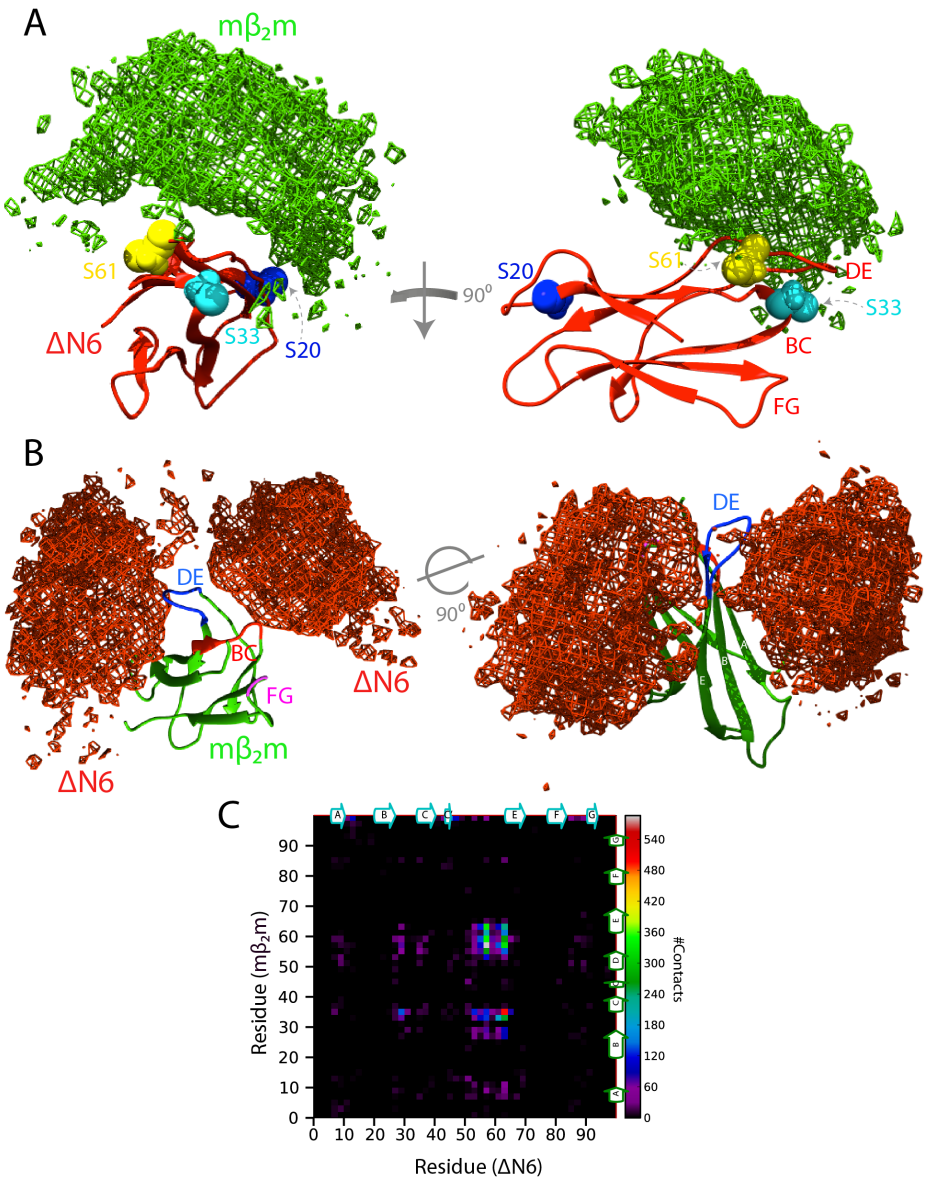

Figure 5.22: Orientational distributions for the association of ΔN6 with mβ₂m. Structures of the top 50 scoring ensembles (N=2, 2*50 structures). (A) ΔN6 is shown as red cartoon and the distribution of mβ₂m molecules around it is shown as green mesh. The positions of the spin label are highlighted in spheres. (B) Mβ₂m is shown as green cartoon and ΔN6 in red mesh. The regions showing increased PRE rates are shown in different colour on the structure of mβ₂m (DE loop-blue, BC loop-red and FG loop-magenta). All density maps are drawn with a cut-off of 40%. (C) Contact map for the 50 best scoring ensembles. Every non-hydrogen atom with an intermolecular distance less than 4Å to any other atom is identified as a hit. The number of contacts for atoms of each residue is colour-coded as shown in the colour-bar.

The same analysis was performed for the ΔN6-hβ₂m interaction (Figure 5.23).

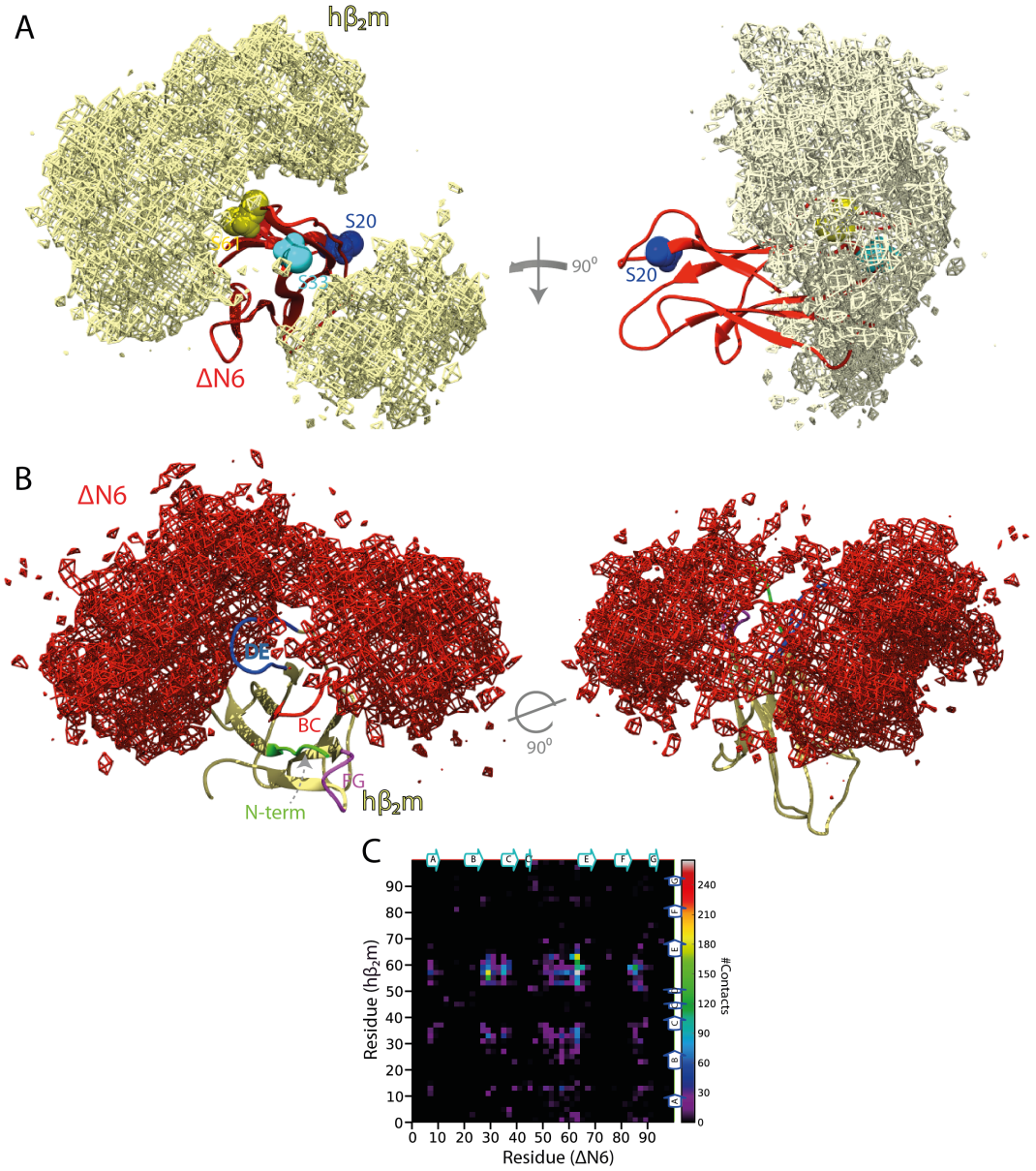

Figure 5.23: Orientational distributions for the association of ΔN6 with hβ₂m. Structures of the top 50 scoring ensembles (N=2, 2*50 structures). (A) ΔN6 is shown as red cartoon and the distribution of hβ₂m molecules around it is shown as grey mesh. The positions of the spin label are highlighted in spheres. (B) Hβ₂m is show as grey cartoon and ΔN6 in red mesh. The regions showing increased PRE rates are shown in different colour on the structure of hβ₂m (DE loop-blue, BC loop-red and FG loop-magenta, N-terminus-green). All density maps are drawn with a cut-off of 40%. The pose of ΔN6 is the same as the one in Figure 5.21 to allow comparison. (C) Contact map for the 50 best scoring ensembles. Every non-hydrogen atom with an intermolecular distance less than 4Å to any other atom is identified as a hit. The number of contacts for atoms of each residue is colour-coded as shown in the colour-bar.

The analysis revealed the DE loops of both molecules as the epicentre of the ΔN6-hβ₂m interaction, in a similar manner as it was observed for the inhibitory

ΔN6-mβ$_2$m collision. However, the angle of hβ$_2$m density in relation to ΔN6 is significantly altered in comparison with the orientation that mβ$_2$m adopts when bound to ΔN6 (compare green and grey meshes in Figures 5.22A, 5.23A). Figure 5.24 shows an overlay of the density maps of mβ$_2$m and hβ$_2$m around ΔN6.

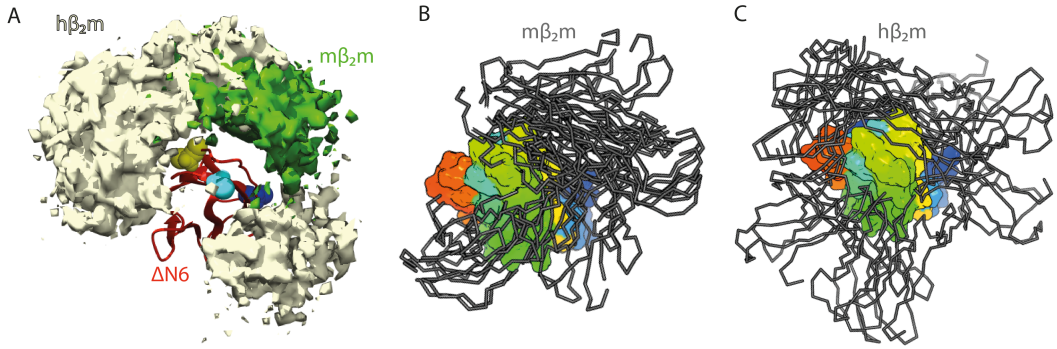

Figure 5.24: Differences in the orientational distributions of the ΔN6-hβ$_2$m and the ΔN6-mβ$_2$m interactions. (A) An overlay of the electron density maps (grey-hβ$_2$m, green-mβ$_2$m) shown in Figures 5.21A and 5.22A. (B) Structures of the 5 best scoring ensembles of the ΔN6-mβ$_2$m association. ΔN6 is shown as rainbow-coloured surface and mβ$_2$m as ribbons. (C) As in (B) but for hβ$_2$m. An ensemble size of 2 was used in all cases.

The amyloid promoting ΔN6-hβ$_2$m interaction seems to be much more heterogeneous, expanding to both sides of the apical region of ΔN6 (around Pro32), while the inhibitory interaction is more localised. The volume of the ΔN6-mβ$_2$m density map is calculated at 7157Å^3, while the ΔN6-hβ$_2$m map is almost twice as big (13670Å^3, a cut-off of 40% was used in both cases). Interestingly, the distributions of mβ$_2$m and hβ$_2$m around ΔN6 do not overlay, an observation, which might be important for the outcome of fibril assembly. New areas showing high intermolecular contacts (BC and FG loops and the A strand of both ΔN6 and hβ$_2$m) appear in the contact map shown in Figure 5.23C. This phenomenon is also depicted in the chemical shift perturbation data, in which the chemical shifts differences upon binding are larger and/or more spread out throughout the sequence of both molecules when ΔN6 interacts with hβ$_2$m compared with mβ$_2$m (Section 5.2.6). The space sampled by ΔN6 molecules around hβ$_2$m does not show the characteristic bimodal distribution observed in Figure 5.22B, but is more uniformly distributed around the apical region of hβ$_2$m (Figure 5.23B).

A striking correlation between the hydrophobic surface of mβ$_2$m and the space sampled by ΔN6 is observed, revealing that the interaction is mainly of hydrophobic nature (Figure 5.25). On the other hand the relation between hβ$_2$m's hydrophobic surface and the molecular distribution of ΔN6 is not obvious. Some of the density locates opposite to the hydrophobic region of the DE loop in hβ$_2$m, created mainly by the contributions of F56 and W60, but some ΔN6 molecules seem to be attracted by the negative charge of residues D34, E36, D38 (Figure 5.26). Importantly, D34 and D38 are not present in mβ$_2$m.

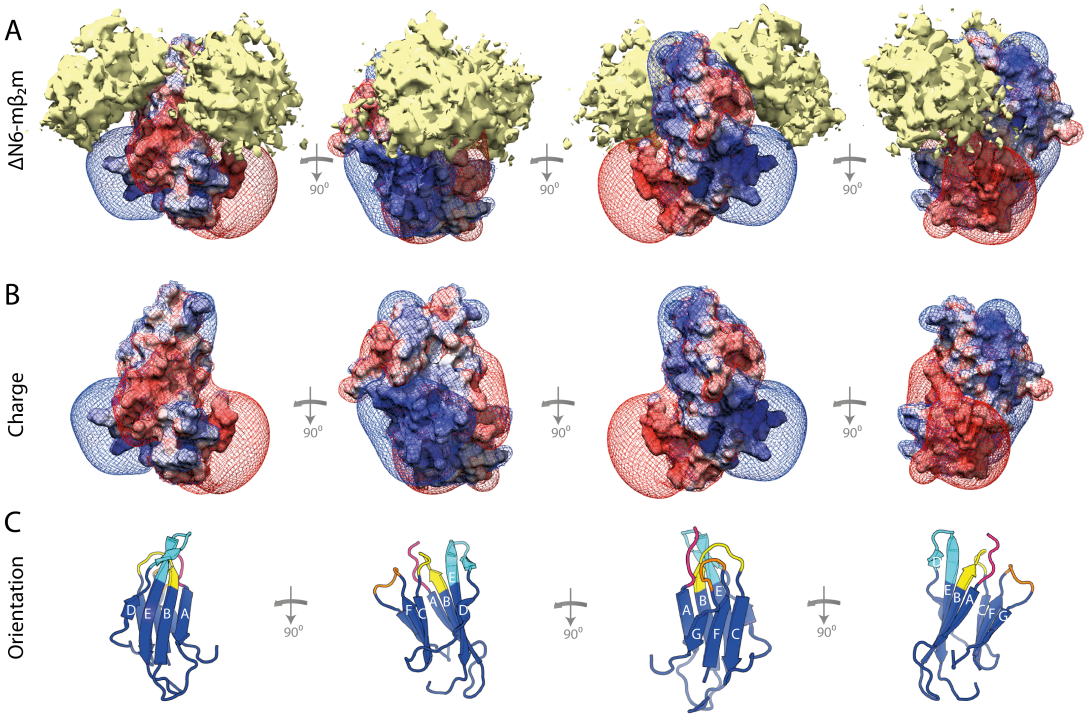

Figure 5.25: Hydrophobic interactions between mβ$_2$m and ΔN6.
(A) The distribution of ΔN6 molecules around mβ$_2$m is shown as yellow surface. The surface of mβ$_2$m is coloured based on the electrostatic potential of the molecule calculated as ±2kT (blue positive, red negative, white no charge). The electrostatic potential isosurface is also shown as mesh (absence of contour lines-no charge). The distribution of ΔN6 molecules calculated using an ensemble of size of two (N=2). (B) Mβ$_2$m alone, coloured according to its electrostatic potential, is shown in the same poses an in A. (C) A cartoon representation of mβ$_2$m is shown in the same pose as in (A) and (B). The DE loop is coloured in cyan, the BC loop in yellow, the FG loop in orange and the N-terminus in pink.

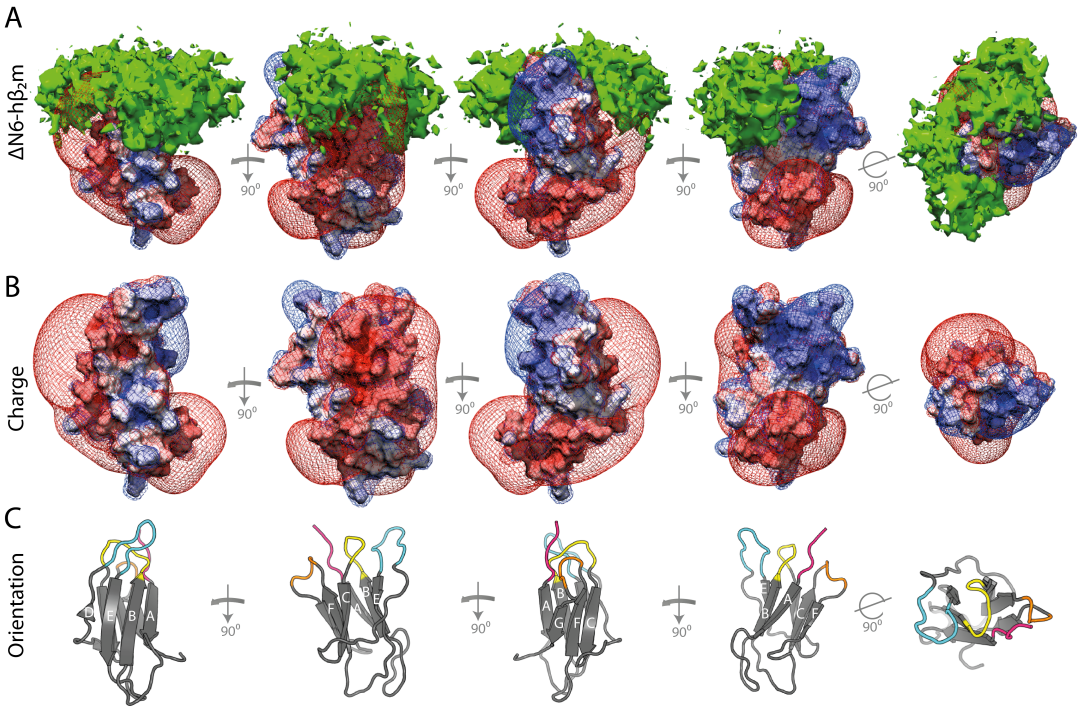

Figure 5.26: Interaction surfaces between hβ2m and ΔN6.
(A) The distribution of ΔN6 molecules around hβ2m is shown as green surface. The surface of hβ2m is coloured based on the electrostatic potential of the molecule calculated as ±2kT (blue positive, red negative, white no charge). The electrostatic potential isosurface is also shown as mesh (absence of contour lines-no charge). The distribution of ΔN6 molecules calculated using an ensemble of size of two (N=2). (B) Hβ2m alone, coloured according to its electrostatic potential, is shown in the same poses an in A. (C) A cartoon representation of hβ2m is shown in the same pose as in (A) and (B). The DE loop is coloured in cyan, the BC loop in yellow, the FG loop in orange and the N-terminus in pink. All the poses of the molecules shown here are exactly the same as those used in Figure 5.24 for mβ2m (plus one added).

Together, the data presented here suggest that inhibition of ΔN6 fibril formation involves a specific head-to-head protein association driven by key hydrophobic interactions through residues in the DE loops of ΔN6 and mβ2m, while the amyloid-promoting ΔN6-hβ2m collision, although also adopting a head-to-head configuration, is more heterogeneous involving a larger interaction surface and potentially a larger (in terms of size and population) ensemble of interacting species. Furthermore, the differences in the charge distribution between mβ2m and hβ2m (Figure 1.24) seem to be crucial for the recognition by ΔN6. Whether these data reflect the formation of a range of encounter complexes between ΔN6-hβ2m that is not observed for the ΔN6-mβ2m interaction or whether they report on the transient oligomerisation between ΔN6-hβ2m, which is suppressed by the presence of mβ2m (as also

shown by AUC in Figure 4.14) remains to be determined. The structural ensembles presented here, reflect the dynamic nature of weak protein association, but to test the uniquiness, additional studies with more PRE spin labels or RDC data will be required.

5.2.4.2 Nucleation of fibril assembly involves multiple binding sites

The PRE data for the self-association of ΔN6 molecules shown in Figure 5.7 showed significant PREs when MTSL is attached at position 20 which cannot be attributed to non-specific, MTSL-driven binding as judged by their amplitude (Figure 5.7). This finding suggests that the loops distal to the BC and DE loops in the apical part of the protein (AB, CD, EF loops) are also taking part in the interface. Fitting of all the PRE data (sites 61 and 20) together typically resulted in good fitting of the data arising from position 61, but poor fitting for the C20 site (Figure 5.27A). This is not unexpected since PREs from position 61 have larger amplitude and thus produce significantly higher forces during the calculation. To resolve this problem and allow a more complete description of the binding event data arising from position 20 were fitted separately from the dataset with MTSL at position 61 (note that this treatment is different than a ensemble calculation) (Figure 5.27B). This analysis showed that a head-to-head dimer with the DE loops in the interface fits the C61 data with good accuracy, while a head-to-tail interaction is adequate to describe the C20 data (Figure 5.27C).

