Solution-state NMR studies of amyloid fibrils

PhD Thesis
Andrew James Baldwin
Trinity College
University of Cambridge

Submitted July 2007

Contents

Declaration

This dissertation is a summary of research carried out in the University Chemical Laboratory, University of Cambridge, between January 2004 and June 2007. The work described in the dissertation is my own and contains nothing which is the outcome of work done in collaboration with others except as specified in the text and acknowledgements. It has not, either in part or as a whole been submitted for a degree, diploma or other qualification at any other university. The length of this dissertation does not exceed the word limit.

Andrew James Baldwin
Cambridge, July 2007.

Acknowledgements

> "For it is esteemed a kind of dishonour unto learning to descend to inquiry or meditation upon matters mechanical, except they be such as may be thought secrets, rarities, and special subtilities, which humour of vain supercilious arrogancy is justly derided in Plato... But the truth is, they be not the highest instances that give the securest information; as may well be expressed in the tale... of the philosopher, that while he gazed upwards to the stars fell into the water; for if he had looked down he might have seen the stars in the water, but looking aloft he could not see the water in the stars. So it cometh often to pass, that mean and small things discover great, better than great can discover the small."

Francis Bacon, The advancement of learning, Book II, London,1605

> "Tact is the knack of making a point without making an enemy."

Isaac Newton

> "What I am going to tell you about is what we teach our physics students in the third or fourth year of graduate school... It is my task to convince you not to turn away because you don't understand it. You see my physics students don't understand it... That is because I don't understand it. Nobody does."

Richard P. Feynman, QED, The Strange Theory of Light and Matter, 1990, p 9

Special thanks to Paul Barker, without whose guidance and support, my PhD study would have been a much bleaker time. Special thanks to Reverend Professor Spencer Anthony-Cahill and John Christodoulou, the latter for introducing me to magnets and the former for showing me that an academic career is a very worthwhile pursuit.

Thanks to Guy Lippens, Xavier Salvatella and Shang-Te Danny Hsu for helping me understand the kinds of spin dynamics required to work out what nuclear spins do after firing some radio waves at them. Thanks to Tuomas Knowles and Chris Waudby for invaluably productive discussions. Thanks to those whose teachings on the artful methods of cultivation of proteins helped me minimise the time I spent in the bug-room; particularly Reto Bader, Spencer and Becky Michael.

Many Dobson group members were kind enough to prepare samples and lend their expertise. The work on αB-crystallin was with Sarah Meehan, TTR work was with Sally Gras, SH3 work was with Neil Birkett, lysozyme work was with Anne Dhulesia, insulin work was with Glynn Devlin and Sarah Shammas and work on Starch was with an old school friend Danielle Egan at King's college, London.

Finally, special thanks, of course, to Professor Christopher Dobson without whose support I would never have had the opportunities and the freedom to explore my own ideas during my PhD. The marvellous thing about the Dobson group is the wealth of expertise it contains; with experts in all aspects of biological studies; from organism work through to biophysics and atomist simulations, answers to almost all things can be obtained in exchange for the cost of a coffee.

Abstract

Amyloid fibrils are highly stable proteinaceous structures associated with many pathological diseases and whose structure is poorly understood. The core structure is based on protofilaments composed of assemblies of proteins, folded into extended beta-sheet structures. By harnessing the remarkable stability of amyloid fibrils and their propensity to self-assemble, there exists great potential to utilize them as the basis for novel bio-materials for technological applications. In this thesis, a protein is designed incorporating an electron transport protein that self-assembles into amyloid fibrils in order to investigate the possibility of producing self-assembling bio-electronic conductors. These fibrils are found to display the electron transport protein and as a consequence, exchange electrons with their surroundings.

The effects of non beta-sheet core regions on the properties of amyloid fibrils have been studied. In particular it is shown that non-core regions have sufficient flexibility to average their local magnetic environments to yield sharp resonances when observed by solution-state NMR. NMR diffusion measurements are essential for reaching this conclusion. The diffusion properties of the non-core regions on the surface of the fibrils reflect that of the underlying fibrils. A theory of how rotational diffusion will contribute to NMR diffusion measurements is derived and is used to estimate the lengths of the fibrils under study. These lengths are in reasonable accord with those observed using imaging techniques.

In a range of different amyloid fibril forming systems, many, but not all were observed to have resonances originating from fibrils. It is shown that in fibril systems where NMR resonances from the non-core regions of fibrils are observed, the tendency for dissociation of the constituent monomers from fibrils is greatly reduced. A systematic study shows that the origin of these effects is non-specific interactions between non-core regions of amyloid fibrils highlighting the importance of these in the process of fibril assembly.

Chapter 1

General Introduction

The questions posed in trying to understand biological systems are among the most fascinating in modern science. The ability of biological systems to perform immensely complex processes, from duplication of genetic material to immunogenic responses and signal transduction, arises through the interplay between biological molecules. Biological molecules are able to exchange information through the non-covalent interactions between juxtaposing, complementary surfaces, a principle elegantly described by Linus Pauling and Max Delbrück [1]. Biological molecules must therefore maintain these juxtaposing surfaces in order to retain their complementary interactions, and functionality.

The most extensively studied class of biological molecules is that of the proteins, originally described by Berzelius in 1839. For example, globular proteins form an essential component of living organisms and perform a wide range of functions, such as enhancing the rate of transcription of DNA, transporting electrons, maintaining structural integrity of component cells and modulating cellular immune responses. From the point of view of physical chemistry, proteins are linear hetero-polymers of a basic set of 20 naturally occurring Lα-amino acids, held together covalently through amide linkages. The amino acid sequences of the proteins found in nature adopt well defined, aperiodic structures under physiological conditions, which represent the frustrated attempt to satisfy the non-covalent interactions of each component amino acid. Owing to thermal motions, the resulting structures can undergo large fluctuations and sample a range of conformations. Some aspects of these fluctuations are collision with solvent molecules, libration of closely packed interior groups, hinge-bending motion between elements of well-defined secondary structure, rolling and sliding of helices and β-sheets, rattling motion of caged atoms within the protein interior, and concerted fluctuations and jumps of neighboring dihedral angles. Such motions may cover a wide range of amplitudes, energies, and time scales [2].

All the information required to describe both the ensemble of conformations proteins adopt, and the pathway that leads a newly synthesised nascent protein chain to fold, is encoded in the primary structure [3, 4]. Armed with a complete understanding of these interactions, the observed structural ensembles could be calculated from knowledge of the primary structure alone. All biological problems, from calculating which drug molecule will shut down a particular cellular process in the treatment of disease, to developing structural materials with the properties of spider's silk could then be solved directly from theory.

The difficulty in calculating these structures lies in the complexity of the non-covalent forces. Other than the electrostatic interactions, the non-covalent forces important for the structures of biological molecules are quantum mechanical phenomena [5, 6]. London dispersion forces, the attractive component of Van der Waals forces, for example, are electron correlation effects [7] requiring time dependent quantum mechanics to be understood. In addition, the hydrogen bond has been experimentally verified to have covalent character, through measurement of ${}^2J_{NN}$ scalar couplings in NMR experiments [8]. A full quantum mechanical treatment of a protein chain will be required to allow us to predict correctly the structures of biomolecules *ab-initio*. Such calculations at the time of writing are far beyond the remit of our current computational power and so semi-classical treatments, parameterised with results from quantum mechanical calculations have been developed [9]. Particularly when further constrained with further experimental data, molecular dynamics simulations provide our most detailed insight into understanding the properties of biopolymers [10].

1.1 Bionanotechnology

Bionanotechnology is the term coined for synthetic technology based on the principles and chemical pathways of living organisms, ranging from genetically-engineered microbes to custom-made organic molecules. It encompasses the study, creation, and illumination of the connections between structural molecular biology and molecular nanotechnology, since the development of useful nano-materials might be guided by studying the structure and function of the natural biopolymers found in living cells.

It is interesting to examine biological molecules in this context. For example, proteins that catalyse chemical changes are particularly adept organic chemists; by selectively stabilising the transition state between the starting reagent and the desired product, through specific atomic contacts, molecules with precisely defined regio- and stereo-chemistry are constructed, under physiological conditions, a feat much of the development of modern synthetic organic chemistry endeavours to emulate.

Clearly, an ability to design molecules with structures capable of performing functions we define would bring about a technological revolution. Although our current understanding of biological systems is not yet up to this task, evolution has supplied us with a vast array of functional molecules. By taking proteins out of their biological context, manipulating them, and employing them in artificial devices, we can create novel and useful bio-electronic devices. One commercially successful example of this approach is in the use of a modified fungal glucose oxidase in modern glucose sensors [11].

1.2 Thesis introduction

The initial project described in this thesis (chapter 2) investigates the feasibility of producing a self-assembling nanoscale conductor. If one could array a path of electron transport proteins between two points, electrons would be able to travel between the two points under an applied potential, with electrons tunnelling between the individual carriers. Arranging matter in bulk, with sub-nm precision is a hugely difficult problem, and is the primary aim of nanotechnology, as originally described by Richard Feynman [12]. One solution to this problem is to array the electron transport proteins onto a scaffold that forms spontaneously, such as a fibril. If the electron transport proteins are held sufficiently closely to interact, the fibril would be able to transport electrons along its length under an applied potential.

Electron transport proteins and the conditions required for their interaction are introduced in section 1.3. By covalently attaching them to a monomer that assembles into a fibril, and then polymerising the monomer, one could obtain a fibril with the required density of electron transport proteins for an interaction. Covalent polymers such as poly-ethylene have little tendency to form defined nano-scale architectures. One suitable set of monomers to polymerise, are protein sequences that non-covalently assemble into ‘amyloid fibrils’, introduced in section 1.4.

The target material is therefore a covalent fusion of an electron transport protein with a protein sequence that readily forms amyloid fibrils. Polymerising this will yield an amyloid fibril displaying the electron transport proteins. If the electron transport proteins on the surface of the fibril are able to interact, then the fibrils will be conducting. The production of this system and its resulting properties are described in chapter 2. The molecular basis of amyloid assembly is, however, poorly understood. The electron transport protein has been shown to perturb the structure of the underlying fibrils in a manner that has allowed a deeper understanding of the generic structure of amyloid fibrils. This is studied in detail in chapter 2 and illustrates a synergy between technological and scientific goals.

1.3 Electron transfer proteins

One particular class of reaction at which evolved proteins are highly adept are the electron transfer reactions, the reactions that are primarily responsible for an organism's ability to harvest energy from its surroundings. Highly energetic reactions, such as the reduction of molecular oxygen by NADH are broken down into stages, with each step governed by different molecules, the combination of which form an electron transport chain[17]. Energy is extracted in discrete increments, at specific, complex catalytic sites embedded within protein molecules, at distinct positions along an electron transport chain. Electrons are transported between these catalytic centres at an almost constant potential. As the net flow of electrons proceeds only between specific redox pairs, biological systems can have many parallel pathways operating in close proximity, that adapt in response to changing environments. This contrasts with conduction in the solid-state, where electrons follow predefined tracks prepared on silicon crystals using lithographic techniques.

Two distinct types of outer-sphere electron transfer (ET) pathways can be identified in biological systems. Intramolecular pathways involve ET between redox pairs that are fixed in space relative to each other, held in place by the surrounding protein structure (figure 1.1 B)[17, 18]. Intermolecular ET involves redox pairs located in separate molecules, and require the prior formation of a protein-protein or protein-ligand complex, that positions the redox pairs in an appropriate conformation for rapid ET [19].

While inter-protein ET has been studied for many decades in the context of small and soluble metalloproteins, our understanding of intra-protein ET pathways has recently been revolutionised by structural insight into multi-cofactor redox enzymes [20]. In the past 10 years, the atomic structures of all of the key respiratory and photosynthetic complexes [21], and many bacterial multi-cofactor redox enzymes [13, 14, 15, 16], have been solved. There are three important, unifying observations.

1. Catalytic redox centres are separated by tens of Ångstroms, and electronically linked by chains of co-factors. The principal co-factors used for this purpose, iron porphyrins and iron-sulphur complexes, have highly delocalised electronic structures that interact to define a specific pathway for electrons to follow, within the complex. The size of these co-factors spans a significant portion of the distance that separates the catalytic centres.

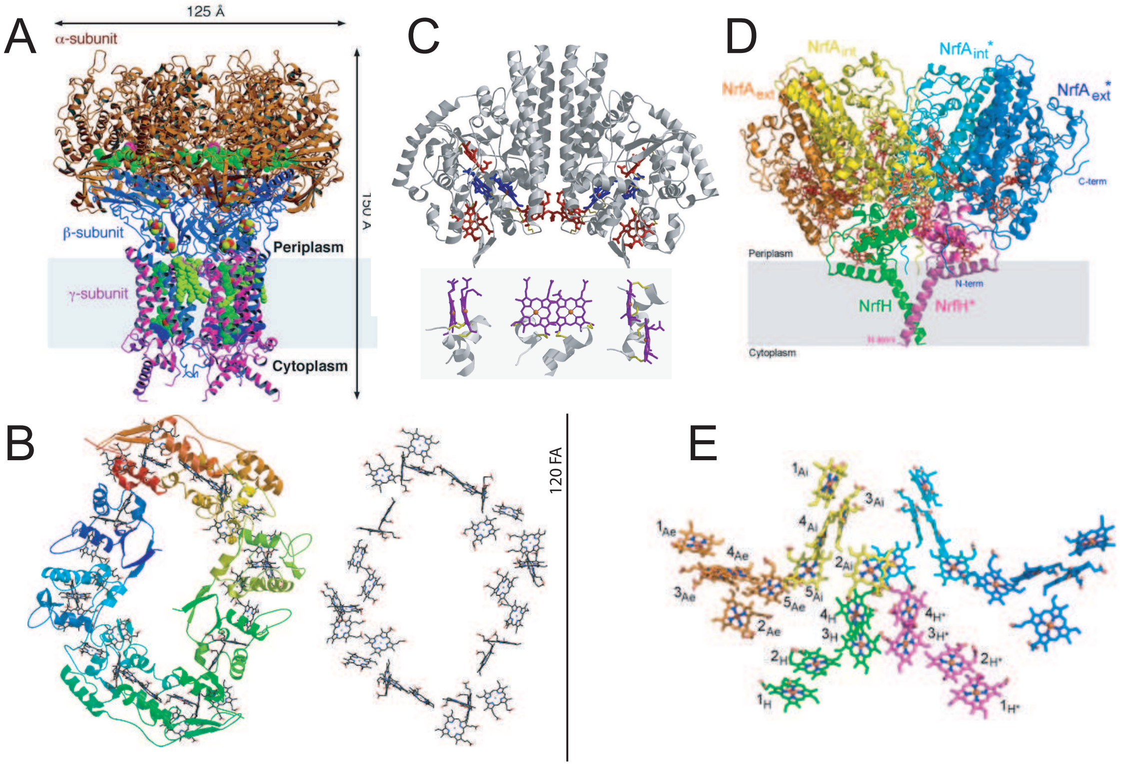


Figure 1.1: The structures of electron transport proteins, shown with their cofactor arrangements, obtained through X-ray diffraction techniques. **A -** The structure of the membrane protein formate dehydrogenase-N (Fdn-N), a major component of *E. coli* nitrate respiration. The structure demonstrates 11 redox centers, including molybdopterin-guanine dinucleotides, five [4Fe-4S] clusters, two haem b groups, and a menaquinone analog. These redox centers are aligned in a single chain, which extends almost 90 Å through the enzyme [13]. **B -** Crystal structure of HMC (high molecular weight cytochrome c) [14] from *D.vulgaris*, a chain of non-covalently linked iron porphyrins, which holds porphyrin molecules no further than 15nm apart, to funnel electrons through the complex. The porphyrin arrangement is also shown. **C -** Cytochrome c nitrite reductase (NrfA) from *E. coli* [15]. The precise, slipped parallel arrangements of the central porphyrins is shown from three orthogonal directions. **D -** Structure of membrane bound NrfH from *D. vulgaris*, complexed with 4 NrfA molecules [16]. Electrons are supplied to the complex from the two cytoplasm reaction centres, and funneled through to the 4 reaction centres on the periplasmic face. **E -** The arrangement of porphyrins in the complex.

2. The relative positions of co-factors are highly conserved throughout all multi-cofactor proteins. These relative geometries reflect orientations that have presumably been optimised for tunnelling electrons through protein complexes. To ensure that electron transfer is not rate limiting, analysis using Marcus' theory [22] has shown that evolution has positioned these cofactors within 14Å of each other [23].

3. The co-factors within a given chain are chemically similar. Consequently, the potential difference between each individual pair of co-factors is small, and so electronic energy is conserved during the transport of electrons.

The overall driving force for ET through an electron transport chain, is provided by the large potential difference of the reaction that the chain catalyses, such as that between molecular oxygen and NADH. The transfer of electrons by inter-protein ET and intra-protein ET occurs at an approximately constant electro-potential, with larger jumps in occurring at the catalytic centres. These centres irreversibly transduce electronic energy to other forms of chemical energy, such as ion gradients, and so the less energy lost while moving electrons between catalytic centres, the more can be transduced into useful, chemical energy. This principle is common to respiratory and photosynthetic systems in all organisms. A good example of a self-contained system of two catalytic sites linked by intra molecular co-factors, is the bacterial formate dehydrogenase shown in figure 1.1A.

The combination of large, fixed cofactor arrays, found in the energy transducing enzyme complexes, with small dynamic carrier proteins and molecules that link these large complexes together, results in a flexible set of components for generating dynamic electronic circuits, quite unlike anything that can be made in solid-state materials. For example, bacteria can rapidly adapt to changing sources of, and sinks for, electrons.

Although we do not currently know exactly how all the components interact in bacteria, we imagine that these multi-haem proteins assemble into multi-protein complexes. The first glimpse of such a system is the NrfA/NrfH complex that assembles 28 c-type haem molecules into a multi-catalytic unit [16] (figure 1.1C, D and E) that delivers electrons fed from the membrane into one of the four catalytic sites in this assembly. Since these systems have evolved to operate in a self-contained, cellular compartment, it is difficult, if not impossible to connect these types of system to external electronic circuits. Hence if the chemistry of these components is to be probed and used in electronic devices, we have to design and build unnatural counterparts of these systems that can be easily connected to solid-state devices.

The evolved components of biological electron transport chains can therefore provide inspiration for assembling metal co-factors into unnatural chains that are tailored for control by external solid state circuits, connected via nanoscale connections, rather than by the chemistry of catalytic centres at either end. The major difficulty in producing such chains is ensuring that the co-factors are arranged for optimal inter-protein ET. By fusing electron transport molecules able to rapidly exchange electrons between themselves to a scaffold such that the molecules are held in their inter-protein ET conformation, intra-molecular ET rates could be obtained. The relatively well ordered and predictable structure of amyloid fibrils lends itself to organisation of porphyrin binding proteins and hence to mimicking the chains of haem molecules observed in natural electronic wires.

1.4 Amyloid fibrils

1.4.1 Protein mis-folding and disease

As with all complex systems, all does not always run smoothly. Thermal fluctuations can lead to protein molecules populating metastable, non-native, 'misfolded' conformations, which if the cell's error-checking machinery does not interfere, can lead the protein molecules concerned to aggregate. Where these aggregates accumulate within an organism, they can disrupt normal biological function, either through inherent bulk or toxicity, loss of function, or as a correlate of an up-stream toxic process. These processes can lead to diseases, many of which are highly debilitating and often fatal.

The term *amyloid*, from the Greek 'amylon', which due to an historical confusion translates as 'starch', is used to describe plaques of aggregated proteinaceous material in fibrillar morphologies. Deposition of these plaques in tissues is associated with a collection of diseases, including the pathogenic Alzheimer's, Parkinson's, Huntington's diseases, Creutzfeldt-Jakob Disease (CJD) and BSE [24]. The various distinct, often loosely termed amyloidoses, are distinguished by the location within the organism where the plaques accumulate, and the protein sequences that dominate the composition of the plaque.

Aggregates *in vivo* are far from unstructured. Individual proteins within these aggregates often assemble into structures based on elongated β-sheets, which in turn assemble into fibres of diameter on the order ca. 10nm, and lengths on the order of microns, which due to their nano-scale diameter are usually termed *fibrils*. Amyloid fibrils induce the mis-folding and aggregation of surrounding natively folded protein, providing the protein sequences are similar [25, 26]. If this were not the case, the occurrence of amyloid plaques would lead to almost immediate

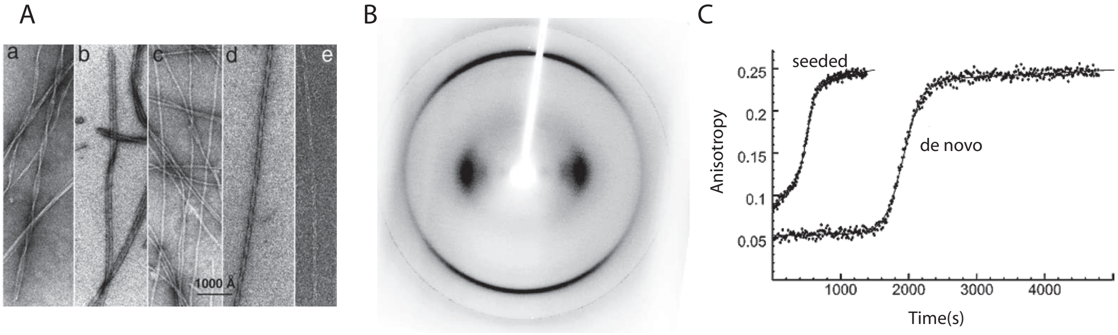


Figure 1.2: General markers of amyloid fibrils. **A -** SH3 Fibrils observed through negatively stained TEM [35]. **B -** Cross-β fibre XRD pattern of SH3 fibrils, courtesy of Dr Reto Bader. **C -** The sigmoidal accumulation of amyloid and the effect of seeding, data from [36].

termination of all biological function as all protein in an organism co-aggregated into a single plaque.

Many unrelated protein sequences form these structures in disease states. More recently, many sequences unrelated to disease have been found to aggregate *in vitro* and form fibrils with all the structural characteristics of amyloid fibrils [27, 28]. This has led to the "generic hypothesis", which states that amyloid fibrils are highly stable structural states for all polypeptide chains [29], due to the generic finding of polypeptide sequences to form stable β-sheet conformations. These states are generally separated from globular, natively folded states by large, condition-dependent kinetic barriers.

Though formed from many different sequences, amyloid fibrils share many structural and kinetic characteristics, discussed in sections 1.4.2 and 1.4.3. Understanding all steps in the assembly process, as well as understanding the determinants of the structural stability is crucial both to our understanding of how to prevent these structures from forming *in vivo*, and in our exploitation of their unusual stability for use as nanoscale materials.

1.4.2 Structural characteristics of amyloid fibrils

Despite extensive efforts, amyloid fibrils have yet to be crystallised, and they tumble too slowly for their core packing to be studied by conventional solution-state NMR structure determination methods. As such, high resolution information on how proteins pack to form an amyloid fibril has yet to be obtained, and the interactions that lead to their stability are unclear. In general, medium resolution biophysical techniques are employed to classify whether a fibrous species has an amyloid structure, summarised in table 1.1. By combining the information each technique brings, structural models have been developed for a number of systems.

Property	Detail
Distinct morphology	When visualised in the atomic force microscope (AFM) or transmission electron microscope (TEM), discrete, long, unbranched fibrils of diameter ca. 10nm, extending up to several microns are observed. Example images are shown in figure 1.2A.
Resistance to degradation and proteolysis	Individual peptide linkages between amino acids are proteolytically inaccessible and are hence resistant to natural proteases. This property allows amyloid to accumulate and persist in tissues *in vivo*, and has been exploited in the purification of amyloid fibrils *in vitro* [30].
β sheet conformation	Circular dichroism (CD) and Fourier transform infra-red (FTIR) spectroscopies show that the monomers within the fibril adopt a β-sheet conformation.
Cross-β fibre X-ray diffraction (XRD) pattern	In non-crystalline samples such as amyloid fibrils, common reflections will be observed where structural features give rise to a common periodicity. Partially aligning the fibrils leads to anisotropy in the diffraction pattern allowing meridional (perpendicular to the fibril) and equatorial (parallel to the fibril) structural features to be distinguished. Amyloid fibrils generically show a 'cross-β' pattern, consisting of two strong reflections; a perpendicular (meridional) reflection around 4.7Å and a parallel (equatorial) reflection around 10Å. An example of a cross-β pattern from SH3 fibrils is shown in figure 1.2B.
Binding to histological dyes	Certain histological dyes such as Thioflavin-T (ThT), Congo Red and ANS yield distinct spectroscopic signatures when added to aggregation mixtures[31, 32]. Despite their exact mode of binding being uncertain [33, 34], their signatures are traditionally associated with the presence of amyloid fibrils.

Table 1.1: Common structural features of amyloid fibrils.

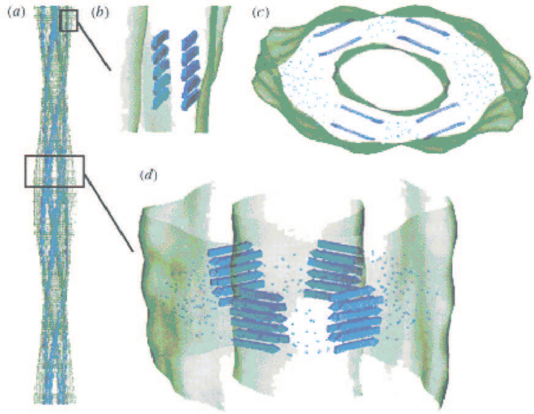


Figure 1.3: Proposed model for the packing of bovine PI3-SH3 into fibrils based, on electron density data obtained by CryoEM [35].

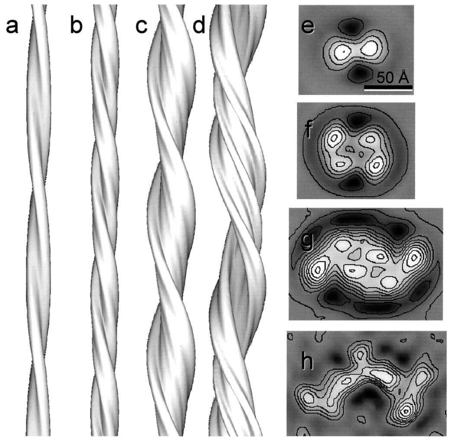


Figure 1.4: CryoEM data from insulin fibrils shows several species with differing higher order protofilament packing, all of which have reduced electron density in the core, adapted from [37].

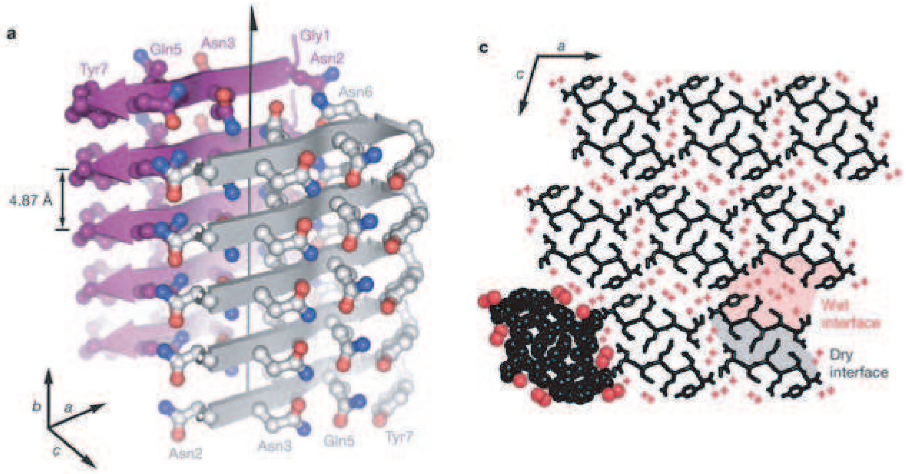


Figure 1.5: Crystal structure of model peptide GNNQQNY from X-ray diffraction data. The structure show the peptides forming extended paired β-sheets. Under alternative conditions, the model peptide forms amyloid fibrils, adapted from [38].

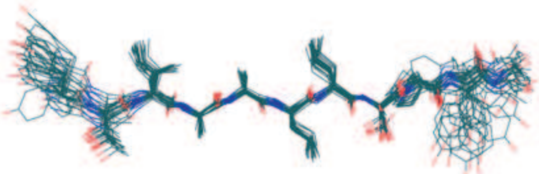


Figure 1.6: The structure of a single $\mathrm{TTR}_{(105-115)}$ peptide molecule in an amyloid fibril, determined by solid-state NMR, adapted from [39].

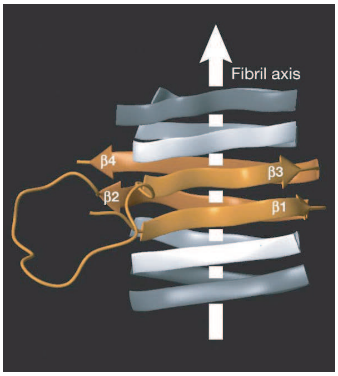


Figure 1.7: The structure of a fragment of a yeast prion protein, HET-$\mathrm{s}_{(218-289)}$, refined using solid-state NMR measurements. Ambiguity exists as to how the four β-sheets pack themselves into the fibril [40].

Although amyloid fibrils can be formed from many difference sequences, they share many structural features, based on a structural motif known as the *protofilament*. The backbone of individual protein molecules adopt a β-sheet conformation, and generate extended β-sheets with the hydrogen bonding axis parallel to the fibril axis. This is shown schematically with an anti-parallel arrangement in figure 1.3B. This model is consistent with the secondary structure observations (FTIR,CD) and the 4.7Å XRD reflection is consistent with the mean spacing between two hydrogen bonded β-sheets. These sheets pair up, giving rise to the 10Å XRD reflection perpendicular to the fibril axis. This twinned β-sheet structure termed the protofilament, is accepted to be the smallest unit of fibril assembly, despite being unstable as an individual unit. Protofilaments further assemble to form the amyloid fibril. While solid-state NMR and diffraction studies on microcrystals have yielded structural information, and identified factors that stabilise the protofilament, very little is known about the factors that stabilise protofilament packing.

Solid-state NMR spectroscopy has been used to measure distance and torsion angle restraints that can give the structure of a single monomer, within a fibril. These have shown which residues are involved in β-sheet core for polypeptides such as TTR$_{(105-115)}$[39] (figure 1.6), HET-s$_{(218-289)}$[40] (figure 1.7), A$\beta_{(11-25)}$[41] and WW2[42]. The main limitation on solid state NMR comes from its low sensitivity; increasing the size of the system increases the overlap of the observed resonances which hinders structure determination. In addition, the measurement of inter-protein interactions has proven difficult.

Short amyloidogenic peptides have been produced that crystallise in conformations that resemble protofilaments, as shown by electron diffraction [43] and X-ray diffraction [38, 44] (figure 1.5). In some cases, the peptides are found to crystallise in paired β-sheet structures, with interactions between interdigitated side chains providing the energy required to hold β-sheets together. The peptide structures are obtained from macroscopic crystals, not directly from amyloid fibrils, and so offer no information on the interactions that might exist between protofilaments.

A further complexity arises when one considers aggregation of longer protein sequences, in that not all of the sequence assembles into the fibril core. Limited proteolysis experiments [45, 46, 47, 48], hydrogen-exchange methods [49] and solution-state NMR [50, 51, 52] have been used to distinguish regions of fibril core from flexible, relatively unstructured regions, on the surface of the fibril. The significance of the flexible regions on the physio-chemical properties of the fibril, is the subject of chapters 4 and 5 of this thesis.

Cryo-EM measurements have been used to provide medium resolution (ca.

25Å) structures of amyloid fibrils. Using image reconstruction techniques, electron density maps can be inferred from images obtains from carefully prepared vitrified samples. One such map for fibrils of SH3, shown in figure 1.3[35], describes four protofilaments associating and twisting around a hollow core. Cryo-EM structures of insulin fibrils [37](figure 1.4) shows fibrils can have some promiscuity in how they pack, with fibrils formed from 2,4 6 and 8 protofilaments being distinguished. In each structure, the protofilaments are twisted around a core of reduced electron density.

Further data describing the variability of protofilament packing comes from AFM measurements. The persistence of correlations between the helical repeats along fibrils, show that SH3, insulin and $\mathrm{TTR}_{(105-115)}$ are highly ordered and that there is greater variability between individual fibrils than within a single fibril suggesting a degree of structural degeneracy is permitted [53] in protofilament packing. The molecular basis for this is poorly understood. These measurements allow an estimate of 310 k_BT per μm of fibril for the energy associated with the lateral association of protofilaments, two orders of magnitude higher than that estimated for sickle haemoglobin fibres [53]. In is interesting to note however, that fibrils are generally found with right-handed twists, consistent with the observation that right-hand twisted β-sheet structures are more stable than those that are twisted in a left-handed fashion [54].

Amyloid fibrils formed from different protein sequences therefore, have similar underlying structures, based on the strong hydrogen bonding interactions between β-sheets within a protofilament. While solid-state NMR and diffraction studies on microcrystals have yielded structural information and have identified factors that stabilise protofilaments, very little is known about the factors that stabilise the packing of the protofilaments. In particular, the molecular basis of why protein sequences require large sequence specificity in order to co-aggregate, and the interactions that stabilise inter protofilament interactions are currently unknown. Chapters 2 and 6 investigate the effects on the properties of the fibril when the packing is perturbed by additional domains on the surface of the fibrils.

Understanding the structure of amyloid fibrils will assist both in the development of therapeutic strategies to combat amyloidoses, and in the exploitation of the desirable physical characteristics of amyloid fibrils in nano-materials.