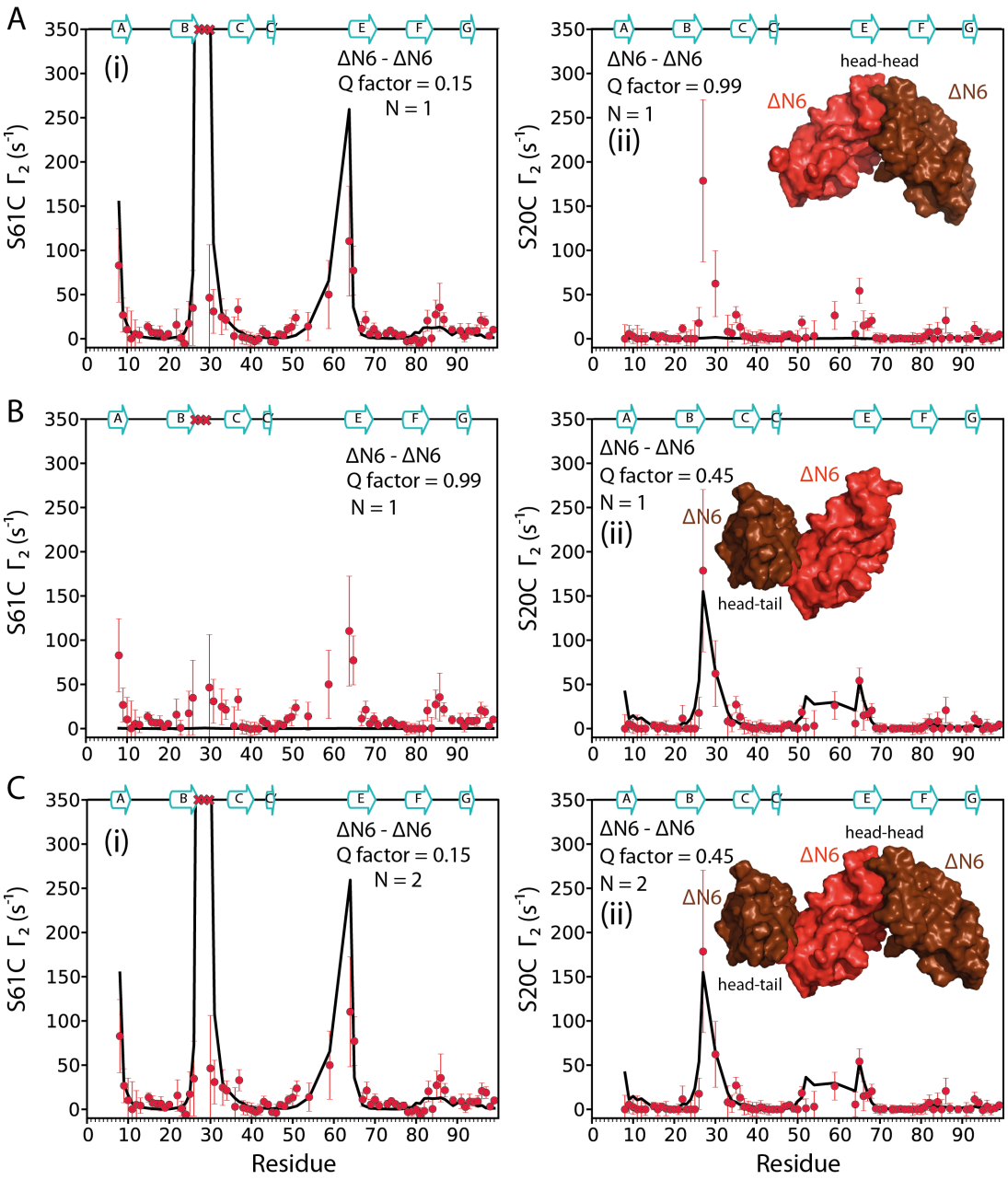

Figure 5.27: Agreement between observed and back-calculated PRE data for the ΔN6 self-association. (A) Experimental (red dots) and back-calculated (black line) PRE rates for the data obtained with MTSL attached at position 61 (panel i) or 20 (panel ii), using a single conformer representation (N=1). The lowest energy homodimer structure (in a head-to-head configuration) is shown as inset. (B) Single conformer representation using only the data arising from position 20. The lowest energy structure, showing a head-to-tail dimer is shown as inset. (C) When put together, these two dimer configurations describe all the PRE data with reasonable accuracy.

Together the data in Figure 5.27 suggest that the self-association of ΔN6 is complex and involves multiple site binding. Therefore, the PRE data possibly give clues on the structure of the ΔN6 oligomers observed by AUC (Figure

4.14). However, due to low convergence, the structures shown in Figure 5.26 are rough models and might deviate from the true structure of a ΔN6 polymer. Further, more detailed and complex studies will be required to validate the models proposed for this interaction.

5.2.5 Controlling the fate of amyloid inhibition by site-directed mutagenesis of interface residues in mβ$_2$m

The data presented in Figure 5.25 revealed that the inhibitory ΔN6-mβ$_2$m interaction is hydrophobic in nature with a number of aromatic residues in the DE loops of both partners locating in the interface. More specifically, residues F56 and W60 of mβ$_2$m showed high PRE rates using spin-labelled ΔN6 at positions 61 and 33 and were always found to form part of, or locate near to, the interface, independently of the parameters used in the structure calculation protocol (Figure 5.28).

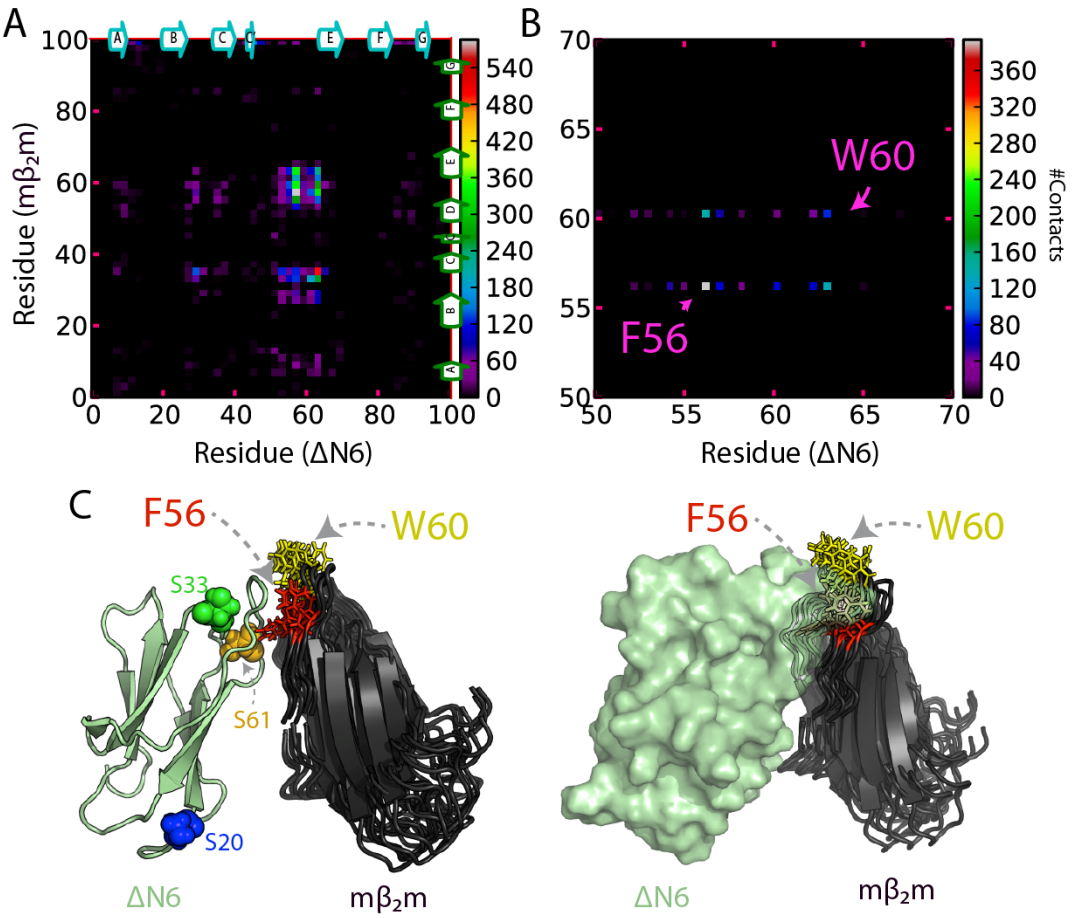

Figure 5.28: Hydrophobic residues in the ΔN6-mβ$_2$m interface.
(A) Contact map between ΔN6 and mβ$_2$m. Each pixel in this 2D-histogram represents a pair of non-hydrogen atoms for which the intermolecular distance

is less than 4Å. The frequency of contacts between each residue is plotted in the Z axis and is coloured coded according to the colourbar shown on the right. A set of 100 independent ensemble calculations with an ensemble size of 2 (2*100 structures in which the orientation of the monomers may differ significantly) was analysed. The secondary structure elements of ΔN6 are shown as blue cartoons on top of the panel and those of mβ$_2$m as green cartoons on the side of the panel. (B) Shows a zoom-in in the residues 50-70 shown in A and only for the residues F56 and W60 of mβ$_2$m. Those two residues participate in the majority of contacts between the DE loops of the monomers (compare the intensity scales of A and B). (C) The 10 lowest energy structures of the ΔN6-mβ$_2$m heterodimer are shown (ΔN6 green cartoon/surface, mβ$_2$m black cartoon). This converged structure could represent the major species in solution. F56 is highlighted in red sticks and W60 in yellow sticks. In these structures F56 is completely buried in the interface, while W60 is partially buried.

Taking into account the prevelance of F56 and W60 (of mβ$_2$m) in the ΔN6-mβ$_2$m interface, these residues were chosen to be subjected to site-specific mutagenesis in order to confirm the validity of the structural models and to highlight the importance of the hydrophobic interactions that lead to inhibition of amyloid formation. Both residues were mutated to glutamic acid (creating the F56E/W60E-mβ$_2$m variant) in an effort to interrupt the hydrophobic nature of the interaction by the introduction of charge in the interface. ^{1}H-NMR spectra showed that the mutations introduced do not interrupt the structure of mβ$_2$m, and that F56E/W60E-mβ$_2$m forms a well-folded protein in solution (Figure 5.29A).

NMR was used next, in order to investigate the interaction of the mutant form of mβ$_2$m with ΔN6. When 2 molar equivalents of ^{14}N-labelled mβ$_2$m are mixed with ^{15}N-labelled ΔN6, chemical shift differences are observed for the latter, indicative of protein association, in line with the PRE data shown in Figure 5.4 and the titration data shown in Figure 5.11, 5.12, 5.13. However when the same amount of ^{14}N-labelled F56E/W60E-mβ$_2$m is mixed with ^{15}N-labelled ΔN6 no significant chemical shift changes were observed as exemplified by residues 8, 86 and 97 shown in Figure 5.29B. The small chemical shift differences still observed in Figure 5.29C suggest that the introduction of charge in the interface has reduced the binding affinity between the proteins substantially and hence the population of the bound species, but the proteins (ΔN6 and F56E/W60E-mβ$_2$m) are likely to be able to interact at least to some extent.

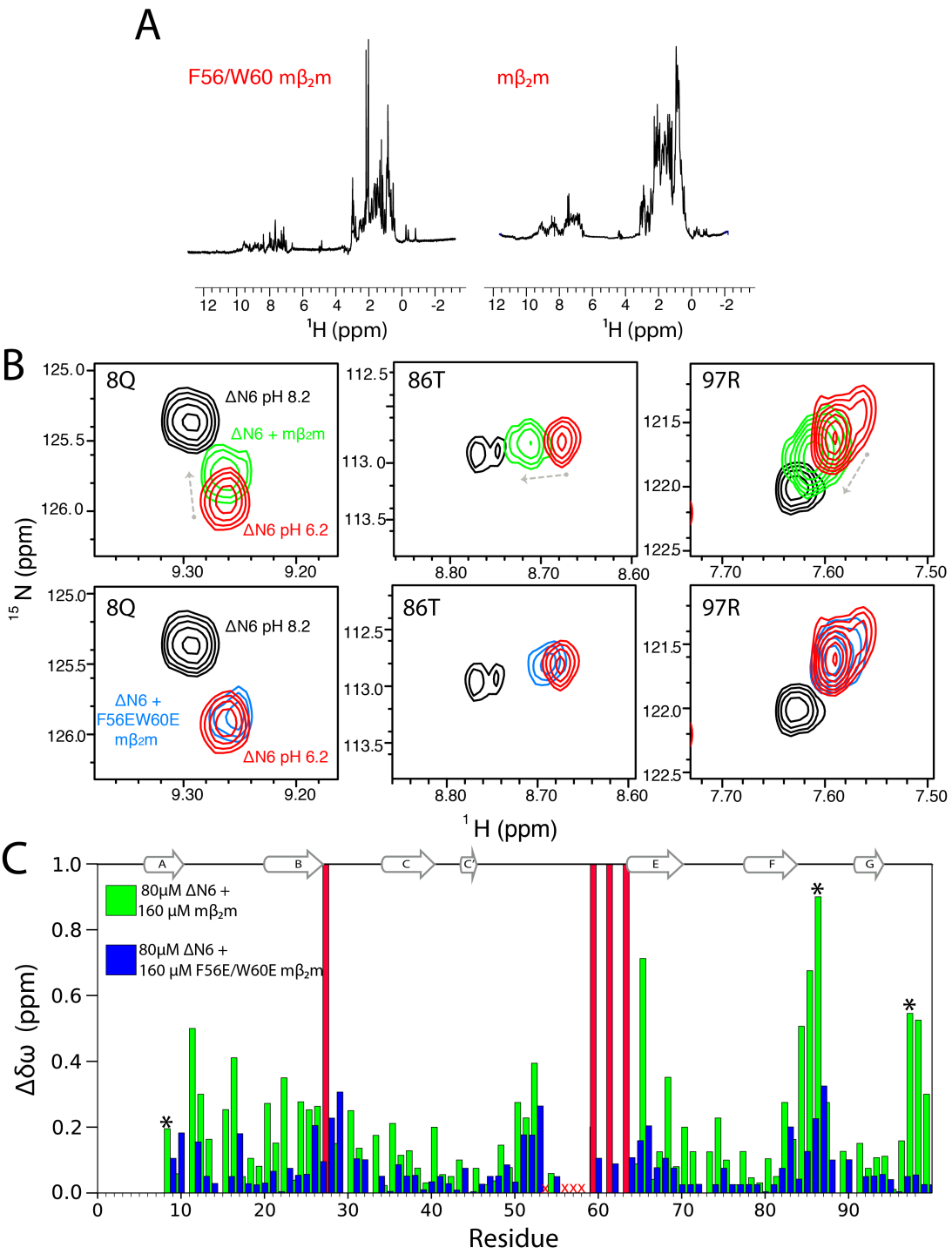

Figure 5.29: Disrupting the mβ$_2$m-ΔN6 interaction by the introduction of charge in the interface. (A) ^{1}H-1D NMR spectra of 200μM F56E/W60E-mβ$_2$m (left) or 500μM mβ$_2$m (right). (B) Zoom-in in areas of the ^{1}H-^{15}N HSQC spectrum of ^{15}N-ΔN6 that show chemical shift changes upon addition of ^{14}N-labelled mβ$_2$m. Addition of ^{14}N-mβ$_2$m (160μM) to 80μM of ^{15}N-ΔN6 shifts the resonances of the latter towards their positions at pH 8.2 where the protein is shown to be non-amyloidogenic (red:80μM ^{15}N-ΔN6, pH 6.2, black:80μM ^{15}N-ΔN6, pH 8.2, green: 80μM ^{15}N-ΔN6 + 160μM ^{14}N-mβ$_2$m, top panels). Addition of equivalent amount of ^{14}N-F56E/W60E mβ$_2$m (blue spectrum) eliminates the observed chemical shift differences – bottom panels. (C) Histograms of the total CSP (calculated using the formula in eq. 21 of resonances of 80μM ^{15}N-

labelled ΔN6 upon addition of 160μM ^{14}N-labelled mβ$_2$m (green) or 160μM ^{14}N-labelled F56E/W60E-mβ$_2$m (blue). Residues that are broadened beyond detection when ΔN6 is mixed with mβ$_2$m are shown in red bars. Residues shown in (A) are marked with an asterisk. Secondary elements of ΔN6 are shown on top of the graph as cartoons.

Data presented in Figure 4.14 showed that the biomolecular collision between mβ$_2$m-ΔN6 is responsible for the destruction of the pre-amyloid ΔN6 oligomers and thus causes the delay of amyloid formation. Since F56E/W60E mβ$_2$m was found less prone to interaction with ΔN6, as shown in Figure 5.29, its effect on the oligomer distribution and the aggregation kinetics of ΔN6 was investigated by AUC and ThT fluorescence, respectively. Addition of F56E/W60E mβ$_2$m in equimolar amounts to ΔN6 (60μM each) did not cause the disappearance of the ΔN6 oligomers, which persist at pH 6.2 (Figure 5.30A). Importantly, F56E/W60E mβ$_2$m itself does not aggregate at pH 6.2 – Figure 5.30B. As discussed in Section 4.2.2.1 the addition of 2-fold molar excess of mβ$_2$m to ΔN6 (40μM mβ$_2$m-20μM ΔN6) results in a retardation of the lag time of fibril formation for more than 120h (green trace in Figure 5.30C). However, incubation of ΔN6 under identical solution conditions but in the presence of a 2-fold molar excess of F56E/W60E mβ$_2$m does not cause any delay in the kinetics of fibril formation as shown by a lag time of ~30h for this mixture of proteins (blue trace in Figure 5.30C). Notably, (shorter) amyloid fibrils were detected in the presence of F56E/W60E mβ$_2$m (Figure 5.30D).

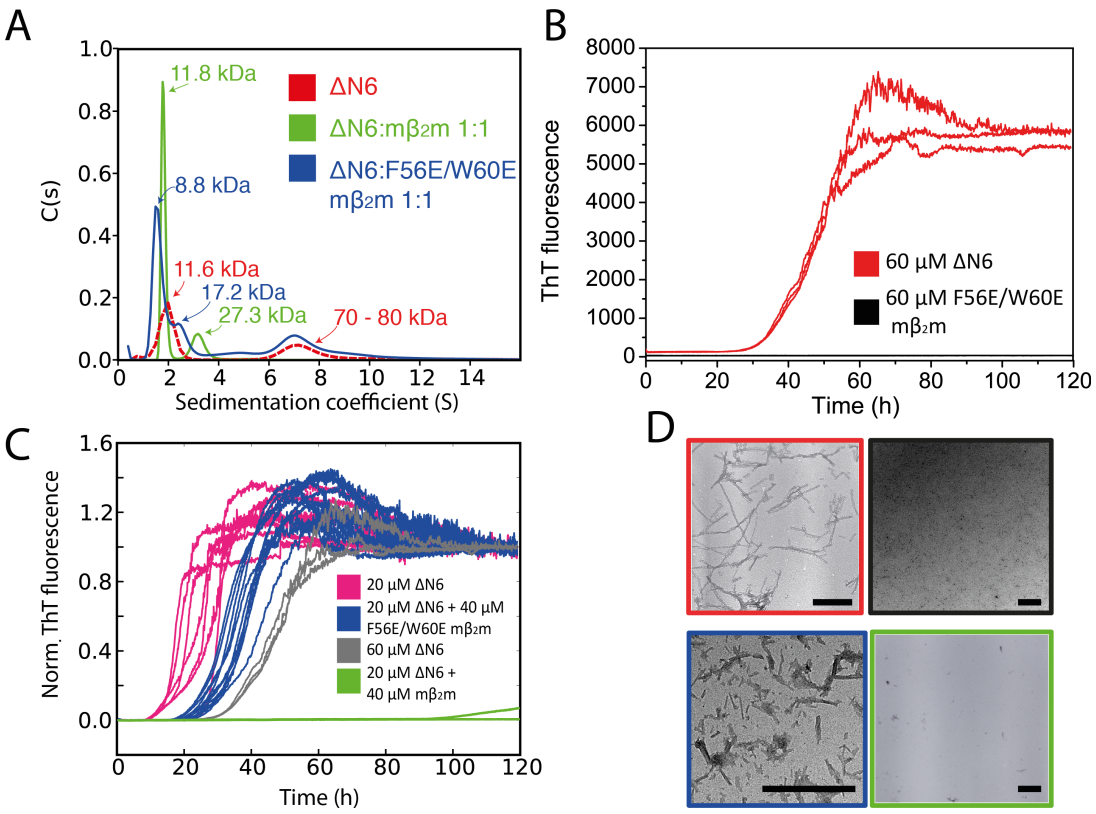

Figure 5.30: Controlling the assembly of ΔN6 amyloid fibrils by mutation. (A) Sedimentation velocity AUC of ΔN6 alone (red), ΔN6 mixed with mβ$_2$m (green) or ΔN6 mixed with F56E/W60E mβ$_2$m (blue) in 1:1 molar ratio (60μM each) in 10mM sodium phosphate buffer pH 6.2, 83mM NaCl, 25°C. The estimated molecular weight of each peak is noted on top, using the same colour code as AUC traces. (B) Aggregation kinetics of 60μM ΔN6 (red) or of 60μM of F56E/W60E mβ$_2$m in 10mM sodium phosphate buffer pH 6.2, 83mM NaCl, 37°C, 600 rpm. (C) Aggregation kinetics of 20μM ΔN6 alone (pink) or of 60μM ΔN6 alone (grey) or of 20μM of ΔN6 mixed with 40μM F56E/W60E mβ$_2$m (blue) or of 20μM of ΔN6 mixed with 40μM mβ$_2$m (green) in 10mM sodium phosphate buffer pH 6.2, 83mM NaCl, 37°C, 600 rpm. (D) Electron micrographs of the end-point of the reactions shown in B and C. Each micrograph is highlighted in a box of the same colour as the correspondig ThT trace (60μM ΔN6 –red, 60μM of F56E/W60E mβ$_2$m – black, 20μM ΔN6 + 40μM F56E/W60E mβ$_2$m - blue, 20μM ΔN6 + 40μM mβ$_2$m – green).

Taking into account the reverse concentration-dependence of the aggregation of ΔN6 (see Section 4.2.2), the observation that the mixture of 20μM of ΔN6 with 40μM F56E/W60E mβ$_2$m aggregates with the same lag time as 60μM ΔN6 (and not with that of 20μM ΔN6 - pink traces in Figure 5.30) suggests that the proteins are still able to interact to some degree, in line with the NMR data shown in Figure 5.29. Figure 4.18 shows that mβ$_2$m constitutes an aggregation incompetent species, that remains associated with the ΔN6 fibrils after their assembly. Since F56E/W60E mβ$_2$m does not alter the kinetics of fibril

formation compared with the same concentration of ΔN6 alone, the possibility of whether the residual interaction with the ΔN6 monomer, or even with its oligomeric species, can result in the co-polymerisation of the proteins, was investigated. In these experiments, fibrils made by the incubation of F56E/W60E mβ$_2$m and ΔN6 in different molar ratios, when the proteins were mixed prior to fibril assembly, were pelleted, re-suspended in 100% *(v/v)* HFIP and the depolymerised products were subjected to ESI-mass-spectrometry analysis. These experiments revealed that a small percentage of F56E/W60E mβ$_2$m is associated with the fibril pellet, the amount being independent of the starting concentration of F56E/W60E mβ$_2$m added to the reaction (Figure 5.31A-C). This finding is indicative of post-assembly association of F56E/W60E mβ$_2$m to the fibrils made by ΔN6, in a similar manner as was observed for mβ$_2$m (Figure 4.18). Indeed, when F56E/W60E mβ$_2$m was added post-assembly to ΔN6 fibrils, a similar amount of the protein was found in the pellet as when the proteins were mixed pre-assembly, as shown by SDS-PAGE (Figure 5.31D). Thus, this mutant variant of mβ$_2$m, even though has abolished its inhibitory capability, is not incorporated into the fibril shaft, as was found for wt-mβ$_2$m itself.

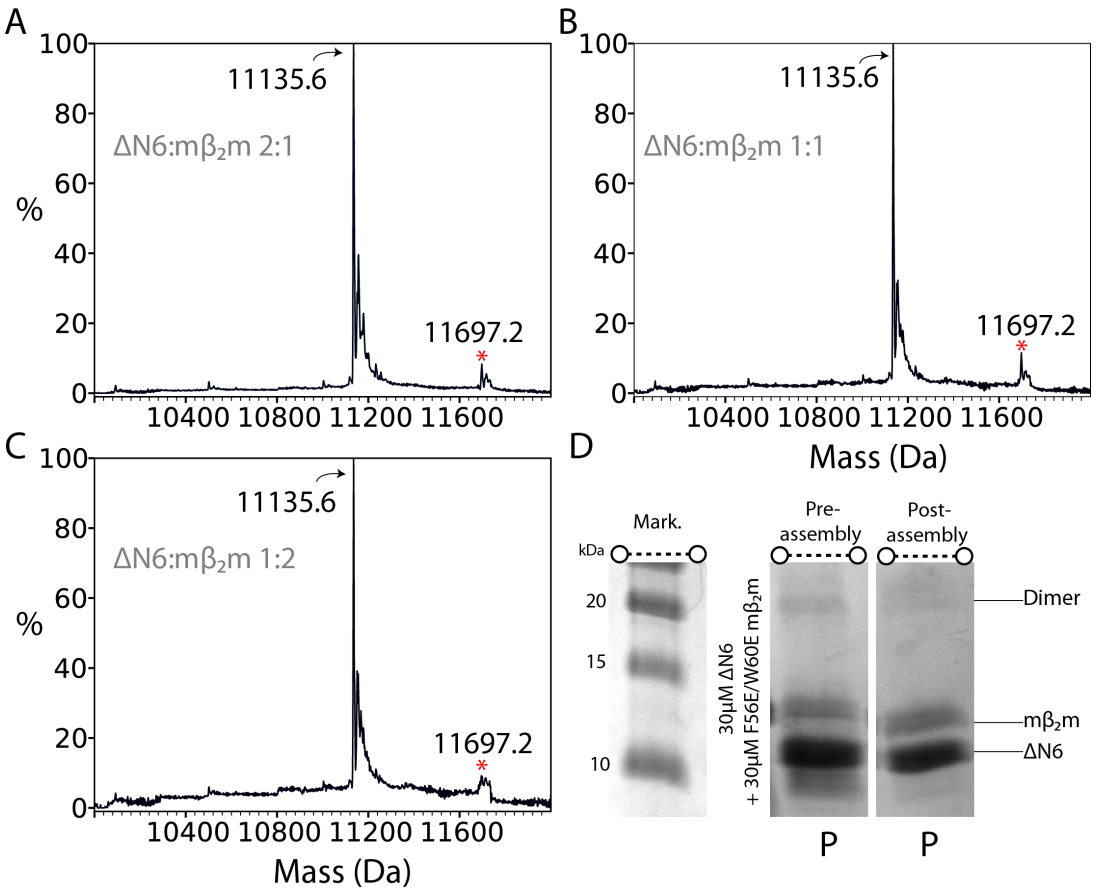

Figure 5.31: F56E/W60E-mβ$_2$m is not incorporated into the ΔN6 fibrils. ESI-mass-spectra of the depolymerised products of 40μM ΔN6 mixed with 20μM F56E/W60E-mβ$_2$m (A), or 30μM ΔN6 mixed with 30μM F56E/W60E-mβ$_2$m (B), or 20μM ΔN6 mixed with 40μM F56E/W60E-mβ$_2$m (C). ΔN6 and F56E/W60E-mβ$_2$m have a molecular weight of 11135.6Da and 11697.2Da respectively. All proteins were mixed pre-fibril assembly. (D) SDS-PAGE analysis of the insoluble fraction of a mixture of 30μM ΔN6 with 30μM F56E/W60E-mβ$_2$m when the proteins were mixed prior to amyloid assembly (left) or post-assembly (right).

5.2.6 Destabilisation upon binding of hβ$_2$m to ΔN6 at the origin of amyloid formation

It has been shown before using different biophysical tools such as ssNMR[341] and EPR[343] that hβ$_2$m adopts a parallel, in-register arrangement of its β-strands in fibrils made from the acid unfolded state at pH 2.0, in marked contrast with the antiparallel arrangement observed in the native state[344]. Recent data obtained by ssNMR (C.J.Sarell, University of Leeds, personal communication) revealed that ΔN6 behaves in a similar manner, showing that a major conformational change of the native fold has to take place on pathway to fibril

formation. The exact time-point of this phenomenon remains to be elucidated and has been a subject of debate in the amyloid field, however the early events en route to fibril assembly, involving the interaction of still vastly monomeric precursors should provide glimpses on its occurrence.

Close examination of the chemical shift changes that occur when ^{14}N-labelled ΔN6 is added to ^{15}N-labelled mβ$_2$m reveals that the residues that undergo significant chemical shift changes also experience increased PRE rates (BC and DE loops), showing that only residues in the interface are affected by the interaction (Figure 5.13A). On the other hand, residues in the AB loop (residues 23-24) of hβ$_2$m show significant chemical shift changes upon binding to ΔN6 (Figure 5.13B) but only minor PREs arising only from position 33, while they are also not included in the interface of the lowest energy structures of the ΔN6-hβ$_2$m heterodimer. In these experiments, the protein concentrations were adjusted such as the population of the bound ^{15}N-labelled protein is the same in both cases (mβ$_2$m-ΔN6, hβ$_2$m-ΔN6) to allow comparison of the data. Even though further PRE experiments will be required in order to prove whether these areas (A strand and AB loop) are part of the interface, the intriguing possibility that this part of hβ$_2$m is allosterically affected upon binding, emerges. The increased T_2 rates of residues in the AB loop upon addition of ΔN6 observed previously[272], are in line with this hypothesis.

To investigate the effect of the biomolecular collision of ΔN6 and hβ$_2$m further, a series of hydrogen-deuterium (H/D) exchange experiments were carried out in which the H/D exchange rates of the free (unbound) versus ΔN6-bound hβ$_2$m or mβ$_2$m monomers were compared. In these experiments the protein concentration was adjusted so as the % ^{15}N-labelled protein bound is the same within error in both cases (~21%), allowing direct comparison of the data. Remarkably, the results of these experiments showed that addition of ΔN6 to hβ$_2$m causes a 2-3 fold increase in the H/D exchange rates of the hβ$_2$m throughout the protein (Figure 5.32A), revealing that the interaction causes a global destabilisation of the hβ$_2$m monomer. By contrast, the inhibitory interaction between ΔN6 and mβ$_2$m has no effect on the stability of mβ$_2$m and its exchange rates remain largely unaffected upon interaction (Figure 5.32B).

The above observations suggest that the binding of ΔN6 to hβ$_2$m provides sufficient energy (ΔG_d^o=16.5 ± 2.9kJ/mol) to destabilise the native fold of hβ$_2$m (ΔG_{un}^o=22.5 ± 0.9kJ/mol, Clare Pashley, personal communication) such that a more amyloidogenic conformation is adopted. The non-amyloidogenic mβ$_2$m, on the other hand, is not affected significantly by binding.

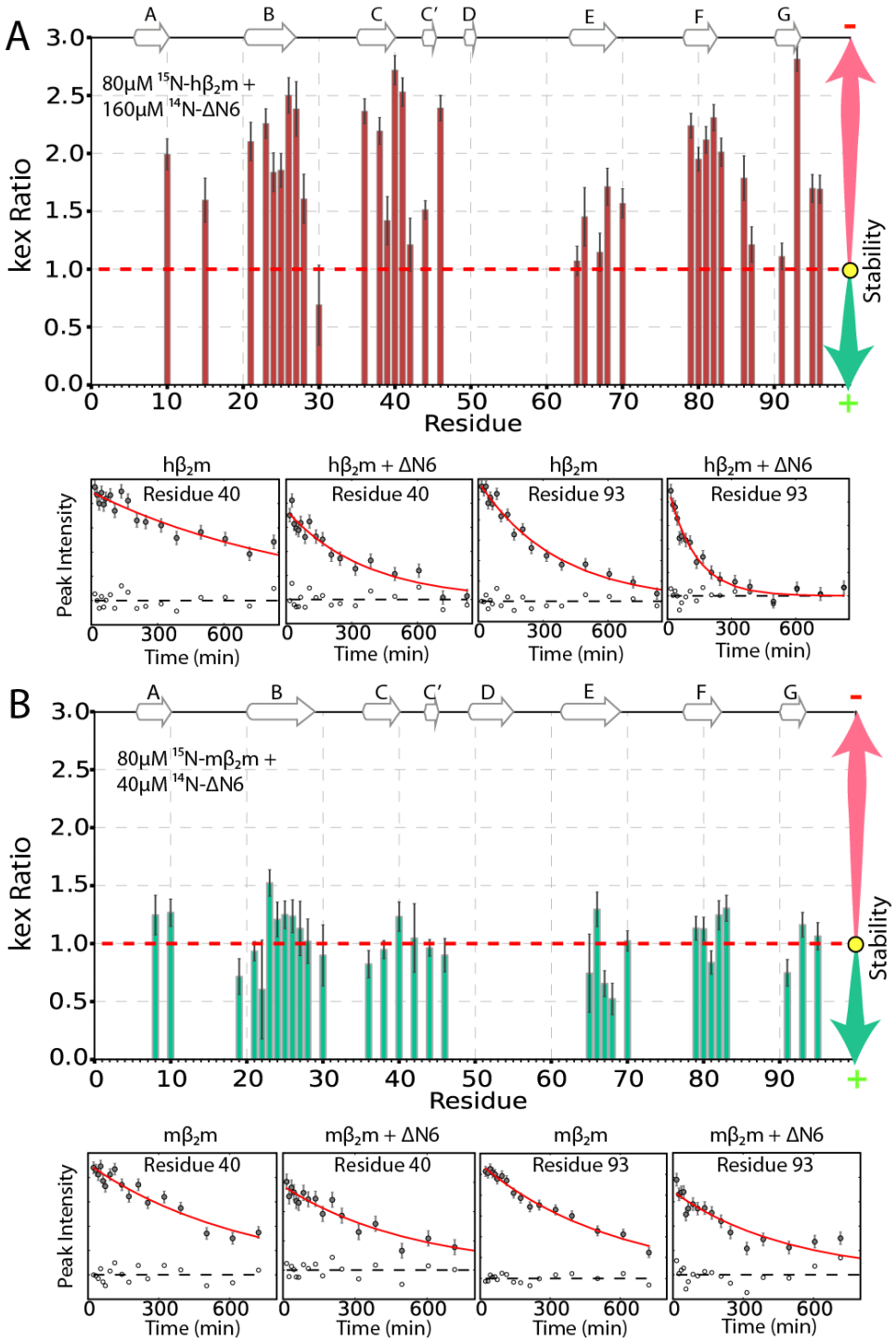

Figure 5.32: Dynamic vs rigid body interactions.
(A) The ratio of the H/D exchange rates of free hβ$_2$m (80μM) versus ~21% bound to ΔN6 (80μM hβ$_2$m + 160μM ΔN6) is plotted against residue number. A ratio greater than 1 (dotted line) indicates that hβ$_2$m in the bound sample

exchanges with solvent faster than when it is incubated alone, suggesting a destabilisation of the native state. Example of raw data for residues 40 and 93 for the free and bound samples is shown underneath. The red solid line represents a fit to single exponential and the residuals of the fit are shown as open dots. Dashed black line denotes 0. (B) As in A but for free (80μM mβ_2m) versus ~21% bound to ΔN6 (80μM mβ_2m + 40μM ΔN6).

The consequences of the different interactions on the conformational properties of ΔN6 were also monitored by following the changes in the chemical shifts of ^{15}N-labelled ΔN6 upon titration with ^{14}N-labelled hβ_2m or ^{14}N-labelled mβ_2m (Figure 5.33). Larger chemical shift differences were observed when ^{14}N-hβ_2m was added to ^{15}N-labelled ΔN6 compared with the addition of ^{14}N-labelled mβ_2m (corrected for the same %bound). These results are in line with the larger interacting surface for the ΔN6-hβ_2m association and potentially suggest that not only hβ_2m is affected by the interaction with ΔN6, but also that ΔN6 itself undergoes a conformational change upon binding. The picture that starts to emerge, therefore, reveals that the interaction between ΔN6-hβ_2m, which promotes fibril formation is dynamic, causing allosteric effects and/or global destabilisation on both interacting partners. By contrast, the inhibitory interaction (ΔN6-mβ_2m), even though very similar at a structural level, leads to a tight, 'specific' inhibitory complex with little effect on the structure or dynamics of the individual interacting partners.

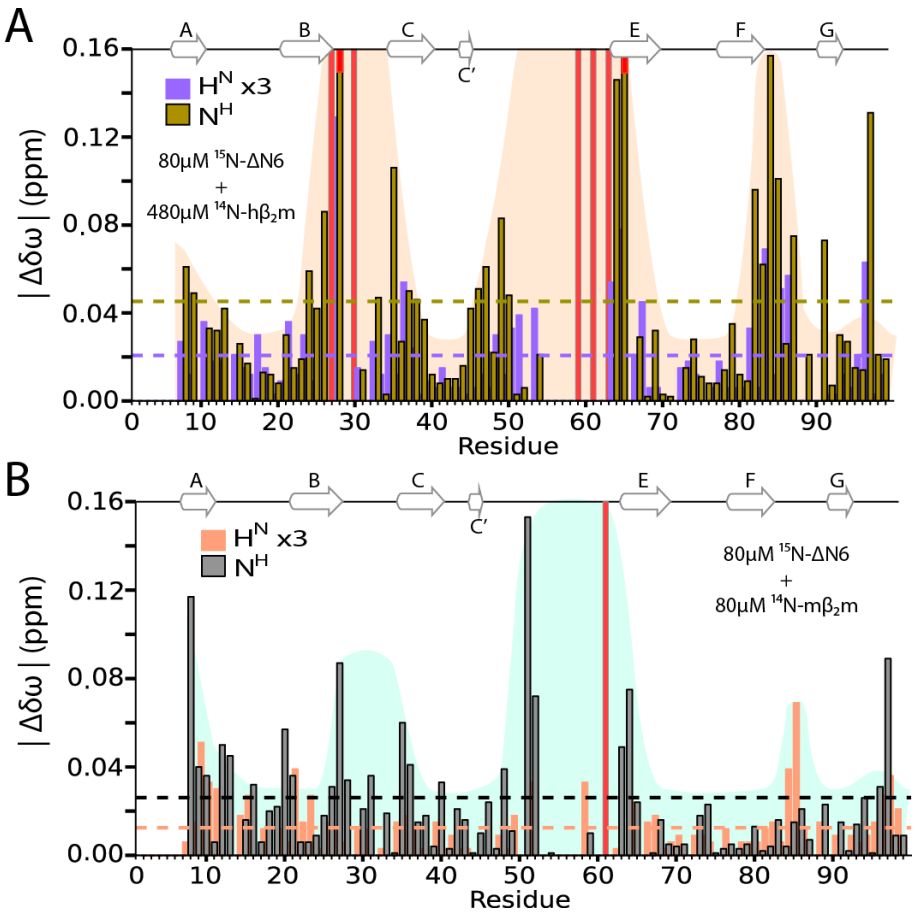

Figure 5.33: Allosteric effects of binding.
(A) ^{1}H (blue) and ^{15}N (brown) chemical shift differences for resonances of ^{15}N-labelled ΔN6 upon interaction with ^{14}N-labelled hβ$_2$m in a 1:6 molar ratio (80μM ΔN6 – 480μM hβ$_2$m~42% ΔN6 bound). (B) ^{1}H (orange) and ^{15}N (grey) chemical shift differences when ^{15}N-labelled mβ$_2$m and ^{14}N-labelled ΔN6 are mixed in a 1:1 molar ratio (80μM each ~42% ΔN6 bound). Dotted lines in (A) and (B) represent 2 standard deviations of the mean over the entire dataset. Red bars represent residues that show significant chemical shift changes in previous titration points, but are broadened beyond detection using the protein concentrations stated here. The shaded areas highlight regions that undergo significant chemical shift perturbations upon binding.

5.3 Discussion

5.3.1 Comparison of transient dimers of β$_2$m with other previously reported oligomeric forms of the protein

Over the years numerous studies have reported on oligomeric assemblies of β$_2$m using mainly X–ray crystallography. To allow comparison of these assemblies with the transient intermolecular interactions reported here, the measured PREs were back-calculated starting from different crystallographic

structures of β_2m oligomers. Miranker and co-workers, have solved the structure of Cu^{++} stabilised hβ_2m variants[399,427]. Cu^{++} has a moderate affinity for hβ_2m (2.7μM), while patients that were treated with Cu^{++}-free dialysis membranes had a >50% reduced possibility of developing DRA[428,429]. Therefore, these oligomers were considered to be clinically relevant and on-pathway to fibril assembly[430]. Importantly, the Cu^{++} stabilized dimer of P32A[399] and the hexameric form of H13F[429] both show a trans A32 or P32, respectively. The hexameric form of H13F contains two different dimer configurations. In the first, the monomers are arranged in a side-to-side conformation (Figure 5.34A), an arrangement which fails completely to describe the PRE data. However, the second dimer conformation (in the H13F hexamer), which involves a head-to-head configuration, shows a similar pattern of predicted PREs to their experimentally measured values. Specifically, the FG, BC and DE loops show increased PRE rates using the C33 and the C61 sites, while when the spin label is attached at position 20 no significant PREs are observed (Figure 5.34B). The amplitude of the back-calculated PRE rates is altered in comparison with the experimentally-determined values, suggesting that this head-to-head conformation might be close to the one observed in solution but it differs in detail. On the other hand the dimeric structure of P32A bound to Cu^{++} shows a vastly different PRE profile, indicating that this dimer is not able to explain the PRE data (Figure 5.34C). As mentioned in Section 1.6.5.3, the structure of a domain-swapped ΔN6 homodimer was reported recently (Figure 1.20)[363]. The authors propose that the hydrophobic residues exposed by the unfolding of the G strand may act as the nucleus for templated self-polymerisation and thus this species might be critical for nucleation of fibril assembly. Back-calculation of the PRE data using the domain swapped structure of ΔN6 yields no significant PREs for the 61 and the 20 sites, while PREs in the C strand and the CD loop are obtained using the C33 site (Figure 5.34D). This scenario is clearly different from the PREs measured for the self-association of ΔN6 (Figure 5.7).

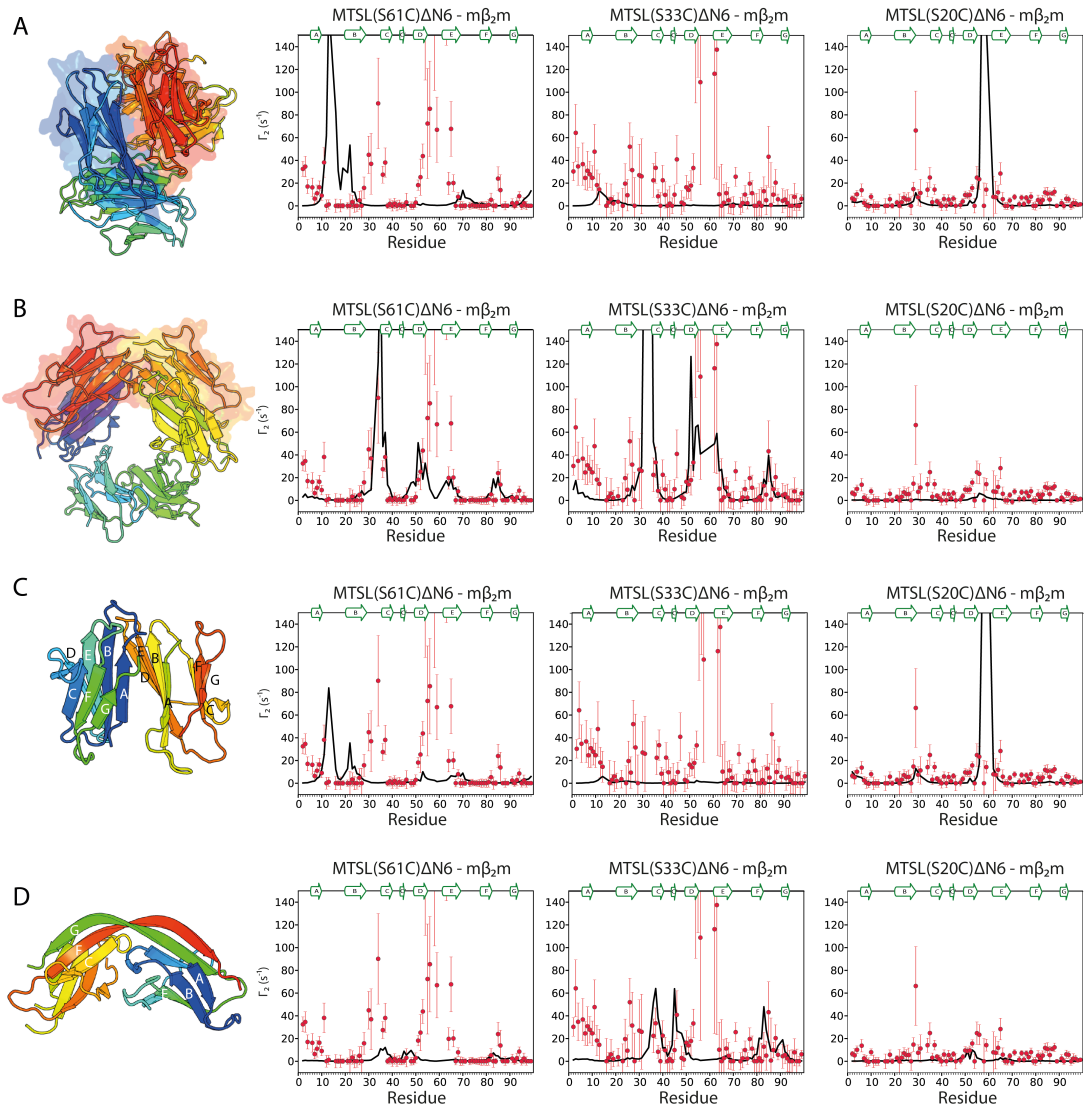

Figure 5.34: Compatibility of the PRE data with oligomeric structures of β₂m. PREs were back-calculated using the hexameric form of H13F hβ2m (A and B, pdbID: 3CIQ)[429], the dimeric form of P32A (C, pdbID: 2F8O)[399], or the domain swapped ΔN6 dimer (D, pdbID: 2X89)[363]. The side-by-side interface of the hexamer is used in (A) and the head-to-head configuration in (B) (shown as surface representation). Back-calculation of the PRE from the C61, C33 and C20 sites (black lines) are shown in each panel. The experimental data are shown as red dots and are for the ΔN6-mβ2m association. To allow back-calculation of the PRE data the X-ray structures were treated as rigid bodies and back-calculation of the PRE values was carried out with the same parameters as in Figures 5.16, 5.17. A short energy minimisation was run in each case to optimise the positions of the MTSL side chains.

5.3.2 Inhibition and promotion of amyloid formation in atomistic details

The data presented in Chapter IV showed that inhibition of amyloid formation by the ΔN6-mβ$_2$m interaction is due to the kinetic stabilisation of off-pathway species (presumably heterodimers) a phenomenon which was not observed by AUC for the biomolecular collision of ΔN6 with hβ$_2$m. In this Chapter, NMR analysis of the different protein-protein interactions showed that the inhibitory ΔN6-mβ$_2$m interaction has a 6-fold higher affinity than the amyloid promoting ΔN6-hβ$_2$m association, which results in the accumulation of non-amyloidogenic ΔN6-mβ$_2$m intermediates. These off-pathway species involve hydrophobic interactions through side-chains of residues in the DE and BC loops of both molecules with F56 and W60 presumably participating in key intermolecular contacts in the interface (Figure 5.35 bottom). Thus, these results showed that a tight control of the affinity of the protein-protein interactions involved in the early stages of fibril assembly is sufficient to determine the fate of the amyloid cascade. Interestingly, the addition of mβ$_2$m to ^{15}N-ΔN6, moves the chemical shifts of the latter towards their positions at pH 8.2 (Figure 5.29), suggesting that the interaction with mβ$_2$m converts ΔN6 into its non-amyloidogenic form.