1.4.3 Mechanistic features of amyloid fibril formation

The mechanism of amyloid fibril formation has similarities to a crystal growth, and has the features of nucleation-dependent polymerisation models developed for the analysis of actin polymerisation [56, 57], and has two distinct phases (figure 1.8).

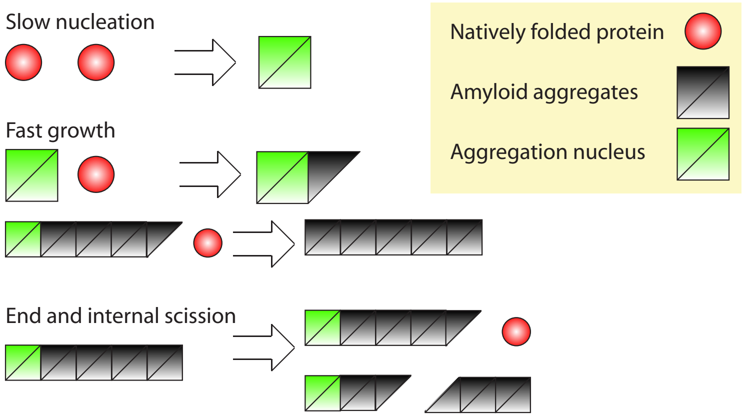


Figure 1.8: Schematic figure illustrating proposed kinetic steps in an aggregation process, adapted from [55].

Nucleation phase In this phase, monomers coalesce into oligomers. The oligomers, once above a threshold size in an appropriate conformation, are capable of inducing fibril elongation. This critical oligomer is usually known as the fibril 'nucleus'. The formation of the fibril nucleus is a stochastic process, depending on the formation of a low probability oligomeric nucleus, and so the time required for this process shows a large degree of variability. Recent work has shown that pre-fibrillar oligomers are likely to be the toxic species in several amyloid diseases [58].

Condensation phase Fibril elongation is usually a significantly faster process, whose rate is proportional to the number of fibril ends. The aggregation rate is therefore further enhanced by breakage processes that further increase the number of fibril ends.

Aggregation will proceed in this way until a steady state is reached, dictated by the relative, system dependent values of the various rate constants [55]. By varying the rate constants of the various processes, a wide range of kinetic behaviour can be observed, from systems that aggregate rapidly but that tend to a state with a high proportion of protein still in monomeric form, to systems that have lag phases of years but effectively aggregate until no monomeric species remain. Remarkably, initiating a fibril forming reaction in the presence of preformed fibrils circumvents the nucleation phase, leading to fibril forming reactions that start directly with elongation. This process is known as 'seeding' [57].

As proteins can form aggregates with different morphologies, as well as persist as oligomers, solutions containing aggregating proteins are highly heterogeneous and so distinguishing experimentally between these species during *in vitro* studies is challenging. Spectroscopic methods tend to average over all protein molecules in the sample and so deconvolving structural information is difficult.

Light scattering, for example, will report on particle size not particle geometry, intrinsic fluorescence emission spectroscopy will report on the local electronic environment of aromatic amino acid chains, and CD spectroscopy will report on the secondary structure averaged over all peptide bonds in the sample. Spectroscopic changes when certain histological dyes, such as thioflavin-T (ThT) [31], Congo red [32] and 1-anilinonaphthalene-8-sulfonate (ANS), are added to aggregation reactions can be measured. However, the mechanism behind these spectroscopic changes, and the species the dyes are binding is not clear [33, 34, 59]. Broadly, three phases are observed when aggregation reactions are monitored, indicated in figure 1.2C:

Lag phase is thought to be associated with the formation and rearrangement of oligomers, that are required for elongation.

Growth phase associated with the elongation and scission of the fibrils.

Saturation phase associated with the depletion of free monomeric protein and the end of the aggregation reaction.

Empirically, generating reproducible data on aggregation rates is technically challenging [27, 60, 61]. This, combined with the difficult of relating spectroscopic measurements to a unique species, has made the determination of the mechanism of amyloid fibril formation exceedingly difficult.

Different polypeptide chains differ greatly in their rates of aggregation, the duration of their lag phases, the elongation rates and the relative proportions of aggregated to non-aggregated protein in the steady state. Thus the individual processes in fibril formation have their own inherent system dependencies. By comparing aggregation rates between mutants of the same protein/peptide system, for a variety of systems, it has been possible to fit the kinetic data to a linear free energy relationship. This allows predictions to be made as to which regions within a protein are more susceptible to aggregation, and as to where mutations will be most likely to cause a given system to aggregate more rapidly [62, 63]. This model has been tested *in vivo*. When *Drosophila* expressing mutants of Aβ were produced, it was found the predicted aggregation rates correlated inversely with the life span (Luheshi *et al.*, unpublished data).

It is interesting to note, and indeed fortunate, that a structure formed from one type of protein will not seed amyloid formation in another protein species [25, 26]. Sequences must be very similar in order to co-aggregate, and, for example, though they both form amyloid fibrils on their own, D form Aβ will not seed L form Aβ [64]. The physical basis of this is not well understood, but is likely to be allied to phase separation and crystallisation.

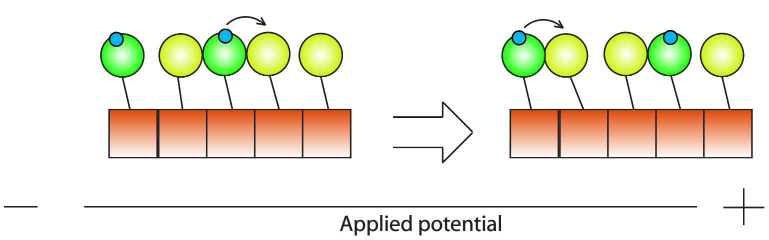


Figure 1.9: An illustration of the benefits of bringing interacting domains together on the surface of an amyloid fibril. The green circles represent reduced functional protein, and yellow an oxidised functional protein. Both are connected to the fibril by a linker. Where the interaction conformation can be found, an electron can be made to hop onto the adjacent group. An applied potential will bias this random exchange, leading to a flow of electrons.

Proteinaceous infections particles, or prions, are a type of infectious agent composed only of protein molecules that induce other protein to form amyloid [65]. They are associated with a number of diseases in a variety of animals including CJD and BSE, and are highly conserved between species. Although their function is not known, it is has been associated with a range of phenomena including organismal memory [66, 67]. Because of high sequence similarity between species, prion diseases have sometimes been able to cross the species barrier. By feeding to younger cattle the meat from older members of the same species, meat relatively high in amyloid deposits, the chance of younger members of the species inheriting the amyloid disease, and the subsequent human consumption of the infected meat increases. Where this happens, humans have been found to develop the pathological human variant Creutzfeld-Jacob (hvCJD) disease [68]. Similarly, cannibalism is associated with the spread of the fatal kuru disease throughout the South Fore people of Papua New Guinea [69].

By combining *in vitro* mechanistic studies with *in vivo* models[70], the mechanisms of amyloid formation are being elucidated. A detailed understanding of the factors that affect protein aggregation is required to develop strategies that will help prevent and alleviate the symptoms of mis-folding diseases.

1.5 Amyloid fibrils as nano-materials

Interestingly, the same properties that make amyloid fibrils potentially hazardous, medically speaking, make them potentially useful materials for a technological application. Primarily, they self-assemble into well defined fibrillar architectures under well defined conditions. In addition, unlike most biological molecules, amyloid fibrils stable structures, resistant to a wide range of temperature, pressure and denaturant conditions. In accordance with such a conclusion, a variety of phenomena in nature, including functional coatings of micro-organisms [76]

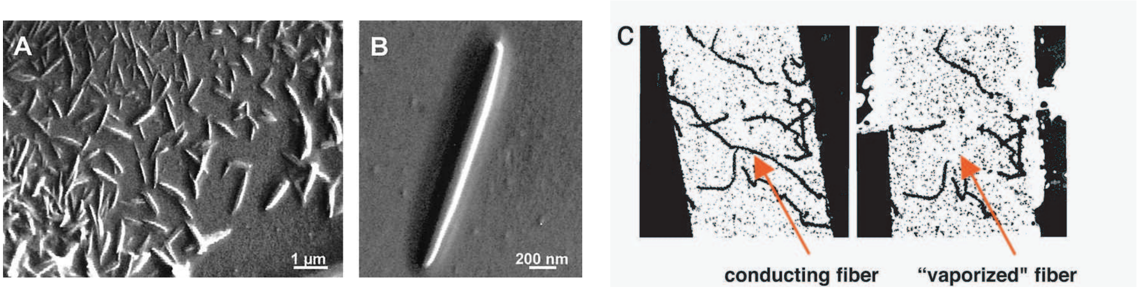


Figure 1.10: A and B show low resolution SEM images of silver wires cast inside Phe-Phe fibrils [71]. C shows SEM images of gold coated amyloid fibils between two electrodes [72].

Modification	Detail
Fluorophore display	Two 10 residue forming peptides TTR(10-19) and TTR(105-115) were covalently fused to fluorophores. Fibrils were then formed and found to display the flurophores [73]. In addition, green fluorescent protein (GFP) was fused to the monomeric yeast prion protein sup35 and then used as an optical marker to track protein through the cell, using confocal microscopy. Under conditions of cellular stress, Sup35 forms aggregates, leading to a change in the phenotype of the cell. Labelling techniques showed that once formed, the aggregates were highly localised within the cells [74].
Protein display	Carbonic anhydrase (CA), barnase, GFP and glutathione S-transferase (GST) were fused to the fibril forming yeast prion Ure2 protein. Reduced, but non-zero, protein activity was found in the resultant fibrils. [75].
Templates to produce metal nanowires	Amyloid fibrils were produced with cysteine mutants of sup35. These were bound to gold nanoparticles. Further metal coats were applied, interspersed with heat treatment to give gold nanowires of average diameter 50nm, templated around the original amyloid fibrils [72], shown in figure 1.10C. In a second study, fibrils from a dipeptide Phe-Phe were prepared in the presence of aqueous silver ions [71]. Subsequent reduction of the metal and proteolysis of the dipeptide yielded discrete silver wires of 100s of nm in length and 20nm diameter as measured by TEM, cast from within the Phe-Phe fibrils, shown in figure 1.10A and B.

Table 1.2: Uses of amyloid fibrils as potential nanomaterials.

and the *in vivo* catalysis of polymerisation reactions [77] have recently been associated with amyloid fibrils.

By covalently fusing a functional group of interest to an amyloid forming protein, it is possible to produce amyloid fibrils displaying the group of interest (figure 1.9). In principle, any functional group can be arrayed on a fibril surface providing the group does not prevent the amyloid protein from forming fibrils. As the two domains are linked covalently, the concentration of fibril forming and functional units within the fibril will be identical. The effective concentration of functional units on the surface of the fibril will greatly exceed that while free in solution. The stability, propensity to self-assemble and the fibrillar geometry of amyloid fibrils can be therefore harnessed in a nano-material and groups have been displayed on amyloid fibrils using this method. Examples are summarised in table 1.2 and figure 1.10.

There are several materials one could envisage to be useful, using amyloid fibrils as a scaffold. For example, by displaying a catalytic protein, a 'homogeneous' catalyst can be produced that has the advantages of a heterogeneous catalyst in that it can be recovered from the reaction products by centrifugation. Where the displayed group is capable of pair-wise interactions in solution, on the surface of the fibril they would form a network of interactions, as the effective concentration of the group is greatly increased. Thus the network would be limited by the interaction rate, and not the collision rate of an equivalent system, in the absence of the fibril.

Possibly the simplest interaction network to detect would be a conducting network. Where the target protein is able to exchange electrons with itself in solution, on the surface of a fibril, it might be able to transmit charge along its length under an applied potential, illustrated in figure 1.9, provided that neighbouring protein has sufficient conformational freedom to reach its interaction conformation, despite the mobility constraints imposed by the fibril. In the absence of an applied potential, electrons would be expected to diffuse throughout the network. Under an applied potential, biased random diffusion of electrons should occur, leading to net flow of electrons, and hence a current.

An additional desirable property of using electron transport proteins as a target functional group is that the currents could be mediated by external molecules. By allosteric inhibition of electron transport, single molecule binding events would be sufficient to shut off the conductance of the fibril. Through monitoring the conductance and its variations, the systems would function, in principle, as highly sensitive sensors.

By localising fibril to two points on a surface, perhaps through selective placement of fibril seed, it is conceivable that amyloid fibrils could be grown between two defined sites. Given a method which discourages the growth of fibrils without two ends fixed on the surface, conducting fibrils could be used to 'grow' circuitry. Contacts labeled with the seed for a fibril of a given sequence would connect. By using several fibril forming systems, circuitry could be spontaneously assembled in this manner, in three dimensions. A consequence of a system being formed in this way is that the system would be automatically self-repairing as well as self-assembling.

1.6 Thesis overview

This thesis focuses on two mutually compatible, but philosophically distinct goals. By preparing and studying amyloid fibrils displaying electron transport proteins the possibility of exploiting such fibrils as a useful material is explored. In this case, the target material is a self-assembling conductor. Fibrils shown to act as electron reservoirs are described in chapter 2.

Perturbing the amyloid forming domain through the incorporation of extra, non-fibril forming sequence perturbs the protofilament packing of the resulting amyloid fibrils. By systematically observing the effects of perturbing the core fibril, structural deductions are made that are interpreted in terms of the fundamental packing of amyloid fibrils. These systems and effects are described in chapter 2.

In the course of the analysis of these fibrils, it was discovered that flexible regions on the surface of the fibrils are sufficiently mobile to yield solution-state NMR resonances, despite the molecular weight of a 2μm fibril being on the order ca. GDa. This discovery led to the work discussed in chapter 3, where NMR diffusion experiments are employed to measure the length distributions of fibrils in solution. NMR diffusion measurements measure displacement of nuclear spins which Stejskal and Tanner related to translational diffusion in 1965. Chapter 3 extends this theory to account for rotational diffusion.

The curious observation that dynamic motion can narrow NMR resonances on the surface of large structures such as the ribosome and amyloid fibrils is discussed further in chapter 6. Simulations are used to estimated the line broadening one might expected from situations of this type.

Chapter 4 uses the theory developed in chapter 3 to study amyloid fibrils formed from a range of different sequences. Two classes of amyloid fibrils are distinguished. One class is in equilibrium with its constituent monomers and the majority of the sequence of each constituent monomer is incorporated in the fibril core. In contrast, the constituent monomers from fibrils in the second class have a significantly reduced tendency to dissociate and significant portions of their sequence are sufficient flexible to be observed by solution-state NMR. This difference is suggested to be a consequence of interactions between the the non-core residues outside the fibril core. This model is systematically tested in chapter 5, where it is demonstrated that the addition of a small number of residues to a fibril forming sequence can destabilise the elongation transition state by 8.6 ± 1 kJ mol^{-1}, with no change in the elongation free energy.

Chapter 2

Engineered amyloid fibrils

Abstract

By covalently fusing an amyloid fibril forming protein to an electron transport protein, amyloid fibrils displaying electron transport proteins have been engineered in order to explore the potential utility of amyloid fibrils as nano-materials. The transfer of electrons into and out of these fibrils is demonstrated. By investigating the perturbations to the fibril's structure caused by the additional group, important insights into the higher order assembly of protofilaments can be inferred. Remarkably, when unfolded, the displayed proteins have sufficient conformational flexibility to be observed by solution-state NMR. From these measurements it is possible to determine the length distribution of these fibrils. With molecular weights of approximately 1 GDa in the case of a 5 μm fibril, these systems are among the largest systems to be investigated by solution state NMR. 50% of the displayed protein molecules are not solvent exposed. Combined with observations of amyloid fibrils having hollow cores from cryo-EM and STEM measurements, it is deduced that protofilament packing will naturally lead to populations of displayed groups on the inside and outside faces of the fibril in the ratio 1:2.

2.1 Design Strategy

In this study, the target material to illustrate the potential utility of amyloid fibrils as a nano-material is a self-assembling conductor, formed from amyloid fibrils displaying electron transport proteins as described in section 1.5. Providing an oxidised electron transport protein is able to reach an appropriate interaction conformation through which transfer of electrons to a reduced protein can be achieved, amyloid fibrils displaying such a protein will be conducting as illustrated in figure 1.9.

The design of this material is focused on covalently fusing two different proteins into a larger fusion protein:

- An electron transport protein (the functional unit).
- An amyloid fibril forming protein (the fibril forming unit).

Under conditions that cause the fibril forming unit to assemble into fibrils, the fibrils will display the functional unit. This would ideally give rise to a material with the stability of amyloid fibrils and the functionality of the group chosen to be displayed. The starting fusion protein can be conveniently synthesised as a single polypeptide, using recombinant techniques. Constraints exist on the choice of fibril forming unit, the functional unit, and the linking group.

- The amyloid forming domain must rapidly aggregate into fibrils.
- Under the same conditions, the electron transport protein must not aggregate.
- The electron transport protein must not prevent the fibril forming unit from assembling into fibrils, illustrated in figure 2.1
- There must be sufficient space for the electron transport protein to fold and reach the required conformation relative to its neighbours for electron transport, illustrated in figure 1.9.

It is important that the functional group does not prevent the fibril formation due to steric hinderance of the aggregation pathway. For example, where the fibril forming domain is significantly smaller than the functional domain, it is likely that the interaction between the two domains will prevent fibril formation. This is illustrated in one dimension in figure 2.1. Unpublished work by Barker *et al.* investigated fusions of a 13kDa protein to a 1.5kDa 10 amino acid domain $\mathrm{TTR}_{(105-115)}$. The construct was found to form non-specific aggregates but not fibrils. To avoid such a clash in three dimensions; the area occupied by a displayed protein on the surface of the fibril must be less than the surface area contributed to the surface of the fibril by a single amyloid forming unit. This geometric constraint is explored more thoroughly in appendix 2.8.3.

2.1.1 Fibril forming unit - SH3 dimer $(SH3)_2$

The aggregation properties of an 85 residue SH3 domain from a regulatory subunit of phosphatidylinositol 3-kinase (referred to herein as SH3) have been well

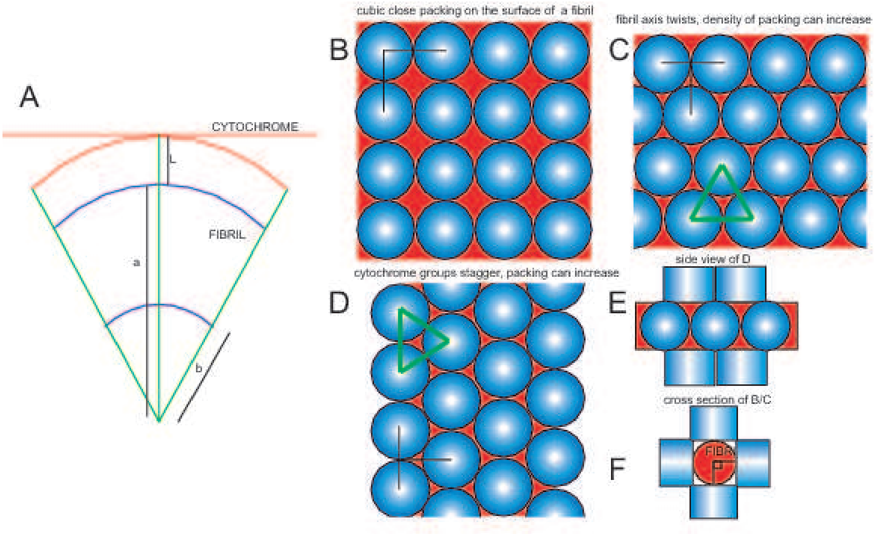


Figure 2.1: A 1D illustration of the size matching principle of displaying groups on the surface of assembling scaffolds. The size of the target functional unit must not be sufficiently large as to prevent the self assembling scaffold from forming. The geometric constraints are further considered in appendix 2.8.3.

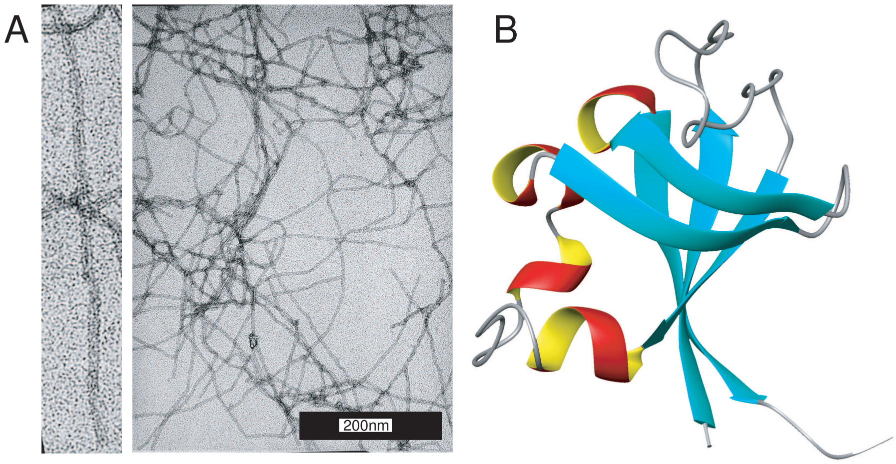


Figure 2.2: **A -** TEM images of negatively stained SH3 dimer fibrils. **B -** NMR structure of bovine PI3-SH3 domain [78] PDB 1PNJ

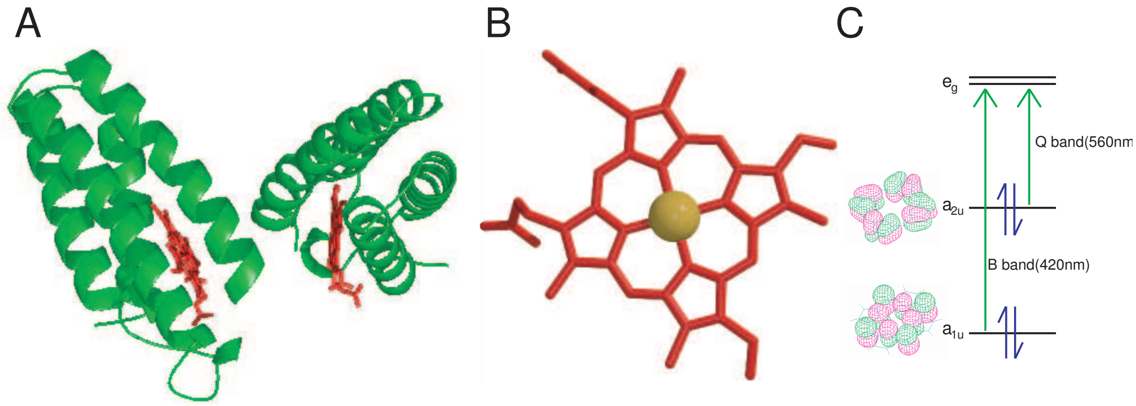


Figure 2.3: **A -** Crystal structure of cytochrome b_{562} showing dimer. **B -** Representation of metalloporphyrin. C- The four orbitals responsible for the distinctive absorption spectrum. The molecular orbitals were calculated with Hyperchem.

studied [28, 60]. Although not associated with any known disease, SH3 forms discrete fibrils at low pH over a period of several weeks. This was a crucial observation that has led to the "generic hypothesis" of amyloid formation [79, 29]. TEM micrographs and a fibre X-ray diffraction pattern of SH3 fibrils are shown in figure 1.2 and a cryoEM reconstruction of the electron density of SH3 fibrils is shown in figure 1.3. The NMR structure of the domain at pH 7 is shown in figure 2.2.

An engineered tandem repeat of this protein forms fibrils at low pH in less than 6 hours after a pH cycling protocol, described in section 8.1.2. TEM images of these fibrils are shown in figure 2.2. The mechanism of rate enhancement has been thoroughly studied [80]: lowering the pH to 2 unfolds the protein. Raising the pH to 3.6 will then protonate groups such as carboxylic acids (glutamic acid and aspartic acid) bringing the overall charge closer to neutral. Under these conditions, the protein is observed to form smaller, less structured aggregates by TEM and analytical ultracentrifugation. These pre-aggregates efficiently seed fibril formation, and further seeding with preformed fibrils does not lead to further enhancements in the aggregation rate.

The conformation adopted by the tandem repeat when prepared under these conditions therefore contains all the information required to nucleate fibril growth. The rapid, controllable fibril formation of this system, combined with its relatively large size (182 residues) makes it an excellent candidate for the fibril forming unit in the fusion protein.

2.1.2 Functional unit - cytochrome b_{562}

Cytochrome b_{562} is a 4-helix bundle *E. coli* electron transport protein of 106 residues (figure 2.3A). It non-covalently binds metallated protoporphyrin IX (PPIX) molecules

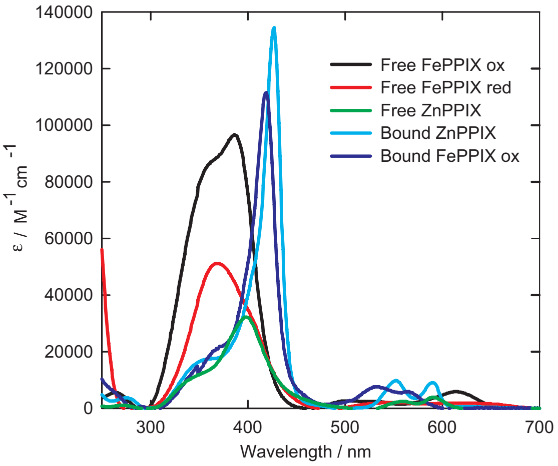


Figure 2.4: Extinction coefficients for cytochrome b$_{562}$ bound and unbound iron and zinc PPIX. Iron porphyrin is shown in both oxidised and reduced states

with nM affinity (figure 2.3B), where the metal atom at the centre can be iron, zinc or cobalt. The natural substrate is FePPIX, more familiarly known as heme. The binding to a metalloporphyrin stabilises the folded protein [81]. In the absence of bound porphyrin, the protein is termed the apo-form and the protein:porphyrin complex is termed the holo-form. Structural and electronic properties of cytochrome b$_{562}$ have been widely studied under a range of conditions [82, 83].

The porphyrin is a highly useful spectroscopic reporter. In the electronic absorption spectrum there are two strong transitions centered on 400 nm (B, or Soret band) and 560nm (Q band). A D_{4h} treatment of the π system of the porphyrin shows these bands can be assigned to $\pi - \pi^*$ transitions - $a_{1u}^2 a_{2u}^2 e_g^0 \rightarrow a_{1u}^1 a_{2u}^2 e_g^1$ for the B transition and $a_{1u}^2 a_{2u}^2 e_g^0 \rightarrow a_{1u}^1 a_{2u}^2 e_g^1$ for the Q transition. These transitions were first assigned by Gouterman [84] (figure 2.3C). For all porphyrins, binding to the cytochrome lowers the energy associated with the B and Q bands and the linewidth of the transition decreases significantly, reflecting the strong binding of the protein, and the defined conformation of the porphyrin within the protein.

The electronic spectrum readily allows distinction between oxidised and reduced states of the cofactor. Figure 2.4 shows the extinction coefficients for the bound and unbound states of zinc porphyrin, and the unbound, oxidised and reduced states of iron porphyrin. The peak maxima are recorded in table 2.1.

When bound to ZnPPIX, the protein becomes fluorescent. When bound to FePPIX the d-levels of the iron are quench the excited state which consequently has a lifetime <ns. FePPIX (heme) and can exist in low spin oxidised spin $\frac{1}{2}$, d^5 Fe^{3+} species and a low spin reduced diamagnetic d^6 Fe^{2+} species. The redox potential at pH 7 is 180mV vs NHE (normal hydrogen electrode) [82]. Importantly for this project, two molecules of cytochrome b$_{562}$ can self-exchange electrons at a rate of 4×10^6 M^{-1}s^{-1} at 315 K and pH 7 [86].

	Zinc porphyrin		Heme (Iron porphyrin)				
			Oxidised		Reduced		Isosbestic
	Free	Bound	Free	Bound	Free	Bound	
δ band	338(10)	343(2.5)	355(75)	374 (33.9)	-	319 (65.4)	350(12.9)
B Band	399(13)	427 (141)	385 (90)	419 (117)	367 (40)	426 (180)	420(110)
Q1 band	558 (1.9)	552 (1.2)	494(2)	564 (9.3)	517 (0.8)	531 (17.6)	565(5.9)
Q2 band	590 (2.5)	590 (1.1)	612(4.3)	531 (11.3)	628 (0.4)	562 (32.1)	533(7.6)

Table 2.1: UV-Vis peak/nm and (extinction/mM^{-1}cm^{-1}) data for FePPIX and ZnPPIX in a variety of states. Between pH 2-4, the porphyrin exists mainly in its oxidised form in the absence of reductant. Isobestic extinction data is used in subsequent analysis. All extinction data ± 5%. FePPIX bound data from [85]

$$Fe_a^{2+} + Fe_b^{3+} \rightleftharpoons Fe_a^{3+} + Fe_b^{2+}$$

By placing and holding these proteins on the surface of the fibril, the effective concentration can be estimated by taking the cytochrome to be a cylinder of radius 1 nm and length 5 nm [83]. By considering the number of molecules that can pack into 1 dm^{-3}, assuming a density of the concentration would be of the order 300mM (assuming a constant density, section 2.4.12). The frequency of electron transport would therefore be of the order 300s^{-1}. Currents from chains of cytochromes can then be estimated. The current I caused by a flux of n electrons per unit volume of charge e moving at a rate of v per second through an area A is given by $I = nAve$. The number density of electrons will be 1 per unit cytochrome volume, $I = ev/L$ where L is the length of a single cytochrome. Thus currents of the order 100 nA per chain could be expected through these systems. Values of this order should be readily detectable with modern conductivity probes. If parallel routes through the fibril are available when its three dimensional structure is taken into account, the current would be expected to be higher. If the concentration of cytochrome was overestimated, the currents will be lower.

While the exact function of cytochrome b$_{562}$ *in vivo* is not known, rapid electron transport is likely to be central to its role. The exact mechanism of electron exchange in biological systems in general is unclear. By analogy with electron transport processes in inorganic systems, two protein molecules must form a transient dimeric encounter complex, the *interaction conformation* (section 1.3). Porphyrin orbitals will have to overlap to allow electrons to tunnel from one complex to another. The cytochrome has been observed as a dimer in both the crystal structure (figure 2.3) [87], and at concentrations >1mM, by NMR [88] which could reflect the geometry of the interaction conformation.

On the surface of a fibril, the linker must therefore ensure that the cytochrome

Abbreviation	Description	Residues
Cyt	Wild type cytochrome b_{562}, both apo- and holo-forms discussed	106
$(SH3)$	The bovine PI3-SH3 monomer	86
$(SH3)_2$	Tandem repeat of bovine PI3-SH3 domain, both monomeric and fibrillar forms discussed	182
$(SH3)_2Cyt$	Tandem repeat of bovine PI3-SH3 domain, fused to wild type cytochrome b_{562}. Apo-,holo- , monomeric and fibrillar forms are discussed.	294
$(SH3)_3Cyt$	Trimeric repeat of bovine PI3-SH3 domain, fused to wild type cytochrome b_{562}. Apo-,holo- , monomeric and fibrillar forms are discussed.	388

Table 2.2: Constructs produced for this investigation and the abbreviations used in the text.

has sufficient mobility to reach its interaction conformation with its neighbour. The study of electron transport between cytochrome molecules held in fixed orientations, could allow details of their interaction conformation to be elucidated. One method of achieving this, for example, might be displaying cytochrome molecules in defined geometries on amyloid fibrils.

Cytochrome b_{562} has been used extensively in protein engineering experiments in the past and does not rapidly aggregate at low pH. Critically, the cytochrome is approximately half the mass of the SH3 dimer, and is predicted to be sufficiently small to avoid perturbing the underlying amyloid fibrils (figure 2.1).

2.2 Design solutions

The proteins designed for this investigation were expressed in *E. coli* with the functional domain as a C-terminal fusion, attached to the SH3 dimer by a widely used 10 residue flexible serine/glycine linker $(GGGGS)_2$ [89, 90]. The expression protocol is described in section 8.1 and sequence information is found in section 8.1.1. Table 2.2 and figure 2.2 further describe the species produced for this project.

Characterisation of the fibrillar state of $(SH3)_2Cyt$ is the primary focus of the following sections. Its aggregation behaviour is compared and contrasted with that of $(SH3)_2$ and its electronic properties are compared to that of wild type Cyt to elucidate the effects of the cytochrome when included in fibrils. $(SH3)_3Cyt$ was produced to distinguish between several structural models raised from working with $(SH3)_2Cyt$. All species apart from Cyt formed amyloid fibrils.

The properties of the monomeric states of the fusion proteins are described,

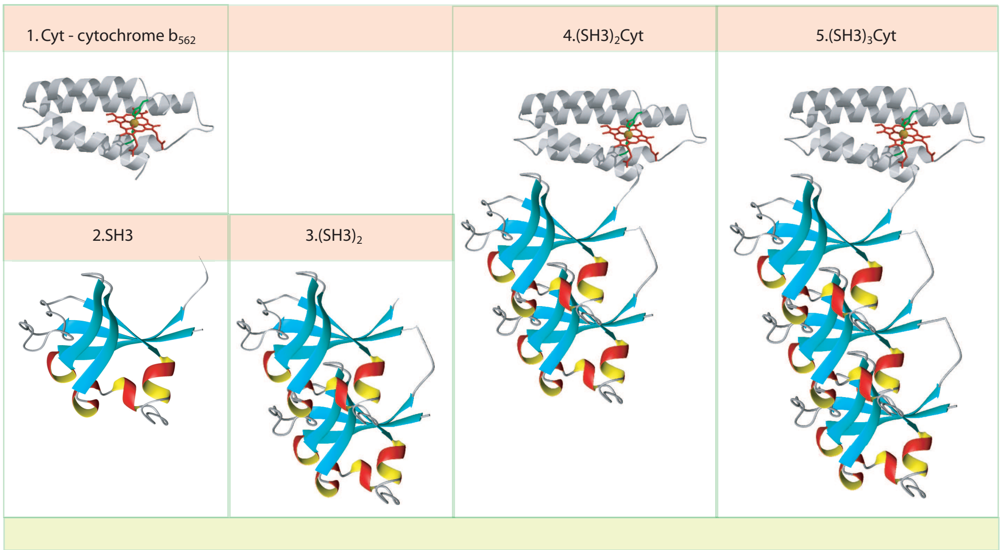


Figure 2.5: Representations of the protein constructs referred to in this chapter. All domains are depicted in their folded state at pH 7, and all cytochromes are shown in holo form with heme. At low pH, both domains are found to unfold under conditions described in section 2.4.

followed by a characterisation of how their behaviour changes at low pH. Finally the fibrillar states were subjected to a biophysical analysis to ascertain whether the design strategy has been successful. Figure 2.6 shows a cartoon of the target structure.