On the other hand, the "catalytic" association of ΔN6-hβ$_2$m although in a similar overall head-to-head configuration, proved to be more heterogeneous, showing extra interaction sites on both molecules, and more transient in nature. These subtle structural differences are translated into a global destabilisation of hβ$_2$m as depicted by the 2-3 fold increase of its hydrogen exchange rates. The energy provided by binding to ΔN6 is able to alter the conformation of the AB loop of hβ$_2$m (as shown by the enhanced chemical shifts in this area upon interaction) which leads to a global loss of H/D exchange protection. Additionally, this interaction causes conformational changes in ΔN6 that potentially act as a trigger for switching this protein to a more aggregation-prone species as shown by the pronounced differences in its chemical shifts upon interaction (Figure 5.35 top). Presumably, both molecules are now able to visit a pool of more amyloidogenic states and eventually to become incorporated into an oligomer before getting co-polymerised[405]. In this mechanism the nucleation event is the formation of an ordered ΔN6 oligomer (as suggested by the PRE data for the

self-association of ΔN6-Figure 5.7), which acts as a precursor of amyloid assembly (Figure 5.35 middle).

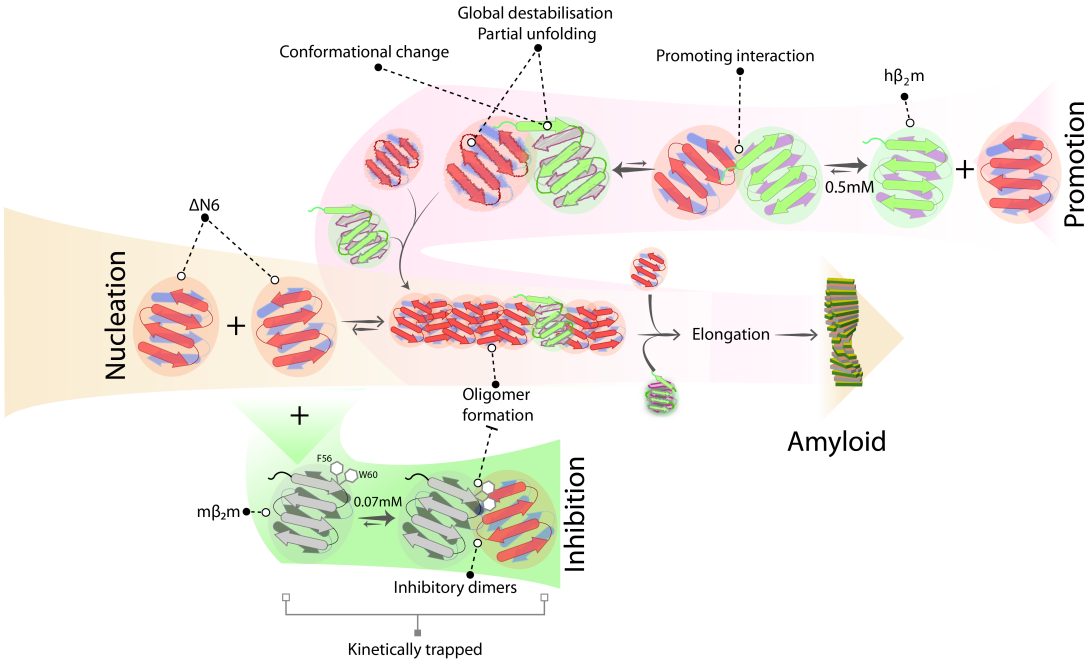

Figure 5.35: Controlling the fate of amyloid assembly.
ΔN6 (red cartoon) self-assembles into amyloid in a reaction that involves the formation of ordered pre-amyloid oligomers (middle). Interaction with mβ2m (grey cartoon) pushes the equilibrium back to monomers and trapped heterodimers, destroying or delaying the formation of the critical nucleus, thus kinetically inhibiting the formation of fibrils. The interaction involves the hydrophobic surfaces of the DE loops of both molecules, where F56 and W60 play crucial role, and results in the accumulation of off-pathway heterodimers (bottom). Hβ2m (green cartoon) interacts with ΔN6 in a similar manner to mβ2m but this interaction causes conformational changes on hβ2m in the AB loop and subsequent destabilisation and/or changes in the structure of both monomers involved. The destabilisation of the native fold generates species with increased amyloidogenic potential that can nucleate and elongate to explain the mechanism by which ΔN6 is able to enhance the amyloid potential of hβ2m.

CHAPTER VI

SUMMARY & CONCLUSIONS

6.1 Success and limitations

Protein aggregation is a complex multi-factorial reaction. We are still far from understanding the underlying principles of *'why proteins aggregate'* but for simplicity, this process can be (naively) seen as two steps: first, the natively folded protein is able to visit an aggregation-prone state through either unfolding or via excursions of the native fold. Second, once this species is formed, it is able to self-assemble into amyloid fibrils through a network of protein-protein interactions. In this study, these two stages were interrogated separately and in combination. After having presented the key findings of this work in Chapters III, IV and V, it is now time to summarise and try to answer the questions set out in Chapter I, highlighting the limitations of our current understanding of the system.

Why is mβ$_2$m not prone to aggregation? Does it fold through an I_T state?

Mβ$_2$m represents one of the least aggregation prone β$_2$m variants known to date. Here, characterisation of the structure and dynamics of mβ$_2$m in solution and under close to physiological conditions showed that the protein lacks the specific characteristics that are hypothesised to be responsible for increased amyloidogenicity (Chapter III). The observation that mβ$_2$m does not show increased T_2 dynamics at pH 6.2 suggests that proton uptake does not cause the conformational changes that are related with the increased amyloidogenicity of ΔN6 under these conditions. Investigation of the folding pathway of mβ$_2$m showed that it populates an intermediate species with properties reminiscent of those of the human I_T state. However, the significantly slower folding kinetics of mβ$_2$m suggest a disturbed folding pathway, possibly due to unfavourable contacts and/or accumulation of additional species that may alter the mechanism of folding and/or increse the activation energy required for the formation of the murine I_T. These phenomena might result in an I_T state with decreased propensity to oligomerise, explaining why mβ$_2$m is not fibrillogenic at neutral pH. The validity of this hypothesis as well as the origin of the

reduced propensity of mβ$_2$m to aggregate from its unfolded state remain to be elucidated.

Are the surfaces involved in promotion and inhibition of amyloid formation different?

Even though ΔN6, mβ$_2$m and hβ$_2$m are similar both in terms of sequence and in terms of structure their effect on the course of amyloid assembly is remarkably different. In principle, this phenomenon could be achieved via a plethora of different mechanisms. For instance, it is easy to imagine that the proteins form different hetero(homo)-polymeric assemblies with different interfaces, of which some are capable of further polymerisation and some others are not. However as shown in Chapter V, this is not the case. Using paramagnetic relaxation enhancement and analysis of chemical shift changes upon binding, the surfaces involved in promotion and inhibition of amyloid assembly were mapped in atomistic details, revealing the apical region of the molecule (BC, DE loops) surrounding P32 as the epicentre of binding in both cases.

What is the mechanism of inhibition?

A brief and concise answer to that question is the following; inhibition occurs by stabilisation of non-amyloidogenic hetero-assemblies. Mβ$_2$m, which as shown in Chapter III, is not able to form amyloid at pH 6.2 binds to ΔN6 monomers tightly and creates a 'specific', rigid complex which is held together by hydrophobic interactions (the word specific is used to describe the difference with the much more heterogeneous ΔN6-hβ$_2$m interaction and does not imply the formation of a stable, 'obligate' dimer). This results in destruction of the oligomeric species of ΔN6 that serve as nuclei for elongation of fibril formation and thus retardation of the kinetics of assembly is observed. Disrupting the ΔN6-mβ$_2$m interaction by mutation of residues in the interface restores the oligomeric nucleus of ΔN6 and fibril formation proceeds readily. Using a very simple 1:1 binding model, a striking linear correlation between the population of ΔN6 molecules bound to mβ$_2$m and the lag time of fibril assembly emerges (Figure 6.1).

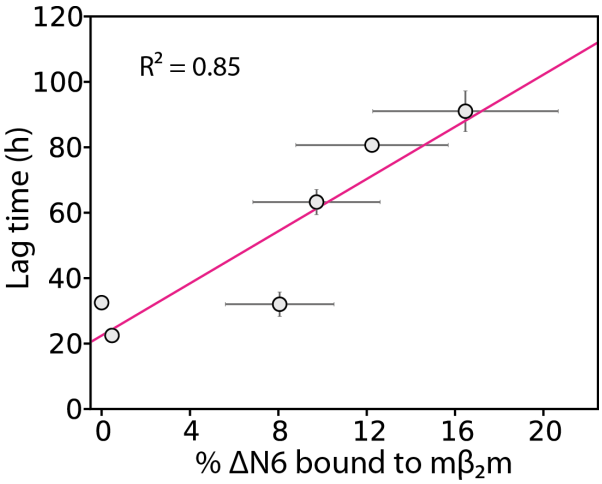

Figure 6.1: Correlation between %ΔN6 bound to mβ$_2$m and the lag time of fibril formation. %ΔN6 bound was calculated based on the K_d derived from the NMR chemical shift titration data and the x error-bars represent its propagated error. Y error-bars denote the standard error of the lag time measured by ThT fluorescence. The red line represents a linear fit to the data.

What is the mechanism of promotion of amyloid assembly?

Even though the sequences of mβ$_2$m and hβ$_2$m are identical in the DE loop, a region found to be the centre of the interface for both the ΔN6-mβ$_2$m and the ΔN6-hβ$_2$m interactions, the binding of hβ$_2$m to ΔN6 is more transient. As visualised by quantitative analysis of the PRE data in terms of the ensemble of species formed, the ΔN6-hβ$_2$m interaction involves a larger interface area and potentially a larger population of alternatively bound species. The heterogeneity of this interaction seems to be important for the promotion of fibril formation, since suppression of the alternative protein states by mβ$_2$m leads to the formation of (a more) specific, more stable inhibitory complex. Whether the observed heterogeneity in the ΔN6-hβ$_2$m association is due to oligomerisation or the formation of a range of encounter complexes is difficult to answer. However, this event causes destabilisation of the native fold of hβ$_2$m, which loses its H/D exchange protection and enforces conformational changes on ΔN6. As a consequence both molecules could access aggregation-prone intermediates and hetero-polymerisation proceeds rapidly to the assembly of mixed fibrils[405].

What about nucleation?

As revealed by the PRE data, nucleation of fibril formation by the self-association of ΔN6 molecules requires a more complex multi-binding event. Interestingly, even though still at preliminary stage, the results presented in Chapter V, revealed a symmetric ordered oligomer that can extend from both apices and can thus act as the building block for elongation of fibril assembly.

6.2 Implications on the β_2m aggregation pathway and disease

So what did we learn about the aggregation of β_2m? The results presented in this thesis further highlight the importance of ΔN6 as a 'promotor' of fibril formation *in vivo*. The exposure of hydrophobic side chains as an outcome of the cis-to-trans isomerisation of P32 increases the oligomerisation potential of this molecule. Interestingly, the apical part of the protein was suggested to be at the interface between $I_{T,}$ or I_T and native hβ_2m molecules, creating a head-to-head dimer during folding[396]. Here, using the power of paramagnetic relaxation enhancement, direct evidence for the oligomerisaion of ΔN6 was obtained in atomic detail. A head-to-head association is prevalent, but not sufficient to explain the interaction of amyloidogenic ΔN6 molecules. Importantly, the PRE approach can yield insights into the oligomeric species formed by ΔN6 that act as precursors of fibril assembly and lead to their structural characterisation. However, what emerges from this study is that it is the outcome of the interaction, not merely the site of interaction, that is responsible for controlling the fate of assembly. ΔN6 has two major characteristics that constitute this molecule a key player in the aggregation of hβ_2m: its ability to destabilise natively folded hβ_2m molecules and its transient binding. The combination of these two properties allows this protein saboteur to interact with numerous copies of hβ_2m which lose the stability of the native fold and are directly primed for aggregation. This mechanism could also explain the absence of circulating ΔN6 in patients suffering from DRA. Cleavage of hβ_2m by specific proteases, either in the plasma or in the sites of aggregation, generates ΔN6 molecules that are quickly consumed by additional proteases as suggested by limited proteolysis experiments[405,431]. This reaction is too rapid to allow detection of ΔN6 in the plasma, but during their short lifetimes these

molecules convey their catastrophic work by transmitting their amyloidogenic properties to non-amyloidogenic hβ$_2$m molecules. Importantly, as shown previously[272], ΔN6 can also nucleate its own fibril formation and ΔN6 fibrils can seed fibril formation of hβ$_2$m, adding another two weapons in its malicious repertoire.

Charge (and its lack) is also important in the initiating events of aggregation. The differences in the charge distribution of mβ$_2$m (compared with ΔN6 and hβ$_2$m) generate a hydrophobic area that acts as a 'magnet' to amyloidogenic ΔN6 molecules. As a consequence the whole protein turns into a safeguard of misfolding. The role of the hydrophobic residues in the DE loop in aggregation was highlighted previously using molecular dynamics simulations[432] or mutagenesis studies[338,433]. Interestingly, the W60G mutant of hβ$_2$m abolishes the ability to nucleate fibril formation at neutral pH and in the presence 20% *(v/v)* TFE[338], while substitution to Phe retains the amyloidogeniciy of the protein[433]. Hydrophobicity, therefore, is important even for the self-association of hβ$_2$m under fibrillogenic conditions. In the light of the findings presented in this thesis, these observations can be extended to include the role of affinity. Introduction of hydrophobic residues in the apical part of hβ$_2$m would potentially increase the affinity of the interaction between hβ$_2$m monomers and stop aggregation at the level of the dimer. On the other hand transient interactions through the same region (which theoretically are abolished in W60G, but preserved in the W60F variant) are an important initiating event for fibril formation, as exemplified here by the transient and heterogeneous nature of the ΔN6-hβ$_2$m association. The results suggest the potential of using protein engineering to convert hβ$_2$m into an inhibitor of its own aggregation or even stopping the aggregation of ΔN6 by introducing hydrophobic residues in the DE loop, mimicing in this manner the role of mβ$_2$m. Or can we force mβ$_2$m to be fibrillogenic? These questions represent new and exciting future perspectives.

6.3 Controlling the interplay between structure and energetics opens new therapeutic avenues

The importance of understanding the balance between kinetics and thermodynamics in amyloid formation is clearly shown in the case of transthyretin (TTR). In this case, a natively folded tetramer has to dissociate to release a less stable monomer with high propensity to aggregate[298,350,434]. Single point mutations (e.g T119M) that slow down the dissociation rate of the monomers from the native tetramer result in an increased kinetic energy barrier and, thus, decreased amyloid propensity[298]. Taking advantage of the (usually) empty thyroxine binding pockets of TTR, Kelly and co-workers managed to design the first commercially available drug against amyloidosis[298,400,435], that acts as a kinetic stabiliser of the native TTR tetramer. In superoxide dismutase-1 (SOD-1) *in silico* design of small molecules that stabilise the protein in its dimeric form via hydrophobic interactions, substantially decrease the propensity to aggregate by preventing the release of highly amyloidogenic monomers[300]. On the other hand, destabilisation of the native fold of human lysozyme, by mutations related with disease (I56T, D67H)[436], leads to hampered folding cooperativity and population of intermediate species with increased amyloidogenicity[268]. In this thesis, kinetic stabilisation and thermodynamic destabilisation were each observed as the outcome of interaction of very similar proteins.

As highlighted in Chapter I, nature has cleverly adjusted the affinity of protein-protein interactions according to their function. Proteins that participate in reaction cascades, where they have to be rapidly turned over to fulfil their duty, normally have low or moderate affinities for their partners, while very high affinities are observed in cases where the stability of the complex is crucial for cell survival (see Section 1.4). Unfortunately, the same phenomenon seems to be true for malfunction. The transient nature of protein-protein interactions in the early stages of assembly provide a plethora of opportunities for the correct encounter complex to form, while a high affinity association at this stage would stop the reaction at this early time-point.

Protein complexes are dynamic entities and the same stands for the individual partners (see Section 1.5.3). Protein dynamics and binding affinity were found interlinked in protein-DNA binding (catabolite activator protein-CAP)[80], where modulation of the conformational entropy manipulates binding affinity. In a similar manner, the data presented in Chapter V suggest an intriguing mechanism that has potentially evolved to fight against aggregation. The balance between folding and misfolding seems to be controlled by the response of protein dynamics to binding. The increased dynamics of $h\beta_2m$ as a consequence of a multivalent binding to ΔN6 is an uncommon observation, as most proteins undergo a disorder-to-order transition upon interaction (Figure 6.2), but the increased protein flexibility in the case of $h\beta_2m$ is enough to tip the balance towards aggregation. Suppression of the additional conformational states by $m\beta_2m$ on the other hand, suppresses dynamics and inhibits fibril assembly.

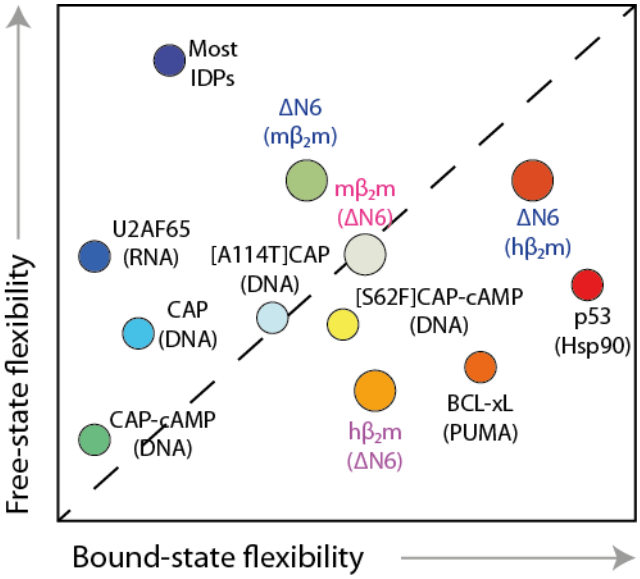

Figure 6.2: Correlation between protein dynamics in the free- and bound-state. Each protein whose dynamics are affected upon binding is shown as a dot and its partner is mentioned in the parenthesis. Proteins above the diagonal show decreased flexibility upon binding (IDPs that gain structure, CAP variants that show decreased conformational entropy[80], the U2A protein that alters its inter-domain motions upon RNA binding[162] and ΔN6 which gets sequestered upon rigib body docking with $m\beta_2m$. Below the diagonal are proteins that show increased dynamics upon interaction, such as the core domain of p53 that adopts a molten globule state upon binding to Hsp90[161], BCL-xL that's shows unfolding of two helices after binding[437] or $h\beta_2m$ which loses its H/D protection upon binding to ΔN6 and ΔN6 itself that (potentially) changes its conformation when bound to $h\beta_2m$. The data shown here cannot be

interpreted in a quantitative manner, but are used to show the range of responses to protein association. Inspired by reference [148].

Recently, Eisenberg and colleagues developed a novel screening approach for small molecule amyloid inhibitors based on the knowledge of the steric-zipper structure of the fibril[402,438]. The identification of the surfaces involved in the formation of early stage oligomers on pathway to fibrils performed here could pave the way to a structure-based design approach to target these species and stop the onset of the disease. Furthermore, modulators of protein dynamics upon binding could also represent a new therapeutic avenue. Building on the results presented here, future studies could potentially lead to a full kinetic, thermodynamic and structural elucidation of the early events on the aggregation pathway of β_2m and open the way of understanding the mechanism of aggregation of this and other proteins *in vitro,* and in the long term, *in vivo*.

REFERENCES

1. Perrett, D. From ‘protein’ to the beginnings of clinical proteomics. *Proteomics Clin. Appl.* **1,** 720–738 (2007).
2. Tanford, C. & Reynolds, J. *Nature's Robots : A History of Proteins*. (Oxford University Press, 2001).
3. Mulder, G. J. Ueber die Zusammensetzung einiger thierischen Substanzen. *Journal für Praktische Chemie* **16,** 129–152 (1839).
4. Meyer, C. E. & Rose, W. C. The Spatial Configuration of α-Amino-β-Hydroxy-n-butyric Acid. *J. Biol. Chem.* **115,** 721–729 (1936).
5. Sanger, F. The terminal peptides of insulin. *Biochemical J* **45,** 563 (1949).
6. Astbury, W. T. & Street, A. X-ray studies of the structure of hair, wool, and related fibres. I. General. *Phil. Trans. R. Soc.* (1932).
7. Astbury, W. T. & Woods, H. J. The molecular weights of proteins. *Nature* **127,** 663–665 (1931).
8. Pauling, L. & Corey, R. B. Configurations of polypeptide chains with favored orientations around single bonds: two new pleated sheets. *Proc. Natl. Acad. Sci. U S A* **37,** 729–740 (1951).
9. Eisenberg, D. Classic Perspective: The discovery of the α-helix and β-sheet, the principal structural features of proteins. *Proc. Natl. Acad. Sci. U S A* **100,** 11207–11210 (2011).
10. Pauling, L. L. & Corey, R. B. R. Atomic coordinates and structure factors for two helical configurations of polypeptide chains. *Proc. Natl. Acad. Sci. U S A* **37,** 235–240 (1951).
11. Kendrew, J. C. *et al.* A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* **181,** 662–666 (1958).
12. Perutz, M. F. M., Muirhead, H. H., Cox, J. M. J. & Goaman, L. C. L. Three-dimensional Fourier synthesis of horse oxyhaemoglobin at 2.8 A resolution: the atomic model. *Nature* **219,** 131–139 (1968).
13. Bernstein, F. C. F. *et al.* The Protein Data Bank: a computer-based archival file for macromolecular structures. *Arch. Biochem. Bioph.* **185,** 584–591 (1978).
14. Fischer, E. Einfluss der configuration auf die Wirkung der enzyme. *Ber. Dtsch. Chem. Ges.* **27,** 2985–2993 (1894).
15. Lichtenthaler, F. W. 100 Years‘Schlüssel-Schloss-Prinzip’: What made Emil Fischer use this analogy? *Angew. Chem. Int. Ed. Engl.* **33,** 2364–2374 (1995).
16. Johnson, L. N. The early history of lysozyme. *Nat. Struct Biol.* **5,** 942–944 (1998).
17. Blake, C. C. C. *et al.* Structure of hen egg-white lysozyme. A three-dimensional Fourier synthesis at 2 Angstrom resolution. *Nature* **206,** 757–761 (1965).
18. Blake, C. C. *et al.* Crystallographic studies of the activity of hen egg-white lysozyme. *Proc. R. Soc. Lond., B, Biol. Sci.* **167,** 378–388 (1967).
19. Monod, J., Wyman, J. & Changeux, J.-P. On the nature of allosteric transitions: a plausible model. *J. Mol. Biol.* **12,** 88–118 (1965).
20. Wüthrich, K. K., Wider, G. G., Wagner, G. G. & Braun, W. W. Sequential resonance assignments as a basis for determination of spatial protein structures by high resolution proton nuclear magnetic resonance. *J. Mol. Biol.* **155,** 311–319 (1982).