2.3 Characterisation of the monomeric states

The spectral characteristics of the SH3 and cytochrome fusions were studied at pH 7. The $(SH3)_2Cyt$ and $(SH3)_3Cyt$ possess the spectral characteristics of both the SH3 and the cytochrome. As each SH3 domain supplies 1 tryptophan and 8 tyrosine residues, the intrinsic fluorescence emission spectra of the fusion proteins are dominated by the SH3 domains (data not shown). The CD spectrum is a sum of the α helical cytochrome, and $\alpha\beta$ SH3 spectra (data not shown). The electronic spectra of holo-$(SH3)_2Cyt$ and holo-$(SH3)_3Cyt$ are identical to that of holo-Cyt.

In order to induce the fibril forming $(SH3)_2$ to aggregate, it has to be taken to pH 2, and so the structural changes in the two domains and in the fusion proteins were studied. Holo-Cyt can be observed to unfold at pH 2.2 by CD and UV-Vis spectroscopy whereas apo-Cyt is found to unfold at pH 2.9. $(SH3)_2Cyt$ was found to have unfolding transitions at comparable pH values for both apo- and holo- forms (appendix 2.32).

At early time points within the first hour at low pH it would appear that there

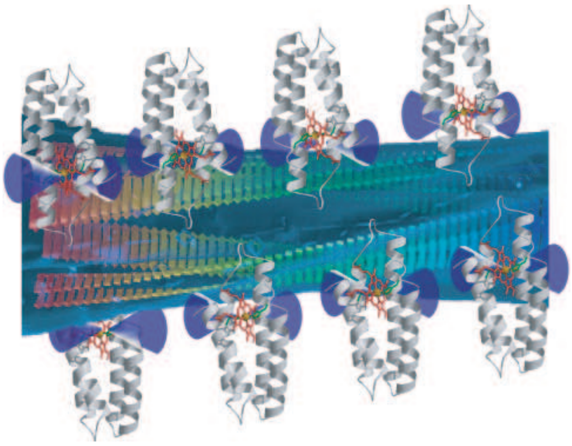


Figure 2.6: Spin fibril. The cones depict the anisotropy of the spin 1/2 in the paramagnetic Fe^{2+} state. The fibril structure is adapted from [35] and the cytochrome structure is adapted from [83]

is little interaction between the SH3 domain and the cytochrome in the fusion proteins $(SH3)_2Cyt$ and $(SH3)_3Cyt$. The aggregation behaviour of these systems was then studied.

2.4 Characterisation of the fibrillar states

2.4.1 $(SH3)_2$ reversibly aggregates at pH 3

From the dispersion of proton resonances in a ^{15}N-HSQC spectrum it can be determined whether or not a given polypeptide chain is folded. ^{15}N labelled protein was prepared according to a protocol described in section 8.1. At pH 3, the dispersion in the ^{15}N-HSQC spectrum of $(SH3)_2$ is ca. 3ppm (figure 2.8A green), whereas at pH 2 the dispersion is ca. 1ppm (figure 2.8A blue), suggesting at pH 3 $(SH3)_2$ is partially structured. In contrast, SH3 is unfolded at both pH 3 and pH 2 [60]. The two SH3 domains of $(SH3)_2$ therefore interact at pH 3, stabilising their structural elements.

At pH 3, despite the residual structure, $(SH3)_2$ aggregates (TEM images of the resulting aggregates are shown in figure 2.7B). The ^{1}H signal intensity decreases with time (figures 2.7A and B, in grey). NMR diffusion measurements confirm that the observed species is monomeric. The larger aggregates observed in the TEM therefore tumble too slowly to be observed by solution-state NMR (discussed further in chapter 6). The NMR signal intensity of monomeric protein is observed to decrease according to a bi-exponential decay (figure 2.7A). 90% of the signal was lost in 100 hours.

It is important to distinguish between the morphology of these aggregates and the morphology of the fibrils. A summary of features that distinguish between

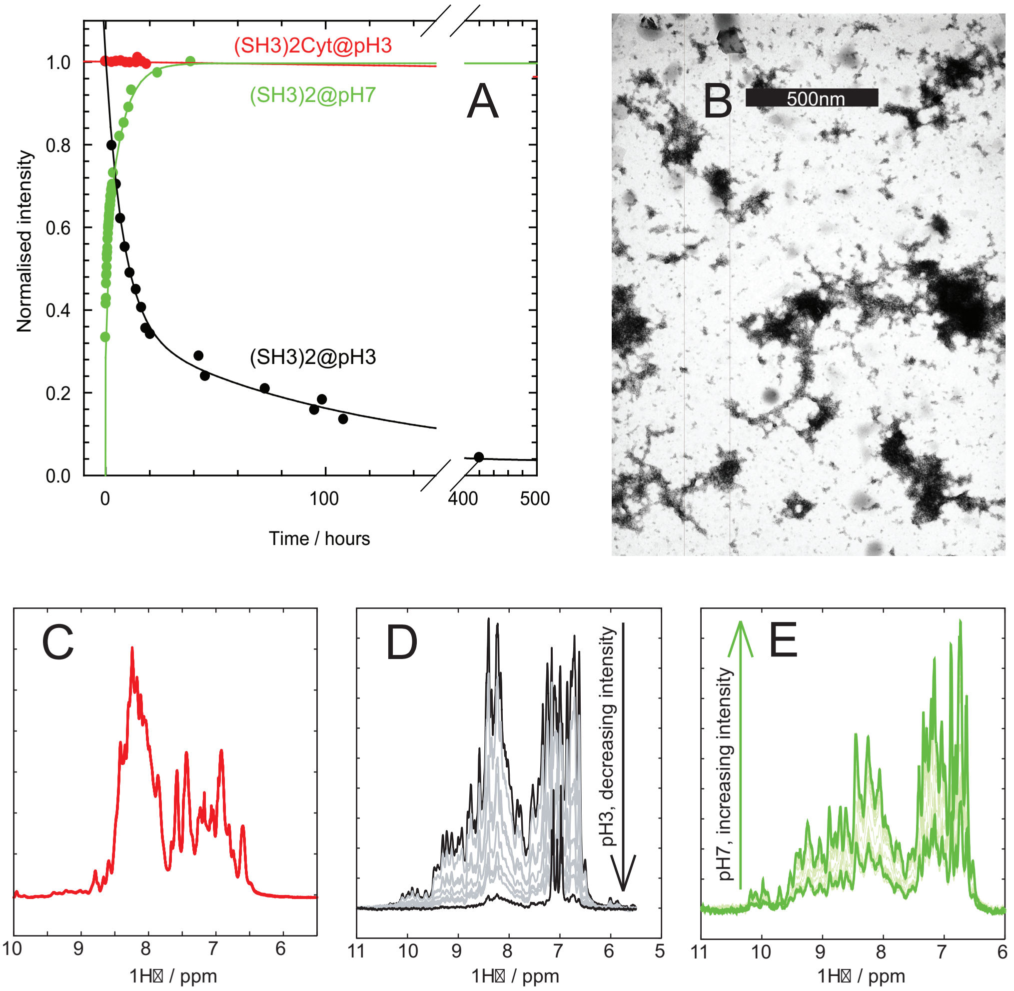


Figure 2.7: $(SH3)_2$ and $(SH3)_2Cyt$ are taken to pH 3 and the intensity of their 1D proton spectra followed with time. **A -** No change in the signal intensity of $(SH3)_2Cyt$ spectra is observed (red), whereas signal intensity from $(SH3)_2$ decays through a bi-exponential process (black), followed here for 14 days. These data were fitted to $y = a_1e^{-k_1t} + a_2e^{-k_2t}$ yielding $a_1 = 0.66, k_1 = 0.12hrs^{-1}, a_2 = 0.34, k_2 = 7.2 \times 10^{-3}hrs^{-1}$. At this pH, fibrillar aggregates of the type shown in panel B are observed. On taking $(SH3)_2$ aggregates back to pH 7 (green), in under 12 hours, the initial intensity is restored and the aggregates seen in the TEM are observed to disappear. These data were fitted to $y = a_1(1 - e^{-k_1t}) + a_2(1 - e^{-k_2t})$ yielding $a_1 = 0.47, k_1 = 7.3hrs^{-1}, a_2 = 0.51, k_2 = 0.20hrs^{-1}$. **C/D/E -** The spectra from which the quantitative analysis in panel A was derived for $(SH3)_2Cyt$@pH 3, $(SH3)_2$@pH 3 and $(SH3)_2$@pH 7 respectively.

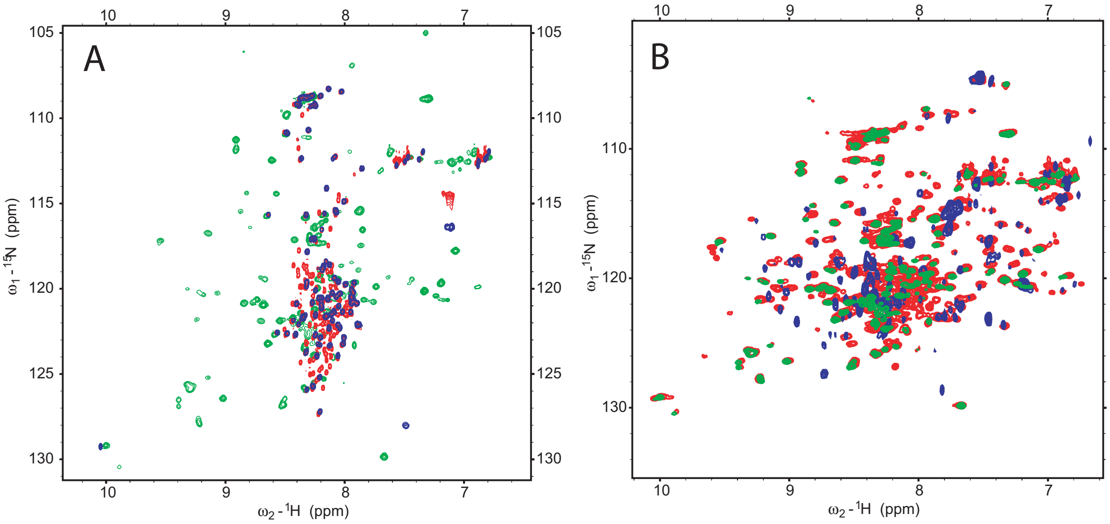


Figure 2.8: ^{15}N-HSQC spectra of different states of $(SH3)_2Cyt$ and $(SH3)_2$. **A -** An unfolded spectrum of $(SH3)_2$ obtained by taking the protein to pH 2 then returning to pH 3 is shown in blue. A pH 3 folded spectrum of $(SH3)_2$ is shown in green, after taking the protein to pH 3 from pH 7, taken during the kinetic experiment, above. A spectrum of apo-$(SH3)_2Cyt$ at pH 3 is shown in red, also acquired during the above kinetic experiment. **B -** The same pH 3 folded spectrum of $(SH3)_2$ in A is shown here, again in green. Holo-Cyt is shown in blue at pH 3. Holo-$(SH3)_2Cyt$ is shown in red, and has resonances that overlay with both the folded $(SH3)_2$ and holo-Cyt, in contrast to that seen for apo-$(SH3)_2Cyt$, where both domains are unfolded.

the two types of aggregates is shown in table 2.3. Fibrillar aggregates apparently share some features of fibrils particularly at their edges, though structurally they are distinct. Aggregated material from many protein systems gives rise to species with this morphology when exposed to denaturing conditions. These species are often termed 'amorphous aggregates', though it is unlikely they are unstructured and will be termed 'fibrillar aggregates' in this thesis. Amyloid fibrils show discrete fibrils of constant width with the observable features spread homogenously over a TEM grid. Fibrillar aggregates in contrast appear to exist in patches - regions of non-covered grid and regions of very high protein density are observed. Representative TEM images of $(SH3)_2Cyt$ fibrils are contrasted with TEM images of $(SH3)_2Cyt$ fibrillar aggregates in figure 2.9B and C.

When $(SH3)_2$ aggregates are returned to pH 7, the fibrillar aggregates are observed to disappear in the TEM, and NMR signal intensity is observed to reappear to 100% of its original value, as monomeric species return to solution (figures 2.7E and 2.7A, green). The time taken for the species to return to monomeric protein is significantly faster than the rate for the species to aggregate at pH 3. In less than 2 hours, over 90% of the signal has returned, whereas 20 hours were required for a similar intensity loss.

The aggregation of $(SH3)_2$ at pH 3 is therefore reversible. At this pH the aggregates are favoured, and at pH 7 the monomeric folded state is the more stable. Amyloid fibrils of $(SH3)_2$ however, are stable at pH 7.

Finally, $(SH3)_2$ has an interesting hysteresis in its folding behaviour. When the protein is taken from pH 7 to pH 2, the protein is unfolded (figure 2.8A, blue). When this state is taken to pH 3, it remains unfolded. Taking $(SH3)_2$ to pH 3 directly from pH 7 results in folded protein (figure 2.8A, green), whereas going to pH 3 via pH 2 results in unfolded protein. The aggregates formed by pH 3 state that has been initially unfolded are crucial to the rate enhancement of the pH cycling protocol used to form fibrils from $(SH3)_2$ (section 8.1.2, [80]).

2.4.2 Cytochrome reduces the tendency of the $(SH3)_2$ to aggregate at pH 3

In contrast to $(SH3)_2$, neither apo nor holo-$(SH3)_2Cyt$ aggregate at pH 3, or do so on a timescale many orders of magnitude slower than that observed for $(SH3)_2$ (figures 2.7C and 2.7A). The ^{15}N-HSQC spectrum at pH 3 of holo-$(SH3)_2Cyt$ (figure 2.8B, red) is a sum of $(SH3)_2$ at pH 3 (figure 2.8B and 2.8A, green) and pH 3 holo-Cyt (figure 2.8B, blue). The SH3 domains are structured, as they are for $(SH3)_2$.

The spectrum of apo-$(SH3)_2Cyt$ (figure 2.8A, red) is a sum of unfolded apo-Cyt and unfolded $(SH3)_2$ (figure 2.8A, blue, section 2.4.7). The unfolded cytochrome therefore destabilises the interactions between the two SH3 domains at pH 3, preventing them from forming the structure observed in $(SH3)_2$.

The presence of either folded or unfolded cytochrome adjacent to the $(SH3)_2$ domain prevents its aggregation at pH 3. Unfolded cytochrome prevents the interactions between the two SH3 domains that lead them to be structured at pH 3. The observations that $(SH3)_2Cyt$ at pH 3 displays folded cytochrome resonances and exists in a 1:1 heme binding stoichiometry in its un-aggregated state are crucial to the conclusions of section 2.4.10.

2.4.3 Cytochrome must be unfolded in order for $(SH3)_2Cyt$ to form fibrils

At pH 2 $(SH3)_2$ readily forms amyloid fibrils, requiring the pH cycling protocol for optimum rate of formation (sections 2.1.1 and section 8.1[80]). Under these conditions, holo-$(SH3)_2Cyt$ was found to form fibrillar aggregates only, (figure 2.9C). Apo-$(SH3)_2Cyt$, however forms fibrils under these conditions (figure 2.9B). The rate of apo-$(SH3)_2Cyt$ fibril formation is slower than the rate of $(SH3)_2$ fibril formation by a factor of ~4. The presence of an unfolded cytochrome

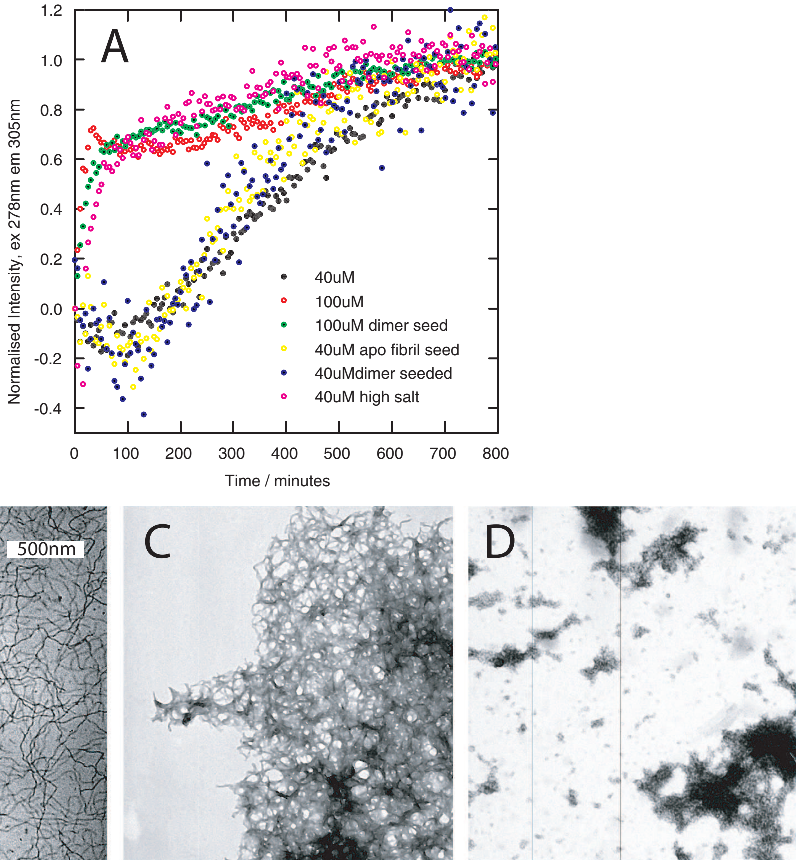


Figure 2.9: Fibril formation followed by intrinsic fluorescence. Under conditions where fibrils form, in the first 13 hours the lower trace is observed and fibrils as shown in B are observed 4 days later. Under conditions that follow the upper trace pattern, fibrillar aggregates, shown in C are observed. When purified fibrils are taken from pH 3 to pH 7, fibrils are observed to decompose and return back to their monomeric state going via states similar to that shown in D.

Measurement	Fibrillar aggregates	Fibrils
TEM morphology	Protein concentrated in large aggregates. Aggregates have fibrillar aspects, particularly at the edges. Distinguishable from unbranched, discrete fibrils.	Discrete, unbranched fibrils.
TEM grid coverage	Patchy	Homogeneous
AFM	Not observed	Morphology identical to that observed by TEM
UV-Vis	Strongly curved baseline due to scattering	Flat baseline

Table 2.3: Methods of distinguishing fibrillar aggregates from fibrils

Species	pH 3	pH 2
$(SH3)_2$	fibrillar aggregates	fibrils
holo-$(SH3)_2Cyt$	stable	fibrillar aggregates
apo-$(SH3)_2Cyt$	stable	fibrils

Table 2.4: Types of aggregate formed under conditions of different pH

domain causes fibrils to form at a reduced rate. A summary of aggregation state obtained for the various systems with varying pH is shown in table 2.4.

Fluorescence spectroscopy can be used to distinguish between non-specific aggregation and aggregation leading to fibrils, as illustrated in figure 2.9A. As the proteins aggregate, the frequency of intrinsic emission of the tyrosine and tryptophan residues is found to shift depending on the hydrophobicity of their environment. When this is followed as a function of time, non-specific aggregation leads to a hyperbolic spectral trace (top traces in figure 2.9A), whereas fibril formation follows a sigmoidal profile (bottom traces in figure 2.9A). When using the pH cycling protocol, the lag-phase is independent of additional seeding, a property of $(SH3)_2$ (section 8.1.2 [80]).

The range of conditions that lead to apo-$(SH3)_2Cyt$ fibrils is narrower than that of $(SH3)_2$. When the concentration of $(SH3)_2Cyt$ is greater than 100μM, at ionic strengths >10mM, and when the sample volume is >1ml, $(SH3)_2Cyt$ aggregates non-specifically, while $(SH3)_2$ forms fibrils. Below these thresholds however, $(SH3)_2Cyt$ readily forms fibrils. High molecular mass fibrils are purified from smaller species by ultracentrifugation at 90k in a TLA-120 rotor for 1.5 hours, corresponding to species with a sedimentation coefficient $> 10S$. Monomeric $(SH3)_2Cyt$ does not detectably sediment under these conditions. Ultracentrifugation is used throughout this thesis to purify amyloid fibrils from smaller materials. Fibrils purified by ultracentrifugation were subject to further biophysical analysis. Both aggregates and fibrils of $(SH3)_2Cyt$ and $(SH3)_3Cyt$ dissociate after 10 minutes in SDS at 70^{o}C. A single band is seen on the corresponding PAGE gels validating that the fibrils are formed from full length protein, and not hydrolysed protein.

The π-π^* transitions of aromatic amino acids lead to strong absorption ca. 280nm that scales linearly with the number of aromatic amino acids [91]. The extinction coefficients of this absorption are insensitive to local electronic environments, and knowledge of the appropriate extinction coefficient allows an estimate of the total protein concentration. For $(SH3)_2Cyt$ this is found to be 33,000M^{-1}cm^{-1}, verified by amino acid analysis. The total absorbance at 280nm is found not to change over the course of a fibril forming reaction. This leads to a useful result again, verified by amino acid analysis;

$$\epsilon_{280nm}^{Freeprotein} = \epsilon_{280nm}^{Fibril} = 33,000M^{-1}cm^{-1}$$

This allows total protein concentration to be determined using electronic spectroscopy independently of the aggregation state. It was found that between 50-90% of the total protein is found in the pellet after a purification by ultracentrifugation. The variance in yield most likely reflects the stochastic nature of the aggregation reaction, varying incubation times (between 4 and 10 days) and the intrinsic variability of individual protein preparations. s. Once purified, the only observed difference between individual fibril preparations was the length of the fibrils in the preparation as observed by TEM and AFM.

Several days after purification, samples were re-pelleted to investigate whether the fibrils were stable on this timescale. Second supernatants from fibrils of $(SH3)_2Cyt$ and $(SH3)_3Cyt$ were found to contain less than 0.1% of the total material, suggesting that at pH 3, the fibril are stable. Interestingly, fibrils from $SH3$ and $(SH3)_2$ were found to have ca. <10% of the total material in the second supernatant. It has been previously observed that fibrillar and monomeric SH3 are in slow exchange [49]. The observed decrease in dissociation rate for $(SH3)_2Cyt$ and $(SH3)_3Cyt$ fibrils led to the project discussed in chapters 4 and 5, where the dissociation rate is correlated to an increased number of flexible regions on the fibril surface. The studies described in the remainder of this chapter were performed on samples purified by ultracentrifugation, unless stated otherwise.

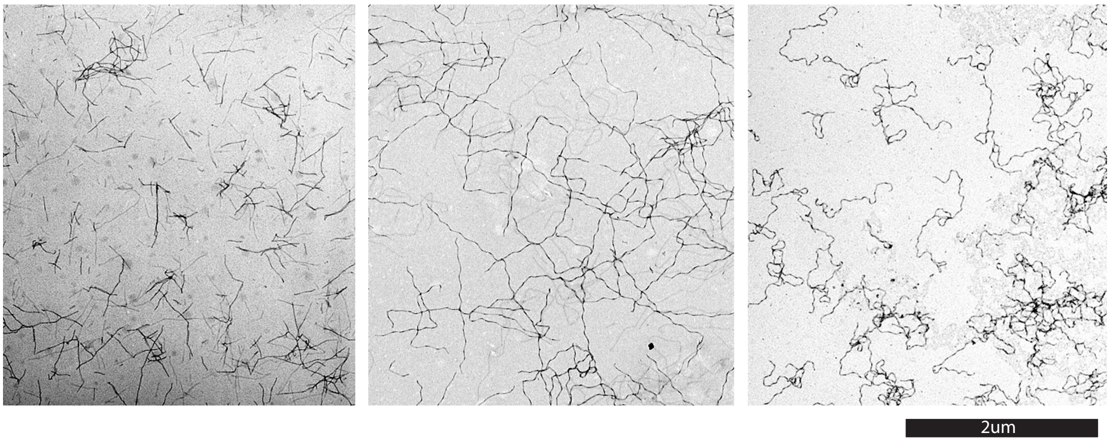


Figure 2.10: Negatively stained bright field TEM micrographs of $(SH3)_2$ fibrils, $(SH3)_2Cyt$ and $(SH3)_3Cyt$ fibrils.

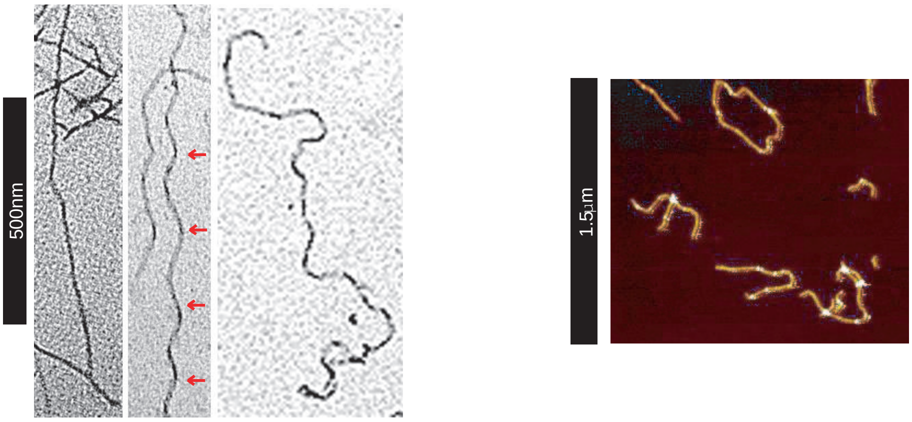


Figure 2.11: Negatively stained bright field TEM micrographs of $(SH3)_2$ fibrils, $(SH3)_2Cyt$ and $(SH3)_3Cyt$ fibrils. An AFM image of rings formed from $(SH3)_3Cyt$ is shown on the right. A height analysis shows the ring to be a closed loop and not a linear fibril curved back on itself.

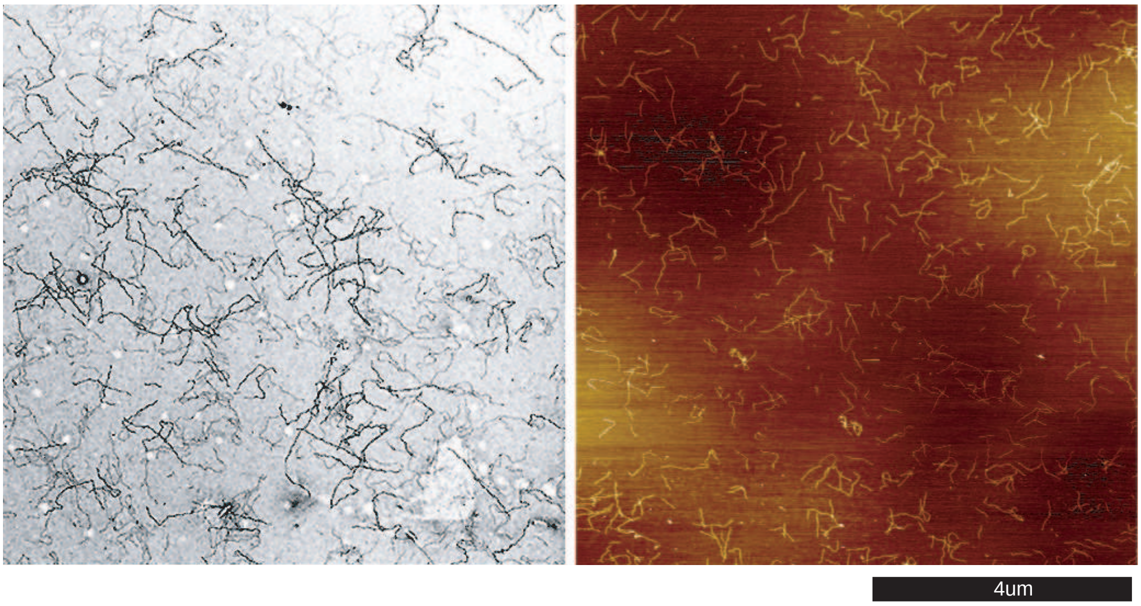


Figure 2.12: Morphology of $(SH3)_2Cyt$ fibrils compared by of AFM and TEM. AFM images courtesy of Tuomas Knowles.

2.4.4 TEM and AFM studies show the higher order packing in the fibrils is perturbed by the presence of the cytochrome

Fibrils formed from $(SH3)_2$, $(SH3)_2Cyt$ and $(SH3)_3Cyt$ were imaged by AFM and TEM, following protocols described in section 8.2.4, and the observed morphologies were compared. Bright field TEM images after negative staining with uranyl acetate are compared to AFM images that were acquired in tapping mode to minimise damaging the structures of interest. For TEM, the sample is placed on carbon coated nickel grids and washed with uranyl acetate. For AFM, sample is deposited on atomically smooth mica substrates. For both techniques, a similar quantity of material (ca. 25 pmoles per cm^2 substrate) was deposited on the substrate and dried. The AFM and TEM protocols are described in section 8.2.4.

Figure 2.10 shows the varying morphology of $(SH3)_2$, $(SH3)_2Cyt$ and $(SH3)_3Cyt$ fibrils, showing that the presence of the cytochrome perturbs the morphology of the fibril. Whereas fibrils of $(SH3)_2$ are rod-like, those from $(SH3)_2Cyt$ and $(SH3)_3Cyt$ have an apparent additional coil, giving rise to a regular 150nm repeat (figure 2.11, red arrows). The fibril axes on average follow linear paths despite this regular helical repeat. Kinks are observed in the body of some of the fibrils that lead to changes in direction of the fibril axes. Before and after kinks, the fibril axes extend linearly. Fibrils from $(SH3)_3Cyt$ are significantly shorter and more kinked than fibrils from $(SH3)_2Cyt$ (figure 2.11). In some cases, $(SH3)_3Cyt$ fibrils are found to have sufficient flexibility to form closed loops (figure 2.11)[92]. Height analysis of the loops confirms that they are closed, and not linear fibrils apparently curled round on themselves. These species are not seen in fibril preparations of $(SH3)$, $(SH3)_2$ or $(SH3)_2Cyt$. Formation of loops requires bidirectional fibril growth and a low bending modulus.

Many amyloid fibril forming systems show several levels of protofilament assembly, manifest as several distinct populations of species with different dimensions observed by AFM [53] and TEM [35, 93]. SH3 fibrils are found to show this behaviour. $(SH3)_2$, $(SH3)_2Cyt$ and $(SH3)_3Cyt$ fibrils yield only one species of height 3 ± 0.5nm, as measured by AFM. This suggests that the core fibril is similar in each case; flexible protein would not be imaged by AFM with the cantilevers that are currently commercially available as softer structures displayed on the surface of a tightly packed fibril would be displaced by the passage of the AFM cantilever. Furthermore, this observation suggests that the additional groups displayed on the surface of the fibrils interfere with the higher order protofilament assembly observed for $(SH3)$ fibrils.

Fibrillar aggregate species formed from holo-$(SH3)_2Cyt$ and $(SH3)_2$ observed by TEM are not observed by AFM under the present experimental conditions. The fibrillar aggregates are considerably less rigid than amyloid fibrils and so are either destroyed or displaced on imaging, or they are too soft to cause deflection in the AFM's cantilever.

This idea is supported when the AFM fibril height is compared to the fibril width as measured by TEM. The agreement is poor. The height by AFM is approximately 3 nm, whereas the width by TEM is approximately 12 nm. The AFM estimate is likely to be an underestimate as the forces applied by the cantilever will compress softer structures. The TEM estimate will be an overestimate as the origin of the stain is in the deposition of the heavy metal precipitate. Softer groups would be more likely to trap staining agents and hence be resolved in the TEM experiment. Thus a discrepancy in the dimensions measured by the two techniques would be expected. Individual cytochrome domains are not resolvable in the TEM suggesting that they are closely packed. This is optimal for any form of charge transfer process one would hope to observe.

The composition of the fibril forming protein directly affects the morphology of the produced fibrils. The structure formed in the absence of perturbing groups is likely to be the most stable fibril arrangement and so the morphologies observed in the modified fibril forming domain reflect an energetic compromise between maximum fibril stability and minimising packing constraints imposed by the perturbing groups. The observed transition from rods to helical rods to highly flexible fibrils is observed moving from $(SH3)_2$ to $(SH3)_2Cyt$ to $(SH3)_3Cyt$ is to be expected. This transition is discussed further in section 2.6.4.

The length distributions measured by TEM and AFM are in close agreement. A novel method for measuring length distributions of $(SH3)_2Cyt$ fibrils using NMR spectroscopy is described in section 3.3. The length distributions measured using NMR are found to be in good agreement with those obtained from TEM and AFM measurements [94].

2.4.5 The fibrils have amyloid structure

Purified fibrils of $(SH3)_2$ and $(SH3)_2Cyt$ were subjected to a biophysical characterisation to verify whether or not the fibrils observed in the TEM and AFM had the structure characteristic of amyloid. The rationale behind these tests is discussed in section 1.1. CD and FTIR spectroscopies were used to estimate the bulk secondary structure in a given sample. Far-UV CD spectra are shown in figure 2.13D. Monomeric apo-$(SH3)_2Cyt$ at pH 7 has an α helical hydrogen bonding pattern (black). At pH 2.3, it unfolds (red). After fibrils are formed, a minimum at 216nm is observed (green), coinciding with that seen for fibrils of $(SH3)_2$ (light blue). Above pH 3, a refolding transition is observed in the apo-$(SH3)_2Cyt$ fibrils, consistent with a refolding event in the cytochrome (dark blue). The morphology of the fibrils at pH 3 was indistinguishable from the morphology of the fibrils at

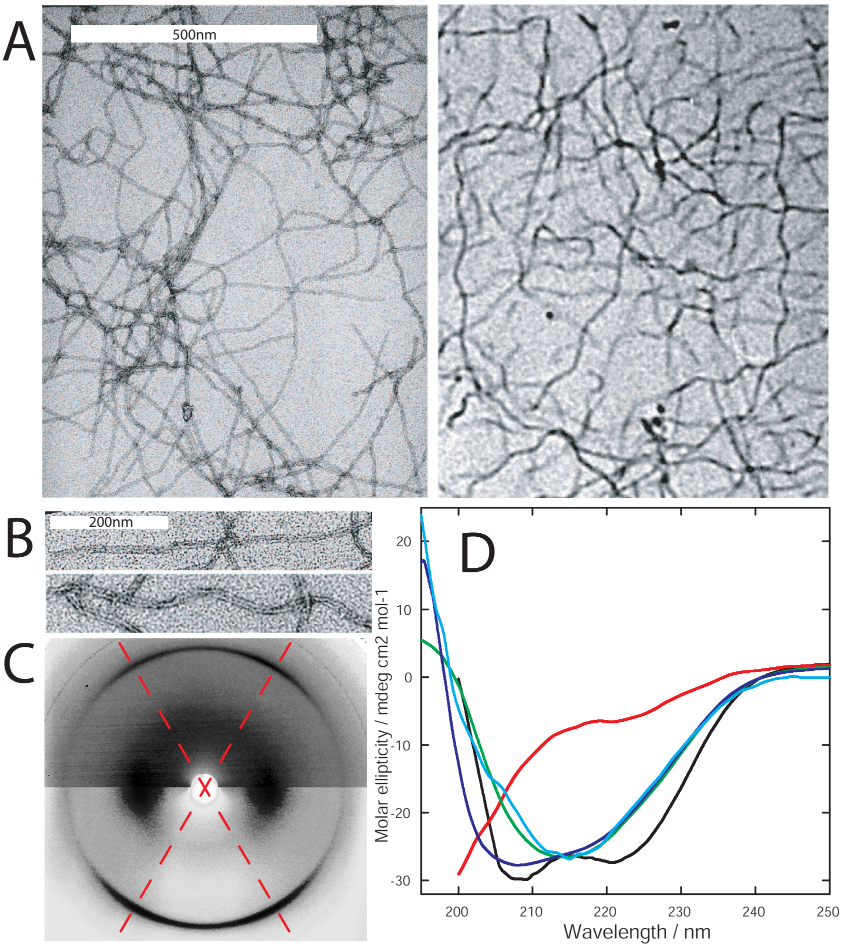


Figure 2.13: Biophysical comparison of CD, XRD and TEM data between fibrils of $(SH3)_2Cyt$ and $(SH3)_2$. **A -** Bright field negatively stained TEM images of $(SH3)_2$ fibrils (left) and holo-$(SH3)_2Cyt$ fibrils (right). Differences in morphology are highlighted in B, where representative $(SH3)_2$ fibrils (top) are compared to $(SH3)_2Cyt$ fibrils (bottom), the latter possessing a regular coil. **C -**Fibre XRD images taken from $(SH3)_2$ fibrils (bottom half) and apo-$(SH3)_2Cyt$ fibrils (top half). Both share meridional and equatorial reflections and so have similar protofilament structure. **D -** CD spectra. Monomeric apo-$(SH3)_2Cyt$ at pH 7 has an α helical hydrogen bonding pattern (black). At pH 2.3, it unfolds (red). After fibrils are formed, a minimum at 216nm is observed (green), coinciding with that seen for fibrils of (SH3)2 (light blue). Above pH 3, a refolding transition is observed in the apo-$(SH3)_2Cyt$ fibrils, consistent with a refolding event in the cytochrome (dark blue). The fibril morphology does not visible change between pH 2 and pH 3.

pH 2 (figure 2.13D).

By carefully drying samples between wax beads, stalks of partially aligned fibrils can be prepared as the fibrils align with the beads during the drying process. A fibre X-ray diffraction pattern of such a sample is anisotropic, yielding a cross-β pattern. Interpretation of this pattern has β sheets hydrogen bond parallel to the fibril axis leading to a reflection parallel to the fibril axis at 4.7Å and paired β strands leading to a reflection at 9.6Å perpendicular to the fibril. Samples from both $(SH3)_2$ and $(SH3)_2Cyt$ fibrils were found to give this distinctive pattern (figure 2.13). It is interesting to compare this pattern to that observed from $SH3$ fibrils (figure 1.2). The 4.7Å reflection is found to be split into two maxima (figure 2.13C). This is expected for systems with helical symmetry.

Two histological dyes, Thioflavin-T and Congo Red give distinctive spectroscopic fingerprints when bound to certain biological structures. Historically this binding has been associated with the presence of amyloid structures. $(SH3)_2Cyt$ and $(SH3)_2$ fibrils were found to bind both dyes.

Thus the fibrillar structures visualised in the TEM and the AFM can be considered to be of amyloid structure. Despite the higher order packing being perturbed, the local packing of the protofilaments is not perturbed by presence of the cytochrome. In addition, a refolding transition is observed in the CD between pH 2 and pH 3 for $(SH3)_2Cyt$ fibrils that is not observed for $(SH3)_2$ fibrils, at a pH where wild type apo-Cyt undergoes a CD folding transition. The cytochrome therefore is still able to fold and refold despite being attached to the surface of a fibril (figure 2.32).

2.4.6 Displayed groups are sufficiently flexible for observation by solution state NMR

Surprisingly, solution state NMR spectroscopy has been found to be exceedingly useful in characterising the fibrils formed from $(SH3)_2Cyt$ and $(SH3)_3Cyt$ (figure 2.14). The linewidth of a solution-state NMR resonance is governed by the rotational diffusion of the molecule under investigation. A 5μm fibril would be expected to have a mass ca. 1GDa (section 2.4.12), which would be expected to tumble far too slowly to give experimentally resolvable resonances. However, where a group within the structure has sufficient mobility, this broadening can be overcome. It is possible therefore, to study resonances of flexible regions attached to large structures by solution-state NMR, though the rigid core of the structure gives rises to resonances far too broad to be experimentally detected. This motion required to give rise to these resonances is the subject of chapter 6.

Species involved in an aggregation reaction can be physically separated using

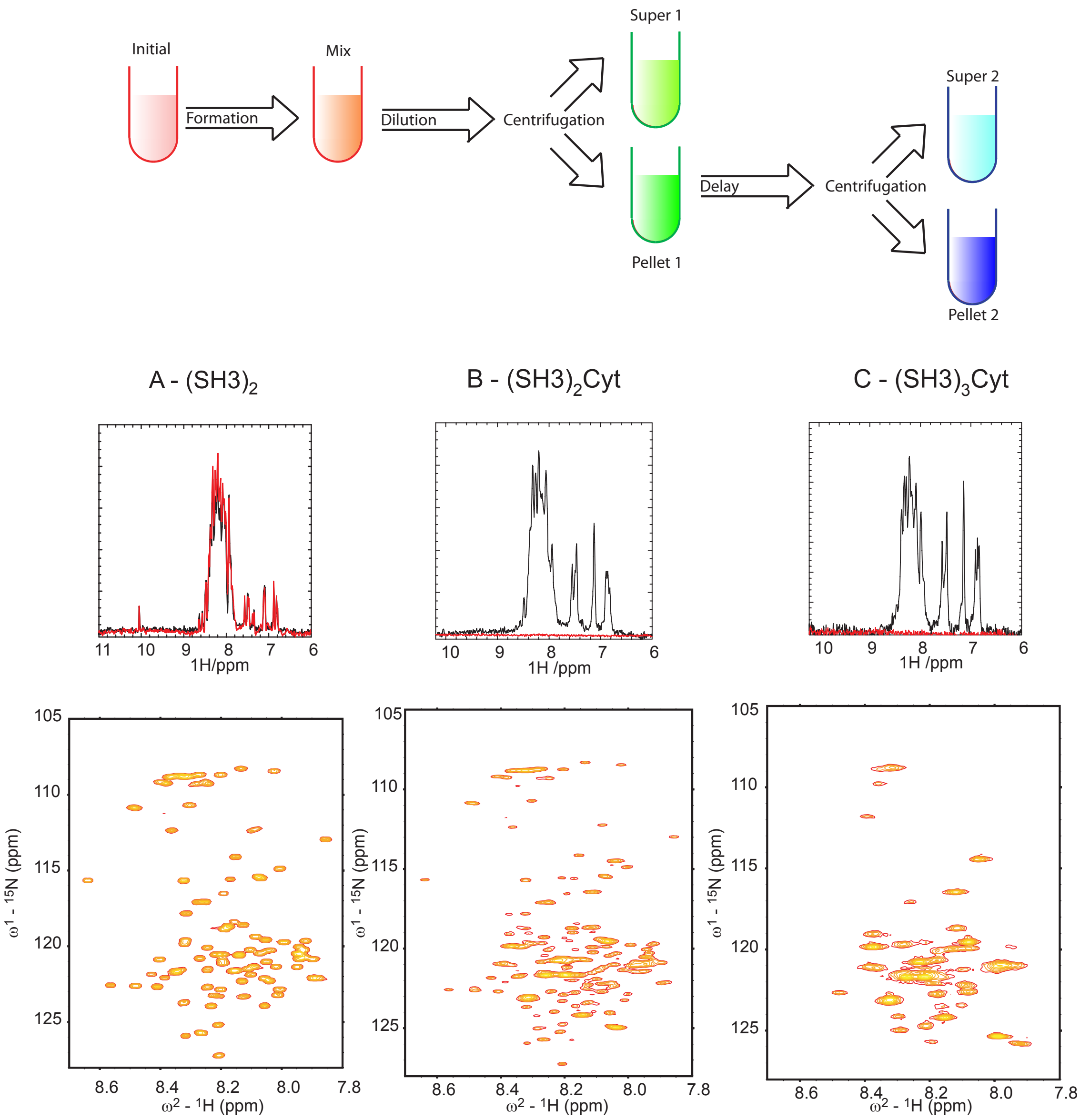


Figure 2.14: Top - experimentally available states of aggregation reactions, showing the relationship between the 'first pellet', the 'second pellet' and the 'second supernatant'. This scheme is discussed more thoroughly in chapter 4. Bottom - ^{15}N-HSQC NMR spectra seen from $(SH3)_2$, $(SH3)_2Cyt$ and $(SH3)_3Cyt$ re-suspended ultracentrifuge pellets of fibril samples, and second supernatant. The dispersion of chemical shifts in the proton dimension is consistent the protein being unfolded in each case. **A -** $(SH3)_2$ fibrils. The pellet 1 and super 2 spectra are essentially identical suggesting the monomeric population re-equilibrates rapidly after re-suspension of a fibril sample. **B/C -** , $(SH3)_2Cyt$ and $(SH3)_3Cyt$ fibrils respectively. Less than 0.1% of the protein found in pellet 1 was found in super 2, showing the NMR signals observed originate from the large species found in pellet1.

ultracentrifugation, and the concentration of proteins in the various fractions can be determined using UV-Vis spectroscopy. This principle gives rise to the scheme described in figure 2.14, and discussed at length in the context of a variety of fibril forming systems in chapter 4. Once a fibril formation reaction is in a steady state, large material is separated from small, yielding purified aggregated material in a re-suspended initial pellet (pellet1). This can be re-centrifuged after a delay of several days and NMR spectra of pellet1 can be compared to the NMR spectra of the second supernatant (super2). If they are identical, the species observed in pellet1 are from small species that have rapidly dissociated from the fibril after purification. If super2 is found to be devoid of material, then the NMR spectra observed in the pellet1 fraction most likely has originated from a larger species in solution. NMR diffusion measurements can further characterise the size of the molecule that gave rise to the observed resonances.

^{15}N labelled protein was prepared according to a protocol described in section 8.1. ^{15}N labelled fibrils were then prepared and purified by ultracentrifugation. The pellet1 fraction of all three fibril samples yielded well defined ^{15}N-HSQC spectra, shown in figure 2.14. The dispersion of peaks in the proton dimension is less than 1ppm in all cases, and so the species from which the resonances originate are unstructured. As already noted, when either $(SH3)_2Cyt$ or $(SH3)_3Cyt$ fibrils are re-centrifuged, less than 0.1% of material is found in super2 after a delay of several days, and the 1D ^{1}H spectrum of pellet1 is equal to that of pellet2. The spectra of the second supernatant are devoid of any appreciable signal for these systems, shown in figure 2.14. In contrast, when a $(SH3)_2$ or an $SH3$ fibril sample is repurified, approximately 10% of the total material is found in the second supernatant after a delay of one hour. The 1D ^{1}H spectra in these cases are indistinguishable from those of the corresponding pellet1. These two types of behaviour are discussed in the context of other amyloid fibril forming systems in chapter 4. The conclusion, as originally observed in $SH3$ fibrils [49] is that fibril samples depleted in monomer rapidly return to equilibrium after monomer is removed. More interesting, is that fibrils from $(SH3)_2Cyt$ and $(SH3)_3Cyt$ do not do this. This is attributed to a drastic reduction in the dissociation rate as a consequence of the incorporation of additional, non-core protein into the fibril (chapter 4).

The centrifugation comparison can be taken as suggestive evidence that the NMR resonances can be observed from $(SH3)_2Cyt$ and $(SH3)_3Cyt$ fibrils. Four further NMR experiments in the following sections reinforce this conclusion. The first comes from an intensity analysis that shows the loss of one of the two SH3 domains from the spectrum (section 2.4.7). The second comes from diffusion experiments (section 2.4.8), the third comes from the effect of heme binding on these

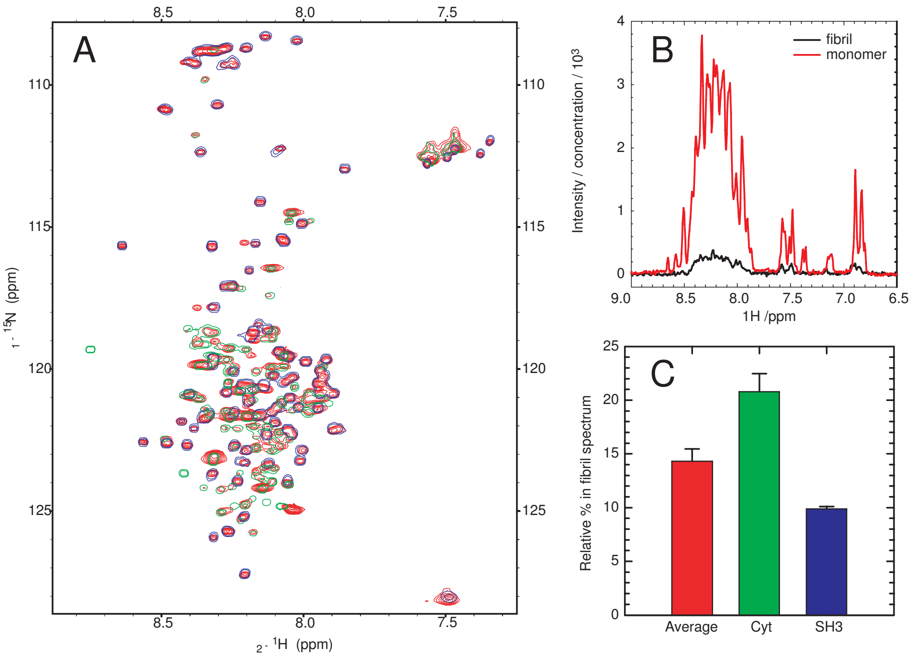


Figure 2.15: Intensity analysis of peaks observed from $(SH3)_2Cyt$ fibrils. **A -** The ^{15}N HSQC spectrum of $(SH3)_2Cyt$ fibrils (red) has peaks that originate from both unfolded apo cytochrome (green) and acid unfolded SH3 (blue). The HSQC spectrum obtained form fibrillar samples is highly reproducible. **B -** When these data are analysed quantitatively, the absolute signal coming from the fibril is ca. 15% of that seen in the monomer. **C -** The relative fibril:monomer signal intensity averaged over 16 discrete peaks from the cytochrome and 22 peaks from SH3 from 3 independent monomeric preparations and 5 independent fibril preparations of $(SH3)_2Cyt$. The SH3 peaks are on average 50% lower in intensity than the cytochrome peaks. This would imply where cytochrome resonances are observable, they are observed associated with just one of the two SH3 domains.

resonances (section 2.4.10) and the final piece of evidence comes from a kinetic experiment, which monitors the evolution of spectra with time in section (section 2.4.15).

2.4.7 For each Cyt observed in $(SH3)_2Cyt$ fibrils, only one of the two SH3 domains is observed

When the ^{15}N-HSQC spectra of fibrillar $(SH3)_2Cyt$ are compared to that of apo-Cyt and unfolded SH3, the spectrum can be entirely understood as a supposition of the two (figure 2.15, fibrillar $(SH3)_2Cyt$ in red, apo-Cyt in green and unfolded SH3 in blue). Ca. 95% of the peaks from the two species are found in spectra from the fibrils. Qualitatively therefore, spectra from monomeric $(SH3)_2Cyt$ are identical to those observed from fibrillar $(SH3)_2Cyt$. This leaves a paradox; if the peaks originate from flexible regions attached to amyloid fibrils, how can all amino acid residues be visible?

A more quantitative intensity analysis can be performed. The NMR intensity of a fibril sample was compared to that of a monomeric species. Figure 2.15 shows a 1D ^{15}N-HSQC concentration normalised spectrum of $(SH3)_2Cyt$ fibrils (black) and a 1D ^{15}N-HSQC concentration normalised spectrum of $(SH3)_2Cyt$ monomer (red), both at pH 3. The total intensity originating from the fibril sample is ca. 15% of the intensity of monomeric $(SH3)_2Cyt$ at pH 3. This figure was confirmed by comparison of several independent samples on several different NMR spectrometers. A significant proportion of the $(SH3)_2Cyt$ molecules present in the fibrils are not observed in the solution-state NMR experiments.

To go further, individual peaks are resolvable outside the highly overlapped central region. As shown in figure 2.15A, these can be assigned to either apo-Cyt (green) or SH3 (blue) from the overlaid spectra of these species at this pH. Their integrated intensity in the monomeric form can be compared to that seen in the fibrillar form. The rationale behind interpreting integrated signal intensities in these quantitative terms is discussed in section 4.5.1.

The signal intensities of 39 distinct, individually assignable resonances (17 from Cyt and 22 from $SH3$) were compared over 8 ^{15}N HSQC spectra (3 from monomeric and 5 form fibrillar $(SH3)_2Cyt$). The linewidths, and hence the apparent T_2 values of both the monomeric sample and the fibrillar sample were found to be similar. Interestingly, on average, it was found that the intensity of the cytochrome peaks in the fibril samples were twice as intense as those originating from SH3. Therefore only 20% of the cytochrome resonances and 10% of the resonances from SH3 are observed in fibrillar $(SH3)_2Cyt$ relative to those observed from monomeric $(SH3)_2Cyt$. Thus 90% of the SH3 domains and 80% of the cytochrome domains are not observable to solution state NMR. The cytochrome peaks therefore, are on average double the intensity of the SH3 peaks in spectrum from $(SH3)_2Cyt$, when compared to spectra from monomeric $(SH3)_2Cyt$. Thus whenever a cytochrome domain is observable, it is seen with one SH3 domain, with the other SH3 domain buried in the fibril.

The difference in absolute signal intensity between NMR spectra of monomeric $(SH3)_2Cyt$ and fibrillar $(SH3)_2Cyt$ could be partially explained if the relaxation properties of the two samples were very different. However, the discrepancy between the relative intensity of the SH3 and cytochrome peaks cannot be explained by this, and it would appear that the SH3 domains are half as visible to the NMR experiment as the cytochrome domains. The original construct possess two SH3 domains, so on average, where resonances are observed from the cytochrome, it is observed in conjunction with one of the two SH3 domains. This discrepancy in signal intensity between monomeric and fibrillar samples is further evidence for the resonances originating from fibrils. Three further NMR experiments are subsequently discussed.

2.4.8 Observed fibrillar resonances originate from a species diffusing significantly slower than the monomeric protein

The translational diffusion coefficient, D_T, of resonances observed in an NMR experiment can be estimated using Pulsed Field Gradient Stimulated Echo (PFGSE) NMR experiments (discussed in section 3.1. Where solution conditions are ideal and the geometry of the particle is spherical, the hydrodynamic radius of the measured species, R_H, can be inferred from the observed translational diffusion coefficient using the Einstein-Stokes equation, where k_B is the Boltzman constant, T is the thermodynamic temperature and η is the viscosity of the sample.

$$D_T = \frac{k_B T}{6\pi\eta R_H} \tag{2.1}$$

And so molecular size can be inferred from the translational diffusion coefficient. PFGSE NMR experiments result in the decay in signal intensity S_i from the intensity in the absence of gradients S_0 as a function of the strength of applied pulsed field gradients G, the duration of the applied gradient δ, a delay Δ known as the diffusion delay, and the gyromagnetic ratio of the observed nuclei γ. The following result was derived by Stejskal and Tanner [95] and is derived in chapter 3.

$$S_i = S_0 e^{-\alpha^2 \beta D_T \% G^2} \tag{2.2}$$

Where $\alpha = \gamma\delta G_{max}$ and $\beta = \Delta - \delta/3$. Typically experiments are recorded with constant δ and Δ, where the ratio $\frac{G}{G_{max}} = \%G$ is varied during the experiment. A plot of $\ln(S_i/S_0)/\alpha^2\beta$ against $\%G^2$ will yield a straight line of gradient $-D_T$, given that translational diffusion is the only process that leads to displacements of protein molecules in solution. Here, D_T will be referred to as an effective diffusion coefficient D_{eff}. Where D_{eff} is found not to vary with Δ, then $D_T = D_{\text{eff}}$. Where D_{eff} does vary with Δ, the situation is more complex than a single species diffusing freely in solution. One factor that leads to such a variance, is rotational diffusion. A theory incorporating rotational diffusion is presented in chapter 3.

Diffusion data acquired on 10 independent ca. 50μM $(SH3)_2Cyt$ samples (red) are compared to monomeric $(SH3)_2Cyt$ (blue) and human lysozyme (green) in figure 2.16A. Using techniques described in section 9.2, these data can be trans-

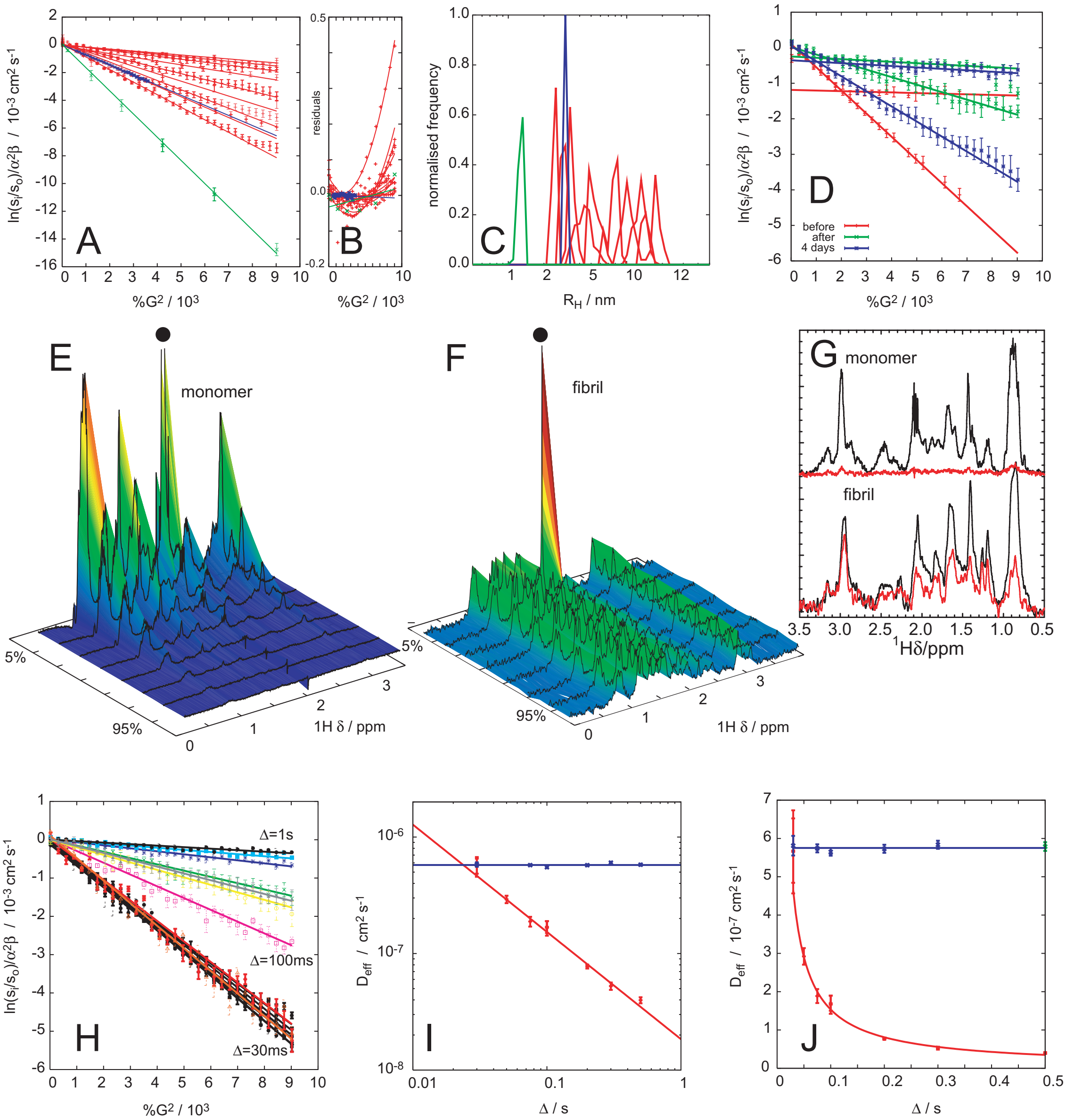


Figure 2.16: **A -** Integrated PFGSE signal intensity data with $\Delta = 0.1s$, plotted such that the gradient of the plot equals $-D_T$, for $(SH3)_2Cyt$ fibril sample (red), monomeric $(SH3)_2Cyt$ (blue) and human lysozyme (green). Data are fit to the Stejskal-Tanner model (residuals shown in B). **C -** NMR diffusion data from A expressed as R_H histograms, whose width reflects the experimental uncertainty of the R_H estimate (see section 9.2). R_H of fibrillar samples was found to vary between 2-30nm when analysed in this way. **D -** Integrated PFGSE signal intensity data from $(SH3)_2Cyt$ fibrils before (red), after (green) and 4 days after (blue) purification. Data are shown with Δ=1s (upper line for each time point) and with Δ=100ms (lower line for each time point). *E/F* - Series of PFGSE spectra (G=5% to 95% of the 32 G cm^{-1} maximum) where $\Delta = 0.5s$ and $\delta = 5.4ms$ for monomer (E) and for fibrils (F). **G -** Monomeric (top) and fibrillar (bottom) $(SH3)_2Cyt$ NMR spectra with $\Delta = 1s$ with 5% (black) and 50% of the maximum gradient. **H -** Diffusion plots for fibrillar and monomeric $(SH3)_2Cyt$ with varying diffusion delays. Data from monomeric $(SH3)_2Cyt$ (black), overlay and all lie on a single line. Those from fibrillar $(SH3)_2Cyt$ (diffusion delays are coloured individually) are found to lie on lines whose gradient depends on Δ. **I/J -** the effective diffusion coefficient D_{eff} as a function of Δ for the data presented in panel H, in double (I) and semi (J) logarithmic plots.

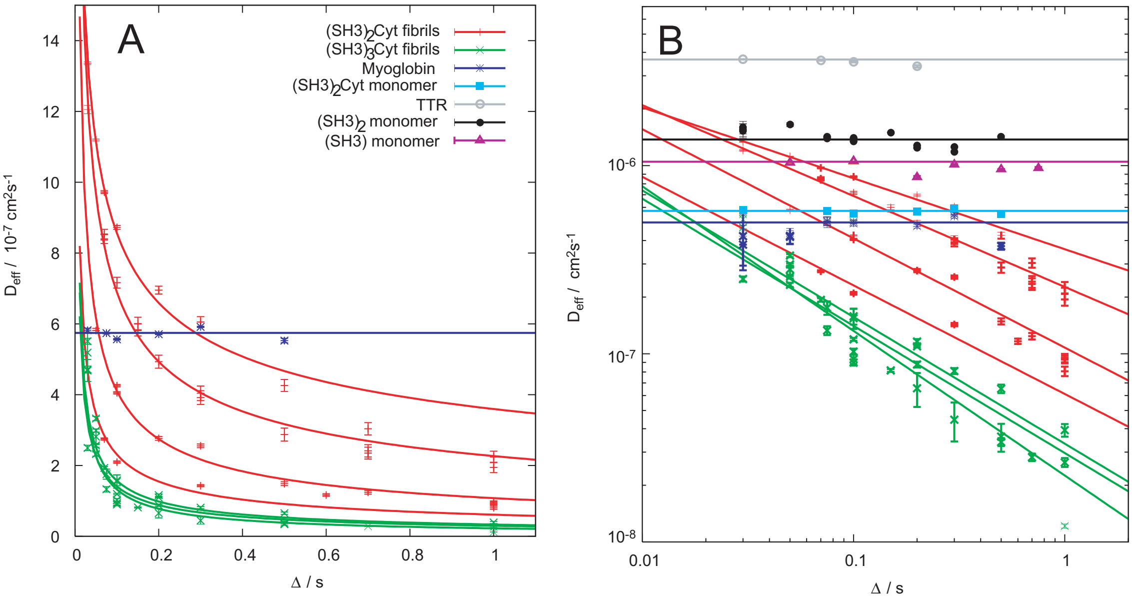


Figure 2.17: D_{eff} versus Δ for monomeric and fibrillar $(SH3)_2Cyt$, fibrillar $(SH3)_3Cyt$, myoglobin and an 11-mer peptide TTR$_{105-115}$. D_{eff} for monomeric protein is found to be invariant with Δ. D_{eff} of different preparations of fibrillar $(SH3)_2Cyt$ and $(SH3)_3Cyt$ are found to have diffusion coefficients that vary strongly with Δ. The data are presented with linear axes (A) and with logarithmic axes (B). The colour scheme defined in the legend of A applies to both A and B.

formed into R_H histograms whose width reflects the experimental uncertainty (figure 2.16C). Interestingly, the effective diffusion coefficients were found to be sample dependent. Interpreting these data using equation 2.1 yields R_H values between 2 and 20 nm. Comparing (figure 2.16 C). The diffusion coefficients were anomalously small in some samples, and very large in others.

PFGSE data with both $\Delta = 0.1s$ and $\Delta = 1s$ from a $(SH3)_2$Cyt sample both before and after purification were compared (figure 2.16D). The increase in the contribution from a larger sample before (red) and after (green) purification is particularly visible in the traces where Δ=1s. Series of PFGSE spectra with G varied between 0 and 32 G cm^{-1}for monomeric and fibrillar $(SH3)_2Cyt$ are shown in figure 2.16E and F, with $\Delta = 500ms$ and individual spectra are compared more closely with $\Delta = 1s$ in figure 2.16G. The spectra from fibrillar $(SH3)_2Cyt$ with $\Delta = 1s$ are well resolved, corresponding to a very slowly diffusion species. Qualitatively the spectra of both monomer and fibril samples are very similar, though PFGSE spectra data from monomeric $(SH3)_2Cyt$ are entirely attenuated by even modest gradients with this exceptionally long diffusion delays whereas those from fibrillar $(SH3)_2Cyt$ are not.

These observations led to the development of the theory discussed in chapter 3. The main prediction of the theory is that D_{eff} should vary with Δ. PFGSE intensity data for fibrillar and monomeric $(SH3)_2Cyt$ data are shown as a function of gradient strength in figure 2.16H with Δ values between 30ms and 1s. Data from monomeric $(SH3)_2Cyt$ were found to lie on a single line corresponding to a single species. In contrast, data from fibrillar $(SH3)_2Cyt$ did not, despite all plots fitting well to a function with a single exponential decay. Corresponding D_{eff} vs Δ plots are shown in figures 2.16I and J.

Exchange effects can contribute to NMR diffusion measurements. A resonance from two species of diffusion coefficients D_1 (mole fraction n) and D_2 (mole fraction $1-n$) in slow exchange would yield PFGSE data that decays according to the following:

$$\frac{1}{\alpha^2\beta}\ln\frac{S_i}{S_0} = \frac{1}{\alpha^2\beta}\ln\left(ne^{-\alpha^2\beta D_1\%G^2} + (1-n)e^{-\alpha^2\beta D_2\%G^2}\right) \tag{2.3}$$

Decaying signal intensity data would therefore be expected to decay according to a bi-exponential function. As shown in figure 2.16H, the decay takes on the form of mono-exponential decays, and so cannot be described by this model. In addition, fibrils of $(SH3)_2Cyt$ and $(SH3)_3Cyt$ were found not to dissociate after purification (section 2.4.6) suggesting exchange between monomeric and fibrillar species is unlikely explanation for the observed NMR behaviour.