21. Williamson, M. P., Havel, T. F. & Wüthrich, K. Solution conformation of proteinase inhibitor IIA from bull seminal plasma by 1H nuclear magnetic resonance and distance geometry. *J. Mol. Biol.* (1985).
22. Wüthrich, K. & Wagner, G. Internal motion in globular proteins. *Trends Biochem. Sci.* **3,** 227–230 (1978).
23. Alberts, B. & Miake-Lye, R. Unscrambling the puzzle of biological machines: the importance of the details. *Cell* **68,** 415–420 (1992).
24. Beadle, G. W. G. & Tatum, E. L. E. Genetic control of biochemical reactions in Neurospora. *Proc. Natl. Acad. Sci. U S A* **27,** 499–506 (1941).
25. Alberts, B. M. The DNA enzymology of protein machines. *Cold Spring Harb. Symp. Quant. Biol.* **49,** 1–12 (1984).
26. Braun, P. & Gingras, A.-C. History of protein-protein interactions: from egg-white to complex networks. *Proteomics* **12,** 1478–1498 (2012).
27. Pauling, L. & Delbruck, M. The nature of the intermolecular forces operative in biological processes. *Science* **92,** 77–79 (1940).
28. Levy, E. D., Pereira-Leal, J. B., Chothia, C. & Teichmann, S. A. 3D Complex: a structural classification of protein complexes. *PLoS Comput. Biol.* **2,** e155 (2006).
29. Uversky, V. N. & Fink, A. L. Conformational constraints for amyloid fibrillation: the importance of being unfolded. *Biochim. Biophys. Acta* **1698,** 131–153 (2004).
30. Kelly, J. W. The alternative conformations of amyloidogenic proteins and their multi-step assembly pathways. *Curr. Opin. Struct. Biol.* **8,** 101–106 (1998).
31. Dobson, C. M. Protein misfolding, evolution and disease. *Trends Biochem. Sci.* **24,** 329–332 (1999).
32. Acuner Ozbabacan, S. E., Engin, H. B., Gursoy, A. & Keskin, O. Transient protein–protein interactions. *Prot. Eng. Des. Sel.* **24,** 635–648 (2011).
33. Klotz, I. M., Langebman, N. R. & Dahnall, D. W. Quaternary structure of proteins. *Annu. Rev. Biochem.* **39,** 25–62 (1970).
34. Wood, W. G. & Stamatoyannopoulos, G. Globin synthesis during erythroid cell maturation in alpha thalassemia. *Hemoglobin* **1,** 135–151 (1976).
35. White, R. J. & Jackson, S. P. The TATA-binding protein: a central role in transcription by RNA polymerases I, II and III. *Trends Genet.* **8,** 284–288 (1992).
36. E M Phizicky, S. F. Protein-protein interactions: methods for detection and analysis. *Microbiol. Rev.* **59,** 94 (1995).
37. Adams, R. L. & Wente, S. R. Uncovering nuclear pore complexity with innovation. *Cell* **152,** 1218–1221 (2013).
38. Watson, J. D. Involvement of RNA in the synthesis of proteins: the ordered interaction of three classes of RNA controls the assembly of amino acids into proteins. *Science* **140,** 17–26 (1963).
39. Steitz, T. A. A structural understanding of the dynamic ribosome machine. *Nat. Rev. Mol. Cell Biol.* **9,** 242–253 (2008).
40. Schmeing, T. M. & Ramakrishnan, V. What recent ribosome structures have revealed about the mechanism of translation. *Nature* **461,** 1234–1242 (2009).
41. Palade, G. E. A small particulate component of the cytoplasm. *J. Bioph. Biochem. Cytol.* **1,** 59 (1955).

42. Lake, J. A. Ribosome structure determined by electron microscopy of Escherichia coli small subunits, large subunits and monomeric ribosomes. *J. Mol. Biol.* **105,** 131–139 (1976).
43. Gao, H. H. *et al.* Study of the Structural Dynamics of the E. coli 70S Ribosome Using Real-Space Refinement. *Cell* **113,** 13–13 (2003).
44. Selmer, M. *et al.* Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* **313,** 1935–1942 (2006).
45. Schuwirth, B. S. *et al.* Structures of the bacterial ribosome at 3.5Å resolution. *Science* **310,** 827–834 (2005).
46. Feng, S., Chen, Y. & Gao, Y.-G. Crystal structure of 70S ribosome with both cognate tRNAs in the E and P sites representing an authentic elongation complex. *PLoS ONE* **8,** e58829–e58829 (2013).
47. Ogle, J. M. *et al.* Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science* **292,** 897–902 (2001).
48. Rodnina, M. V. & Green, R. *Ribosomes Structure, Function, and Dynamics*. (Springer, 2011).
49. Junge, W., Sielaff, H. & Engelbrecht, S. Torque generation and elastic power transmission in the rotary FOF1-ATPase. *Nature* **459,** 364–370 (2009).
50. Gibbons, C., Montgomery, M. G., Leslie, A. G. & Walker, J. E. The structure of the central stalk in bovine F(1)-ATPase at 2.4 A resolution. *Nat. Struct. Biol.* **7,** 1055–1061 (2000).
51. Stock, D., Gibbons, C., Arechaga, I., Leslie, A. G. & Walker, J. E. The rotary mechanism of ATP synthase. *Curr. Opin. Struct. Biol.* **10,** 672–679 (2000).
52. Hershko, A. The ubiquitin system for protein degradation and some of its roles in the control of the cell division cycle*. *Cell Death Differ.* **12,** 1191–1197 (2005).
53. Zwickl, P., Kleinz, J. & Baumeister, W. Critical elements in proteasome assembly. *Nat. Struct Biol.* **1,** 765–770 (1994).
54. Groll, M., Berkers, C. R., Ploegh, H. L. & Ovaa, H. Crystal structure of the boronic acid-based proteasome inhibitor bortezomib in complex with the Yeast 20S proteasome. *Structure* **14,** 451–456 (2006).
55. Glickman, M. H. *et al.* A Subcomplex of the proteasome regulatory particle required for ubiquitin-conjugate degradation and related to the COP9-signalosome and eIF3. *Cell* **94,** 9–9 (1998).
56. Groll, M. M. *et al.* Structure of 20S proteasome from yeast at 2.4 A resolution. *Nature* **386,** 463–471 (1997).
57. Marques, A. J., Palanimurugan, R., Matias, A. C., Ramos, P. C. & Dohmen, R. J. Catalytic Mechanism and Assembly of the Proteasome. *Chem. Rev.* **109,** 1509–1536 (2009).
58. Hill, C. P. *et al.* Structural basis for the activation of 20S proteasomes by 11S regulators. *Nature* **408,** 115–120 (2000).
59. Rossmann, M. G. Structure of viruses: a short history. *Quart. Rev. Biophys.* **46,** 133–180 (2013).
60. Hirokawa, N., Noda, Y., Tanaka, Y. & Niwa, S. Kinesin superfamily motor proteins and intracellular transport. *Nat. Rev. Mol. Cell Biol.* **10,** 682–696 (2009).
61. Tucker, P. A. & Sallai, L. The AAA+ superfamily—a myriad of motions. *Curr. Opin. Struct. Biol.* **17,** 641–652 (2007).

62. Vuignier, K., Schappler, J., Veuthey, J.-L., Carrupt, P.-A. & Martel, S. Drug-protein binding: a critical review of analytical tools. *Anal. Bioanal. Chem.* **398,** 53–66 (2010).
63. Piehler, J. New methodologies for measuring protein interactions *in vivo* and in vitro. *Curr. Opin. Struct. Biol.* **15,** 4–14 (2005).
64. Chothia, C. The nature of the accessible and buried surfaces in proteins. *J. Mol. Biol.* **105,** 1–12 (1976).
65. Lee, B. & Richards, F. M. The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* **55,** 379–400 (1971).
66. Chothia, C. & Janin, J. Principles of protein–protein recognition. *Nature* **256,** 705–708 (1975).
67. Miller, S., Lesk, A. M., Janin, J. & Chothia, C. The accessible surface area and stability of oligomeric proteins. *Nature* **328,** 834–836 (1987).
68. Kastritis, P. L. *et al.* A structure-based benchmark for protein-protein binding affinity. *Protein Sci.* **20,** 482–491 (2011).
69. Tang, C., Iwahara, J. & Clore, G. M. Visualization of transient encounter complexes in protein-protein association. *Nature* **444,** 383–386 (2006).
70. Harel, M., Spaar, A. & Schreiber, G. Fruitful and futile encounters along the association reaction between proteins. *Biophys J.* **96,** 4237–4248 (2009).
71. Schreiber, G. & Fersht, A. R. Rapid, electrostatically assisted association of proteins. *Nat. Struct Biol.* **3,** 427–431 (1996).
72. Norel, R. R., Petrey, D. D., Wolfson, H. J. H. & Nussinov, R. R. Examination of shape complementarity in docking of unbound proteins. *Proteins* **36,** 307–317 (1999).
73. Kastritis, P. L. & Bonvin, A. M. J. J. On the binding affinity of macromolecular interactions: daring to ask why proteins interact. *J. R. Soc. Interface* **10,** 20120835–20120835 (2013).
74. Kleanthous, C. *Protein-Protein Recognition*. (Oxford University Press, 2000).
75. Rajamani, D. Anchor residues in protein-protein interactions. *Proc. Natl. Acad. Sci. U S A* **101,** 11287–11292 (2004).
76. Bogan, A. A. & Thorn, K. S. Anatomy of hot spots in protein interfaces. *J. Mol. Biol.* **280,** 1–9 (1998).
77. Clackson, T. & Wells, J. A hot spot of binding energy in a hormone-receptor interface. *Science* **267,** 383–386 (1995).
78. Cunningham, B. C. & Wells, J. A. Comparison of a structural and a functional epitope. *J. Mol. Biol.* **234,** 554–563 (1993).
79. Tzeng, S.-R. & Kalodimos, C. G. Dynamic activation of an allosteric regulatory protein. *Nature* **462,** 368–372 (2009).
80. Tzeng, S.-R. & Kalodimos, C. G. Protein activity regulation by conformational entropy. *Nature* 488, 236-240, (2012)
81. Jones, S. & Thornton, J. M. Protein-protein interactions: A review of protein dimer structures. *Prog. Bioph. Mol. Biol* **63,** 31–65 (1995).
82. Jones, J. M. T. Principles of protein-protein interactions. *Proc. Natl. Acad. Sci. U S A* **93,** 13 (1996).
83. S J Smerdon, *et. al.*. Structure of the binding site for nonnucleoside inhibitors of the reverse transcriptase of human immunodeficiency virus type 1. *Proc. Natl. Acad. Sci. U S A* **91,** 3911 (1994).
84. Clore, G. M., Appella, E., Yamada, M., Matsushima, K. & Gronenborn, A. M. Three-dimensional structure of interleukin 8 in solution. *Biochemistry* **29,** 1689–1696 (1990).

85. Davies, D. R. & Cohen, G. H. Interactions of protein antigens with antibodies. *Proc. Natl. Acad. Sci. U S A* **93,** 7–12 (1996).
86. Read, R. J., Fujinaga, M., Sielecki, A. R. & James, M. N. Structure of the complex of Streptomyces griseus protease B and the third domain of the turkey ovomucoid inhibitor at 1.8-Å resolution. *Biochemistry* **22,** 4420–4433 (1983).
87. Parsiegla, G. *et al.* Crystal structures of the cellulase Cel48F in complex with inhibitors and substrates give insights into its processive action. *Biochemistry* **39,** 11238–11246 (2000).
88. Nooren, I. M. A. & Thornton, J. M. Structural characterisation and functional significance of transient protein–protein interactions. *J. Mol. Biol.* **325,** 991–1018 (2003).
89. Kuriyan, J. & Eisenberg, D. The origin of protein interactions and allostery in colocalization. *Nature* **450,** 983–990 (2007).
90. Ispolatov, I. Binding properties and evolution of homodimers in protein-protein interaction networks. *Nucleic Acids Res.* **33,** 3629–3635 (2005).
91. Nussinov, R., Xu, D. & Tsai, C.-J. Mechanism and evolution of protein dimerization. *Protein Sci.* **7,** 533–544 (1998).
92. Bhattacharyya, R. P., Reményi, A., Yeh, B. J. & Lim, W. A. Domains, Motifs, and Scaffolds: The Role of modular interactions in the evolution and wiring of cell signaling circuits. *Annu. Rev. Biochem.* **75,** 655–680 (2013).
93. Vetter, I. R. The guanine nucleotide-binding switch in three dimensions. *Science* **294,** 1299–1304 (2001).
94. Linn, S. M. & Lloyd, R. S. *Nucleases*. (Cold Spring Harbor Laboratory Pr, 1993).
95. D'Alessio, G. & Riordan, J. F. *Ribonucleases*. (Academic Press, 1997).
96. Yang, W. Nucleases: diversity of structure, function and mechanism. *Quart. Rev. Biophys.* **44,** 1–93 (2010).
97. Gaur, D. D., Swaminathan, S. S. & Batra, J. K. J. Interaction of human pancreatic ribonuclease with human ribonuclease inhibitor. Generation of inhibitor-resistant cytotoxic variants. *J. Biol. Chem.* **276,** 24978–24984 (2001).
98. Mauguen, Y. *et al.* Molecular structure of a new family of ribonucleases. *Nature* **297,** 162–164 (1982).
99. Martínez, J. C. J., Filimonov, V. V. V., Mateo, P. L. P., Schreiber, G. G. & Fersht, A. R. A. A calorimetric study of the thermal stability of barstar and its interaction with barnase. *Biochemistry* **34,** 5224–5233 (1995).
100. Papadakos, G., Wojdyla, J. A. & Kleanthous, C. Nuclease colicins and their immunity proteins. *Quart. Rev. Biophys.* **45,** 57–103 (2011).
101. Buckle, A. M. A., Schreiber, G. G. & Fersht, A. R. A. Protein-protein recognition: crystal structural analysis of a barnase-barstar complex at 2.0-A resolution. *Biochemistry* **33,** 8878–8889 (1994).
102. Guillet, V., Lapthorn, A., Hartley, R. W. & Mauguen, Y. Recognition between a bacterial ribonuclease, barnase, and its natural inhibitor, barstar. *Structure* **1,** 165–176 (1993).
103. Buckle, A. M. A. & Fersht, A. R. A. Subsite binding in an RNase: structure of a barnase-tetranucleotide complex at 1.76-A resolution. *Biochemistry* **33,** 1644–1653 (1994).
104. Lemmon, M. A. & Schlessinger, J. Cell Signaling by Receptor Tyrosine Kinases. *Cell* **141,** 1117–1134 (2010).

105. Ullrich, A. & Schlessinger, J. Signal transduction by receptors with tyrosine kinase activity. *Cell* **61,** 203–212 (1990).
106. Musacchio, A., Noble, M., Pauptit, R., Wierenga, R. & Saraste, M. Crystal structure of a Src-homology 3 (SH3) domain. *Nature* **359,** 851–855 (1992).
107. Yu, H. *et al.* Solution structure of the SH3 domain of Src and identification of its ligand-binding site. *Science* **258,** 1665–1668 (1992).
108. Viguera, A. R., Arrondo, J. L. R., Musacchio, A., Saraste, M. & Serrano, L. Characterisation of the interaction of natural proline-rich peptides with five different SH3 domains. *Biochemistry* **33,** 10925–10933 (1994).
109. Musacchio, A., Saraste, M. & Wilmanns, M. High-resolution crystal structures of tyrosine kinase SH3 domains complexed with proline-rich peptides. *Nat. Struct Biol.* **1,** 546–551 (1994).
110. Feng, S., Chen, J., Yu, H., Simon, J. & Schreiber, S. Two binding orientations for peptides to the Src SH3 domain: development of a general model for SH3-ligand interactions. *Science* **266,** 1241–1247 (1994).
111. Wolfenson, H., Lavelin, I. & Geiger, B. Dynamic regulation of the structure and functions of integrin adhesions. *Dev. Cell* **24,** 447–458 (2013).
112. Dong, F. & Zhou, H.-X. Electrostatic contribution to the binding stability of protein-protein complexes. *Proteins* **65,** 87–102 (2006).
113. Berg, O. G. & Hippel, von, P. H. Diffusion-controlled macromolecular interactions. *Annu. Rev. Biophys. Biophys. Chem.* **14,** 131–158 (1985).
114. Northrup, H. P. E. Kinetics of protein-protein association explained by Brownian dynamics computer simulation. *Proc. Natl. Acad. Sci. U S A* **89,** 3338 (1992).
115. Schreiber, G. & Fersht, A. R. Interaction of barnase with its polypeptide inhibitor barstar studied by protein engineering. *Biochemistry* **32,** 5145–5150 (1993).
116. Zhou, H. X. Brownian dynamics study of the influences of electrostatic interaction and diffusion on protein-protein association kinetics. *Biophys. J.* **64,** 1711 (1993).
117. Kim, Y. C., Tang, C., Clore, G. M. & Hummer, G. Replica exchange simulations of transient encounter complexes in protein-protein association. *Proc. Natl. Acad. Sci. U S A* **105,** 12855–12860 (2008).
118. Kotrba, P., Inui, M. & Yukawa, H. Bacterial phosphotransferase system (PTS) in carbohydrate uptake and control of carbon metabolism. *J. Biosci. Bioeng.* **92,** 502–517 (2001).
119. Deutscher, J., Francke, C. & Postma, P. W. How phosphotransferase system-related protein phosphorylation regulates carbohydrate metabolism in bacteria. *Microbiol. Mol. Biol. Rev.* **70,** 939–1031 (2006).
120. Ubbink, M. The courtship of proteins: understanding the encounter complex. *FEBS lett.* **583,** 1060–1066 (2009).
121. Schwieters, C. D. *et al.* Solution Structure of the 128 kDa enzyme I dimer from Escherichia coli and its 146 kDa complex with HPr using residual dipolar couplings and small- and wide-angle X-ray scattering. *J. Am. Chem. Soc.* **132,** 13026–13045 (2010).
122. Hu, J. *et al.* Solution NMR structures of productive and non-productive complexes between the A and B domains of the cytoplasmic subunit of the mannose transporter of the Escherichia coli phosphotransferase system. *J. Biol. Chem.* **283,** 11024–11037 (2008).

123. Cao, Y. *et al.* Crystal structure of a phosphorylation-coupled saccharide transporter. *Nature* **473,** 50–54 (2011).
124. Garrett, D. S., Seok, Y. J. & Peterkofsky, A. Solution structure of the 40,000 Mr phosphoryl transfer complex between the N-terminal domain of enzyme I and HPr. *Nat. Struct. Biol.* 1999).
125. Henzler-Wildman, K. & Kern, D. Dynamic personalities of proteins. *Nature* **450,** 964–972 (2007).
126. Palmer, A. G., III. NMR Probes of Molecular Dynamics: Overview and comparison with other techniques. *Annu. Rev. Biophys. Biomol. Struct.* **30,** 129–155 (2013).
127. Eisenmesser, E. Z., Bosco, D. A., Akke, M. & Kern, D. Enzyme dynamics during catalysis. *Science* **295,** 1520–1523 (2002).
128. Gelis, I. *et al.* Structural basis for signal-sequence recognition by the translocase motor SecA as determined by NMR. *Cell* **131,** 756–769 (2007).
129. Tokuriki, N. & Tawfik, D. S. Protein dynamism and evolvability. *Science* **324,** 203–207 (2009).
130. Williamson, M. P. & Potts, J. R. Intrinsically disordered proteins: administration not executive. *Biochem. Soc. Trans.* **40,** 945–949 (2012).
131. Dunker, A. K., Silman, I., Uversky, V. N. & Sussman, J. L. Function and structure of inherently disordered proteins. *Curr. Opin. Struct. Biol.* **18,** 756–764 (2008).
132. Mittag, T. & Forman-Kay, J. D. Atomic-level characterisation of disordered protein ensembles. *Curr. Opin. Struct. Biol.* **17,** 3–14 (2007).
133. Platt, G. W., McParland, V. J., Kalverda, A. P., Homans, S. W. & Radford, S. E. Dynamics in the unfolded state of beta2-microglobulin studied by NMR. *J. Mol. Biol.* **346,** 279–294 (2005).
134. Mittag, T., Kay, L. E. & Forman-Kay, J. D. Protein dynamics and conformational disorder in molecular recognition. *J. Mol. Recognit.* **23,** 105–116 (2010).
135. Dyson, H. J. & Wright, P. E. Coupling of folding and binding for unstructured proteins. *Curr. Opin. Struct. Biol.* **12,** 54–60 (2002).
136. Agalarov, S. C. Structure of the S15,S6,S18-rRNA complex: assembly of the 30S ribosome central domain. *Science* **288,** 107–112 (2000).
137. Spolar, R. S. & Record, M. T. Coupling of local folding to site-specific binding of proteins to DNA. *Science* **263,** 777–784 (1994).
138. Dyson, H. J. & Wright, P. E. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* **6,** 197–208 (2005).
139. Breidenbach, M. A. & Brunger, A. T. Substrate recognition strategy for botulinum neurotoxin serotype A. *Nature* **432,** 925–929 (2004).
140. Tokuriki, N., Oldfield, C. J., Uversky, V. N., Berezovsky, I. N. & Tawfik, D. S. Do viral proteins possess unique biophysical features? *Trends Biochem. Sci.* **34,** 53–59 (2009).
141. Romero, P. R. *et al.* Alternative splicing in concert with protein intrinsic disorder enables increased functional diversity in multicellular organisms. *Proc. Natl. Acad. Sci. U S A* **103,** 8390–8395 (2006).
142. Ward, J. J., Sodhi, J. S., McGuffin, L. J., Buxton, B. F. & Jones, D. T. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J. Mol. Biol.* **337,** 635–645 (2004).

143. Feldman, R. A Complex of Cdc4p, Skp1p, and Cdc53p/Cullin catalyzes ubiquitination of the phosphorylated CDK inhibitor Sic1p. *Cell* **91,** 221–230 (1997).
144. Nash, P. *et al.* Multisite phosphorylation of a CDK inhibitor sets a threshold for the onset of DNA replication. *Nature* **414,** 514–521 (2001).
145. Orlicky, S., Tang, X., Willems, A., Tyers, M. & Sicheri, F. Structural basis for phosphodependent substrate selection and orientation by the SCFCdc4 ubiquitin ligase. *Cell* **112,** 243–256 (2003).
146. Mittag, T. *et al.* Dynamic equilibrium engagement of a polyvalent ligand with a single-site receptor. *Proc. Natl. Acad. Sci. U S A* **105,** 17772–17777 (2008).
147. Tang, X. *et al.* Composite low affinity interactions dictate recognition of the cyclin-dependent kinase inhibitor Sic1 by the SCFCdc4 ubiquitin ligase. *Proc. Natl. Acad. Sci. U S A* **109,** 3287–3292 (2012).
148. Marsh, J. A., Teichmann, S. A. & Forman-Kay, J. D. Probing the diverse landscape of protein flexibility and binding. *Curr. Opin. Struct. Biol.* **22,** 643–650 (2012).
149. Hammes, G. G., Chang, Y. C. & Oas, T. G. Conformational selection or induced fit: a flux description of reaction mechanism. *Proc. Natl. Acad. Sci. U S A* **106,** 13737–13741 (2009).
150. Henkels, C. H., Kurz, J. C., Fierke, C. A. & Oas, T. G. Linked folding and anion binding of the Bacillus subtilis ribonuclease P protein †. *Biochemistry* **40,** 2777–2789 (2001).
151. Csermely, P., Palotai, R. & Nussinov, R. Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends Biochem. Sci.* **35,** 539–546 (2010).
152. Forneris, F., Wu, J. & Gros, P. The modular serine proteases of the complement cascade. *Curr. Opin. Struct. Biol.* **22,** 333–341 (2012).
153. Gros, P., Milder, F. J. & Janssen, B. J. C. Complement driven by conformational changes. *Nat. Rev. Immunol.* **8,** 48–58 (2008).
154. Rühlmann, A., Kukla, D., Schwager, P., Bartels, K. & Huber, R. Structure of the complex formed by bovine trypsin and bovine pancreatic trypsin inhibitor. *J. Mol. Biol.* **77,** 417–436 (1973).
155. Bode, W., Schwager, P. & Huber, R. The transition of bovine trypsinogen to a trypsin-like state upon strong ligand binding. *J. Mol. Biol.* **118,** 99–112 (1978).
156. Noel, J. P., Hamm, H. E. & Sigler, P. B. The 2.2 Å crystal structure of transducin-alpha complexed with GTP gamma S. *Nature* **366,** 654–663 (1993).
157. Lambright, D. G., Noel, J. P., Hamm, H. E. & Sigler, P. B. Structural determinants for activation of the α-subunit of a heterotrimeric G protein. *Nature* **369,** 621–628 (1994).
158. Fraser, J. S. *et al.* Accessing protein conformational ensembles using room-temperature X-ray crystallography. *Proc. Natl. Acad. Sci. U S A* **108,** 16247–16252 (2011).
159. Marsh, J. A. & Teichmann, S. A. Relative solvent accessible surface area predicts protein conformational changes upon binding. *Structure* **19,** 859–867 (2011).
160. Kastritis, P. L. & Bonvin, A. M. Molecular origins of binding affinity: seeking the Archimedean point. *Curr. Opin. Struct. Biol.* ***23,*** *1-10,* (2013).

161. Park, S. J., Borin, B. N., Martinez-Yamout, M. A. & Dyson, H. J. The client protein p53 adopts a molten globule–like state in the presence of Hsp90. *Nat. Struct. Mol. Biol.* **18,** 537–541 (2011).
162. Mackereth, C. D. *et al.* Multi-domain conformational selection underlies pre-mRNA splicing regulation by U2AF. *Nature* **475,** 408–411 (2011).
163. Karplus, M. & Kushick, J. N. Method for estimating the configurational entropy of macromolecules. *Macromolecules* **14,** 325–332 (1981).
164. Palmer, A. G. & Massi, F. Characterisation of the dynamics of biomacromolecules using rotating-frame spin relaxation NMR spectroscopy. *Chem. Rev.* **106,** 1700–1719 (2006).
165. Akke, M. & Palmer, A. G. Monitoring macromolecular motions on microsecond to millisecond time scales by R1ρ-R1 constant relaxation time NMR Spectroscopy. *J. Am. Chem. Soc.* **118,** 911–912 (1996).
166. Dobson, C. M. & Karplus, M. Internal motion of proteins: nuclear magnetic resonance measurements and dynamic simulations. *Methods Enzymol.* **131,** 362–389 (1986).
167. Chapman, H. N. *et al.* Femtosecond X-ray protein nanocrystallography. *Nature* **470,** 73–77 (2011).
168. Frauenfelder, H. H., McMahon, B. H. B. & Fenimore, P. W. P. Myoglobin: the hydrogen atom of biology and a paradigm of complexity. *Proc. Natl. Acad. Sci. U S A* **100,** 8615–8617 (2003).
169. Schotte, F. F., Soman, J. J., Olson, J. S. J., Wulff, M. M. & Anfinrud, P. A. P. Picosecond time-resolved X-ray crystallography: probing protein function in real time. *J Struct. Biol.* **147,** 12–12 (2004).
170. Baldwin, A. J. & Kay, L. E. NMR spectroscopy brings invisible protein states into focus. *Nat. Chem. Biol.* **5,** 808–814 (2009).
171. Joo, C., Balci, H., Ishitsuka, Y., Buranachai, C. & Ha, T. Advances in single-molecule fluorescence methods for molecular biology. *Annu. Rev. Biochem.* **77,** 51–76 (2008).
172. Craggs, T. D. T. & Kapanidis, A. N. A. Six steps closer to FRET-driven structural biology. *Nat. Methods* **9,** 1157–1158 (2012).
173. Uphoff, S. S. *et al.* Monitoring multiple distances within a single molecule using switchable FRET. *Nat. Methods* **7,** 831-836 (2010).
174. Pugh, S. D. S., Gell, C. C., Smith, D. A. D., Radford, S. E. S. & Brockwell, D. J. D. Single-Molecule Studies of the Im7 Folding Landscape. *J. Mol. Biol.* **398,** 14–14 (2010).
175. Michalet, X., Weiss, S. & Jäger, M. Single-molecule fluorescence studies of protein folding and conformational dynamics. *Chem. Rev.* **106,** 1785–1813 (2006).
176. Carr, H. & Purcell, E. Effects of diffusion on free precession in nuclear magnetic resonance experiments. *Phys. Rev.* **94,** 630–638 (1954).
177. Meiboom, S. & Gill, D. Modified spin-echo method for measuring nuclear relaxation times. *Rev. Sci. Instr., Vol. 29 (1958), pp. 688-691* **29,** 688–691 (1958).
178. McConnell, H. M. Reaction rates by nuclear magnetic resonance. *J. Chem. Phys.* **28,** 430–431 (1958).
179. Loria, J. P., Rance, M. & Palmer, A. G. A TROSY CPMG sequence for characterizing chemical exchange in large proteins. *J. Biomol. NMR* **15,** 151–155 (1999).

180. Vallurupalli, P., Hansen, D. F. & Kay, L. E. Structures of invisible, excited protein states by relaxation dispersion NMR spectroscopy. *Proc. Natl. Acad. Sci. U S A* **105,** 11766–11771 (2008).
181. Hansen, D. F., Vallurupalli, P. & Kay, L. E. Using relaxation dispersion NMR spectroscopy to determine structures of excited, invisible protein states. *J. Biomol. NMR* **41,** 113–120 (2008).
182. Zeeb, M. & Balbach, J. NMR spectroscopic characterisation of millisecond protein folding by transverse relaxation dispersion measurements. *J. Am. Chem. Soc.* **127,** 13207–13212 (2005).
183. Sekhar, A. & Kay, L. E. NMR paves the way for atomic level descriptions of sparsely populated, transiently formed biomolecular conformers. *Proc. Natl. Acad. Sci. U S A* **110,** 12867–12874 (2013).
184. Rosenzweig, R., Moradi, S., Zarrine-Afsar, A., Glover, J. R. & Kay, L. E. Unraveling the Mechanism of Protein Disaggregation Through a ClpB-DnaK Interaction. *Science* **339,** 1080–1083 (2013).
185. Shen, Y. *et al.* Consistent blind protein structure generation from NMR chemical shift data. *Proc. Natl. Acad. Sci. U S A* **105,** 4685–4690 (2008).
186. Korzhnev, D. M., Religa, T. L., Banachewicz, W., Fersht, A. R. & Kay, L. E. A transient and low-populated protein-folding intermediate at atomic resolution. *Science* **329,** 1312–1316 (2010).
187. Korzhnev, D. M. *et al.* Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR. *Nature* **430,** 586–590 (2004).
188. Bouvignies, G. *et al.* Solution structure of a minor and transiently formed state of a T4 lysozyme mutant. *Nature* **477,** 111–114 (2011).
189. Neudecker, P. *et al.* Structure of an Intermediate State in Protein Folding and Aggregation. *Science* **336,** 362–366 (2012).
190. Korzhnev, D. M., Religa, T. L. & Kay, L. E. Transiently populated intermediate functions as a branching point of the FF domain folding pathway. *Proc. Natl. Acad. Sci. U S A,* **109,** 17777-17782 (2012)
191. Whittier, S. K., Hengge, A. C. & Loria, J. P. Conformational motions regulate phosphoryl transfer in related protein tyrosine phosphatases. *Science* **341,** 899–903 (2013).
192. LundstrÖm, P. & Akke, M. Off-resonance rotating-frame amide proton spin relaxation experiments measuring microsecond chemical exchange in proteins. *J. Biomol. NMR* **32,** 163–173 (2005).
193. Bouvignies, G., Hansen, D. F., Vallurupalli, P. & Kay, L. E. Divided-evolution-based pulse scheme for quantifying exchange processes in proteins: powerful complement to relaxation dispersion experiments. *J. Am. Chem. Soc.* **133,** 1935–1945 (2011).
194. Tomlinson, J. H., Craven, C. J., Williamson, M. P. & Pandya, M. J. Dimerization of protein G B1 domain at low pH: A conformational switch caused by loss of a single hydrogen bond. *Proteins* **78,** 1652–1661 (2010).
195. Solomon, I. Relaxation Processes in a System of Two Spins. *Phys. Rev.* **99,** 559–565 (1955).
196. Xu, X. *et al.* Dynamics in a pure encounter complex of two proteins studied by solution scattering and paramagnetic NMR spectroscopy. *J. Am. Chem. Soc.* **130,** 6395–6403 (2008).
197. Battiste, J. L. & Wagner, G. Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold

determination of large proteins with limited nuclear Overhauser effect data †. *Biochemistry* **39,** 5355–5365 (2000).
198. Ikegami, T. T. *et al.* Novel techniques for weak alignment of proteins in solution using chemical tags coordinating lanthanide ions. *J. Biomol. NMR* **29,** 339–349 (2004).
199. Gottstein, D., Reckel, S., Dötsch, V. & Güntert, P. Requirements on paramagnetic relaxation enhancement data for membrane protein structure determination by NMR. *Structure* **20,** 1019–1027 (2012).
200. Liang, B., Bushweller, J. H. & Tamm, L. K. Site-directed parallel spin-labeling and paramagnetic relaxation enhancement in structure determination of membrane proteins by solution NMR spectroscopy. *J. Am. Chem. Soc.* **128,** 4389–4397 (2006).
201. Singh, S. K. *et al.* Mapping the interaction between the cytoplasmic domains of HIV-1 viral protein U and human CD4 with NMR spectroscopy. *FEBS J.* **279,** 3705–3714 (2012).
202. Wang, S. S. *et al.* Paramagnetic relaxation enhancement reveals oligomerization interface of a membrane protein. *J. Am. Chem. Soc.* **134,** 16995–16998 (2012).
203. Bertini, I., Luchinat, C. & Piccioli, M. Paramagnetic probes in metalloproteins. *Methods enzymol.* **339,** 314–340 (2001).
204. Bertini, I. From The Cover: Experimentally exploring the conformational space sampled by domain reorientation in calmodulin. *Proc. Natl. Acad. Sci. U S A* **101,** 6841–6846 (2004).
205. Bertini, I. *et al.* Paramagnetism-based NMR restraints provide maximum allowed probabilities for the different conformations of partially independent protein domains. *J. Am. Chem. Soc.* **129,** 12786–12794 (2007).
206. Barbieri, R. *et al.* Paramagnetically induced residual dipolar couplings for solution structure determination of lanthanide binding proteins. *J. Am. Chem. Soc.* **124,** 5581–5587 (2002).
207. Nilges, M. Calculation of protein structures with ambiguous distance restraints. Automated assignment of ambiguous NOE crosspeaks and disulphide connectivities. *J. Mol. Biol.* **245,** 645–660 (1995).
208. Rieping, W. *et al.* ARIA2: Automated NOE assignment and data integration in NMR structure calculation. *Bioinformatics* **23,** 381–382 (2007).
209. Iwahara, J., Schwieters, C. D. & Clore, G. M. Ensemble approach for NMR structure refinement against 1H paramagnetic relaxation enhancement data arising from a flexible paramagnetic group attached to a macromolecule. *J. Am. Chem. Soc.* **126,** 5879–5896 (2004).
210. Clore, G. M. & Iwahara, J. Theory, practice, and applications of paramagnetic relaxation enhancement for the characterisation of transient low-population states of biological macromolecules and their complexes. *Chem. Rev.* **109,** 4108–4139 (2009).
211. Tang, C., Schwieters, C. D. & Clore, G. M. Open-to-closed transition in apo maltose-binding protein observed by paramagnetic NMR. *Nature* **449,** 1078–1082 (2007).
212. Anthis, N. J. N., Doucleff, M. M. & Clore, G. M. G. Transient, sparsely populated compact states of apo and calcium-loaded calmodulin probed by paramagnetic relaxation enhancement: interplay of conformational selection and induced fit. *J. Am. Chem. Soc.* **133,** 18966–18974 (2011).

213. Iwahara, J., Schwieters, C. D. & Clore, G. M. Characterisation of nonspecific protein-DNA interactions by 1H paramagnetic relaxation enhancement. *J. Am. Chem. Soc.* **126,** 12800–12808 (2004).
214. Takayama, Y. & Clore, G. M. PNAS Plus: Intra- and intermolecular translocation of the bi-domain transcription factor Oct1 characterized by liquid crystal and paramagnetic NMR. *Proc. Natl. Acad. Sci. U S A* **108,** E169–E176 (2011).
215. Iwahara, J., Zweckstetter, M. & Clore, G. M. NMR structural and kinetic characterisation of a homeodomain diffusing and hopping on nonspecific DNA. *Proc. Natl. Acad. Sci. U S A* **103,** 15062–15067 (2006).
216. Wu, K. P. & Baum, J. Detection of transient interchain interactions in the intrinsically disordered protein alpha-synuclein by NMR paramagnetic relaxation enhancement. *J. Am. Chem. Soc.* **132,** 5546–5547
217. Dedmon, M. M., Lindorff-Larsen, K., Christodoulou, J., Vendruscolo, M. & Dobson, C. M. Mapping long-range interactions in α-Synuclein using spin-label NMR and ensemble molecular dynamics simulations. *J. Am. Chem. Soc.* **127,** 476–477 (2005).
218. Qu, B. H. & Thomas, P. J. Alteration of the cystic fibrosis transmembrane conductance regulator folding pathway. *J. Biol. Chem.* **271,** 7261–7264 (1996).
219. Welch, W. J. Role of quality control pathways in human diseases involving protein misfolding. *Sem. Cell Dev. Biol.* **15,** 31–38 (2004).
220. Chiti, F. & Dobson, C. M. Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem.* **75,** 333–366 (2006).
221. Dobson, C. M. Principles of protein folding, misfolding and aggregation. *Semin. Cell Dev. Biol.* **15,** 3–16 (2004).
222. Dobson, C. M. Protein folding and misfolding. *Nature* **426,** 884–890 (2003).
223. Sipe, J. D. & Cohen, A. S. Review: history of the amyloid fibril. *J. Struct. Biol.* **130,** 88–98 (2000).
224. Eichner, T. & Radford, S. E. A diversity of assembly mechanisms of a generic amyloid fold. *Mol. Cell* **43,** 8–18 (2011).
225. Sunde, M. & Blake, C. The structure of amyloid fibrils by electron microscopy and X-ray diffraction. *Adv. Protein Chem.* **50,** 123–159 (1997).
226. Kreplak, L. & Aebi, U. From the polymorphism of amyloid fibrils to their assembly mechanism and cytotoxicity. *Adv. Protein Chem.* **73,** 217–233 (2006).
227. Glenner, G. G. Amyloid deposits and amyloidosis. The beta-fibrilloses (first of two parts). *N. Engl. J. Med.* **302,** 1283–1292 (1980).
228. Glenner, G. G. Amyloid deposits and amyloidosis: the beta-fibrilloses (second of two parts). *N. Engl. J. Med.* **302,** 1333–1343 (1980).
229. Koo, E. H., Lansbury, P. T. & Kelly, J. W. Amyloid diseases: abnormal protein aggregation in neurodegeneration. *Proc. Natl. Acad. Sci. U S A* **96,** 9989–9990 (1999).
230. Butler, A. E. A. *et al.* Diabetes due to a progressive defect in beta-cell mass in rats transgenic for human islet amyloid polypeptide (HIP Rat): a new model for type 2 diabetes. *Diabetes* **53,** 1509–1516 (2004).
231. Sousa, M. M., Cardoso, I., Fernandes, R., Guimarães, A. & Saraiva, M. J. Deposition of transthyretin in early stages of familial amyloidotic

polyneuropathy: evidence for toxicity of nonfibrillar aggregates. *Am. J. Pathol.* **159,** 1993–2000 (2001).
232. Verchere, C. B. C. *et al.* Islet amyloid formation associated with hyperglycemia in transgenic mice with pancreatic beta cell expression of human islet amyloid polypeptide. *Proc. Natl. Acad. Sci. U S A* **93,** 3492–3496 (1996).
233. Dodart, J.-C. *et al.* Immunization reverses memory deficits without reducing brain Abeta burden in Alzheimer's disease model. *Nat. Neurosci.* **5,** 452–457 (2002).
234. Kotilinek, L. A. *et al.* Reversible memory loss in a mouse transgenic model of Alzheimer's disease. *J. Neurosci.* **22,** 6331–6335 (2002).
235. Giannakopoulos, P. *et al.* Tangle and neuron numbers, but not amyloid load, predict cognitive status in Alzheimer's disease. *Neurology* **60,** 1495–1500 (2003).
236. Terry, R. D. R. *et al.* Physical basis of cognitive alterations in Alzheimer's disease: synapse loss is the major correlate of cognitive impairment. *Ann. Neurol.* **30,** 572–580 (1991).
237. Katzman, R. *et al.* Clinical, pathological, and neurochemical changes in dementia: a subgroup with preserved mental status and numerous neocortical plaques. *Ann. Neurol.* **23,** 138–144 (1988).
238. Lesné, S. *et al.* A specific amyloid-beta protein assembly in the brain impairs memory. *Nature* **440,** 352–357 (2006).
239. Rahimi, F. & Bitan, G. *Non-fibrillar Amyloidogenic Protein Assemblies - Common Cytotoxins Underlying Degenerative Diseases*. (Springer, 2012).
240. Reixach, N. N., Deechongkit, S. S., Jiang, X. X., Kelly, J. W. J. & Buxbaum, J. N. J. Tissue damage in the amyloidoses: Transthyretin monomers and nonnative oligomers are the major cytotoxic species in tissue culture. *Proc. Natl. Acad. Sci. U S A* **101,** 2817–2822 (2004).
241. Klein, W. L. W., Krafft, G. A. G. & Finch, C. E. C. Targeting small Abeta oligomers: the solution to an Alzheimer's disease conundrum? *Trends Neurosci* .**24,** 219–224 (2001).
242. Teng, M. H. M. *et al.* Amyloid and nonfibrillar deposits in mice transgenic for wild-type human transthyretin: a possible model for senile systemic amyloidosis. *Lab Invest.* **81,** 385–396 (2001).
243. Janson, J. J., Ashley, R. H. R., Harrison, D. D., McIntyre, S. S. & Butler, P. C. P. The mechanism of islet amyloid polypeptide toxicity is membrane disruption by intermediate-sized toxic amyloid particles. *Diabetes* **48,** 491–498 (1999).
244. Bucciantini, M. *et al.* Toxic effects of amyloid fibrils on cell membranes: the importance of ganglioside GM1. *FASEB J.* **26,** 818–831 (2012).
245. Xue, W. F. *et al.* Fibril fragmentation enhances amyloid cytotoxicity. *J. Biol. Chem.* **284,** 34272–34282 (2009).
246. Maji, S. K. *et al.* Functional amyloids as natural storage of peptide hormones in pituitary secretory granules. *Science* **325,** 328–332 (2009).
247. Badtke, M. P., Hammer, N. D. & Chapman, M. R. Functional amyloids signal their arrival. *Sci Signal.* **2,** pe43 (2009).
248. Iconomidou, V. A., Cordopatis, P., Hoenger, A. & Hamodrakas, S. J. The silkmoth eggshell as a natural amyloid shield for the safe development of insect oocyte and embryo: Insights from studies of silkmoth chorion

protein peptide-analogues of the B family. *Biopolymers* **96,** 723–733 (2011).
249. Kyle, S., Aggeli, A., Ingham, E. & McPherson, M. J. Recombinant self-assembling peptides as biomaterials for tissue engineering. *Biomaterials* **31,** 9395–9405
250. Westermark, P. *et al.* A primer of amyloid nomenclature. *Amyloid* **14,** 179–183 (2007).
251. Sunde, M. M. *et al.* Common core structure of amyloid fibrils by synchrotron X-ray diffraction. *J. Mol. Biol.* **273,** 11–11 (1997).
252. Geddes, A. J., Parker, K. D., Atkins, E. D. & Beighton, E. 'Cross-beta' conformation in proteins. *J. Mol. Biol.* **32,** 343–358 (1968).
253. Chan, P. K. *et al.* Structural similarity of wild-type and ALS-mutant superoxide dismutase-1 fibrils using limited proteolysis and atomic force microscopy. *Proc. Natl. Acad. Sci. U S A* **110,** 10934–10939 (2013).
254. Zhang, R. *et al.* Interprotofilament interactions between Alzheimer's Aβ1–42 peptides in amyloid fibrils revealed by cryoEM. *Proc. Natl. Acad. Sci. U S A* **106,** 4653–4658 (2009).
255. Sarell, C. J., Stockley, P. G. & Radford, S. E. Assessing the causes and consequences of co-polymerisation in amyloid formation. *Prion* **7,** 0––1 (2013).
256. Sawaya, M. R. *et al.* Atomic structures of amyloid cross-beta spines reveal varied steric zippers. *Nature* **447,** 453–457 (2007).
257. Chiti, F. & Dobson, C. M. Amyloid formation by globular proteins under native conditions. *Nat. Chem. Biol.* **5,** 15–22 (2009).
258. Ferrone, F. in *Methods Enzymol.* **309,** 256–274 (Elsevier, 1999).
259. Naiki, H. & Gejyo, F. in *Methods Enzymol.* **309,** 305–318 (Elsevier, 1999).
260. Bucciantini, M. M. *et al.* Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases. *Nature* **416,** 507–511 (2002).
261. Kirkitadze, M. D., Bitan, G. & Teplow, D. B. Paradigm shifts in Alzheimer's disease and other neurodegenerative disorders: the emerging role of oligomeric assemblies. *J. Neurosci. Res.* **69,** 567–577 (2002).
262. Stroud, J. C. J., Liu, C. C., Teng, P. K. P. & Eisenberg, D. D. Toxic fibrillar oligomers of amyloid-β have cross-β structure. *Proc. Natl. Acad. Sci. U S A*, 7717–7722 (2012).
263. Xue, W. F., Homans, S. W. & Radford, S. E. Systematic analysis of nucleation-dependent polymerisation reveals new insights into the mechanism of amyloid self-assembly. *Proc. Natl. Acad. Sci. U S A* **105,** 8926–8931 (2008).
264. Knowles, T. P. J. *et al.* An analytical solution to the kinetics of breakable filament assembly. *Science* **326,** 1533–1537 (2009).
265. Cohen, S. I. A. *et al.* Proliferation of amyloid- 42 aggregates occurs through a secondary nucleation mechanism. *Proc. Natl. Acad. Sci. U S A* **110,** 9758–9763 (2013).
266. Raffen, R. R. *et al.* Physicochemical consequences of amino acid variations that contribute to fibril formation by immunoglobulin light chains. *Protein Sci.* **8,** 509–517 (1999).
267. Canet, D. D. *et al.* Local cooperativity in the unfolding of an amyloidogenic variant of human lysozyme. *Nat. Struct. Mol. Biol.* **9,** 308–315 (2002).