Figures 2.17A and B show D_{eff} vs Δ data for a range of fibrillar $(SH3)_2Cyt$ and $(SH3)_3Cyt$ samples. These are found to be sample dependent, and are compared to monomeric $(SH3)_2Cyt$, myoglobin and the 15mer polypeptide $\text{TTR}_{105-115}$, whose D_{eff} does not vary with Δ.

The five following, apparently contradictory observations need to be reconciled.

1. The bulk of the material in pellet1 has been pelleted by ultracentrifugation and must therefore be large.

2. The fibrils do not dissociate to give monomeric protein on the experimental timescale (comparison between super2 and pellet1, section 2.4.6).

3. For short diffusion delays (30ms-100ms), the diffusion coefficient appears comparable to and often smaller than that of monomeric protein

4. For longer diffusion delays (500ms-1s), well resolved spectra are observed, corresponding slowly diffusing species.

These observations are reconciled by the model in chapter 3, where the observed NMR resonances are assumed to originate predominantly from the surface of amyloid fibrils. From the variance of D_{eff} with Δ, the length distribution of the underlying fibrils can be determined (section 3.3). Calculated length dis-

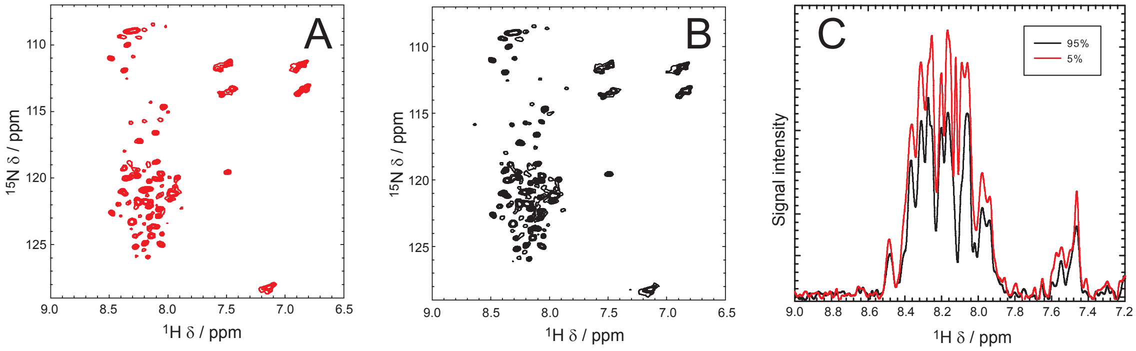


Figure 2.18: ^{1}5N DOSY-HMQC spectra from $(SH3)_2Cyt$ fibrils. **A -** HMQC spectrum at 95% of the maximum. **B -** HMQC spectrum at 5% of the maximum. **C -** The ^{1}H projection of the 5% and 95% DOSY-HMQC spectra.

tributions from NMR measurements are found to be in good accord with those measured using AFM and TEM. Two further NMR experiments support this surprising conclusion; the reduction in signal intensity as a function of bound heme (section 2.4.10) and the evolution of the PFGSE signal intensity as a function of time, as the fibrils grow (section 2.4.15).

2.4.9 SH3 and cytochrome residues diffuse at the same rate

To investigate whether there was sequence variation in the diffusion coefficient, an HMQC/PFG pulse sequence [96] was used (described in section 8.2.3). The principle is identical to that of the 1D pulse sequence. In this way, residue specific diffusion coefficients are obtained. The diffusion coefficients from cytochrome resonances were identical to those measured from the SH3 resonances, illustrated in figure 2.18.

2.4.10 Heme binding shows that only 50% of total cytochrome molecules are solvent exposed in fibrils

As introduced in section 2.1.2, metalloporphyrins have several useful spectroscopic properties. The two $\pi - \pi^*$ transitions in the electronic spectrum can be used to diagnose the binding state of the porphyrin, and the nature and oxidation state of the metal atom at the core of the porphyrin. The spectrum of iron porphyrin (heme) in its bound state is very different from that of its unbound state and so UV-Vis spectroscopy can be used to deduce the quantity of bound to unbound porphyrin in solution. UV-vis spectra from an oxidised heme titration with fibrillar $(SH3)_2Cyt$ are shown in 2.4.10A, showing an increase in the porphyrin binding, though the protein absorption at 280nm is unchanged. With heme addition, no change is seen in the FarUV CD spectrum of the fibrils, however heme binding is observed to yield a trough at 420nm associated with the

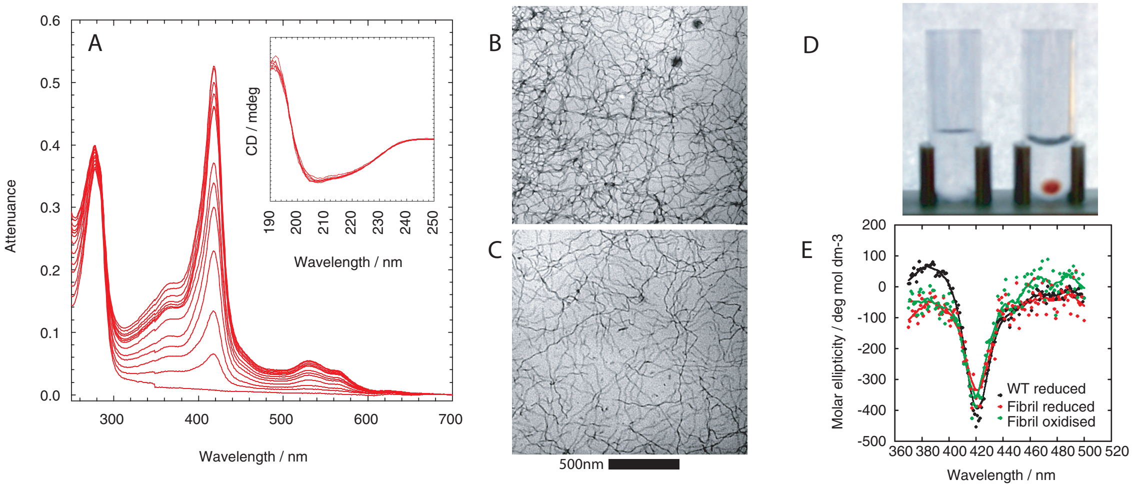


Figure 2.19: UV-Vis absorption increase with heme addition. **A -** Data from a sample titration, with UV-Vis data and concurrently recorded CD spectra shown in the inset. The morphology of the fibrils is not observed to change with heme addition. **B -** TEM micrograph at 25% heme binding. **C -** TEM micrograph before heme addition. **D -** A photograph of pelleted holo-$(SH3)_2Cyt$ fibrils. **E -** Although no change is observed in the far-UV CD on heme addition, heme binding does give rise to a peak in the CD at 420nm on heme addition.

chirality of the binding site of the metal, shown in figure 2.4.10E. The morphology of the fibrils is not visibly changed by heme binding (figures 2.4.10B and C). When a heme bound sample is repurified by ultracentrifugation, the red material was found to accumulate in the pellet, as shown in figure 2.4.10D. Monomeric holo-Cyt does not pellet under these conditions.

The extinction coefficient of oxidised wild type cytochrome b_{562} (figure 2.20A, red) is compared to that of fibrillar $(SH3)_2Cyt$ (figure 2.20A, black). There is no significant difference between the two. The difference between an oxidised spectrum and a reduced spectrum is shown in figure 2.20B. These too are found to be essentially identical, and represent the first demonstration of the possibility of electrons to be reversibly stored in an amyloid fibril. The cytochromes also bind zinc porphyrin non-covalently, producing a fluorescent species. Normalised emission spectrum of zinc bound $(SH3)_2Cyt$ fibrils is compared to that of wild type Cyt in figure 2.20A (inset). The two are found to be identical.

$(SH3)_2Cyt$ and $(SH3)_3Cyt$ fibrils however, have a lower porphyrin binding capacity than Cyt. Difference spectra, normalised at 420nm between an equivalence of 0-1 porphyrin:protein are shown in figure 2.20B. The spectra for Cyt (red) are identical, whereas higher equivalences of heme in fibrillar $(SH3)_2Cyt$ lead to an increased contribution at 400 nm, corresponding to an increasing proportion of unbound heme. Given the extinction coefficients are well known, and the protein concentration can be quantified by measuring the absorbance at 280nm, heme binding can be followed stoichiometrically. The results of eight such bind-

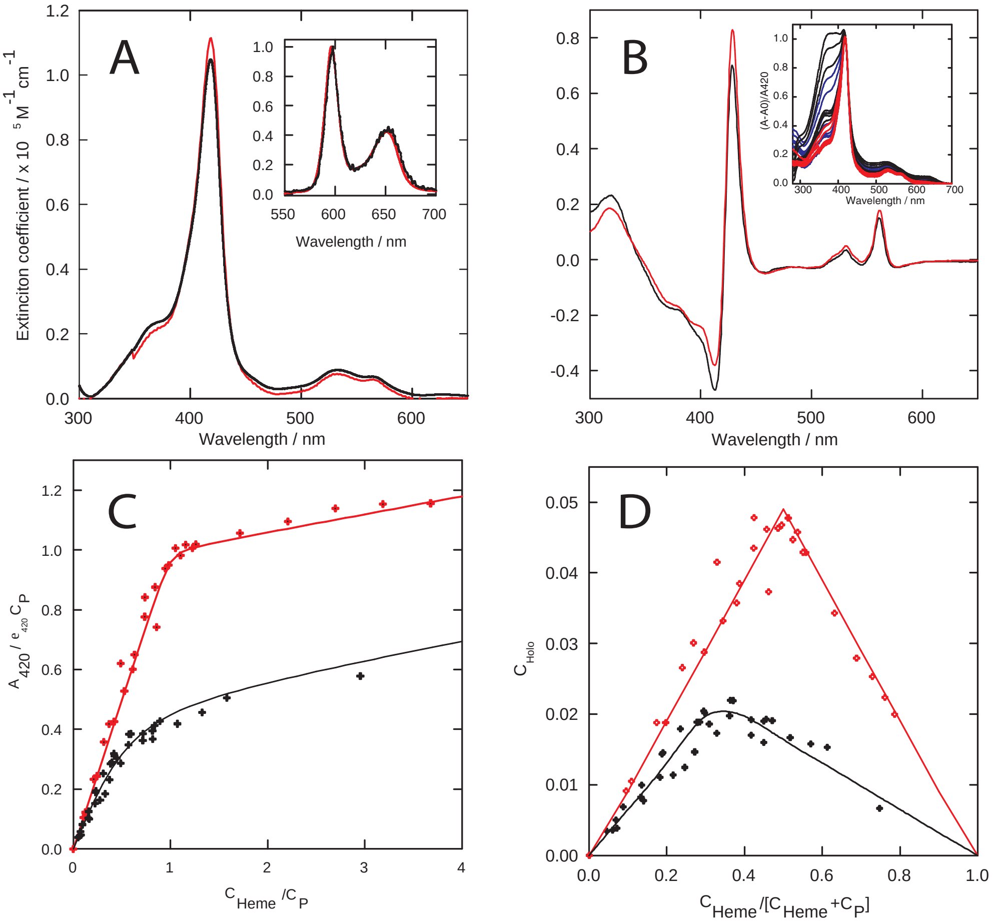


Figure 2.20: UV-Vis comparison between $(SH3)_2Cyt$ fibril sample and wild type Cyt. Monomeric protein is shown in red, and $(SH3)_2Cyt$ fibril samples are shown in black. **A -** The extinction coefficient of holo-cytochrome bound to iron PPIX at pH 3. The inset shows the normalised emission spectrum of cytochrome bound to zinc PPIX, excited at 430nm. **B -** The oxidised - reduced difference spectrum. Inset B shows sample data that alludes to a reduced population of binding. The normalised difference plots shown are from between 0 and 1 molar equivalents of heme to protein. After the 0.5 mark is passed, the contribution of free heme to the spectra increases for the $(SH3)_2Cyt$ fibril samples. **C/D -** Quantitative representations of heme titration data. By following the appearance of the peak in the UV-Vis at 418nm, the quantity of bound protein can be quantified. C shows a hyperbolic plot and D shows a Job plot of the data, both illustrating only half of the available cytochrome sites are accessible to heme.

ing assays are shown in figure 2.20C and D, and analysed as described in section 2.8.2.

Monomeric $(SH3)_2Cyt$ at pH 7 binds heme with a nM dissociation constant, and reaches saturation binding at a 1:1 heme:protein ratio. As more heme is added, the observed increase in absorbance at 420nm is due to residual absorbance of free heme. The binding affinity for $(SH3)_2Cyt$ fibrils was estimated to be 180 $\pm$ 30 μM, significantly lower than that of monomeric $(SH3)_2Cyt$. Remarkably however, both $(SH3)_2Cyt$ and $(SH3)_3Cyt$ fibril samples reach saturation at a bound heme concentration 50±4% of that of the total cytochrome known to be present (section 2.8.2).

That is, 50% of the available cytochrome molecules are surface exposed and are able to refold and bind heme but 50% cannot. This is visualised in the hyperbolic plot and the Job plot in figures 2.20C and D. From denaturing the fibrils in SDS then running PAGE gels, it can be shown that the only component of the fibril is full length protein (figure 8.1). The primary motivation for the creation of $(SH3)_3Cyt$ was to investigate the heme binding properties and the point at which it saturates when binding to free heme. The question of the location of the non-solvent exposed 50% of cytochrome residues and its significant is addressed in section 2.6.3.

2.4.11 Electrons can be reversibly transferred between $(SH3)_2Cyt$ fibrils and species in solution

The midpoint potential E_0 of a redox system can be determined using a technique known as potentiometric titration. By titrating in redox reagents and measuring the potential of the solution relative to a standard E_{exp}, electronic spectra are acquired as a function of the potential of the sample. Deconvolving the spectra into oxidised and reduced components then allows a titration curve that correlates the ratio of oxidised to reduced species with the measured potential. By fitting the data to the Nernst equation, $E_{exp} = E_0 + nF \ln\left(\frac{[Cyt]_{red}}{[Cyt]_{ox}}\right)$ where F is Faraday's constant, the midpoint potential E_0 and n, the number of electrons involved in the transfer can be deduced.

A fibrillar $(SH3)_2Cyt$ sample and a Cyt sample at pH 3 both gave identical values for E_0 at pH 3 (100 $\pm$ 45 mV), and showed the process to be a 2 electron process (figure 2.21). The cytochrome molecules on the surface of $(SH3)_2Cyt$ fibrils are therefore electrochemically identical to those of Cyt and electrons can be transferred reversibly between $(SH3)_2Cyt$ fibrils and species free in solution.

At pH 7 E_0 was determined to be 206 $\pm$ 10mV, consistent with the published value [97, 82]. It was not possible to measure a comparable value for $(SH3)_2Cyt$

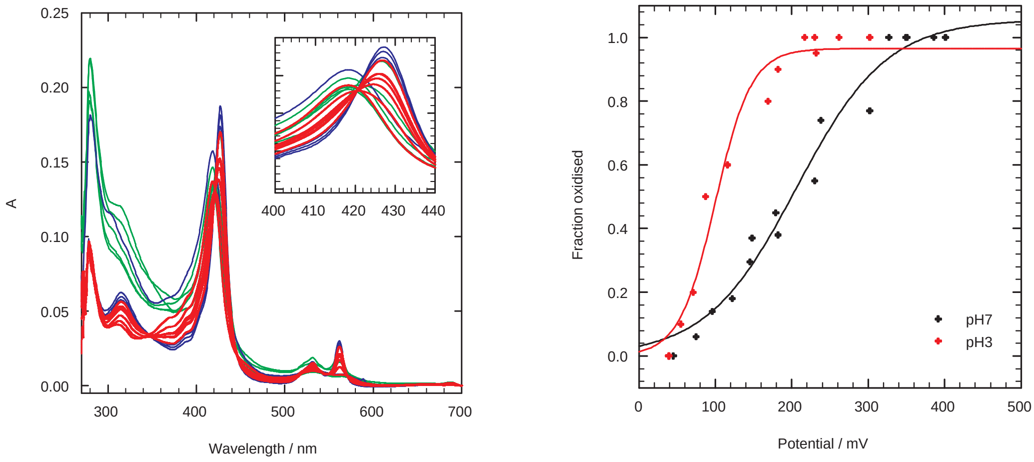


Figure 2.21: UV-Vis spectra of Cyt obtained during potentiometric titration at pH 7. Oxidised Cyt was reduced with DTT (red series), oxidised with ferricyanide (blue series), then re-reduced with DTT (blue series). The data are fitted to a basis set of oxidised and reduced spectra. DTT and ferricyanide are similarly subtracted to estimate the ratio of oxidised to reduced cytochrome residues. Plotting the ratio of oxidised to reduced cytochrome against the electro-potential yields the curves shown on the right which can be fit to the Nernst equation to yield the mid point potential.

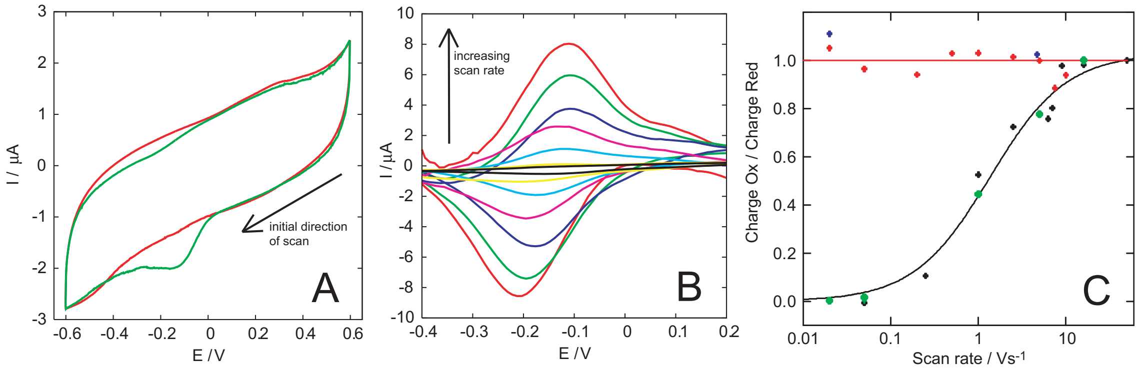


Figure 2.22: Thin film cyclic voltammetry curves. **A -** Voltammagrams from $(SH3)_2Cyt$ fibrils (green) and the electrode baseline (red) at a scan rate of 0.02 V s^{-1}. **B -** Baselined voltammagrams from $(SH3)_2Cyt$ fibrils for scan rates between 0.01 V s^{-1} and 16 V s^{-1}. **C -**The ratio of the oxidised to the reduced charge from voltammagrams as a function of the scan rate for $(SH3)_2Cyt$ fibrils pH 3 (black, E_0 100 mv), Cyt pH 3 (green, E_0 100 mV), Cyt pH 7 (blue, E_0 205 mV) and free heme (red, E_0 33 mV).

fibrils as they are unstable at pH 7 (section 2.4.14). The electrochemical behaviour of Cyt has been studied as a function of pH, down to pH 4. Between pH 7 and pH 4, the redox process involves the transfer of one electron and E_0 increases with decreasing pH. At pH 3 the E_0 measured here is significantly lower than that measured at pH 4 and is a two electron process. The reasons for this are unclear at present.

Cyclic voltammagrams from samples fixed to an electrode on a thin film were recorded for $(SH3)_2Cyt$ fibrils at pH 3 and compared to those from Cyt at pH 3 and pH 7 and free heme at pH 7 (figure 2.21) using methodology described elsewhere [98]. For scan rates >1 V s^{-1}, all species showed reversible electron transfer. Interestingly, below scan rates of 1 V s^{-1}, $(SH3)_2Cyt$ fibrils and Cyt at pH 3 were readily reduced, no evidence for re-oxidation was observed on the oxidising scan (figure 2.21B). The reasons for this are not clear, though it is likely to be related to surprisingly low mid-point potential observed for both species at pH 3. In conclusion therefore, the electrochemical behaviour of $(SH3)_2Cyt$ fibrils and Cyt are identical.

2.4.12 Cytochrome molecules are within conduction distance

The number of iron atoms expected per nm of fibril can be estimated using a geometric argument. The cores of packed globular proteins, in general, have a relatively constant density 1.37×10^3kg m^{-3}, being derived from the partial specific volume, $\overline{v} = 1/\rho = 0.73$ cm^{-3} g^{-1}. Assuming the fibrils have this density, and taking the fibrils as cylinders, the mass M per unit length L of a fibril of radius r can be estimated. Given the fibril is a polymer of monomers of a known molecular mass M_w, this can be related to the number of monomers N per unit length, where N_a is Avogadro's constant:

$$\begin{aligned} \frac{M}{L} &= \frac{\pi r^2}{\overline{v}} \\ \frac{N}{L} &= \frac{N_a \pi r^2}{M_w \overline{v}} \end{aligned} \tag{2.4}$$

This equation can be used to estimate the mass of a fibril of known radius and length. A 10μm fibril of radius 6 nm will have a mass of ca. 1 GDa. On the surface of the fibril, only half of the cytochromes are able to bind heme. The length of fibril occupied on average by 1 iron atom in nm can be estimated to be;

$$L_1 = \frac{2 M_w \overline{v}}{N_a \pi r^2} \times 10^9 \tag{2.5}$$

The molecular mass of a single monomer in a cytochrome fibril is 33 kDa and so the number of iron atoms becomes purely a function of radius with this model. The radius measured by TEM is 6 nm. With this model, L_1 = 0.71 nm. The Marcus limit for domain separation for rapid electron transfer is 1.4nm [23](section 1.3). Thus this model predicts the spacing of domains to be below the resolution of both AFM and TEM, and that the domains are within the physical limits imposed on them for electron transport.

2.4.13 NMR signal intensity from $(SH3)_2Cyt$ fibrils drops to zero when 50% of cytochromes bound to heme

The changes in the NMR spectra of fibrillar $(SH3)_2Cyt$ were investigated as a function of heme addition, in conjunction with UV-Vis spectroscopy (figure 2.23). The methodology of this titration is discussed in more detail in section 8.2.3. As heme is added to a $(SH3)_2Cyt$ fibril sample, no new peaks are observed in ^{15}N-HSQC spectra (figure 2.23D). This is surprising as resonances due to folded $holo-cyt$ would have been expected. As heme is added, all peaks are observed to decrease linearly in intensity (figures 2.23A and C). No NMR signal is observed when a $(SH3)_2Cyt$ fibril sample is at its 50% bound saturation point. When individual peaks are analysed it is found that they all decrease at the same rate, thus heme binding effects the SH3 resonances as well as the Cyt resonances equally (section 2.4.7).

This loss in signal intensity is not caused by free oxidised heme acting as a paramagnetic relaxant (figure 2.24B), where spectra of holo-Cyt are compared as a function of added heme. Monomeric $(SH3)_2Cyt$ at pH 3 can be taken to 100% binding. In this situation, resonances are observed from both folded Cyt and SH3 in ^{15}N-HSQC spectra (figure 2.8B). The fibril morphology is not observed to change in binding heme between 0 and 50% stoichiometry (figure 2.4.10 B and C).

Folded cytochrome is observed in $(SH3)_2Cyt$ fibril samples using electronic spectroscopy. However, it is not observed by solution-state NMR. 50% of the cytochrome domains are solvent exposed and ca. 20% can be detected using solution state NMR. Refolding the cytochrome removes the conformational flexibility required to reduce the linewidths sufficiently for detection by solution-state NMR. Interestingly, refolding the cytochrome has the consequence of also reducing the conformational freedom of the SH3 domain it is attached to, causing this to also be lost from the spectrum. An analysis of the dynamics required to narrow the solution-state NMR resonances of residues attached to larger bodies is the subject of chapter 6.

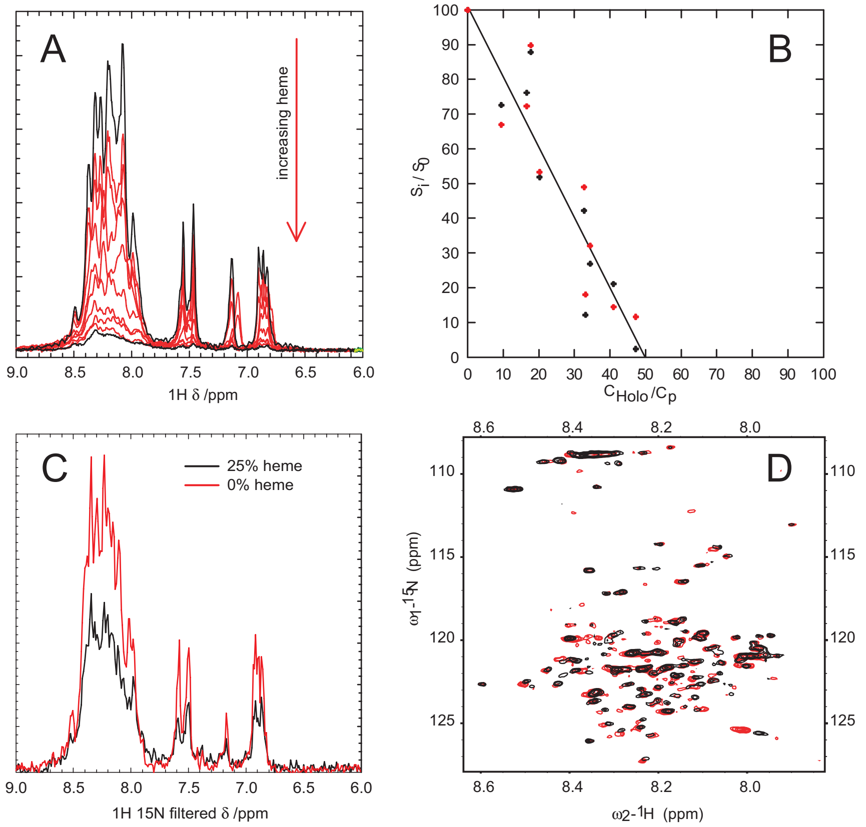


Figure 2.23: The effect of heme addition on NMR spectra of fibrillar $(SH3)_2Cyt$. **A -** As heme is added to a fibril sample, the NMR intensity is observed to decrease. **B -** The signal intensity decreases linearly with bound protein. In red are data from 1D spectra, and black shows data from 1D HSQC spectra, from 3 independent titrations. The intensity is observed to decrease to zero at 50% heme binding. Panel A shows the raw spectra from one such titration. **C/D -** By comparing 2D data at 0 (red) and 25% (black) heme loadings, the effects of heme binding on individual peaks can be followed. The intensities of individual peaks are found to decrease at the same rate.

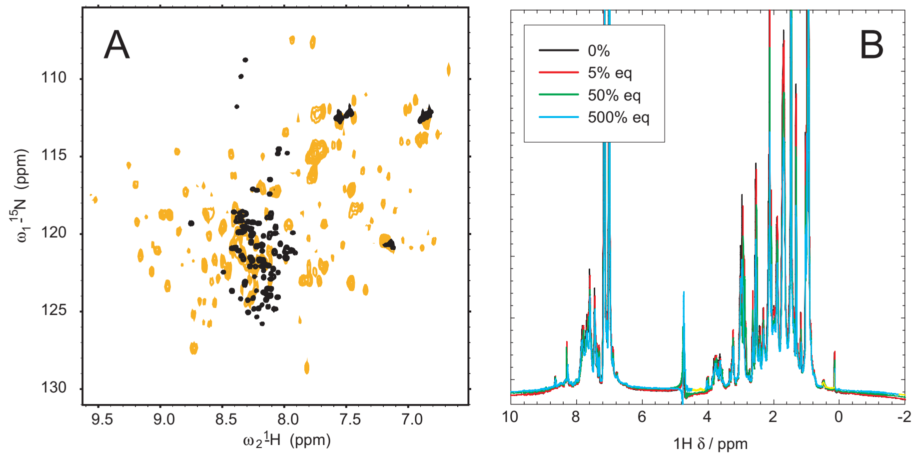


Figure 2.24: **A -** ^{15}N-HSQC spectra of $apo-Cyt$ at pH 3 (black) and $holo-Cyt$ (yellow). $Apo-Cyt$ is unfolded (proton dispersion 0.9 ppm) whereas $holo-Cyt$ is folded (proton dispersion 3 ppm). **B -** Free heme does not act as a paramagnetic relaxant at pH 3. 1D ^{1}H spectra of $holo - cyt$ at pH 3 are shown with successively increasing concentrations of free heme (5%, 50% and 500% molar ratios of free heme to protein). Protein resonances are not observed to decrease in intensity, even at exceedingly high concentrations of free heme.

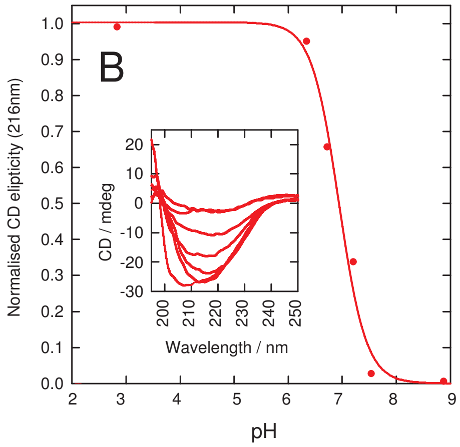


Figure 2.25: pH dependence of the CD spectrum of fibrillar $(SH3)_2Cyt$. The CD spectra decrease to zero as the material aggregates and sinks to the bottom of the cuvette. After several days at pH 7, native CD spectra are returned.

2.4.14 $(SH3)_2Cyt$ and $(SH3)_3Cyt$ fibril formation is reversible

When fibrils of $(SH3)_2Cyt$ or $(SH3)_3Cyt$ are taken above pH 4 and incubated at room temperature for several days, the fibrils dissociate, monomeric material returns to solution and CD scans consistent with native folded protein are observed (figure 2.13). CD data for the decomposition of $(SH3)_2Cyt$ fibrils are shown in figure 2.25. Above pH 4 the β sheet signal initially decays to zero as material visibly precipitates from solution. After several days at elevated pH, native scans are observed to return. At pH 3 however, fibrils are the most stable state. In addition, a similar effect is observed with free heme. When a large excess (ca.100 times) of heme is added to the solution at pH 3, fibrils decompose on a timescale of months. This observation is discussed in the context of the 50% rule (section 2.6.3). On moving to conditions that favour the folding of the cytochrome, driven by the re-folding free energy of the cytochrome, the 'hidden' population of cytochrome molecules leaves the fibril. Subsequently, the fibrils decompose.

2.4.15 Evolution of NMR intensity and fibril elongation observed during fibril formation

$(SH3)_2Cyt$ fibril formation was followed by solution-state NMR. Three types of spectra were followed with time (figure 2.26). These are listed below, with the type of information they would be expected to provide. These experiments were correlated quantitatively with the types of species as observed by TEM in figure 2.27. At all times, diffusion experiments were ran on free water in the sample, showing the viscosity of the sample remained essentially constant at all times. It is important to note that no hydrolysis was observed in the reaction vessel over time, as monitored by SDS-PAGE.

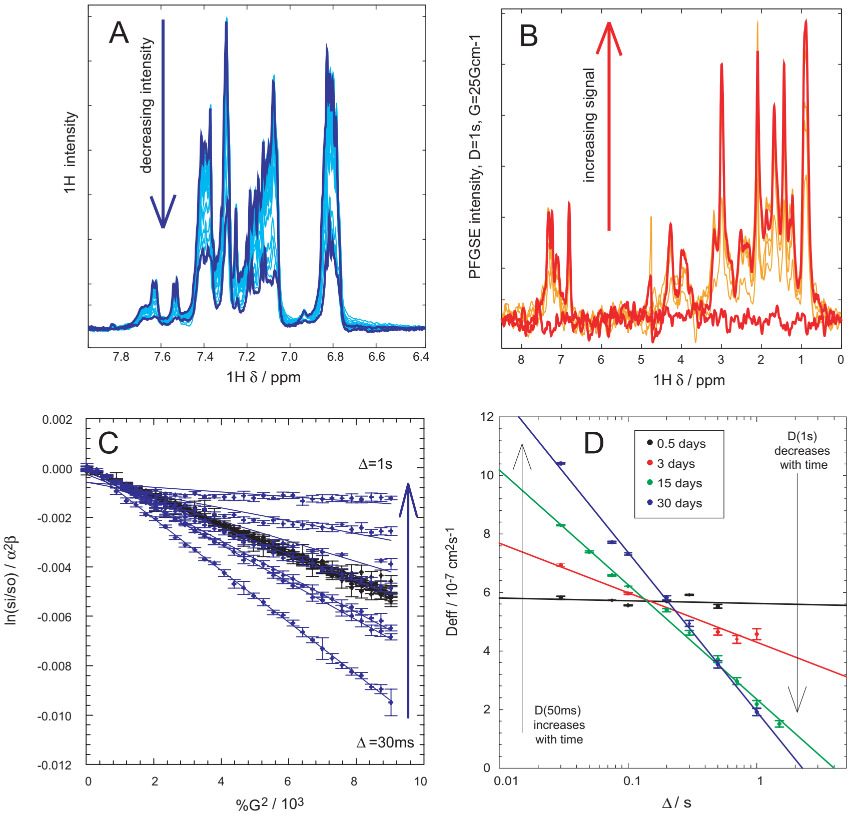


Figure 2.26: $(SH3)_2Cyt$ fibril formation followed by NMR as a function of time. **A -** Evolution of 1D^{1}H spectra. Intensity is observed to decrease in time, with no obvious changes in the spectral envelope as it does so. This is shown quantitatively in figure 2.27B. **B -** evolution of 1D^{1}H diffusion filtered spectra, with 25Gcm^{-1} gradient strength and diffusion delay of 1s. Spectra are observed appear suddenly. **C -** Sample diffusion data acquired at two time points. Initially, the gradient of the plots acquired at different values of Δ are identical (black). 30 days later, this is not the case (blue). These data are followed with time in figure 2.27C. **D -** Fitting the diffusion data to a single exponential allows the apparent diffusion coefficient to be followed as a function of the diffusion delay Δ. The diffusion coefficient measured at the larger times decreases with time, whereas the diffusion coefficient at the shorter times increases. When the diffusion experiment is performed with a diffusion delay of 100ms, no change is observed in the apparent diffusion coefficient. These data are followed with time in figure 2.27D. Phenomenologically these data can be fitted to a function of the form $D_{\text{eff}} = A\ln(-\Delta) + \lim_{\Delta=1} D_{\text{eff}}$. A and $\lim_{\Delta=1} D_{\text{eff}}$ can therefore be followed with time.