268. Dhulesia, A. *et al.* Local co-operativity in an amyloidogenic state of human lysozyme observed at atomic resolution. *J. Am. Chem. Soc.* **132,** 15580–15588 (2010).
269. Marcon, G. G. *et al.* Amyloid formation from HypF-N under conditions in which the protein is initially in its native state. *J. Mol. Biol.* **347,** 13–13 (2005).
270. Soldi, G. *et al.* Amyloid formation of a protein in the absence of initial unfolding and destabilization of the native state. *Biophys. J.* **89,** 4234–4244 (2005).
271. Lim, K. H., Dyson, H. J., Kelly, J. W. & Wright, P. E. Localized structural fluctuations promote amyloidogenic conformations in transthyretin. *J. Mol. Biol.* **425,** 977–988 (2013).
272. Eichner, T., Kalverda, A. P., Thompson, G. S., Homans, S. W. & Radford, S. E. Conformational conversion during amyloid formation at atomic resolution. *Mol. Cell* **41,** 161–172 (2011).
273. Chiti, F. F. *et al.* Mutational analysis of the propensity for amyloid formation by a globular protein. *EMBO J.* **19,** 1441–1449 (2000).
274. Belli, M., Ramazzotti, M. & Chiti, F. Prediction of amyloid aggregation in vivo. *EMBO Rep.* **12,** 657–663 (2011).
275. Nelson, R. & Eisenberg, D. Structural models of amyloid-like fibrils. *Adv. Protein Chem.* **73,** 235–282 (2006).
276. Nelson, R. *et al.* Structure of the cross-beta spine of amyloid-like fibrils. *Nature* **435,** 773–778 (2005).
277. Tycko, R. Solid-state NMR studies of amyloid fibril structure. *Annu. Rev. Phys. Chem.* **62,** 279–299 (2011).
278. Bertini, I., Gonnelli, L., Luchinat, C., Mao, J. & Nesi, A. A new structural model of Aβ40 fibrils. *J. Am. Chem. Soc.* **133,** 16013–16022 (2011).
279. Petkova, A. T. *et al.* A structural model for Alzheimer's beta -amyloid fibrils based on experimental constraints from solid state NMR. *Proc. Natl. Acad. Sci. U S A* **99,** 16742–16747 (2002).
280. Wasmer, C. *et al.* Amyloid fibrils of the HET-s(218-289) prion form a beta solenoid with a triangular hydrophobic core. *Science* **319,** 1523–1526 (2008).
281. Fitzpatrick, A. W. P. *et al.* Atomic structure and hierarchical assembly of a cross-β amyloid fibril. *Proc. Natl. Acad. Sci. U S A* **110,** 5468–5473 (2013).
282. Demuro, A. & Parker, I. Cytotoxicity of intracellular aβ42 amyloid oligomers involves Ca2+ release from the endoplasmic reticulum by stimulated production of inositol trisphosphate. *J. Neurosci.* **33,** 3824–3833 (2013).
283. Kayed, R. Common structure of soluble amyloid oligomers implies common mechanism of pathogenesis. *Science* **300,** 486–489 (2003).
284. Chiou, A. *et al.* Probing neuroserpin polymerisation and interaction with amyloid-β peptides using single molecule fluorescence. *Biophys. J.* **97,** 2306–2315 (2009).
285. Cremades, N. N. *et al.* Direct observation of the interconversion of normal and toxic forms of α-synuclein. *Cell* **149,** 1048–1059 (2012).
286. Campioni, S. *et al.* A causative link between the structure of aberrant protein oligomers and their toxicity. *Nat. Chem. Biol.* **6,** 140–147 (2010).
287. Laganowsky, A. *et al.* Atomic View of a Toxic Amyloid Small Oligomer. *Science* **335,** 1228–1231 (2012).

288. Liu, C. C. *et al.* Out-of-register β-sheets suggest a pathway to toxic amyloid aggregates. *Proc. Natl. Acad. Sci. U S A* **109,** 20913–20918 (2012).
289. Härd, T. & Lendel, C. Inhibition of amyloid formation. *J. Mol. Biol.* **421,** 441–465 (2012).
290. Porat, Y., Abramowitz, A. & Gazit, E. Inhibition of Amyloid Fibril Formation by Polyphenols: Structural Similarity and Aromatic Interactions as a Common Inhibition Mechanism. *Chem. Biol. Drug Design* **67,** 27–37 (2006).
291. Ono, K. *et al.* Effects of grape seed-derived polyphenols on amyloid -protein self-assembly and cytotoxicity. *J. Biol. Chem..* **283,** 32176–32187 (2008).
292. Bieschke, J. *et al.* EGCG remodels mature α-synuclein and amyloid fibrils and reduces cellular toxicity. *Proc. Natl. Acad. Sci. U S A* **107,** 7710–7715 (2010).
293. Conway, K. A., Rochet, J.-C., Bieganski, R. M. & Lansbury, P. T. Kinetic Stabilization of the α-Synuclein Protofibril by a Dopamine-α-Synuclein Adduct. *Science* **294,** 1346–1349 (2001).
294. Woods, L. A. *et al.* Ligand binding to distinct states diverts aggregation of an amyloid-forming protein. *Nat. Chem. Biol.* **7,** 730–739 (2011).
295. Ferreira, N., Saraiva, M. J. & Almeida, M. R. Natural polyphenols inhibit different steps of the process of transthyretin (TTR) amyloid fibril formation. *FEBS lett.* **585,** 2424–2430 (2011).
296. Bieschke, J. *et al.* Small-molecule conversion of toxic oligomers to nontoxic β-sheet–rich amyloid fibrils. *Nat. Chem. Biol.* **8,** 93–101 (2011).
297. Montagna, G. *et al.* Benefit of doxycycline treatment on articular disability caused by dialysis related amyloidosis. *Amyloid* **20,** 173–178 (2013).
298. Johnson, S. M. S., Connelly, S. S., Fearns, C. C., Powers, E. T. E. & Kelly, J. W. J. The transthyretin amyloidoses: from delineating the molecular mechanism of aggregation linked to pathology to a regulatory-agency-approved drug. *J. Mol. Biol.* **421,** 185–203 (2012).
299. Connelly, S., Choi, S., Johnson, S. M., Kelly, J. W. & Wilson, I. A. Structure-based design of kinetic stabilizers that ameliorate the transthyretin amyloidoses. *Curr. Opin. Struct. Biol.* **20,** 54–62 (2010).
300. Ray, S. S., Nowak, R. J., Brown, R. H. & Lansbury, P. T. Small-molecule-mediated stabilization of familial amyotrophic lateral sclerosis-linked superoxide dismutase mutants against unfolding and aggregation. *Proc. Natl. Acad. Sci. U S A* **102,** 3639–3644 (2005).
301. Dumoulin, M. *et al.* A camelid antibody fragment inhibits the formation of amyloid fibrils by human lysozyme. *Nature* **424,** 783–788 (2003).
302. Powers, E. T., Morimoto, R. I., Dillin, A., Kelly, J. W. & Balch, W. E. Biological and chemical approaches to diseases of proteostasis deficiency. *Annu. Rev. Biochem.* **78,** 959–991 (2009).
303. Balch, W. E., Morimoto, R. I., Dillin, A. & Kelly, J. W. Adapting proteostasis for Disease Intervention. *Science* **319,** 916–919 (2008).
304. Dedmon, M. M., Christodoulou, J., Wilson, M. R. & Dobson, C. M. Heat shock protein 70 inhibits alpha-synuclein fibril formation via preferential binding to prefibrillar species. *J. Biol. Chem..* **280,** 14733–14740 (2005).
305. Saper, M. A., Bjorkman, P. J. & Wiley, D. C. Refined structure of the human histocompatibility antigen HLA-A2 at 2.6 Å resolution. *J. Mol. Biol.* **219,** 277–319 (1991).

306. Germain, R. N. Immunology. The ins and outs of antigen processing and presentation. *Nature* **322,** 687–689 (1986).
307. Borbulevych, O. Y. O., Do, P. P. & Baker, B. M. B. Structures of native and affinity-enhanced WT1 epitopes bound to HLA-A*0201: Implications for WT1-based cancer therapeutics. *Mol. Immunol.* **47,** 6–6 (2010).
308. Cascio, P., Hilton, C., Kisselev, A. F., Rock, K. L. & Goldberg, A. L. 26S proteasomes and immunoproteasomes produce mainly N-extended versions of an antigenic peptide. *EMBO J.* **20,** 2357–2366 (2001).
309. Goldberg, A. L. A., Cascio, P. P., Saric, T. T. & Rock, K. L. K. The importance of the proteasome and subsequent proteolytic steps in the generation of antigenic peptides. *Mol. Immunol.* **39,** 147–164 (2002).
310. Paulsson, K. M. K. *et al.* Distinct differences in association of MHC class I with endoplasmic reticulum proteins in wild-type, and beta 2-microglobulin- and TAP-deficient cell lines. *Int. Immunol.* **13,** 1063–1073 (2001).
311. Cresswell, P., Ackerman, A. L., Giodini, A., Peaper, D. R. & Wearsch, P. A. Mechanisms of MHC class I-restricted antigen processing and cross-presentation. *Immunol. Rev.* **207,** 145–157 (2005).
312. Floege, J. *et al.* Clearance and synthesis rates of beta 2-microglobulin in patients undergoing hemodialysis and in normal subjects. *J. Lab. Clin. Med.* **118,** 153–165 (1991).
313. Kay, J. Beta 2-microglobulin amyloidosis in renal failure: understanding this recently recognized condition. *Cleve Clin. J. Med.* **66,** 145–147 (1999).
314. Otsubo, S. *et al.* Characteristics of dialysis-related amyloidosis in patients on haemodialysis therapy for more than 30 years. *Nephrol. Dial. Transplant.* **24,** 1593–1598 (2009).
315. Linke, R. P. *et al.* Production of recombinant human beta2-microglobulin for scintigraphic diagnosis of amyloidosis in uremia and hemodialysis. *Eur. J. Biochem.* **267,** 627–633 (2000).
316. Gejyo, F. *et al.* A new form of amyloid protein associated with chronic hemodialysis was identified as beta 2-microglobulin. *Biochem. Biophys. Res. Com.* **129,** 701–706 (1985).
317. Esposito, G. *et al.* Removal of the N-terminal hexapeptide from human beta2-microglobulin facilitates protein aggregation and fibril formation. *Protein Sci..* **9,** 831–845 (2000).
318. Bellotti, V., Gallieni, M., Giorgetti, S. & Brancaccio, D. Dynamics of beta(2)-microglobulin fibril formation and reabsorption: the role of proteolysis. *Semin. Dial.* **14,** 117–122 (2001).
319. Heegaard, N. H. H., Roepstorff, P., Melberg, S. G. & Nissen, M. H. Cleaved beta 2-microglobulin partially attains a conformation that has amyloidogenic features. *J. Biol. Chem..* **277,** 11184–11189 (2002).
320. Miyata, T. *et al.* Identification of pentosidine as a native structure for advanced glycation end products in beta-2-microglobulin-containing amyloid fibrils in patients with dialysis-related amyloidosis. *Proc. Natl. Acad. Sci. U S A* **93,** 2353–2358 (1996).
321. Niwa, T. *et al.* Amyloid beta 2-microglobulin is modified with imidazolone, a novel advanced glycation end product, in dialysis-related amyloidosis. *Kidney Int.* **51,** 187–194 (1997).

322. Morgan, C. J., Gelfand, M., Atreya, C. & Miranker, A. D. Kidney dialysis-associated amyloidosis: a molecular role for copper in fiber formation. *J. Mol. Biol.* **309,** 339–345 (2001).
323. Myers, S. L. *et al.* A systematic study of the effect of physiological factors on β 2-microglobulin amyloid formation at neutral pH †. *Biochemistry* **45,** 2311–2321 (2006).
324. Relini, A. *et al.* Collagen plays an active role in the aggregation of beta2-microglobulin under physiopathological conditions of dialysis-related amyloidosis. *J. Biol. Chem..* **281,** 16521–16529 (2006).
325. Yamamoto, S. *et al.* Low concentrations of sodium dodecyl sulfate induce the extension of beta 2-microglobulin-related amyloid fibrils at a neutral pH. *Biochemistry* **43,** 11075–11082 (2004).
326. Relini, A. *et al.* Heparin strongly enhances the formation of beta2-microglobulin amyloid fibrils in the presence of type I collagen. *J. Biol. Chem.* **283,** 4912–4920 (2008).
327. Athanasou, N. A., Puddle, B. & Sallie, B. Highly sulphated glycosaminoglycans in articular cartilage and other tissues containing beta 2 microglobulin dialysis amyloid deposits. *Nephrol. Dial. Transplant.* **10,** 1672–1678 (1995).
328. Yamamoto, S. *et al.* Kinetic analysis of the polymerisation and depolymerisation of beta(2)-microglobulin-related amyloid fibrils in vitro. *Biochim. Biophys. Acta* **1753,** 34–43 (2005).
329. Ookoshi, T. *et al.* Lysophospholipids induce the nucleation and extension of beta2-microglobulin-related amyloid fibrils at a neutral pH. *Nephrol. Dial. Transplant.* **23,** 3247–3255 (2008).
330. Hasegawa, K. *et al.* Growth of beta(2)-microglobulin-related amyloid fibrils by non-esterified fatty acids at a neutral pH. *Biochem. J.* **416,** 307–315 (2008).
331. Drüeke, T. B. Beta2-microglobulin and amyloidosis. *Nephrol. Dial. Transplant.* **15 Suppl 1,** 17–24 (2000).
332. Morinière, P. *et al.* Destructive spondyloarthropathy with beta 2-microglobulin amyloid deposits in a uremic patient before chronic hemodialysis. *Nephron* **59,** 654–657 (1991).
333. Ullian, M. E., Hammond, W. S., Alfrey, A. C., Schultz, A. & Molitoris, B. A. Beta-2-microglobulin-associated amyloidosis in chronic hemodialysis patients with carpal tunnel syndrome. *Medicine (Baltimore)* **68,** 107–115 (1989).
334. Porter, M. Y., Routledge, K. E., Radford, S. E. & Hewitt, E. W. Characterisation of the response of primary cells relevant to dialysis-related amyloidosis to β2-microglobulin monomer and fibrils. *PLoS ONE* **6,** e27353 (2011).
335. Valleix, S. *et al.* Hereditary systemic amyloidosis due to Asp76Asn variant β2-microglobulin. *N Engl J Med.* **366,** 2276–2283 (2012).
336. Smith, D. P., Jones, S., Serpell, L. C., Sunde, M. & Radford, S. E. A systematic investigation into the effect of protein destabilisation on beta 2-microglobulin amyloid formation. *J. Mol. Biol.* **330,** 943–954 (2003).
337. Kad, N. M. *et al.* Hierarchical assembly of β2-Microglobulin amyloid in vitro revealed by atomic force microscopy. *J. Mol. Biol.* **330,** 785–797 (2003).

338. Esposito, G. *et al.* The controlling roles of Trp60 and Trp95 in beta2-microglobulin function, folding and amyloid aggregation properties. *J. Mol. Biol.* **378,** 887–897 (2008).
339. Kad, N. M., Thomson, N. H., Smith, D. P., Smith, D. A. & Radford, S. E. Beta(2)-microglobulin and its deamidated variant, N17D form amyloid fibrils with a range of morphologies in vitro. *J. Mol. Biol.* **313,** 559–571 (2001).
340. Smith, D. P., Radford, S. E. & Ashcroft, A. E. Elongated oligomers in beta2-microglobulin amyloid assembly revealed by ion mobility spectrometry-mass spectrometry. *Proc. Natl. Acad. Sci. U S A* **107,** 6794–6798
341. Debelouchina, G. T., Platt, G. W., Bayro, M. J., Radford, S. E. & Griffin, R. G. Intermolecular alignment in β2-microglobulin amyloid fibrils. *J. Am. Chem. Soc.* **132,** 17077–17079 (2013).
342. Debelouchina, G. T., Platt, G. W., Bayro, M. J., Radford, S. E. & Griffin, R. G. Magic angle spinning NMR analysis of beta2-microglobulin amyloid fibrils in two distinct morphologies. *J. Am. Chem. Soc.* **132,** 10414–10423 (2010).
343. Ladner, C. L. *et al.* Stacked sets of parallel, in-register beta-strands of beta2-microglobulin in amyloid fibrils revealed by site-directed spin labeling and chemical labeling. *J. Biol. Chem..* **285,** 17137–17147 (2010)
344. Trinh, C. H., Smith, D. P., Kalverda, A. P., Phillips, S. E. & Radford, S. E. Crystal structure of monomeric human beta-2-microglobulin reveals clues to its amyloidogenic properties. *Proc. Natl. Acad. Sci. U S A* **99,** 9771–9776 (2002).
345. White, H. E. *et al.* Globular tetramers of beta(2)-microglobulin assemble into elaborate amyloid fibrils. *J. Mol. Biol.* **389,** 48–57 (2009).
346. Sheynis, T. *et al.* Aggregation modulators interfere with membrane interactions of β2-Microglobulin fibrils. *Biophys J.* **105,** 745–755 (2013).
347. Milanesi, L. *et al.* Direct three-dimensional visualization of membrane disruption by amyloid fibrils. *Proc. Natl. Acad. Sci. U S A* **109,** 20455–20460 (2012).
348. Dumoulin, M. *et al.* Reduced global cooperativity is a common feature underlying the amyloidogenicity of pathogenic lysozyme mutations. *J. Mol. Biol.* **346,** 773–788 (2005).
349. Plakoutsi, G., Taddei, N., Stefani, M. & Chiti, F. Aggregation of the acylphosphatase from Sulfolobus solfataricus: the folded and partially unfolded states can both be precursors for amyloid formation. *J. Biol. Chem..* **279,** 14111–14119 (2004).
350. Colon, W. & Kelly, J. W. Partial denaturation of transthyretin is sufficient for amyloid fibril formation in vitro. *Biochemistry* **31,** 8654–8660 (1992).
351. McCutchen, S. L., Colon, W. & Kelly, J. W. Transthyretin mutation Leu-55-Pro significantly alters tetramer stability and increases amyloidogenicity. *Biochemistry* **32,** 12119–12127 (1993).
352. Chiti, F. *et al.* Detection of two partially structured species in the folding process of the amyloidogenic protein β2-microglobulin1. *J. Mol. Biol.* **307,** 379–391 (2001).
353. Kameda, A. *et al.* Nuclear magnetic resonance characterisation of the refolding intermediate of β2-Microglobulin trapped by non-native prolyl peptide bond. *J. Mol. Biol.* **348,** 383–397 (2005).

354. Sakata, M. *et al.* Kinetic coupling of folding and prolyl isomerization of β2-microglobulin studied by mutational analysis. *J. Mol. Biol.* **382,** 1242–1255 (2008).
355. Lu, K. P., Finn, G., Lee, T. H. & Nicholson, L. K. Prolyl cis-trans isomerization as a molecular timer. *Nat. Chem. Biol.* **3,** 619–629 (2007).
356. Chiti, F. *et al.* A partially structured species of beta 2-microglobulin is significantly populated under physiological conditions and involved in fibrillogenesis. *J. Biol. Chem..* **276,** 46714–46721 (2001).
357. Jahn, T. R., Parker, M. J., Homans, S. W. & Radford, S. E. Amyloid formation under physiological conditions proceeds via a native-like folding intermediate. *Nat. Struct. Mol. Biol.* **13,** 195–201 (2006).
358. Schanda, P., Forge, V. & Brutscher, B. Protein folding and unfolding studied at atomic resolution by fast two-dimensional NMR spectroscopy. *Proc. Natl. Acad. Sci. U S A* **104,** 11257–11262 (2007).
359. Rennella, E. *et al.* Real-Time NMR Characterisation of Structure and Dynamics in a Transiently Populated Protein Folding Intermediate. *J. Am. Chem. Soc.* **134,** 8066–8069 (2012).
360. Corazza, A. *et al.* Native-unlike long-lived intermediates along the folding pathway of the amyloidogenic protein β2-Microglobulin revealed by real-time two-dimensional NMR. *J. Biol. Chem.* **285,** 5827–5835 (2010).
361. Eichner, T. & Radford, S. E. A generic mechanism of β2-Microglobulin amyloid assembly at neutral pH involving a specific proline switch. *J. Mol. Biol.* **386,** 1312–1326 (2009).
362. Eichner, T. & Radford, S. E. Understanding the complex mechanisms of β2-microglobulin amyloid assembly. *FEBS J.* **278,** 3868–3883 (2011).
363. Domanska, K. *et al.* Atomic structure of a nanobody-trapped domain-swapped dimer of an amyloidogenic 2-microglobulin variant. *Proc. Natl. Acad. Sci. U S A* **108,** 1314–1319 (2011).
364. Tessier, P. M. P. & Lindquist, S. S. Prion recognition elements govern nucleation, strain specificity and species barriers. *Nature* **447,** 556–561 (2007).
365. Ivanova, M. I. An amyloid-forming segment of 2-microglobulin suggests a molecular model for the fibril. *Proc. Natl. Acad. Sci. U S A* **101,** 10584–10589 (2004).
366. Schwieters, C. D., Kuszewski, J. J., Tjandra, N. & Clore, G. M. The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.* **160,** 65–73 (2003).
367. Keeler, J. *Understanding NMR Spectroscopy*. (John Wiley & Sons, 2013).
368. Jacobsen, N. E. *NMR Spectroscopy Explained*. (John Wiley & Sons, 2007).
369. Levitt, M. H. *Spin Dynamics*. (John Wiley & Sons, 2013).
370. John, M., Headlam, M. J., Dixon, N. E. & Otting, G. Assignment of paramagnetic ^{15}N-HSQC spectra by heteronuclear exchange spectroscopy. *J. Biomol. NMR* **37,** 43–51 (2006).
371. Bertini, I., Jiménez, B. & Piccioli, M. 13C Direct detected experiments: Optimization for paramagnetic signals. *J. Magn. Res.* **174,** 125–132 (2005).
372. Lipari, G. & Szabo, A. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc.* **104,** 4546–4559 (1982).
373. Kunkel, T. A. Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc. Natl. Acad. Sci. U S A* **82,** 488–492 (1985).