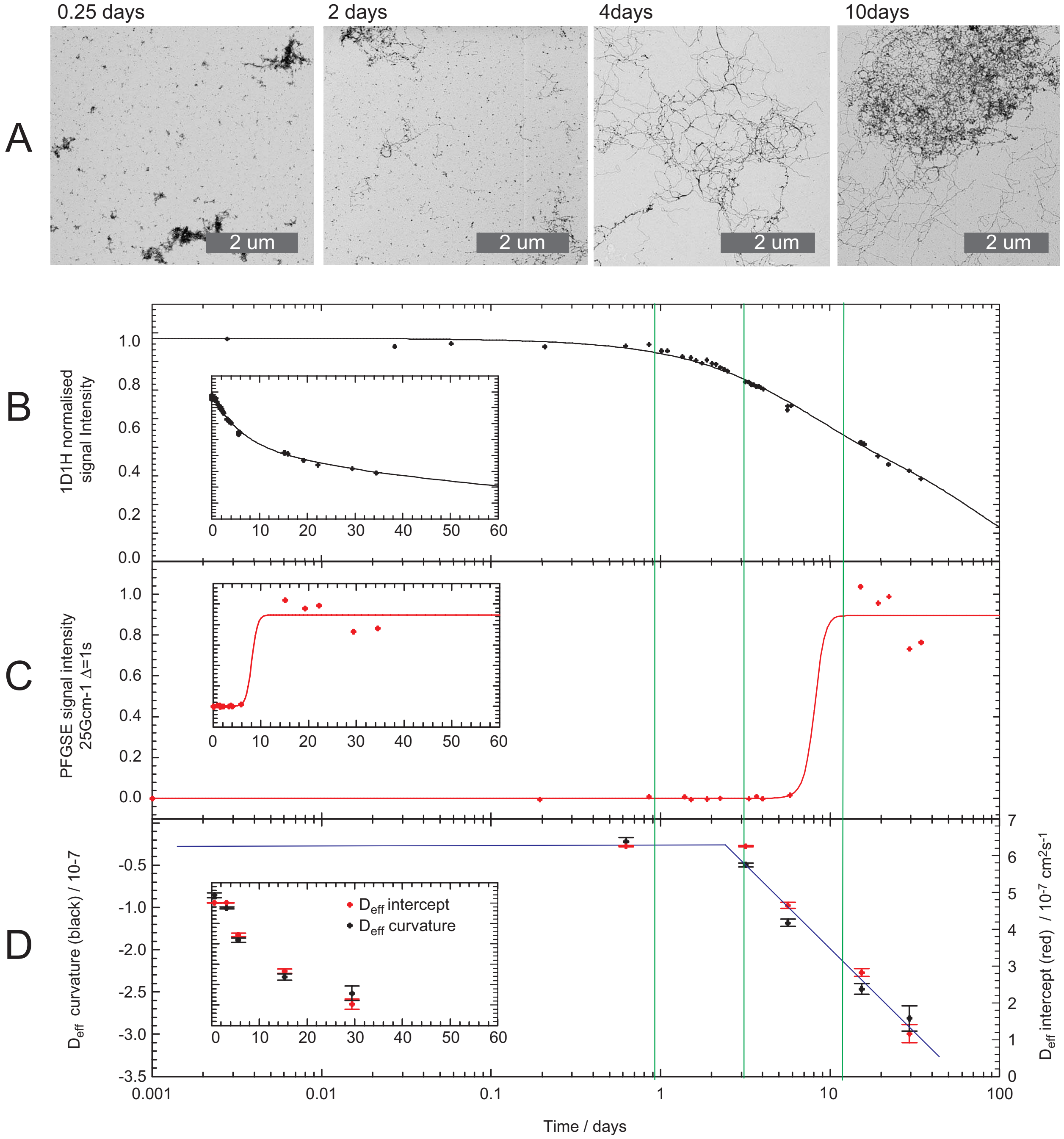


Figure 2.27: $(SH3)_2Cyt$ fibril formation followed by NMR and TEM as a function of time. **A -** TEM images taken at the specified time points to illustrate the dominant species present at the times specified. B, C and D are quantitative integrals of the NMR spectra shown in 2.26, shown against a logarithmic time scale in the main plots, and against a linear time scale in the insets. The green lines mark phase boundaries. **B -** Evolution of $1D^1H$ spectra with time. Intensity is observed to decrease bi-exponentially. A lag phase of one day precedes the exponential loss of signal with a half life of 6.3 days. **C -** Evolution of $1D^1H$ acquired through $25Gcm^{-1}$ with a 1s diffusion delay. Signals observed therefore originate from species diffusing very slowly. These are observed to appear in high concentration at day 10. This is well after about 4 days. This is well after 1D signal has been observed to start to decrease, in A. **D -** Evolution of the diffusion profiles shown in 2.26D. Fitting the diffusion data to $D_{\text{eff}} = A\ln(-\Delta) + \lim_{\Delta=1} D_{\text{eff}}$ allows A and $\lim_{\Delta=1} D_{\text{eff}}$ to be followed with time. These plots reflect the evolution of the size of the species observed. Thus despite the trend in C of the intensity of the large species decreasing with time, the observed signals still originate from species that are getting larger with time.

1D ^{1}H spectra The signal intensity is expected to decrease with time as monomeric material leaves solution and joins the fibrils, as observed in other systems [60, 80] (figures 2.26A and 2.27B).

PFGSE spectra, 25 G cm^{-1}, Δ=1s Resonances observed under these condition can originate only from very slowly diffusing species. If the effective R_H of the species under study were 9 nm, then this spectra would be 95% smaller than a spectra acquired with G = 2 G cm^{-1}, and so only barely detectable. Only large species (effective R_H of 40 nm) are observed under such conditions and so these experiments were used to estimate the point at which larger species emerge (figures 2.26B and 2.27C).

PFGSE spectra, G varied from 0-100%G_{max}, Δ varied between 30ms-1s Acquiring a full data set of ca. 32 gradient points for 5-7 diffusion delays takes approximately 10 hours. These data however, can be used to calculate the mean length distribution of the fibrils over this time period. The sample was prepared in D_2O solution such that weak signal in intensity in the PFGSE experiments would not be obscured by baseline distortions induced by poor water suppression (figures 2.26C and D). In all cases the plots of D_{eff} against Δ can be fitted to a function of the form $D_{\text{eff}} = A\ln(-\Delta) + \lim_{\Delta=1} D_{\text{eff}}$. The coefficients A, the curvature, and $\lim_{\Delta=1} D_{\text{eff}}$, the intercept are shown evolving with time in figure 2.27D.

The 1D intensity was found to decrease after a lag phase of one day (figure 2.27B). Over this initial period, few aggregates were seen in the TEM. Signal intensity was then observed to decay, with the decay fitting well to a function with a single exponential decay of half-life 6.3 days. During this period of initial decrease, larger fibrillar aggregates were observed to appear in the TEM. As for $(SH3)_2$ (section 2.4), these species are not NMR observable due to the slow tumbling and lack of flexible regions. PFGSE experiments show that all observed signals originate from small species over these time periods (figure 2.27D).

After three days the dominant species observed in the TEM was fibrils. At this point, significant signal intensity was observed in the Δ=1s, 25 G cm^{-1} PFGSE experiments (figure 2.27C) and the effective diffusion coefficients were observed to vary with the diffusion delay (figure 2.27D) reflected by the trend in both the intercept and the curvature parameters.

Interestingly, as shown in figure 2.26D and inferred by the increasing curvature term in figure 2.27D, although D_{eff} was observed to decrease for longer diffusion delays, for short diffusion delays it was observed to increase with time. This is consistent with the effects of rotational diffusion; for short diffusion delays, rotational diffusion leads to faster apparent diffusion and for longer delays, signals from the centre of the fibrils diffuse more slowly reflecting an increase in the translational diffusion as the fibrils become longer. At intermediate diffusion times around 100ms, the delay commonly used to study protein systems, very little variation was observed over time. A PFGSE experiment using conventional delays could not observe this trend.

This experiment allows insight into the fibril formation and elongation mechanism. Of particular interest is the lag between the emergence of the fibrillar species in the TEM and the decay of the 1D signal intensity. The larger fibrillar aggregates that cannot be observed by solution-state NMR clearly form before the fibrils. There would appear to be two nucleated processes that occur during the formation of $(SH3)_2Cyt$ fibrils. The first of duration ca. 1 day requires nucleation of the fibrillar aggregates. The second requires internal rearrangements within the fibrillar aggregates to form a species capable of seeding fibril elongation a process observed to take ca. 3 days. This is consistent to the conclusions of fibril formation from the $(SH3)_2$ unit [80], and those of Lund et al. [99] from ESR experiments on a 7-mer polypeptide derived from prion protein.

Thus the time dependent concentrations of three types of species can be distinguished using experiments of this type, monomeric protein, fibrillar aggregates and amyloid fibrils. This experiment in itself validates the diffusion methodology - the time dependence of the diffusion coefficients is correlated with elongating fibrils. Experiments of this type could be used to demonstrate whether the fibrillar aggregates are the fibril forming nucleus or whether they are an off pathway species removed in the course of fibril formation. These results are particularly exciting in light of recent observations that imply the toxic species of amyloid diseases are the fibrillar aggregates [58].

2.5 On-going work

2.5.1 Doping experiments

In order to test further universality of the 50% rule, experiments to co-aggregate $(SH3)_2Cyt$ with $(SH3)_2$ were performed. These were attempted over a range of rations from 1:1 to 1:100. In all cases, mixtures of fibrillar aggregates and fibrils were formed. When fibrils were purified by ultracentrifugation, it was not possible to estimate the relative proportions of $(SH3)_2$ to $(SH3)_2Cyt$ with sufficient accuracy to calculate the ratio of exposed residues.

2.5.2 Limited proteolysis

Limited proteolysis experiments have shown, in other systems, populations of protected full length protein in amyloid fibril forming systems. This population has been estimated to be approximately 50% in two studies [48],(unpublished, Gras *et al.*). The only readily available protease functional at pH 3 is pepsin. Incubating monomeric $(SH3)_2Cyt$ at pH 3 results in complete and rapid proteolysis. Incubation of fibrillar $(SH3)_2Cyt$ with pepsin yielded a complex intermediate with 7 heavily staining bands in an SDS gel, that persist for several hours on incubation. A similar result is seen when $(SH3)_2$ is incubated with pepsin at pH 3. The results from these experiments are currently inconclusive and would require proteolysis combined with HPLC to carefully identify the generated fragments.

2.5.3 Assignment of spectra from $(SH3)_2Cyt$ fibrils

To further investigate the intensity trends described in section 2.4.7 and to allow relaxation experiment to investigate the dynamics of the displayed groups on the fibril, 3D assignment spectra (HNCACO, HNCO, CBCA(CO)NH and HNCACB) were acquired on doubly labelled (^{13}C, ^{15}N) $(SH3)_2Cyt$ fibril samples, prepared as described in section 8.1.2. The spectral overlap, particularly in the carbon dimension make this an exceedingly challenging assignment, allowing few clear assignments from inspection of the peaklist.

To simplify the assignment, an auto-assignment program was developed, described in section 9.3. Once a peaklist has been compiled, the program looks for all possible i+1 and i-1 residues from the peaklist within user specified tolerances within the spectrum. The Cα and Cβ chemical shifts are then compared to random coil shifts. The allowed amino acid combinations from the spectra are then compared to the known sequence of the protein in question and impossible combinations are rejected. The program then looks for all the possible combinations by walking to the next i+1 or i-1, and compares those to the sequence. This continues until the walk reaches a unique solution. This process is repeated by the program reducing the tolerances until it finds a unique assignment. The program also takes additional constraints to assist with the search. For example, as shown in figure 2.4.7, residues associated from the cytochrome can be distinguished from residues associated with the SH3, which further narrows the possible assignments. At the time of writing, ca. 50% of the residues were assigned using the program, approximately equally spaced over the SH3 and cytochrome domains.

The primary motivation of assigning resonances from $(SH3)_2Cyt$ fibrils would be for the acquisition of residue specific relaxation information. Information of

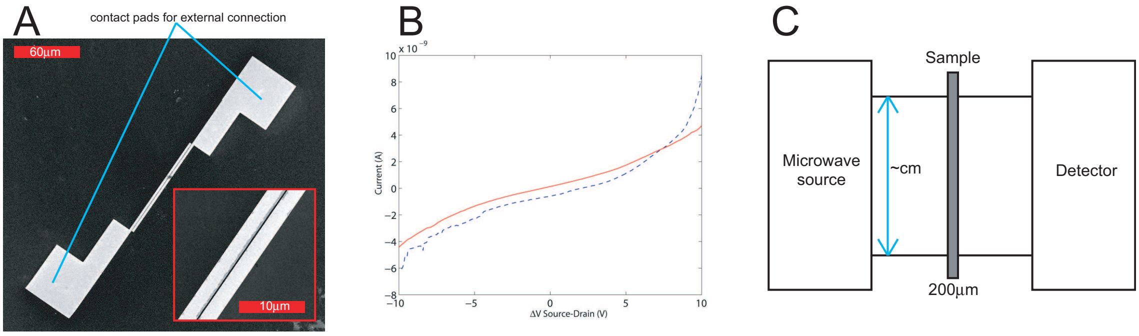


Figure 2.28: **A -** SEM image of electrodes produced using electron lithography for conduction measurements. The 3 μm wide, 50 nm high gold electrodes, connected to 60 μm square for connection to the measurement electronics were fabricated on a mica surface. **B -** Current versus applied voltage (IV) curves for $(SH3)_2$ fibrils (red) and $holo-(SH3)_2Cyt$ fibrils dried in air on the electrodes shown in panel A. There are no significant differences between the curves. **C -** A schematic of the apparatus used for microwave impedance measurements. Microwave radiation is emitted from a source, and allowed to pass through a sample. The loss of microwave intensity is related to the motion of free electrons in a sample.

this type could provide insight into the dynamics of flexible chains that lead to sharp resonances despite their being attached to a much larger body.

2.5.4 Can $(SH3)_2Cyt$ fibrils conduct?

The primary goal of this project is to produce conducting fibrils. The ideal experiment would place a single fibril between two electrodes to measure the current under an applied potential. Cyclic voltammetry studies have shown that it is possible for cytochrome domains can exchange electrons with solid state electrodes [97]. Using this knowledge, one experimental approach is simply to place two electrodes approximately 200 nm apart, deposit fibrils over their surface them and connect them to electronics of sufficient sensitivity to determine the current as a function of the applied potential.

Using standard techniques, gold electrodes such as those shown in figure 2.5.4 were fabricated on a mica surface using electron lithography. The precision of the process ensured that there was no conducting material between the electrodes, spaced 200 nm apart. Samples of $(SH3)_2$ and $(SH3)_2Cyt$ fibrils were deposited on the electrodes and allowed to dry and current vs voltage curves (IV curves) were acquired (figure 2.5.4). No significant difference between the two curves were observed. A current of 5 nA at 10V corresponds to a material with resistivity of 6.0×10^5 Ohm cm consistent with a very good insulator, 10 orders of magnitude more resistive than a metal such as copper. The sensitivity of the measurement was found to be ca. 10 pA, three orders of magnitude smaller than the currents predicted for a single fibril in section 2.1.2.

In addition, microwave impedance measurements were used to assess the conductivity of $holo-(SH3)_2Cyt$ fibrils. This technique is used routinely in microelectronics to non-invasively characterise the conductivity of thin films [100]. In a microwave impedance measurement, microwaves are passed through a dried sample and the loss in power is recorded. The charge on free electrons interacts with the radiation, which leads to collective circular motion. The kinetic energy of this motion is then dissipated, for example through electrons coupling to phonons or through isotropic radiative processes. The more energy dissipated, the more kinetic energy is passed to the free electrons from the microwave beam and hence the greater the number or mobility of free electrons, and hence the conductivity of the sample. Samples are prepared by drying samples to a substrate almost transparent to microwaves, such as glass. A 10 nm gold film gave a 95% attenuation of the microwaves. No difference between the loss in microwave intensity of a clean glass substrate and a substrate prepared with a $holo-(SH3)_2Cyt$ fibril film was detected. It is not clear however that even if individual cytochrome molecules could find appropriate interaction conformations on the surface of the fibril, that electrons would would have sufficient mobility in the plane of the sample to undergo the motion required to strongly interact with the microwaves.

Both experiments showed that under conditions where $(SH3)_2Cyt$ fibrils are dried onto a surface, they are not conducting, using the most sensitive probes developed for microelectronics. One possible explanation for this is that in the absence of solvent, the structure and mobility of the cytochrome domains is restricted, preventing electron transfer. Another explanation for this is that the cytochrome molecules are not able to interact on the surface of the fibrils, preventing electron transfer. Unfortunately it is not possible to perform such conductivity measurements in water to distinguish between these two cases, as it becomes very difficult to separate the effects of ionic conductance in water from that of the structures of interest.

2.6 Discussion

The original target of this project was to produce a self-assembling conductor through producing fibrils from a fusion between an amyloid forming protein to an electron transport protein, demonstrating the potential utility of amyloid fibrils as material for nanotechnology. This projected has demonstrated that amyloid fibrils can be readily used as a scaffold to displaying fully folded protein, providing care is taken in the design of such a system. Furthermore, it has been shown that fibrils formed from $(SH3)_2Cyt$ can reversible exchange electrons with their surroundings showing that fibrils can be functionalised. However, as discussed in section 2.5.4, it has not been possible to detect the ability of the fibrils to transport electrons down their length, despite estimates predicting the surface cytochromes should be packed sufficiently closely for interaction (section 2.4.12). This is partly due to the difficult in making such measurements; sufficiently sensitive conductivity measurements performed in a solvated environment are current not technically possible (section 2.5.4). It is not possible to determine whether or not the cytochrome molecules are able to interact on the surface of the fibril.

Study of these fibrils has however, led to interesting insights into the structure of amyloid fibrils and the mechanism of its formation. Particularly, the observation of solution-state NMR resonances originating from the surface of amyloid fibrils is discussed in section 2.6.1 and evidence inferring information on protofilament packing is described in section 2.6.4.

2.6.1 Flexible regions of amyloid fibrils can be observed by solution-state NMR

The observation of solution-state NMR resonances from GDa protein complexes is a remarkable observation. This conclusion is the simplest that reconciles the following observations;

- Well resolved NMR spectra are observed from pellet1 samples and less than <0.1% of material is transferred to a super2 (fibrils are not dissociating on the timescale of the experiment).

- The chemical shifts of the resonances observed occur at or near random coil shifts ([101, 102]), consistent with the view that displayed groups must be conformationally flexible in order to yield sharp NMR resonances.

- The ^{15}N HSQC signal intensity shows that spectra from $(SH3)_2Cyt$ fibrils are missing one of the two SH3 domains.

- Resonances from folded cytochrome are not observed by solution-state NMR. The NMR signal intensity decreases linearly with heme binding, and decreases to zero on heme saturation at 0.5 equivalents.

- D_{eff} data from the NMR resonances is best explained by a theory that incorporates rotational diffusion. Length distributions from AFM and TEM are in good agreement with those calculated from NMR diffusion measurements.

There is a precedent for this type of behaviour. The L7/L12 domain on the surface of the MDa ribosome is sufficiently flexible to be studied by solution-state NMR [103]. On binding an elongation factor, this domain binds to the L10 domain on the surface of the ribosome, loses its flexibility and becomes invisible to solution-state NMR. The L17/L12 domain in the absence of the elongation factor are observed to diffuse very slowly ($D_{\text{eff}} \sim 10^{-8}$ cm^{2}s^{-1}), but do not show the anomalous diffusion, as the diameter of the ribosome is less than 500 nm, rotational effects are insufficiently small to lead to variance of D_{eff} with Δ. A more thorough discussion of the NMR theory is found in chapter 3, and of relaxation effects in chapters 4 and 6.

2.6.2 Where are the cytochromes in $(SH3)_2Cyt$ fibrils?

50% of the cytochrome domains in $(SH3)_2Cyt$ and $(SH3)_3Cyt$ fibrils are solvent exposed, able to fold and bind heme. Of these, 40% of the cytochrome domains of fibrillar $(SH3)_2Cyt$ are observed by solution-state NMR. As these disappear on heme binding and are completely invisible at the 50% binding stoichiometry, these must originate from a subset of the 50% of solvent exposed cytochromes. Of the observable SH3 molecules, it is interesting to note that the SH3 peaks decrease in intensity at the same rate as the cytochrome peaks with added heme, and using HMQC/PFG experiments, the SH3 resonances and the Cyt resonances diffuse at the same rate. Thus each observed Cyt domain is observed in conjunction with a single $SH3$ domain.

There are several arguments to support the absence of 60% of the exposed cytochrome resonances from the NMR spectrum. The first comes from the observation that experimental agreement between the length distributions calculated from NMR measurements and that from AFM/TEM measurements is improved by the assumption that the signal primarily originates from resonances towards on the ends of the fibrils (section 3.3). This proposal is analogous to the observed decrease in linewidth of NMR resonances at the end of flexible chains [104]. The second argument centres of defects, and the number of SH3 residues in the fibril core. If both SH3 residues of a monomer were incorporated into the fibril, this might be expected to reduce the conformational flexibility required to narrow the resonances from the flexible cytochrome domain. Thus if 60% of the surface cytochrome residues were expected to be held into the fibril by two SH3 domains and 40% were held by only one of the two potential SH3 domains the data would be explained. A third proposal stems from the higher order aggregation of fibrils. This would lead to the suppression of the flexibility of surface residues on the binding interface. Limited proteolysis experiments, combined with a careful intensity analysis on assigned resonances could lead to a distinction between these

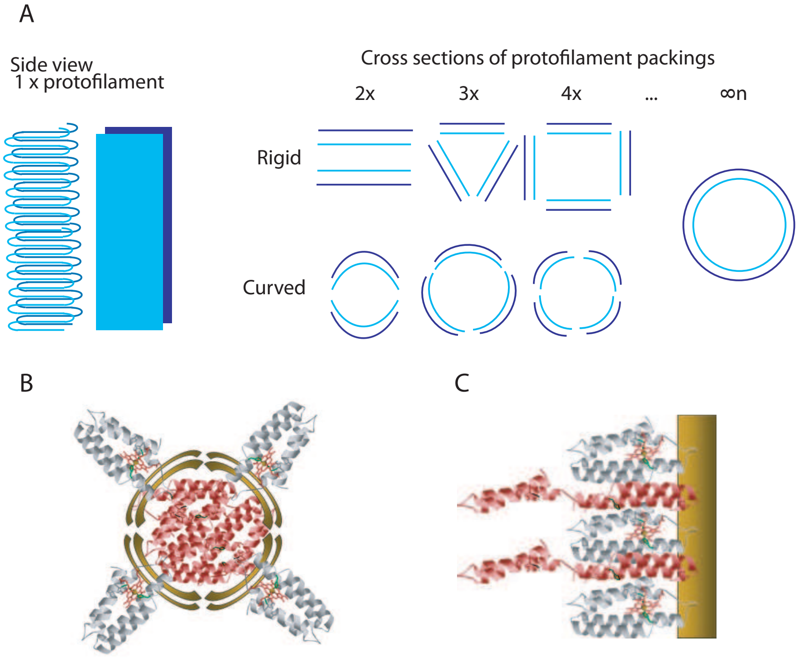


Figure 2.29: Interpretations of the 50% rule as consequences of protofilament packing. Accepting the protofilament as a structure of paired β sheets as the basic packing unit, when trying to construct higher order assemblies, an inside and an outside arise naturally **A -** A representation of a protofilament. Assemblies of n protofilaments with the inside (light blue) and outside (dark blue) faces. **B -** a representation of the cytochrome fibril. SH3 fibrils have been shown to be based on a 4× protofilament arrangement [35]. Cytochrome molecules would then find themselves distributed equally between the inside (red) and the outside (grey) of the fibril. **C -** A representation of how steric effects might lead to a population of cytochrome molecules to be unable to fold.

models.

2.6.3 Insights into inter-protofilament packing - the 50% rule

The 50% rule states that the fraction of surface exposed groups, to the total number of groups within an amyloid fibril will be 50%. There are few techniques allowing this measure to be rapidly and quantitatively made. However, this ratio has been accurately measured for $(SH3)_2Cyt$ and $(SH3)_3Cyt$ fibrils. There have been several other instances where this ratio has been reported for a range of amyloid fibril forming systems. In each case the figure was found to be ca. 50%. To date, there are four references in the literature, and two additional unpublished instances of this rule from within the group from 8 different amyloid fibril systems, summarised in table 2.5.

If such a rule is generic, then its origin would have to be structural. In the case of the cytochrome fibrils, it is unlikely the missing cytochrome groups would be directly involved in the fibril axis. The $(SH3)_3Cyt$ fibrils were originally created

System	Method of determination
Sup35	Cysteine mutants of a Sup35 peptide were made and assembled into fibrils. The resulting amyloid fibril displayed 50% of the cysteine residues, quantified using a biotin assay [105].
Insulin	Insulin fibrils were deamidated, the fibrils were hydrolysed and the relative populations assayed by mass spectrometry [106].
$SH3_3Cyt$ and $SH3_3Cyt$	Of the total available cytochrome molecules, 50% are solvent exposed, measured by following heme binding with electronic spectroscopy.
TTRRGD	TTR fibrils displaying and RGD group have been produced. Limited proteolysis has revealed that 50% of the RGD groups are surface exposed (unpublished, Gras et al).
αB crystallin	Discussed further in section 4.2.2, an 11 amino acid stretch displayed on the surface of αB crystallin complexes is found to be half buried on converting the crystallin complexes to fibrils
Human lysozyme	$\sim$40% of lysozyme residues within amyloid fibrils are protected from pepsin hydrolysis, whereas the remaining $\sim$60% are partially digested [48].
$A\beta_{(1-40)}$	$\sim$50% of amide protons in the backbone found to be highly resistant to hydrogen exchange [107].

Table 2.5: In several amyloid fibril forming systems, the fraction of surface exposed groups to total groups has been measured. When compared to the total protein present in the fibril, the ratio of the two has been found to be consistently ca. 50%.

in order to distinguish between possible explanations for this ratio.

Two other plausible explanations are discussed below. One explanation is that all cytochrome domains were present on the surface of the fibril but 50% were sterically prevented from folding and so $(SH3)_3Cyt$ fibrils would be expected to bind more porphyrin molecules with the additional conformational flexibility (figure 2.29C). An additional hypothesis is that 50% of the cytochrome molecules were involved in the fibril core. An alternative explanation is that missing 50% of the protein is folded into the core of the fibril. It has been shown however that proteins need a high degree of sequence similarity in order to form co-fibrils [25, 26]. Mixed fibrils have been reported, but the relative concentration of the additional group is found to be 1:100 or less [73]. As described in section 2.5.1 for example, co-aggregation of $(SH3)_2$ and $(SH3)_2Cyt$ has proved currently intractable, making this explanation highly unlikely.

A 50% rule arises naturally from consideration of how protofilaments must pack in order to produce a fibril. Protofilaments can be represented as pairs of β-sheet structures, as shown in figures 2.29A and B. In order to produce a structure consistent with those observed by AFM and TEM, several such protofilaments must pack in order to produce a fibril. Thus protofilaments must pack together, in a manner that lowers the free energy. Packing pairs of β sheets leads to a 50% rule as shown in figure 2.29. If the individual β-sheet strands are indistinguishable, then 50% would be expected on the inside of the fibril, and 50% would be expected on the outside of the fibril.

Structures from solid-state NMR such as those from WW2 [42] and HET-s$_{(218-289)}$[40] where a single polypeptide is proposed to occupy both adjacent β-sheet segments of a protofilament would give rise to a 50% rule if such protofilaments had no preference for which side formed the inside, and which side formed the outside of the amyloid fibril.

This would suggest that 50% of the cytochrome molecules are buried within the fibril. This requires sufficient space within the fibril core to accommodate the additional mass of the non-core fibril regions. Cryo-EM measurements and STEM measurements (sections 2.6.3 and 2.6.3) yield electron density maps and mass per unit length measurements show that amyloid fibrils have a reduced density in the core of the fibril, consistent with this conclusion.

As noted in section 2.4.14, raising the pH of $(SH3)_2Cyt$ and $(SH3)_3Cyt$ fibrils to 7 from 3 causes the fibrils to dissociate. In addition, it should be noted that the fibrils could not be formed starting with holo-$(SH3)_2Cyt$. This is elegantly explained if the missing 50% of non-core groups is taken to be inside the fibril. Thus the protofilaments perhaps are unable to pack in a conformation that allows the cytochrome to remain folded inside the fibril and conditions that favour the refolding of the cytochrome particularly stabilising the 50% of residues held inside the fibril. Therefore the cytochrome must be unfolded during fibril formation and elongation.

Cryo-EM measurements

As shown in figure 1.3, where Cryo-EM measurements have been taken on amyloid fibrils, the image reconstructions have yielded models with reduced electron density in their core. SH3 fibrils [35] and insulin fibrils [37] have follow cores. Recent unpublished data by Fitzpatrick et al show $\text{TTR}_{(105-115)}$ to have a hollow core.

STEM measurements

Scanning Transmission Electron Microscopy (STEM) measurements have been performed on other amyloid fibrils, giving mass per unit length (MPL) estimates. Assuming the density model in section 2.4.12 a theoretical MPL can be derived. Comparing the estimated MPL to the experimental value, there is a large discrepancy between the two values as shown in table 2.6. If the fibrils are assumed not to be cylinders, but cylindrical shells of outer radius a and inner radius b, the inner radius can be estimated (table 2.6). This simple geometric analysis and comparison was not performed in the original publications [108, 109, 110, 111].

$$\begin{aligned} \frac{M}{L}_{solid} &= \frac{\pi r^2}{\overline{v}} \\ \frac{M}{L}_{hollow} &= \frac{\pi(a^2-b^2)}{\overline{v}} \end{aligned} \tag{2.6}$$

If the fibrils are assumed to be cylinders, their densities must be far lower than that predicted by the partial specific volume (table 2.6), inconsistent with the picture of amyloid fibrils as dense, tightly packed, stable structures. Thus it is likely that these measurements describe fibrils as shells, not cylinders (nonzero values of b) (table 2.6). From the ratio of a/b, the total radius would be anything between 1.2 and 2.2 times the radius of the hollow core.

In conclusion therefore, Cryo-EM and STEM measurements both support the general picture of amyloid fibrils based on protofilament packing around hollow cores, which in turn leads naturally to a 50% rule.

	Experiment		Predicted				Ref
	MPL / kDa nm^{-1}	Radii / nm	r / nm	b / nm	a/b	ν for b=0	
Amylin	10.0	5.0	2.0	0.5	9.4	1.2	[108]
TTR WT	14.0	8.0	2.3	1.7	4.8	2.2	[109]
TTR L55P	13.7	8.0	2.3	1.7	4.7	2.2	[109]
Calcitonin	7.8	2.76	2.8	1.1	6.8	1.5	[110]
$A\beta_{10-35}$	6.6	2.23	2.2	1.1	6.2	1.6	[111]

Table 2.6: STEM measurements allow an experimental determination of the mass per unit length of fibrils and TEM measurements allow an estimate of the radii of a fibril formed from a specified system. Several physical parameters can be estimated from these measurements using equations 2.6. From this measurement, an effective density ν can be calculated. If the fibrils are assumed to have a density equal to the partial specific volume of globular protein, then using a cylindrical shell model, the inner radius can be estimated.

Figure 2.30: Characteristic conformations formed by twisted chiral ribbons. **A -** helicoid, **B/C -** spiral ribbon [112].

2.6.4 Insights into inter-protofilament packing - morphology changes

By adding additional, non-core regions to amyloid fibrils, the morphology of the fibrils are observed to change. Fibre XRD and CD data show structural features defining the core protofibril are unaffected. The distortion in the fibril structure must therefore be a consequence of a change in the manner through which the protofilaments interact.

Like other fibrous biological assemblies such as collagen [113] and cellulose [114], amyloid fibrils are twisted assemblies of chiral micro-fibres, or protofilaments. Where the protofilaments are twisted with respect to each other along a direction perpendicular to the fibril axis, they form a *helicoid* structure. Another characteristic chiral motif is the *spiral ribbon* [115] (figure 2.6.4A) or barber's pole ribbons (figure 2.6.4B). These two structures are topologically related and of similar energy, separated by a first order phase transition [112]. The elastic properties of the underlying protofibrils, the torsional rigidity and the bending rigidity are crucial parameters in determining which conformation is favoured [116].

In general, amyloid fibrils tend to assemble into structures resembling helicoid geometries. The width all amyloid fibrils are similar, in agreement with the prediction from theory that the energy of helicoid structure is at a minimum at a critical ribbon width [112]. Helicoid structures are observed for SH3 fibrils [35] and insulin fibrils [53]. Fibrils from $(SH3)_2Cyt$ and $(SH3)_3Cyt$ however appear to be better described by a spiral ribbon structure (section 2.4.4). This conclusion is further supported by the XRD data from SH3 fibrils (figure 1.2B) and $(SH3)_2Cyt$ fibrils (figure 2.13C). The distances corresponding to axial and equatorial symmetry in the fibrils are comparable in the two systems. However, fibrils from $(SH3)_2Cyt$ have an additional X-shaped symmetry in their diffraction pattern, consistent with helical symmetry of pitch 80-150 nm, consistent with the observed pitch observed by TEM/AFM.

Similar transitions from a helicoid to a spiral ribbon structure have been observed for fibrils form from modified Ure2 fibrils [46] and for modified $TTR_{(105-115)}$ fibrils (chapter 5). Given that helicoid structures and spiral ribbons are related energetically, it is likely that the flexible regions on the surface of the fibril lead to steric pressures that disfavour the more closely packed helicoid structures, favouring the spiral ribbon.

2.7 Conclusion

Many of the engineering goals of this project have been fulfilled. Amyloid fibrils can be readily employed as a scaffold on which even large functional groups such as proteins can be arranged with sub-nm precision on a macroscopic scale.

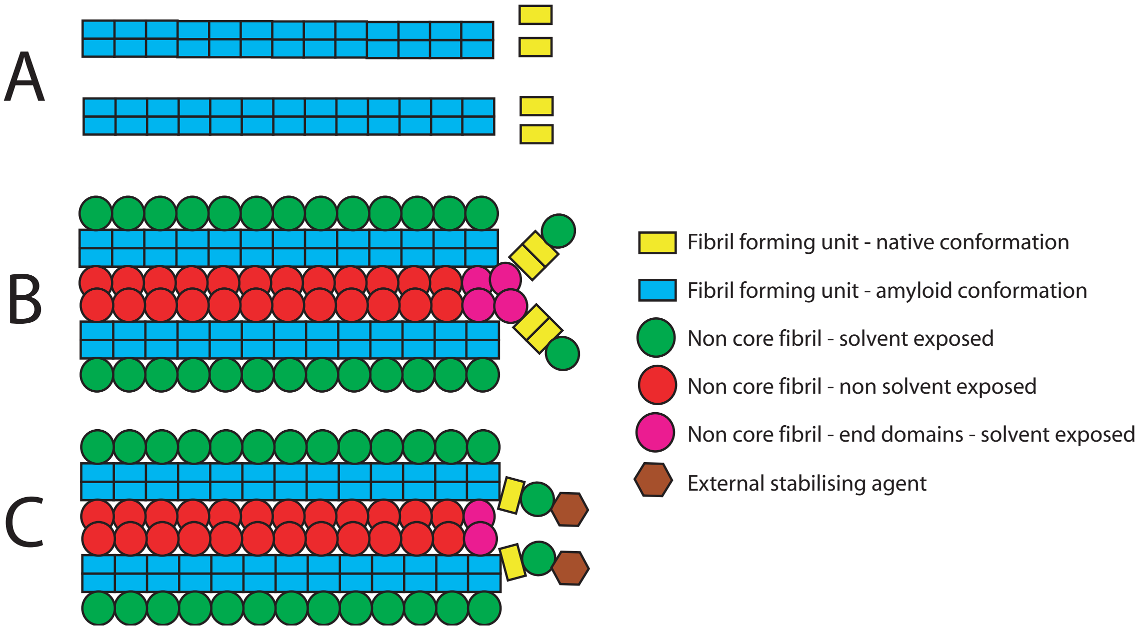


Figure 2.31: Importance of the non-core fibril groups. **A -** Where they are not present, as a monomer detaches from a fibril it could be expected to leave the fibril as a single unit. **B -** Where there are significant non-core regions, additional forces act to keep the fibril forming unit in the vicinity of the fibril, making it less likely to get away and so a significantly slower off rate. **C -** Where an independent agent that encourages refolding of the non-core groups is present, degradation of the fibrils from the ends would be expected.

Amyloid fibrils formed using these methods are functional; $(SH3)_2Cyt$ fibrils can exchange electrons with their surroundings. Utilisation of these fibrils as batteries is obviously hampered by the expense of expressing $(SH3)_2Cyt$ and the low quantity of charge that these fibrils can store (for 50 μM of total protein, the fibrils can store 4.16×10^{-24} C dm^{-3}). The benefit is that the fibrils can be easily separated from a reaction mixture of soluble molecules or protein. These systems could perhaps find use in organic chemistry where separating enzymes from mixtures of small molecules can be difficult. However, as discussed in section 2.5.4, it has not been possible to detect currents passing along the length of the fibril, despite estimates predicting the surface cytochromes should be packed sufficiently closely for interaction (section 2.4.12). Partly due to the difficulty in unambiguously measuring currents on such a small scale, it has not been possible to determine whether or not the cytochrome molecules are able to interact on the surface of the fibril.

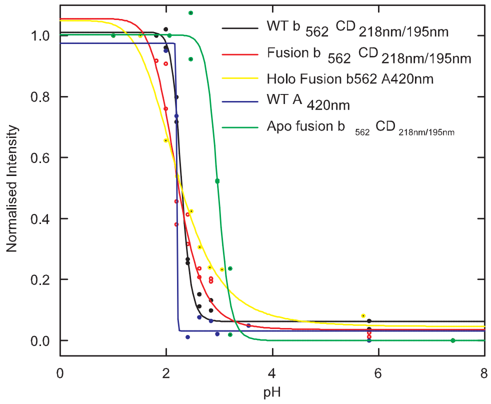
Reflecting the synergy between technology and science, these systems have led to interesting observations that perhaps shed light on the factors that underpin protofilament packing in amyloid fibrils. Although fibrils can be produced displaying folded protein, the local folding of protein into β sheets to form protofilaments does not appear to be significantly altered. However, the higher order packing of the protofilaments into amyloid fibrils are clearly perturbed by the additional groups, giving rise to twisted ribbon morphologies that presumably relieve the additional steric pressures on the inter protofilament interactions. A deeper understanding of the underlying fibril structure is necessary to allow groups to be arranged on their surface with the precision required to ensure that are in the correct geometry to interact.

Many studies have been made of interactions that stabilise the β-sheets within structures such as amyloid fibrils, however less is understood of the determinants of the interactions between the protofilaments. The experimental observations described in this chapter are reconciled by a single generic model for protofilament packing (figure 2.31). This model reconciles experimental data from a range of amyloid fibrils; the 50% surface packing rule, the dissociation of certain fibrils under conditions that favour the non-core regions re-folding and the observations of hollow cores in amyloid fibrils from cryoEM and STEM measurements.

Novel NMR methodology has been developed to investigate the flexible groups displayed on the surface of the fibrils. Through studying the NMR behaviour of $(SH3)_2Cyt$ fibrils, the diffusion theory of chapter 3 was developed and considerations of the dynamics required to narrow the NMR linewidths of flexible regions on the surface of slowly tumbling structures led to the simulations discussed in chapter 6. In addition, applying the methodology developed in this chapter to a range of amyloid fibril forming systems led to the studies discussed in chapter 4 and chapter 5 where it is demonstrated that the flexible regions on the surface of amyloid fibrils have a previously unstudied, but important role in the mechanism of fibril formation, consistent with the model described in figure 2.31.

2.8 Appendix

2.8.1 Unfolding of monomeric protein at low pH



Fitting to

$$f(x) = y_0 + \frac{a}{1 + \frac{x}{x_0}^6}$$

	midpoint pH	Uncertainty
Holo A420	2.21 ± 0.07	3%
WT A420	2.20 ± 0.19	8.4%
WT CD	2.28 ± 0.01	0.6%
Holo CD	2.19 ± 0.06	2.7%
Apo CD	2.97 ± 0.03	1%
Holo b562	2.22 ± 0.05	2.1%
Apo b562	2.97 ± 0.03	1%

Figure 2.32: Structural transitions in wild type cytochrome b_{562}, holo and apo fusion protein as a function of pH followed by UV-Vis and CD. Normalised transitions fitted to sigmoidal function shown for midpoint determination. Good agreement is found for the transition observed in the wild type cytochrome and the holo fusion protein by both UV-Vis and Far UV CD.

2.8.2 Porphyrin titrations

M-PPIX solutions of 1 mM were prepared in 10 mM NaOH. These were diluted by a factor of 10 for the heme titrations. Free heme oxidises and aggregates over a period of 10 hours and so fresh stocks were prepared for each set of experiments. Free metalloporhpyrins are unstable at low pH and so titrations at low pH were mixed extensively to enhance the rate of binding. Once bound to protein, the porphyrin and the protein are both stabilised.

The binding data were analysed as an equilibrium between two species, H and A (heme and apo-cyt), combining to give a single bound species B (holo-cyt), with a 'hidden' population of A that cannot bind. This leads to the following scheme and expression for the binding affinity:

$$\begin{gathered} H + (1+\alpha)A \rightleftharpoons B + \alpha A \\ K_A = \frac{[B_{eq}]}{[A_{eq}][H_{eq}]} \end{gathered} \tag{2.7}$$

Where $n = \frac{\alpha}{1+\alpha} = A_{hidden}/P_0$, the fraction of the molecules of A that are effectively 'hidden'. The equilibrium expressions are related to the total heme concentration H_0 and the total protein concentration P_0 by $H_0 = H_{eq} + B_{eq}$ and $P_0 = B_{eq} + A_{eq} + nP_0$. The analysis makes use of the following parameters, $X = H_0/P_0$, $Y = B_{eq}/P_0$ and $g = K_A \cdot P_0$ where X is under experimental control and Y will be related to the experiment observable. Substituting these into the affinity constant expression and solving the quadratic for Y, discarding the non-physical solution, yields;

$$\begin{aligned} Y &= (1-Y-n)(X-Y)g \\ Y &= \frac{1}{2}\left(X+1+\frac{1}{g}-n-\sqrt{\left(X+1+\frac{1}{g}-n\right)^2-4X(1-n)}\right) \\ Y &= \frac{1}{2}\left(X+1+\frac{1}{g}-n-\sqrt{\left(1+\frac{2}{g}-2X-n\right)(1-n)+\left(X+\frac{1}{g}\right)^2}\right) \end{aligned} \tag{2.8}$$

In the limit where $x \rightarrow \infty$, $Y \rightarrow (1-n)$, as expected as $1-n=\frac{1}{1+\alpha}$. As free heme contributes marginally to the absorbance at 420nm, the effective measured absorbance at 420nm will be a linear sum of both bound and free heme, leading to the following expression:

$$\frac{A_{420}}{P_0} = \epsilon_{420}^{Holo}Y + \epsilon_{420}^{Heme}(X-Y)\frac{A_{420}}{\epsilon_{420}^{Holo}P_0} = \left(1-\frac{\epsilon_{420}^{Heme}}{\epsilon_{420}^{Holo}}\right)Y + \frac{\epsilon_{420}^{Heme}}{\epsilon_{420}^{Holo}}X \tag{2.9}$$

The absorbance data at 420nm can be understood in terms of the equilibrium constant, the fraction of the population competent to bind (n) and the experimental extinction coefficients of the free heme and the holo-cytochrome. A least squares analysis performed in sigmaplot 9.0 with this scheme yield the parameters referred to in the text. The plots of data analysed with this method are shown in figure 2.20C. An alternative expression to determine the stoichiometry of binding is that of the Job analysis, based on the following scheme:

$$\begin{aligned} A + mB &\rightleftharpoons AB_m \\ K_A &= \frac{[AB_m]}{[A][B]^m} \\ [A_0]-[A_{eq}] &= [AB_m] \\ [B_0]-[B_{eq}] &= m[AB_m] \end{aligned} \tag{2.10}$$

Let $C = [A_0]+[B_0]$, $x = \frac{B_0}{A_0+B_0} = \frac{B_0}{C}$ and $y = [AB_m]/C$. The subsequent Job plot of y vs x has the following form;

$$y = K_A c^m \left(1 - x - y\right)\left(x - my\right)^m \tag{2.11}$$

Taking the derivative of the logarithm yields the highly useful result that at the maximum value of y, $x = \frac{m}{1+m}$. Thus the maximum point allows determination of m, the binding stoichiometry and K_A, the binding affinity. The Job plots for heme binding to monomer and fibrillar $(SH3)_2Cyt$ are shown in figure 2.20D. The value of y at the maximum point is given by solving the following polynomial.

$$\begin{aligned} y &= K_A(cm)^m\left(\frac{1}{1+m} - y\right)^{(m+1)} \\ 0 &= y^b + yd^b - bd^b \end{aligned} \tag{2.12}$$

Where $b = \frac{1}{1+m}$ and $d = K_A(cm)^m$.

2.8.3 Twisting the fibril, and staggering displayed groups will relieve steric pressure exerted on the fibril

The degree of steric clashing occurring between groups on the surface of an amyloid fibril can be estimated with the following geometric model. From the PDB structure cytochrome b$_{562}$ can be considered a cylinder of diameter L_c=2nm and height 5nm. These are attached to the surface of a cylinder of radius $r_{fib} = a$,suspended by a linker of length L (figure 2.33A). Let the fibril have a symmetry of M, which denotes the number of cytochrome molecules packing around it in cross section, where the angle subtended is $2\pi/M$. The larger the radius, or the longer the linker, the more domains of a given size can pack in without clashing with their neighbours.

Figure 2.33B shows this for M=4 where the displayed groups are just touching on the surface of the fibril. Twisting the fibril allows a closer packing, such as that shown in figure 2.33C. The cylinders move from a cubic closed packed arrangement to a hexagonal close packed arrangement which allows the underlying amyloid forming group pack closer together. Figure 2.33 D shows another packing mode, obtained through staggering the spacing of the displayed group. Physically this would correspond to the linkers distorting, or a distortion introduced at the level where fibrils are assembling.

There exist two constraints on the size of the displayed group. If the monomer takes up very little of the fibril length, as will happen for fibrils of large radius, then steric clashing will occur along the fibril axis, relieved by twisting. From section 2.4.12, the length occupied by a given monomer in the fibril is $L_1 = \overline{v}M_W/\pi N_a(a-b)^2$, where M_W is the molecular weight of only the core fibril

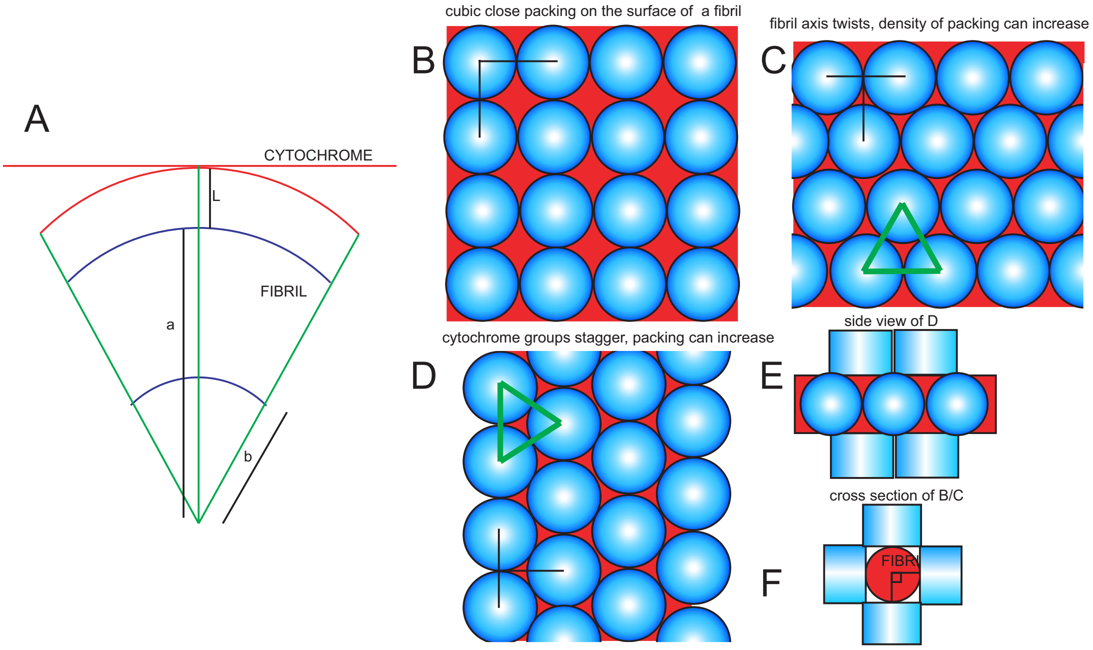


Figure 2.33: **A -** The geometry of the model. **B-F -** Arrangements of displayed groups on the opened surface of a fibril, illustrating how re-organising domains on the surface of the fibril can relieve steric pressures.

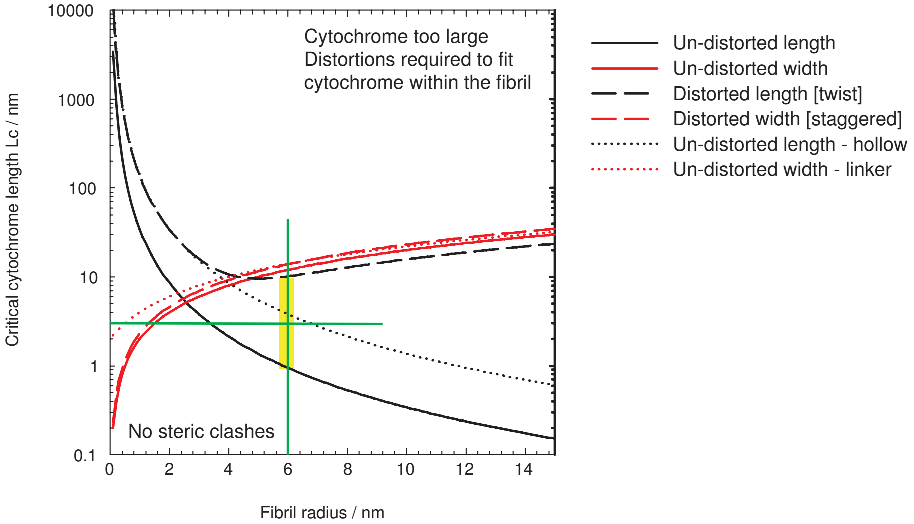


Figure 2.34: A summary of the packing model. The minimum cylinder diameter is shown as a function of supporting fibril radius. For fibrils of small radius, width clashes are the biggest source of steric hinderance (red). For longer fibrils, length clashes dominate (black). Fibrils of radius 6 nm are predicted to be unable to support cylindrical cytochromes of radius 2nm. If the fibril twists, then the space constraint is removed. The solid black line is the length constraint for untwisted, the medium dashed line is that for twisted fibrils. With twisting therefore, displayed groups between 1nm and 10nm can be accommodated on the surface of fibrils of radius 6 nm whereas only groups smaller than 1nm can fit onto the surface of an untwisted fibril. Allowing the fibrils to have hollow cores, or allowing the fibrils longer linkers relaxes both the length and the width constraints.

forming molecule. Where ϕ is a measure of the twist the fibril makes, valued from zero to 45^o, a clash between displayed groups will occur if;

$$L_C^{twist} < \sqrt{\left(\frac{N\bar{v}M_W}{\pi(a-b)^2N_a}\right)^2 + (\phi(a+L))^2} \tag{2.13}$$

The fully twisted packing is shown in figure 2.33C (ϕ=45^o), and the triangle for the limit is shown in green. By allowing the fibril to be hollow, or by twisting the fibril, larger groups can be displayed on the fibril. If the radius is small, then displayed proteins stand the risk of clashing around the circumference of the fibril. This 'width' clash is relieved by the linkers allowing a staggering of the displayed groups. Letting z be an offset parameter ranging from 0 to 0.5:

$$L_C^{stagger} < \frac{2(a+L)\tan\frac{\pi}{N}}{\sqrt{1-z^2}} \tag{2.14}$$

Where figure 2.33D shows the fully staggered geometry ($z = 0.5$), and shows the triangle for the clashing limit in green. Thus offsetting or having a longer linker reduced this type of steric clash. Figure 2.34 shows plots for several values of this two limits with $M = 4$. For a fibril to be stable, L_C must be lower than the critical value calculated from the model. For short fibrils, width clashes are significantly for fibrils of radii 2nm or greater, the length clashing becomes more important. Interestingly, the $(SH3)_2$ fibrils have a radius of 6 nm are predicted to not be able to form with the displayed cytochromes whose L_C=2nm (marked by the green intersecting lines in figure 2.34) as the critical L_C for fibrils of radius 6 nm is ca. 1nm. However, by twisting the fibrils to ϕ=45^o, the critical L_C is increased to 10 nm. Thus the cytochrome can only fit into $(SH3)_2$ fibrils if the underlying protofilament structure twists.

In conclusion, where groups are tightly packing the surface of a fibril, twisting of the underlying protofilaments can relieve steric pressure. A range of protofilament twist angles were found in fibrils in the absence of steric pressures [53]. Thus in the presence of steric pressure, the conformations with higher twist would be expected to be significantly lower in energy that those without twist. The observed geometry of the $(SH3)_2Cyt$ fibrils therefore represents a frustrated structure where the reduction in steric pressure is offset by the reduction in protofilament binding energy resulting in the spiral ribbon structures of helical pitch ca. 100nm.

Chapter 3

Contributions of rotational diffusion to PFGSE NMR measurements

Abstract

NMR diffusion experiments employing pulsed field gradients (PFG) are well established as sensitive probes of the displacement of individual nuclear spins in a sample. Conventionally such measurements are used as a measure of translational diffusion, but here we demonstrate that under certain conditions rotational motion will contribute very significantly to the experimental data. This situation occurs when at least one spatial dimension of the species under study exceeds the root mean square displacement associated with translational diffusion, and leads to anomalously large apparent diffusion coefficients when conventional analytical procedures are employed. We show that in such a situation the effective diffusion coefficient is a function of the duration of the diffusion delay used, and that this dependence provides a means of characterising the dimensions of the species under investigation.

3.1 Introduction

NMR experiments employing pulsed field gradients (PFG) have been used for many years to measure the diffusion coefficients of species in solution [117]. PFG NMR experiments directly measure the net displacement of nuclear spins during a defined delay, and such displacements can be related to the translational diffusion coefficients of the spins that give rise to the resonances in question [118]. As a non-invasive solution-state technique that yields information on the size of the species under study, this is a highly versatile experiment whose uses in-

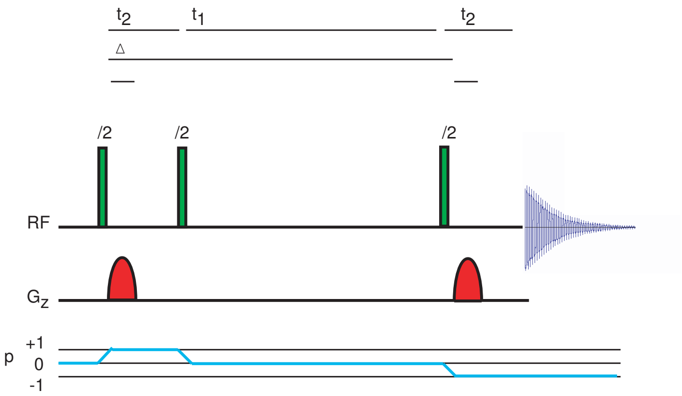


Figure 3.1: The gradient spin-echo pulse sequence with the required CTP. The sequence is broken into two relaxation periods, τ_1 and τ_2, and two gradient periods Δ and δ. The resulting spectrum has an attenuation factor that depends on the displacement of the molecules under study over time. In order to minimise artifacts and to ensure that the second gradient pulse acts in the opposite direction to the first, the following phase cycle can be used: ψ 0000222233334444, ϕ 1302, receiver 0022220033111133.

clude probing ligand-receptor interactions [119, 120], dissociation constants [121], membrane interactions [122, 123] and the folding and aggregation of protein molecules and peptide systems [124, 125, 126, 127].

The most common experimental approach to measuring such diffusion coefficients involves the use of a spin-echo (SE)[118]. Such PFGSE experiments, whose pulsed sequence is shown in figure 3.1, have been widely applied. By the end of the PFGSE pulse sequence, observable magnetisation acquires a phase term, $\widehat{I}_{-}e^{-i\alpha'\delta z}$, where δz is the net displacement of a spin during the experiment. The length scale is set by $\alpha' = G\gamma\delta$, the inverse of the pitch length generated by the pulsed gradient of field strength G and duration δ, where γ is the gyromagnetic ratio of the observed nucleus. To obtain an expression for the observed signal intensity of a sample, this term must be integrated over all spins in the system, where $P(\delta z)$ is a function that describes the probability of a given displacement occurring within the sample during the experiment. The diffusion weighted signal intensity, S_i, is then given by:

$$S_i = \int_{-\infty}^{+\infty} \widehat{I}_{-} P(\delta z) e^{-i\alpha'\delta z} d\delta z \tag{3.1}$$

For a single species diffusing under Brownian motion for a time Δ, from the central limit theorem $P_{\text{trans}}(\delta z) = \frac{1}{2\sqrt{\pi D_T \Delta}} e^{-\delta z^2/4D_T\Delta}$, where the mean square displacement is related to the translational diffusion coefficient D_T, through $<\delta z^2>_{\text{trans}} = 2D_T\Delta$. Δ is the diffusion delay, during which magnetisation is stored on the z axis, while the nuclear spins diffuse. Integration of equation 3.1 and setting $\alpha = G_{max}\gamma\delta$ yields the Stejskal-Tanner equation, which relates the PFGSE signal intensity S_i, to the signal intensity obtained in the absence of the gradients, S_0.

$$S_i = S_0 e^{-\alpha^2 D_T \Delta (G/G_{max})^2} \tag{3.2}$$

This is the Stejskal-Tanner (ST) equation, and a plot of $\ln(S_i/S_0)/\alpha^2\Delta$ against $(G/G_{max})^2$ yields a straight line of gradient $-D_T$ [95, 128], where G, δ and Δ are experimental variables. When the duration of the applied gradients are considered in more detail, $\Delta = \Delta' - \delta/3$, where Δ is the effective delay between the two applied gradients. This relationship will be explicitly assumed in this work.

Diffusion coefficients of proteins in the 10-30kDa range, typical of the systems that are subject to studies by NMR spectroscopy, are on the order of 10^{-6}cm^{2}s^{-1}, with an R_H of several nm [129]. For an experimental diffusion delay of 50ms, the corresponding root mean square displacement, $\sqrt{<\delta z^2>}$, is 4.4 μm and so is entirely dominated by translational diffusion. For larger molecular systems, where R_H is comparable with the net displacement due to translational diffusion, rotational diffusion will contribute to the displacement function $P(\delta z)$ however, and so contribute to the measured signal attenuation in the NMR experiment. This in more detail in the chapter appendix, section 3.6.1.

By considering the effects of combined rotational and translational diffusion on $P(\delta z)$, through equation 3.1, we arrive in the present study at novel expressions (equation 3.11) that describe the signal decay in NMR diffusion measurements. These expressions deviate from the form predicted in equation 3.2 and indicate that the apparent diffusion coefficient becomes a function of Δ. In this chapter we describe the analysis that gives rise to this conclusion and consider its significance for the characterisation of macromolecular dimensions.

3.2 Results

3.2.1 Calculation of $P(\delta z)$ for rotating systems

Consider two spins placed within a hard sphere of radius r, one at the centre and the other on the surface of the sphere. If during the diffusion time in a PFGSE experiment, the sphere rotates freely, then the measured displacement along the z axis for a spin on the surface of the sphere will differ from that of a spin at the centre, as illustrated in figure 3.2A. For the centre of mass, the displacement will be given by $P_{\text{trans}}(\delta z) = \frac{1}{\sqrt{2\pi<\delta z^2>}} e^{-\delta z^2/2<\delta z^2>}$. The function describing the behaviour of a spin on the surface of the sphere requires further consideration. The effect of a single rotation from θ_1 to θ_2, where the angles are defined with re-

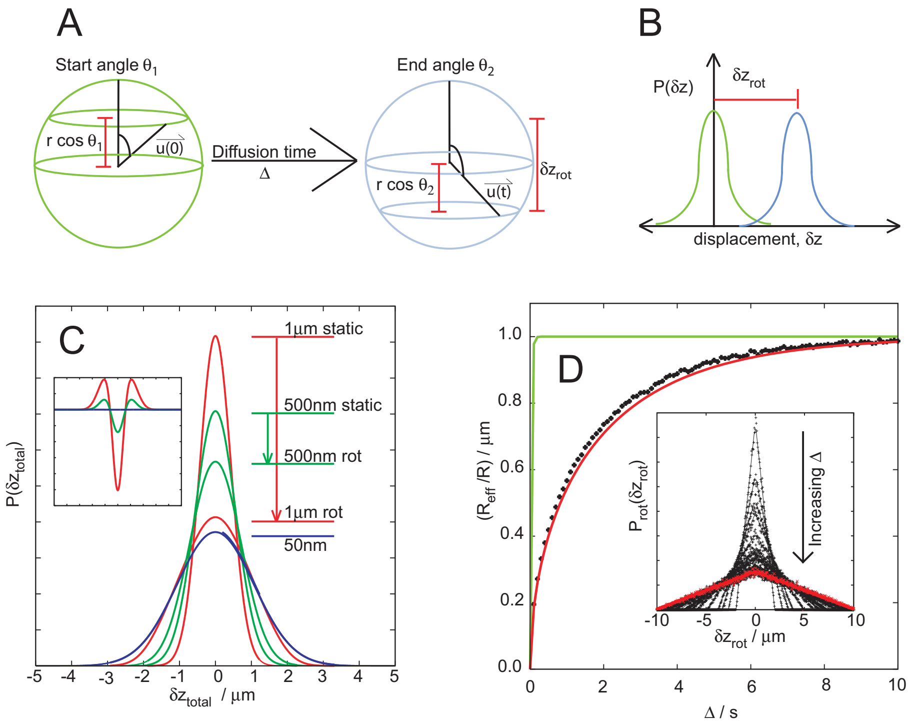


Figure 3.2: **A -** Illustration of how rotational motion contributes to the displacement function of a point a distance r from the centre of a sphere. **B -** When this single rotation is considered in addition to a translational diffusion, it can be seen that it acts to shift the Gaussian translational displacement distribution by a constant distance δz_{rot}. **C -** $P_{\mathrm{total}}(\delta z)$ for spheres of radii 1μm(red), 500nm(green) and 50nm(blue) for $\Delta = 100ms$ using the explicit form of equation 3.5 given in section 3.6.2. For each sphere, the distribution in the absence of rotation is shown, together with the distribution obtained in the limit of unrestricted diffusion. Remarkably, the displacement function for the rotating 1μm sphere closely resembles that of the 50nm sphere. The inset shows the difference between the displacement distribution functions in the static and rotating cases. For spheres of radii 50nm, there is no difference between these two cases. For larger spheres, rotational motion promotes an increase in larger displacements at the expense of shorter displacements. **D -** Inset - Simulated $P_{\mathrm{rot}}(\delta z_{\mathrm{rot}})$ restricted distribution functions for rods of length 5μm and radius 10nm, calculated as described in Methods. Distribution curves are drawn for Δ from 100ms to 10s (black) and compared to an unrestricted diffusion simulation (red). These values are fitted to equation 3.3 to obtain an effective radius, and plotted against Δ in the main figure (black). These values are compared with the theoretical estimate for r_{eff} from equation 3.7 (red) and that for a rod of length 1μm (green) which reaches its unrestricted diffusion limit much more rapidly than a 5μm rod.

spect to the z axis, is to shift the translational distribution function by a constant value $\delta z_{\mathrm{rot}} = r(\cos\theta_2 - \cos\theta_1)$, illustrated in figure 3.2B. The width of this distribution function will not be affected by rotation, just the position of its maximum. For each displacement of the centre of mass in the Brownian ensemble, the point on the sphere will undergo every possible rotation from every possible starting position, providing the time of the diffusion experiment Δ is large. Only the displacement along the z axis is measured in the PFGSE experiment, so only the contribution of rotation to the displacement along the z axis need be considered in this context.