374. Gill, S. C. & Hippel, von, P. H. Calculation of protein extinction coefficients from amino acid sequence data. *Anal. biochem.* **182,** 319–326 (1989).
375. Brown, P. H. & Schuck, P. Macromolecular size-and-shape distributions by sedimentation velocity analytical ultracentrifugation. *Biophys. J.* **90,** 4651–4661 (2006).
376. Delaglio, F. *et al.* NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6,** 277–293 (1995).
377. Vranken, W. F. *et al.* The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **59,** 687–696 (2005).
378. Shen, Y., Delaglio, F., Cornilescu, G. & Bax, A. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* **44,** 213–223 (2009).
379. Yang, D. & Kay, L. E. Contributions to conformational entropy arising from bond vector fluctuations measured from NMR-derived order parameters: application to protein folding. *J. Mol. Biol.* **263,** 369–382 (1996).
380. Englander, S. W. & Mayne, L. Protein folding studied using hydrogen-exchange labeling and two-dimensional NMR. *Annu. Rev. Biophys. Biomol. Struct.* **21,** 243–265 (1992).
381. Englander, S. W. Protein folding intermediates and pathways studied by hydrogen exchange. *Annu. Rev. Biophys. Biomol. Struct.* **29,** 213–238 (2000).
382. Schanda, P., Kupče, Ē. & Brutscher, B. SOFAST-HMQC Experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds. *J. Biomol. NMR* **33,** 199–211 (2005).
383. Johnson, B. A. & Blevins, R. A. NMR View: A computer program for the visualization and analysis of NMR data. *J. Biomol. NMR* **4,** 603–614 (1994).
384. Bieri, M. & Gooley, P. R. Automated NMR relaxation dispersion data analysis using NESSY. *BMC Bioinformatics* **12,** 421 (2011).
385. Schanda, P. & Brutscher, B. Very Fast Two-Dimensional NMR spectroscopy for real-time investigation of dynamic events in proteins on the time scale of seconds. *J. Am. Chem. Soc.* **127,** 8014–8015 (2005).
386. Tang, C., Ghirlando, R. & Clore, G. M. Visualization of transient ultra-weak protein self-association in solution using paramagnetic relaxation enhancement. *J. Am. Chem. Soc.* **130,** 4048–4056 (2008).
387. Clore, G. M. & Schwieters, C. D. Docking of protein−protein complexes on the basis of highly ambiguous intermolecular distance restraints derived from 1H N/ 15N chemical shift mapping and backbone $^{15}N-{}^{1}H$ residual dipolar couplings using conjoined rigid body/torsion angle dynamics. *J. Am. Chem. Soc.* **125,** 2902–2912 (2003).
388. Kuszewski, J., Gronenborn, A. & Clore, M. Improving the packing and accuracy of NMR structures with a pseudopotential for the radius of gyration. *J. Am. Chem. Soc.* **121,** 2337–2338 (1999).
389. Baldwin, A. J. *et al.* Metastability of native proteins and the phenomenon of amyloid formation. *J. Am. Chem. Soc.* **133,** 14160–14163 (2011).
390. Tartaglia, G. G. *et al.* Prediction of aggregation-prone regions in structured proteins. *J Mol Biol* **380,** 425–436 (2008).
391. Hodkinson, J. P., Jahn, T. R., Radford, S. E. & Ashcroft, A. E. HDX-ESI-MS reveals enhanced conformational dynamics of the amyloidogenic protein beta(2)-microglobulin upon release from the MHC-1. *J. Am. Soc. Mass Spectrom.* **20,** 278–286 (2009).

392. Santambrogio, C. C. *et al.* DE-loop mutations affect beta2 microglobulin stability, oligomerization, and the low-pH unfolded form. *Protein Sci.* **19,** 1386–1394 (2010).
393. Esposito, G. *et al.* Solution structure of β2-microglobulin and insights into fibrillogenesis. *Dialysis-related amyloidosis: from molecular mechanisms to therapies* **1753,** 76–84 (2005).
394. Verdone, G. G. *et al.* The solution structure of human beta2-microglobulin reveals the prodromes of its amyloid transition. *Protein Sci.* **11,** 487–499 (2002).
395. Wang, Y. Y., Li, C. C. & Pielak, G. J. G. Effects of proteins on protein diffusion. *J. Am. Chem. Soc.* **132,** 9392–9397 (2010).
396. Rennella, E. *et al.* Oligomeric states along the folding pathways of β2-microglobulin: kinetics, thermodynamics, and structure. *J. Mol. Biol.* **425,** 2722–2736 (2013).
397. Dominguez, C., Boelens, R. & Bonvin, A. M. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* **125,** 1731–1737 (2003).
398. Jahn, T. R., Tennent, G. A. & Radford, S. E. A common β-sheet architecture underlies in vitro and in vivo β2-microglobulin amyloid fibrils. *J. Biol. Chem.* **283,** 17279–17286 (2008).
399. Eakin, C. M., Berman, A. J. & Miranker, A. D. A native to amyloidogenic transition regulated by a backbone trigger. *Nat. Struct. Mol. Biol*. **13,** 202–208 (2006).
400. Bulawa, C. E. *et al.* Tafamidis, a potent and selective transthyretin kinetic stabilizer that inhibits the amyloid cascade. *Proc. Natl. Acad. Sci. U S A* **109,** 9629–9634 (2012).
401. Hoyer, W., Grönwall, C., Jonsson, A., Ståhl, S. & Härd, T. Stabilization of a β-hairpin in monomeric Alzheimer's amyloid-β peptide inhibits amyloid formation. *Proc. Natl. Acad. Sci. U S A* **105,** 5099–5104 (2008).
402. Sievers, S. A. *et al.* Structure-based design of non-natural amino-acid inhibitors of amyloid fibril formation. *Nature* **275**, 96-101, (2011).
403. Chien, P., DePace, A. H., Collins, S. R. & Weissman, J. S. Generation of prion transmission barriers by mutational control of amyloid conformations. *Nature* **424,** 948–951 (2003).
404. Tessier, P. M. & Lindquist, S. Unraveling infectious structures, strain variants and species barriers for the yeast prion [PSI+]. *Nat. Struct. Mol. Biol.* **16,** 598–605 (2009).
405. Sarell, C. J. *et al.* Expanding the repertoire of amyloid polymorphs by co-polymerisation of related protein precursors. *J. Biol. Chem..* **288,** 7327–7337 (2013).
406. Qin, Z., Hu, D., Zhu, M. & Fink, A. L. Structural characterisation of the partially folded intermediates of an immunoglobulin light chain leading to amyloid fibrillation and amorphous aggregation †. *Biochemistry* **46,** 3521–3531 (2007).
407. Bermel, W., Felli, I. C., Kümmerle, R. & Pierattelli, R. 13C Direct-detection biomolecular NMR. *Concepts Magn. Reson.* **32A,** 183–200 (2008).
408. Bermel, W. *et al.* Protonless NMR experiments for sequence-specific assignment of backbone nuclei in unfolded proteins. *J. Am. Chem. Soc.* **128,** 3918–3919 (2006).

409. Hsu, S.-T. D., Bertoncini, C. W. & Dobson, C. M. Use of protonless NMR spectroscopy to alleviate the loss of information resulting from exchange-broadening. *J. Am. Chem. Soc.* **131,** 7222–7223 (2009).
410. Bertini, I., Jiménez, B., Pierattelli, R., Wedd, A. G. & Xiao, Z. Protonless 13C direct detection NMR: Characterisation of the 37 kDa trimeric protein CutA1. *Proteins* **70,** 1196–1205 (2007).
411. Meinhardt, J., Sachse, C., Hortschansky, P., Grigorieff, N. & Fändrich, M. Aβ(1-40) Fibril Polymorphism implies diverse interaction patterns in amyloid fibrils. *J. Mol. Biol.* **386,** 869–877 (2009).
412. Eisenberg, D. & Jucker, M. The amyloid state of proteins in human diseases. *Cell* **148,** 1188–1203 (2012).
413. Platt, G. W. & Radford, S. E. Glimpses of the molecular mechanisms of beta2-microglobulin fibril formation in vitro: aggregation on a complex energy landscape. *FEBS lett.* **583,** 2623–2629 (2009).
414. Prusiner, S. Novel proteinaceous infectious particles cause scrapie. *Science* **216,** 136–144 (1982).
415. Prusiner, S. B. Prions. *Proc. Natl. Acad. Sci. U S A* **95,** 13363–13383 (1998).
416. Chien, P., Weissman, J. S. & DePace, A. H. Emerging principles of conformation-based prion inheritance. *Annu. Rev. Biochem.* **73,** 617–656 (2004).
417. Chien, P. & Weissman, J. S. Conformational diversity in a yeast prion dictates its seeding specificity. *Nature* **410,** 223–227 (2001).
418. Tanaka, M., Chien, P., Yonekura, K. & Weissman, J. S. Mechanism of cross-species prion transmission: an infectious conformation compatible with two highly divergent yeast prion proteins. *Cell* **121,** 49–62 (2005).
419. Chen, B., Newnam, G. P. & Chernoff, Y. O. Prion species barrier between the closely related yeast proteins is detected despite coaggregation. *Proc. Natl. Acad. Sci. U S A* **104,** 2791–2796 (2007).
420. Greenwald, J. *et al.* The Mechanism of Prion Inhibition by HET-S. *Mol. Cell* **38,** 889–899 (2010).
421. Cao, P., Meng, F., Abedini, A. & Raleigh, D. P. The ability of rodent islet amyloid polypeptide to inhibit amyloid formation by human islet amyloid polypeptide has important implications for the mechanism of amyloid formation and the design of inhibitors. *Biochemistry* **49,** 872–881 (2010).
422. Seeliger, J. *et al.* Cross-amyloid interaction of Aβ and IAPP at lipid membranes. *Angew. Chem. Int. Ed. Eng.* **51,** 679–683 (2012).
423. Yan, L.-M. L., Tatarek-Nossol, M. M., Velkova, A. A., Kazantzis, A. A. & Kapurniotu, A. A. Design of a mimic of nonamyloidogenic and bioactive human islet amyloid polypeptide (IAPP) as nanomolar affinity inhibitor of IAPP cytotoxic fibrillogenesis. *Proc. Natl. Acad. Sci. U S A***103,** 2046–2051 (2006).
424. Yan, L.-M., Velkova, A., Tatarek-Nossol, M., Andreetto, E. & Kapurniotu, A. IAPP mimic blocks Abeta cytotoxic self-assembly: cross-suppression of amyloid toxicity of Abeta and IAPP suggests a molecular link between Alzheimer's disease and type II diabetes. *Angew. Chem. Int. Ed. Engl.* **46,** 1246–1252 (2007).
425. Yan, L.-M. *et al.* Selectively N-methylated soluble IAPP mimics as potent IAPP receptor agonists and nanomolar inhibitors of cytotoxic self-assembly of both IAPP and Aβ40. *Angew. Chem. Int. Ed. Engl.* **52,** 10378–10383 (2013).

426. Schwieters, C. D. & Clore, G. M. Reweighted atomic densities to represent ensembles of NMR structures. *J. Biomol. NMR* **23,** 221–225 (2002).
427. Calabrese, M. F. M., Eakin, C. M. C., Wang, J. M. J. & Miranker, A. D. A. A regulatable switch mediates self-association in an immunoglobulin fold. *Nat. Struct. Mol. Biol.* **15,** 965–971 (2008).
428. Blaho, D. V. D. & Miranker, A. D. A. Delineating the conformational elements responsible for Cu(2+)-induced oligomerization of beta-2 microglobulin. *Biochemistry* **48,** 6610–6617 (2009).
429. Calabrese, M. F. & Miranker, A. D. Metal binding sheds light on mechanisms of amyloid assembly. *Prion* **3,** 1–4 (2009).
430. Srikanth, R., Mendoza, V. L., Bridgewater, J. D., Zhang, G. & Vachet, R. W. Copper Binding to beta-2-microglobulin and its pre-amyloid oligomers. *Biochemistry* **48,** 9871–9881 (2009).
431. Monti, M., Amoresano, A., Giorgetti, S., Bellotti, V. & Pucci, P. Limited proteolysis in the investigation of beta2-microglobulin amyloidogenic and fibrillar states. *Biochim. Biophys. Acta* **1753,** 44–50 (2005).
432. Fogolari, F. *et al.* Molecular dynamics simulation suggests possible interaction patterns at early steps of beta2-microglobulin aggregation. *Biophys. J.* **92,** 1673–1681 (2007).
433. Kihara, M. *et al.* Conformation of amyloid fibrils of beta2-microglobulin probed by tryptophan mutagenesis. *J. Biol. Chem..* **281,** 31061–31069 (2006).
434. Hammarström, P., Schneider, F. & Kelly, J. W. Trans-suppression of misfolding in an amyloid disease. *Science* **293,** 2459–2462 (2001).
435. Penchala, S. C. *et al.* AG10 inhibits amyloidogenesis and cellular toxicity of the familial amyloid cardiomyopathy-associated V122I transthyretin. *Proc. Natl. Acad. Sci. U S A* **110,** 9992–9997 (2013).
436. Pepys, M. B. *et al.* Human lysozyme gene mutations cause hereditary systemic amyloidosis. *Nature* **362,** 553–557 (1993).
437. Follis, A. V. *et al.* PUMA binding induces partial unfolding within BCL-xL to disrupt p53 binding and promote apoptosis. *Nat. Chem. Biol.* **9,** 163–168 (2013).
438. Jiang, L. *et al.* Structure-based discovery of fiber-binding compounds that reduce the cytotoxicity of amyloid beta. *Elife* **2,** e00857–e00857 (2013).

APPENDIX I

#Residue	Residue	Nucleus	CS(ppm)
1	I	C	175.249
1	I	CA	61.28
1	I	CB	38.476
2	Q	C	175.465
2	Q	CA	55.913
2	Q	CB	30.482
2	Q	H	8.459
2	Q	N	123.603
3	K	C	175.547
3	K	CA	55.975
3	K	H	9.252
3	K	N	124.213
4	T	CA	59.462
4	T	CB	70.96
4	T	H	8.312
4	T	N	118.995
5	P	C	175.952
5	P	CA	62.555
6	Q	CA	64.105
6	Q	CB	29.276
6	Q	H	8.893
6	Q	N	122.296
7	I	C	174.665
7	I	CA	61.371
7	I	CB	33.335
7	I	H	8.439
7	I	N	124.197
8	Q	C	174.105
8	Q	CA	54.784
8	Q	CB	33.144
8	Q	H	8.985
8	Q	N	126.737
9	V	C	175.096
9	V	CA	60.629
9	V	CB	33.091
9	V	H	9.108
9	V	N	125.325
10	Y	C	173.387
10	Y	CA	56.254
10	Y	CB	41.159
10	Y	H	8.31
10	Y	N	122.197
11	S	C	174.622
11	S	CA	56.613
11	S	CB	64.971
11	S	H	9.45
11	S	N	117.936
12	R	C	175.517
12	R	CB	32.124
12	R	H	8.981
12	R	N	122.417
13	H	C	173.999
13	H	CA	52.258
13	H	CB	30.403
13	H	H	8.153
13	H	N	114.166
15	P	C	175.894
15	P	CA	63.067
16	E	C	175.146
16	E	CA	55.317
16	E	CB	32.768
16	E	H	8.907
16	E	N	124.468
17	N	C	177.494
17	N	CA	54.607
17	N	CB	37.16
17	N	H	8.858
17	N	N	123.616
18	G	C	173.517
18	G	CA	45.493
18	G	H	8.951
18	G	N	109.272
19	K	CA	52.384
19	K	CB	33.037
19	K	H	8.029
19	K	N	122.393
20	P	C	175.714
20	P	CA	63.689
20	P	CB	32.442
21	N	C	173.99
21	N	CA	51.116
21	N	CB	42.561
21	N	H	9.071
21	N	N	124.746
22	I	C	173.642
22	I	CA	60.155
22	I	CB	41.909
22	I	H	10.235
22	I	N	123.642
23	L	C	173.479
23	L	CA	53.006
23	L	H	9.119
23	L	N	130.318
24	N	C	173.379
24	N	CA	51.869
24	N	CB	41.496
24	N	H	8.263
24	N	N	123.155

25	C	C	171.004	38	Q	CA	53.774
25	C	CA	54.003	38	Q	H	9.085
25	C	H	9.456	38	Q	N	124.682
25	C	N	120.743	39	M	C	173.757
26	Y	C	174.342	39	M	CA	55.15
26	Y	CA	55.338	39	M	CB	35.136
26	Y	CB	41.897	39	M	H	9.131
26	Y	H	9.672	39	M	N	120.966
26	Y	N	129.588	40	L	CA	53.498
27	V	C	174.844	40	L	CB	35.182
27	V	CA	61.229	40	L	H	8.952
27	V	CB	31.213	40	L	N	121.46
27	V	H	9	41	K	C	176.813
27	V	N	131.352	41	K	CA	54.003
28	T	C	174.179	41	K	CB	34.568
28	T	CA	59.682	41	K	H	8.839
28	T	H	8.916	41	K	N	121.716
28	T	N	115.354	42	N	C	175.712
29	Q	C	173.344	42	N	CA	54.422
29	Q	CA	57.025	42	N	CB	37.157
29	Q	CB	26.344	42	N	H	9.537
29	Q	H	8.047	42	N	N	128.539
29	Q	N	113.566	43	G	C	173.187
30	F	C	174.488	43	G	CA	45.597
30	F	CA	54.179	43	G	H	8.63
30	F	H	7.296	43	G	N	101.864
30	F	N	109.54	44	K	C	175.768
31	H	CA	57.765	44	K	CA	54.254
31	H	H	8.254	44	K	CB	34.573
31	H	N	118.756	44	K	H	7.847
33	P	C	174.245	44	K	N	121.526
33	P	CA	63.779	45	K	C	176.872
34	H	C	173.558	45	K	CA	58.036
34	H	CA	56.643	45	K	CB	32.644
34	H	CB	30.005	45	K	H	8.665
34	H	H	7.573	45	K	N	124.132
34	H	N	119.031	46	I	CA	59.668
35	I	C	171.342	46	I	H	8.841
35	I	CA	60.452	46	I	N	130.029
35	I	H	8.356	47	P	C	176.992
35	I	N	126.986	47	P	CA	64.605
36	E	C	174.547	47	P	CB	32.734
36	E	CA	55.063	48	K	C	174.559
36	E	CB	32.571	48	K	CA	55.928
36	E	H	7.845	48	K	CB	31.488
36	E	N	124.106	48	K	H	7.736
37	I	C	174.601	48	K	N	117.663
37	I	CA	60.521	49	V	C	175.192
37	I	CB	41.551	49	V	CA	62.06
37	I	H	8.647	49	V	CB	34.515
37	I	N	126.144	49	V	H	7.933
38	Q	C	174.566	49	V	N	124.882

50	E	C	174.645	64	I	H	9.273
50	E	CA	55.217	64	I	N	119.822
50	E	CB	31.446	65	L	C	174.484
50	E	H	8.51	65	L	CA	53.779
50	E	N	127.672	65	L	CB	46.015
51	M	C	176.67	65	L	H	8.228
51	M	CA	53.718	65	L	N	127.062
51	M	CB	35.35	66	A	C	175.882
51	M	H	8.563	66	A	CA	49.558
51	M	N	120.289	66	A	CB	22.381
52	S	C	173.345	66	A	H	9.094
52	S	CA	58.331	66	A	N	131.012
52	S	CB	65.814	67	H	C	172.59
52	S	H	8.95	67	H	CA	54.601
52	S	N	119.795	67	H	CB	32.506
53	D	C	177.224	67	H	H	9.085
53	D	CA	54.932	67	H	N	116.265
53	D	CB	40.488	68	T	C	171.82
53	D	H	8.486	68	T	CA	60.194
53	D	N	118.006	68	T	CB	71.575
54	M	C	175.891	68	T	H	8.43
54	M	CA	56.261	68	T	N	113.408
54	M	H	8.941	69	E	C	175.959
54	M	N	125.317	69	E	CA	57.265
55	S	CA	57.597	69	E	CB	30.183
55	S	H	8.377	69	E	H	8.398
55	S	N	122.262	69	E	N	127.789
56	F	CA	57.653	70	F	C	171.871
56	F	H	8.197	70	F	CA	55.615
56	F	N	119.811	70	F	CB	41.198
57	S	C	174.664	70	F	H	8.74
58	K	C	176.485	70	F	N	123.65
58	K	CA	58.525	71	T	C	172.852
58	K	H	7.452	71	T	CA	58.59
58	K	N	113.545	71	T	CB	69.791
59	D	CA	54.694	71	T	H	8.243
59	D	H	7.754	71	T	N	118.012
59	D	N	116.744	72	P	C	175.001
61	S	C	174.132	72	P	CA	63.312
61	S	CA	58.513	72	P	CB	31.777
61	S	CB	64.249	73	T	C	175.448
62	F	C	174.086	73	T	CA	60.265
62	F	CA	56.375	73	T	CB	73.205
62	F	CB	37.694	73	T	H	8.144
62	F	H	8.408	73	T	N	110.918
62	F	N	121.528	74	E	C	178.102
63	Y	C	173.474	74	E	CA	59.136
63	Y	CA	55.953	74	E	CB	30.431
63	Y	H	8.234	74	E	H	9.19
63	Y	N	112.464	74	E	N	118.087
64	I	C	170.35	75	T	C	174.957
64	I	CA	61.486	75	T	CA	61.94

75	T	CB	69.919	86	S	CA	60.291
75	T	H	7.651	86	S	CB	63.658
75	T	N	103.738	86	S	H	8.009
76	D	C	176.244	86	S	N	112.245
76	D	CA	55.117	87	M	C	175.282
76	D	CB	43.674	87	M	CA	54.672
76	D	H	7.207	87	M	CB	35.765
76	D	N	121.015	87	M	H	8.113
77	T	C	173.2	87	M	N	122.229
77	T	CA	60.465	88	A	C	177.57
77	T	CB	71.236	88	A	CA	54.444
77	T	H	8.745	88	A	CB	18.944
77	T	N	115.442	88	A	H	8.758
78	Y	C	174.85	88	A	N	125.923
78	Y	CA	57.027	89	E	C	173.791
78	Y	CB	44.491	89	E	CA	52.876
78	Y	H	9.415	89	E	CB	31.156
78	Y	N	122.128	89	E	H	7.713
79	A	C	174.383	89	E	N	113.928
79	A	CA	50.567	90	P	C	175.574
79	A	CB	24.406	90	P	CA	63.073
79	A	H	8.744	90	P	CB	31.818
79	A	N	120.704	91	K	C	174.881
80	C	C	171.078	91	K	CA	54.607
80	C	CA	53.045	91	K	CB	35.8
80	C	CB	43.614	91	K	H	8.736
80	C	H	9.079	91	K	N	125.29
80	C	N	119.996	92	T	C	173.405
81	R	C	174.293	92	T	CA	62.52
81	R	CA	54.263	92	T	CB	69.156
81	R	CB	34.014	92	T	H	8.837
81	R	H	9.392	92	T	N	123.642
81	R	N	129.512	93	V	C	175.067
82	V	C	173.476	93	V	CA	61.418
82	V	CA	60.723	93	V	CB	34.365
82	V	H	9.008	93	V	H	8.965
82	V	N	127.754	93	V	N	128.541
83	K	C	174.409	94	Y	C	176.146
83	K	CA	54.778	94	Y	CA	58.925
83	K	CB	35.8	94	Y	CB	39.542
83	K	H	8.728	94	Y	H	8.983
83	K	N	126.46	94	Y	N	126.539
84	H	C	174.74	95	W	C	174.784
84	H	CA	57.641	95	W	CA	57.175
84	H	H	8.454	95	W	CB	29.051
84	H	N	125.321	95	W	H	8.895
85	A	C	177.053	95	W	N	121.783
85	A	CA	54.891	96	D	C	175.11
85	A	CB	18.445	96	D	CA	52.909
85	A	H	8.258	96	D	CB	41.58
85	A	N	127.755	96	D	H	8.604
86	S	C	174.106	96	D	N	131.371

97	R	CA	56.976	98	D	N	120.303
97	R	CB	29.744	99	M	C	180.835
97	R	H	7.52	99	M	CA	57.482
97	R	N	120.844	99	M	CB	34.25
98	D	C	175.253	99	M	H	7.542
98	D	CA	54.656	99	M	N	125.1
98	D	CB	41.203				
98	D	H	8.228				

Table A: Chemical shift assignments of the backbone (^{1}H, ^{13}C, ^{15}N) atoms of mβ_2m. Experimental details and data analysis are described in Section 2.3.10.1.