Defining $\overrightarrow{u}$ as the vector connecting the origin and the point on the sphere in the laboratory reference frame as shown in figure 3.2A, the rotational autocorrelation function $< \overrightarrow{u(0)} \cdot \overrightarrow{u(t)} >$ will tend to zero in the limit of $t \to \infty$, indicating that θ_1 and θ_2 are completely independent. In this limit, the degeneracy of a given value of θ on the sphere is proportional to the circumference of the circle defined by the intersection of the sphere and the cone defined by θ. The probability of each starting angle is therefore $P(\theta_1) = \sin\theta_1 d\theta_1$ and each ending position $P(\theta_2) = \sin\theta_2 d\theta_2$. We can write the probability of a given $\theta_1 \to \theta_2$ as $P(\theta_1, \theta_2) = P(\theta_2|\theta_1).P(\theta_1) = P(\theta_2).P(\theta_1)$. In this limit, we can calculate the rotational displacement probability function for a point on a sphere of radius r by imposing simply that the rotation leads to the required δz_{rot}:

$$\begin{aligned} P_{\mathrm{rot}}(\delta z_{\mathrm{rot}}) &= \int_{\theta_1=0}^{\pi}\int_{\theta_2=0}^{\pi} \sin\theta_1 \sin\theta_2 \delta(r(\cos\theta_2 - \cos\theta_1) - \delta z_{rot}) d\theta_1 d\theta_2 \\ &= \int_{-1}^{1}\int_{-1}^{1} \delta(r(u-v) - \delta z_{\mathrm{rot}}) du dv \\ &= \frac{1}{2r}\left(1 - \frac{|\delta z_{\mathrm{rot}}|}{2r}\right) \end{aligned} \tag{3.3}$$

where we have integrated the delta function. This triangular probability function can also be obtained by a simple Monte Carlo simulation, where we draw random start and end vectors on a unit sphere, and construct a histogram of δz_{rot} (figure 3.2D, inset, red). The average squared displacement of this distribution is given by:

$$< \delta z_{\mathrm{rot}}^2 > = \frac{\int_{-2r}^{2r} \delta z^2 P_{\mathrm{rot}}(\delta z) d\delta z}{\int_{-2r}^{2r} P_{\mathrm{rot}}(\delta z) d\delta z} = \frac{2}{3} r^2 \tag{3.4}$$

The combined probability of a given displacement is therefore the convolution of the translational and rotational displacement probability distributions:

$$
\begin{aligned}
P_{\text{total}}(\delta z) = P_{\text{trans}}(\delta z) * P_{\text{rot}}(\delta z) &= \int_{-\infty}^{\infty} P_{\text{rot}}(\tau) P_{\text{trans}}(\delta z - \tau) d\tau \\
&= \frac{1}{8\sqrt{\pi D_T \Delta}} \int_{\delta\tau=-\infty}^{\infty} \int_{\theta_1=0}^{\pi} \int_{\theta_2=0}^{\pi} \sin\theta_1 \sin\theta_2 \delta\left(r(\cos\theta_2 - \cos\theta_1) - \tau\right) e^{-(\delta z-\tau)^2/4D_T\Delta} d\theta_1 d\theta_2 d\tau \\
&= \frac{1}{8\sqrt{\pi D_T \Delta}} \int_{\theta_1=0}^{\pi} \int_{\theta_2=0}^{\pi} \sin\theta_1 \sin\theta_2 e^{-\frac{(\delta z - r(\cos\theta_2 - \cos\theta_1))^2}{4D_T\Delta}} d\theta_1 d\theta_2
\end{aligned}
\tag{3.5}
$$

This result shows, importantly, that the distribution function is no longer Gaussian as it was for purely translational diffusion. The integral form of the above equation can be expressed in a closed form in terms of error functions, the result of which is given in supplementary section 3.6.2. The mean squared displacement of a given spin undergoing both rotational and translational motion then becomes:

$$
< \delta z_{\text{total}}^2 > = \frac{\int_{-\infty}^{\infty} \delta z^2 P_{\text{total}}(\delta z) d\delta z}{\int_{-\infty}^{\infty} P_{\text{total}}(\delta z) d\delta z} = \frac{2}{3} r^2 + 2 D_T \Delta \tag{3.6}
$$

As both translational and rotational probability functions are independent, we obtain the linear sum of second moments. The displacement functions for spheres with radii of 1μm, 500nm and 50nm are shown in figure 3.2C, which compares the normalised functions in the freely rotating and the static limits, and figure 3.2D shows the difference between the two. The distribution functions for the 50nm spheres are essentially unaffected by the incorporation of rotational motion, but those for the 500nm and 1μm spheres are found to have a significant contribution from such motion. Remarkably, the 1μm sphere has a displacement function with a root mean squared displacement similar to that of the 50nm sphere. Figure 3.2B shows quantitatively that the probability of zero displacement is lowered, and that the probability of longer displacements increases. The displacement distributions when rotational effects become significant are no longer Gaussian, and thus rotational motion acts to increase the apparent diffusion of a nuclear spin.

The effect of restricted diffusion on $P(\delta z)$

For Brownian rotation over finite times, the rotational autocorrelation function $< \overrightarrow{u(0)} \cdot \overrightarrow{u(t)} >$ can be expressed analytically [130]as $< \cos\alpha >= e^{-2D_R\Delta}$ where α is the angle subtended between the initial and final vectors, Δ is the diffusion delay and D_R is the rotational diffusion coefficient. At finite times, the final angle θ_2 is no longer independent of the start angle θ_1, as illustrated in figure 3.8 in section 3.6.3. Only when the diffusion delay becomes large do we recover the previously described situation of independent starting and ending positions.

Although an analytical form of the displacement probability function in this regime of finite diffusion delays is not easily obtained, we can on the basis of a simple geometrical model (proof in section 3.6.3 in the Supplementary Material) calculate the average displacement due to rotational motion for the general case as:

$$< \delta z_{\text{rot}}^2 > = \frac{2}{3} r^2 \left(1 - e^{-2D_R\Delta}\right) \tag{3.7}$$

Interestingly, this average does not depend on a distribution function for the angle α separating the starting and ending orientations of the vector. Calculating the probability displacement function for combined rotational and translational motion in the limit $< \overrightarrow{u(0)} \cdot \overrightarrow{u(t)} >=< \cos\alpha >= e^{-2D_R\Delta}$ is complicated by the fact that $P_{\text{rot}}\left(\delta z_{\text{rot}}\right)$ is a function of time in this regime. By comparing equation 3.4 and equation 3.7 we can incorporate the time dependent rotation by using a reduced radius, $r_{\text{eff}} = r(1 - e^{-2D_R\Delta})^{1/2}$, and so θ_1 and θ_2 can still be considered to be independent.

We can simulate $P_{\text{rot}}(\delta z_{\text{rot}})$ for restricted diffusion using a Monte Carlo method, and by drawing rotation angles α that satisfy the distribution $< \cos\alpha >= e^{-2D_R\Delta}$ a histogram of δz_{rot} can be calculated (see Methods). The simulated distribution functions for $P_{\text{rot}}(\delta z_{\text{rot}})$ (Figure 3.2D inset, black) show the same triangular distribution as for the case of unrestricted rotation, but the limits no longer span the interval (-2r,2r). Fitting the distributions to equation 3.3, we can estimate r_{eff} from the maximal displacement (figure 3.2D, black) and this estimate is in excellent agreement with the theoretical value for r_{eff} from equation 3.7 (figure 3.2D, red). Thus, the substitution of r for the reduced radius $r_{\text{eff}} = r\left(1 - \exp(-2D_R\Delta)\right)$ into equation 3.5 provides a convenient method of estimating the effects of restricted diffusion. The restricted rotation factor $(1 - \exp(-2D_R\Delta))$ is shown in figure 3.3A for spheres (green) and rods (blue). For short times and large particles this factor deviates from unity, as rotation becomes restricted.

$P_{\text{total}}(\delta z)$ for any arbitrary rigid geometry can now be considered. All points on the object under investigation will have the same translational and rotational diffusion coefficients. Each point that defines the system is considered to be a single point on a sphere. Summing over the N point spheres that map out the geometry will then give the displacement function for the system as a whole such that $P_{\text{total}}^{\text{arb}}(\delta z) = \frac{1}{N} \sum^{N} P_{\text{total}}(\delta z)$. A common and important geometry that is often of experimental interest is that of the rigid rod of length L. The rod here is considered as a collection of spheres of radii $-L/2$ to $L/2$. The displacement probability distribution is therefore given by $P^{\text{rod}}_{\text{total}}(\delta z) = \frac{2}{L}\int_{r=0}^{L/2} P_{\text{total}} dr$. The mean squared displacement can be calculated by integrating over equation 3.7:

$$< \delta z^2_{\text{total}} >_{\text{rod}} = \frac{2}{L}\int_{r=0}^{L/2} < \delta z^2_{\text{total}} > dr = \frac{1}{18}L^2\left(1 - e^{-2D_R\Delta}\right) + 2D_T\Delta \tag{3.8}$$

This displacement due to translational diffusion will be constant for all points on the rod. Points near the centre of the rod will experiment less displacement due to a given rotation than a point near the end of the rod. The ratio of the rotational root mean square displacements of the ends of a rod (a sphere of radius $L/2$) and the full length rod hence is:

$$\frac{< \delta z^2_{\text{rot}} >_{\text{rod}}}{< \delta z^2_{\text{rot}} >_{\text{sphere}}} = \frac{< \delta z^2_{\text{rot}} >_{\text{rod}}}{< \delta z^2_{\text{rot}} >_{\text{rodend}}} = \frac{1}{3} \tag{3.9}$$

This approach can be generalized to any arbitrary geometry. A treatment applied to spheroidal geometry is presented in supplementary section 3.6.4.

3.2.2 Derivation of NMR observables from rotational $P(\delta z)$ functions

Application to a point sphere

The effect of rotational diffusion and its resulting non-Gaussian displacement probability functions on the NMR experiment can be assessed by substituting the rotating $P(\delta z)$ expressions into equation 3.1 and integrating over the ensemble. For a point on a sphere, using equation 3.5 and noting $D_T\Delta$ is positive, the NMR intensity after a gradient echo S_i, is related to the signal seen in the absence of a gradient echo S_0, by:

$$\begin{aligned} S_i &= S_0 \int_{-\infty}^{+\infty} P_{\text{total}}(\delta z) e^{-i\alpha d\delta z} d\delta z \\ &= \frac{S_0}{4}\int_{\theta_1=0}^{\pi}\int_{\theta_2=0}^{\pi} \sin(\theta_1)\sin(\theta_2) e^{-D_T\alpha^2 - ir_{\text{eff}}\alpha(\cos\theta_2 - \cos\theta_1)} d\theta_1 d\theta_2 \end{aligned} \tag{3.10}$$

The imaginary part of the integral vanishes leading to the conclusion that size effects will not add gradient dependent phase terms to the expression. Evaluation

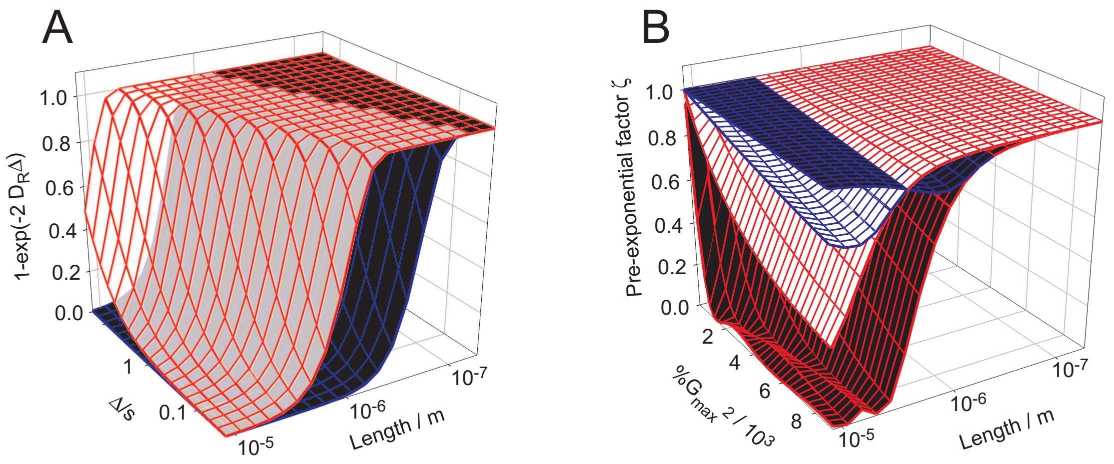


Figure 3.3: Rotational diffusion factors for spheres (black surfaces) and rods (white surfaces). **A -** The restricted diffusion factor $(1 - e^{-2D_R\Delta})$ from equation 3.7. When this factor equals unity, rotations are unrestricted and θ_1 is independent of θ_2. For short diffusion delays and larger particles, the factor is less than unity, and rotations are restricted. **B -** The behaviour of the pre-exponential factor ζ. When ζ equals unity, the ST equation is being obeyed, and rotational motion will not contribute to NMR diffusion measurements. The surfaces outlined in red are calculated in the freely rotating limit, those outlined in blue are calculated for restricted diffusion with Δ=5s. Increasing Δ causes the surface in the restricted case to tend towards that of the freely rotating limit, and decreasing Δ causes it to tend to the static limit of unity. When $\zeta < 1$, ie, where species are larger than 1μm, rotational motion will significantly contribute to the observed signal attenuation measured in a PFGSE experiment.

of this integral yields the following important result:

$$S_i = S_0 \frac{\sin^2\left(\alpha r\left(1-e^{-2D_R\Delta}\right)^{1/2}\right)}{\alpha^2 r^2\left(1-e^{-2D_R\Delta}\right)} e^{-D_T\alpha^2\Delta} \tag{3.11}$$

This function is the product of a Stejskal-Tanner factor, $\sigma(G,\delta,\Delta,L) = e^{-D_T\alpha^2\Delta}$, and a pre-exponential factor, a geometry dependent term $\zeta(G,\delta,\Delta,L)$. ζ is shown in figure 3.3 for spheres (black surface) in the freely rotating limit (red lines) and in the restricted diffusion limit for Δ=5s (blue lines). As ζ deviates from unity, rotational effects will contribute to the PFGSE measurement. When the NMR signal is normalised to that acquired in the absence of gradients, S_0, we obtain:

$$\ln\frac{S_i}{S_0} = -D_T\alpha^2\Delta + \ln\frac{\sin^2\left(\alpha r\left(1-e^{-2D_R\Delta}\right)^{1/2}\right)}{\alpha^2 r^2\left(1-e^{-2D_R\Delta}\right)} \tag{3.12}$$

Two limiting cases are particularly interesting. When the rotational diffusion coefficient or the diffusion delay is large, the effective radius of a sphere increases to its exact radius, leading to the following expression for the NMR data in the freely rotating limit:

$$\lim_{\Delta D_r\to\infty} S_i = S_0\frac{\sin^2\alpha r}{\alpha^2 r^2}e^{-D_T\alpha^2\Delta} \tag{3.13}$$

Here, the pre-exponential factor is no longer a function of the diffusion delay, but the ratio of the radius of the sphere and the length scale set by the pitch of the magnetization helix plays the dominant role. Where the product αr is small, the pre-exponential factor reduces to unity and equation 3.13 reduces to equation 3.2:

$$\lim_{\alpha r\to 0} S_i = S_0 e^{-D_T\alpha^2\Delta} \tag{3.14}$$

This expression reflects the fact that rotational motion will make no contribution to the PFGSE experiment when the species under consideration is small, and we recover the ST equation. In the limit where αr is small but not negligible, the signal intensity can be approximated by:

$$\begin{aligned}\lim_{\alpha r\to 0}\ln\frac{S_i}{S_0} &= \left(-D_T-\frac{r^2(1-e^{-2D_R\Delta})}{3\Delta}\right)\alpha^2\Delta = -D_{\text{eff}}^{\text{sphere}}\alpha^2\Delta \\ D_{\text{eff}}^{\text{sphere}} &= D_T+\frac{r^2(1-e^{-2D_R\Delta})}{3\Delta}\end{aligned} \tag{3.15}$$

Thus intensity will decay will a single exponential, with the *effective diffusion coefficient*, D_{eff}, not simply the translational diffusion coefficient as predicted by the ST equation, but a function of Δ when $\frac{r^2}{3}\Delta > D_T$. The decay deviates from monoexponential with a term in $\alpha^4 R^4$. The consequence of this effect is that spins at a large distance from the centre of mass will appear to diffuse much more quickly than in the case when diffusion depends solely on the translational motion of the particle under consideration. A similar result can be derived for rigid rods as we now show.

Application to a rigid rod

By integrating equation 3.11 we arrive at the following result for the total NMR signal intensity from a rigid rod of length L:

$$\begin{aligned}T_i &= 2S_0\int_{r=0}^{L/2}\frac{\sin^2\left(\alpha r\left(1-e^{-2D_R\Delta}\right)^{1/2}\right)}{\alpha^2 r^2\left(1-e^{-2D_R\Delta}\right)}e^{-D_T\alpha^2\Delta}dr \\ &= S_0\frac{\cos\left(\alpha L(1-e^{-2D_R\Delta})^{1/2}\right)-1+\alpha L(1-e^{-2D_R\Delta})^{1/2}\int_{t=0}^{\alpha L(1-e^{-2D_R\Delta})^{1/2}}\frac{\sin(t)}{t}dt}{\frac{1}{2}\alpha^2 L(1-e^{-2D_R\Delta})}e^{-D_T\alpha^2\Delta}\end{aligned} \tag{3.16}$$

As in the case of the sphere, this expression takes the form of the product of an ST factor σ and a pre-exponential factor ζ, as shown in figure 3.3B (white) in the freely rotating limit (red lines) and in the restricted diffusion limit for Δ=5s (blue lines). For the non-rotating case, when either $D_R\Delta$ or αr is small, ζ tends to unity and equation 3.16 reduces to equation 3.2. As a resonance from the centre of a given rod will not undergo displacements due to rotational diffusion, the rod ζ function is seen to decay more slowly with field strength and length than is the case of a point from the centre of mass as shown in figure 3.3. Rods are hence intrinsically less sensitive to the effects of rotation than spheres. In the limit where $\alpha \to 0$, we can expand the trigonometric functions with their Taylor series, $T_0 = \lim_{\alpha L\to 0} T_i = LS_0$, yielding:

Geometry	f_T	f_R	ζ	$\lim_{\alpha r\to 0}\zeta$
Sphere	$6\pi\eta r$	$8\pi\eta r^3$	$\frac{\sin^2 \alpha r}{\alpha^2 r^2}$	1
Rod	$3\pi\eta L\frac{1}{\ln(L/r)-0.3}$ [131]	$\frac{\pi\eta L^3}{3\ln\frac{L}{2r}}$ [132]	$\frac{\cos(\alpha L)-1+\alpha L\int_{t=0}^{\alpha L}\frac{sin(t)}{t}dt}{\frac{1}{2}\alpha^2 L}$	L

Table 3.1: Translational, f_T, and rotational, f_R, friction factors, pre-exponential factors ζ, and limiting pre-exponential functions $\lim_{\alpha r\to 0}\zeta$ for idealised geometries. η is the viscosity, r is the radius and L is the rod length

$$\ln\frac{T_i}{T_0} = -D_T^{\text{rod}}\Delta\alpha^2 + \ln\frac{\cos\left(\alpha L(1-e^{-2D_R\Delta})^{1/2}\right) - 1 + \alpha L(1-e^{-2D_R\Delta})^{1/2}\int_{t=0}^{\alpha L(1-e^{-2D_R\Delta})^{1/2}}\frac{\sin(t)}{t}dt}{\frac{1}{2}\alpha^2 L^2(1-e^{-2D_R\Delta})} \tag{3.17}$$

In the limit where αL is small, we can expand this expression to give:

$$\begin{aligned}\lim_{\alpha r\to 0}\ln\frac{T_i}{T_0} &= \left(-D_T^{\text{rod}} - \frac{L^2(1-e^{-2D_R\Delta})}{36\Delta}\right)\Delta\alpha^2 = -D_{\text{eff}}^{\text{rod}}\alpha^2\Delta \\ D_{\text{eff}}^{\text{rod}} &= D_T^{\text{rod}} + \frac{L^2(1-e^{-2D_R\Delta})}{36\Delta}\end{aligned} \tag{3.18}$$

and D_{eff} is a function of Δ when $\frac{L^2}{36\Delta} > D_T^{rod}$.

3.2.3 Calculation of NMR diffusion data

Using theoretical models for D_T and D_R we are now in a position to calculate NMR diffusion data under conditions where these models are valid. Calculated friction factors, f, are related to diffusion coefficients through the Einstein relation $D = \frac{kT}{f}$, where T is the thermodynamic temperature and k is the Boltzman constant. Friction factors derived from theory for hard spheres, rigid rods and spheroids, and the derived ζ functions for the appropriate geometries are summarised in figure 3.1, using the parameters stated in the Methods section.

Simulated NMR intensity data, scaled as $\frac{1}{\alpha^2\Delta}\ln S_i/S_0$ with $\alpha = G_{\max}\gamma\delta$, for rods of 2.4$\mu$m are shown in figure 3.4A, plotted against the squared percentage of the maximum gradient $G_{\max}$ that can be applied, the experimental parameter that is explicitly varied during typical experiments. When plotted in this way, all data will lie on a straight line of gradient $-D_T$ if the ST equation is obeyed, a situation observed in the non-rotating limit for rigid rods (figure 3.4, green). When combined rotational and translational diffusion are considered, the ST equation is no longer sufficient, and the data no longer fall on a single line. D_{eff} can be taken as the slope of each plot (figure 3.4B). This slope is found to be a function

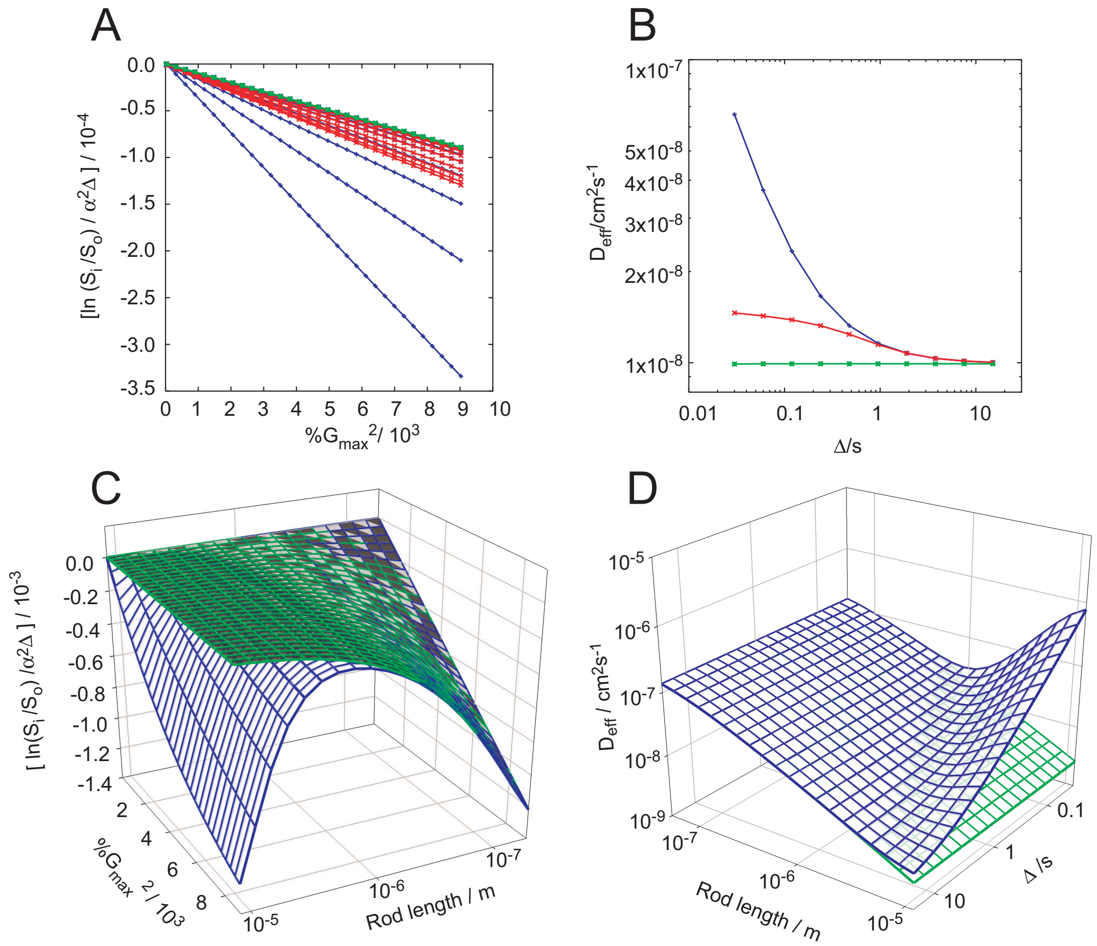


Figure 3.4: Simulated NMR intensity data and D_{eff} for rigid rods. **A -** Intensity data as a function of $\%G^2_{\text{max}}$ for rods of length 2.4μm, and Δ ranging from 30ms to 15s in the freely rotating limit (blue), under restricted rotation (red), and in the static limit (green). As plotted, the gradient of the intensity data is independent of Δ in the static regime. **B -** D_{eff} against Δ obtained from taking the gradients of the plots in A. D_{eff} is a function of the experimental diffusion delay. **C -** Simulated data for Δ=100ms of increasing length in the freely rotating limit (blue) and static limit (green). **D -** Variation of effective diffusion coefficients with rod length and Δ. As the rod length exceeds 1μm, D_{eff} in the freely rotating limit is significantly larger than that expected from the effects of translational diffusion alone.

of the diffusion delay when rotational effects are incorporated in the model.

Diffusion data with Δ of 100ms, a typical delay employed in PFGSE measurements, are shown in figure 3.4C. When rotation is neglected, the diffusion data scale as expected (green), with the slope decreasing with the length. When rotations in the freely rotating limit are considered (blue), for rods of length greater than ca. 1μm, the slope increases as the rotational term starts to dominate and D_{eff} increases with rod length. The value of D_{eff} as a function of rod length and Δ is shown in figure 3.4D. For longer rods, and shorter diffusion delays, D_{eff} is significantly larger than D_T. As the displacement due to translational motion increases with time, at large values of Δ, the ST factor dominates and D_{eff} tends to D_T. This is equivalent to noting that at longer times, NMR signals from the rapidly diffusing rod ends will have been significantly attenuated and so the only signals observed will be from the centre of the rod. In addition, when the system under study has a heterogeneous size distribution, account for this can be taken by multiplying equation 3.16 by a size distribution function and then summing over the distribution. This procedure is described in supplementary section 3.6.5.

3.3 Experimental

These effects were initially considered to explain unexpected results derived from HMQC/PFG experiments where measurements obtained by integrating signal intensity over individual resonances can yield diffusion coefficients that varied by up to an order of magnitude over the different residues of SH3. In considered possible reasons for this, a residue specific diffusion model was derived in order to describe variation in motion within a molecule. It was shown that the observed diffusion coefficients could only vary by one part in 10^6 using this model as the radius of gyration of SH3 is ca. 2 nm. The origin of the variance was subsequently determined to be an error in the pulse program. However, the concept was resurrected in studies of $(SH3)_2Cyt$ fibrils.

TEM analysis of fibrils of apo-cytochrome b_{562} fused to a tandem repeat of an SH3 domain ($apo-(SH3)_2Cyt$), prepared and purified as described in chapter 2, reveals structures typical of amyloid assemblies, with widths of ca. 10nm, and a range of lengths in the vicinity of 2μm (figure 3.5A). In addition these fibrillar species yield well resolved solution-state NMR spectra as a result of the fast dynamics of the cytochrome component in the fibrils that is substantially unfolded (figure 3.5B). The dynamics required for such a group to yield solution state NMR resonances are described more fully in chapter 6.

The values of D_T obtained using PFGSE experiments and the Stejskal-Tanner equation, with Δ=100ms were found to be sample dependent (figure 3.6C) and

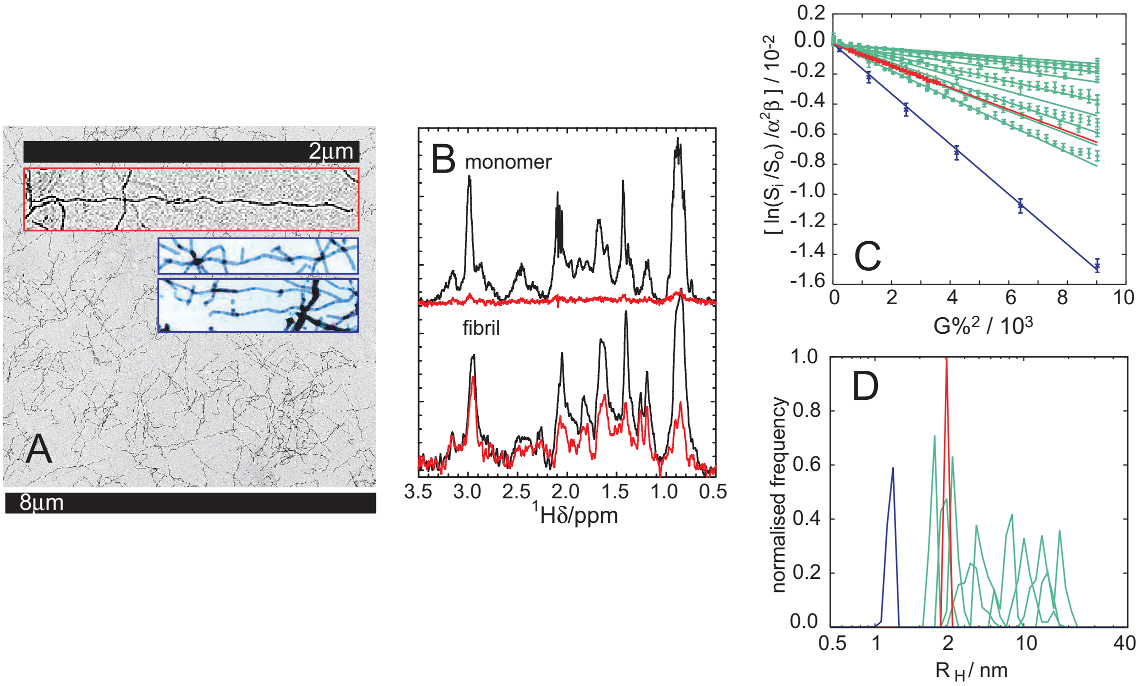


Figure 3.5: **A -** Images of $apo-(SH3)_2Cyt$ fibrils. Inset - TEM (red) and AFM (blue) images of individual fibrils. **B -** 1H PFGSE NMR spectra from monomeric (top) and fibrillar (bottom) $apo-(SH3)_2Cyt$. In both cases Δ=1s and δ=5.4ms with 10% (black) and 50% (red) of the maximum gradient (32 Gcm^{-1}). **C -** integrated PFGSE profiles recorded with Δ=100ms using samples of $apo-(SH3)_2Cyt$ fibrils at various incubation times (green), monomeric $apo-(SH3)_2Cyt$ (red), and human lysozyme (blue). Data are fitted to the ST equation. **D -** an alternative representation of the same data, converting D_T into R_H, with the width of the trace indicating the experimental uncertainty.

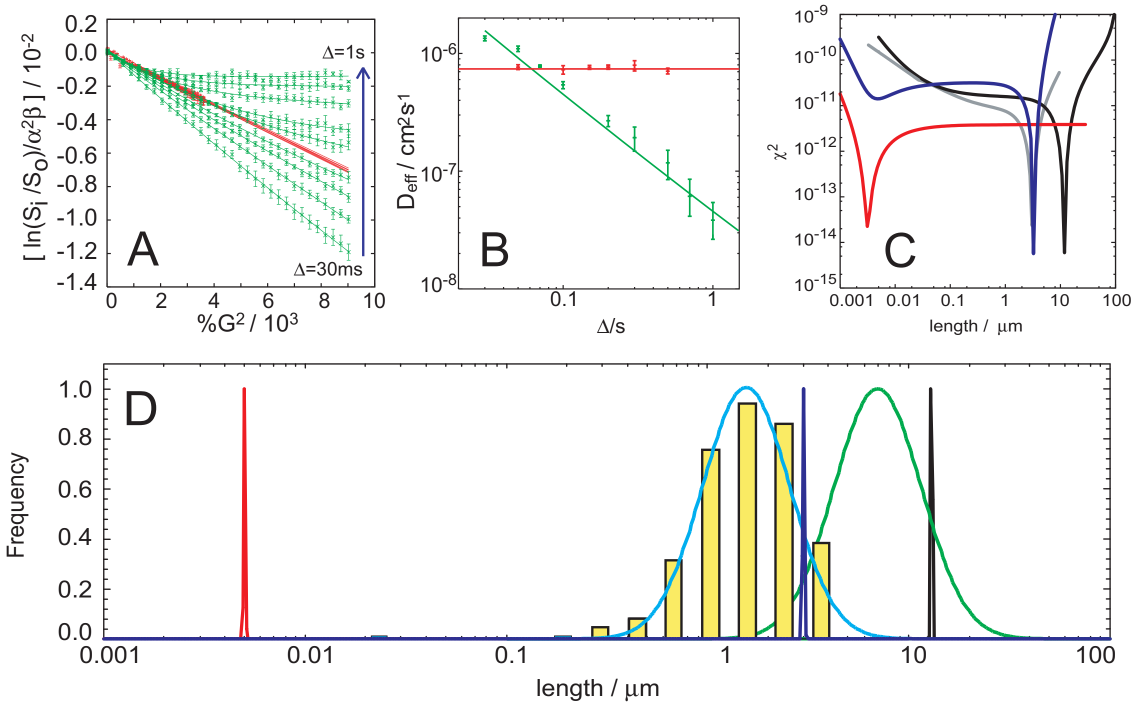


Figure 3.6: **A -** Integrated PFGSE profiles as a function of field strength for a range of values of Δ. Data from monomeric $apo-(SH3)_2Cyt$ (red) fit well to the ST model, but those from fibril samples show marked deviations (green). **B -** The decay constants of the datasets plotted against the diffusion time. For the fibrils, the larger of the two coefficients (blue) is of similar magnitude to that of the monomer (red) and is effectively independent of diffusion delay. The smaller decay constant (green) for the fibrils is however strongly dependent on this delay; the fit is to equation (1) as described in the text. **C -** χ^2 values fitting to the fibril D_{eff} vs Δ dataset, as a function of length, for full length (black) and rod end (dark blue) models (using D_T^{rod}) and for the rod end model (grey) using D_T^{sphere}. As the fit is most sensitive to the gradient of the D_{eff} vs Δ plot, the fitted length estimate with the lowest χ^2 is insensitive to the model used for D_T. These contrast with the χ^2 values obtained from fitting monomeric $apo-(SH3)_2Cyt$ data (red) using D_T^{sphere}. **D -** Histogram (yellow) showing the length distribution obtained from statistical analyses of AFM/TEM images. Fits using the method described here to single and multiple length distributions are shown for the full length (black, green) and rod end (dark blue, light blue) models (see text). Fitting the NMR data using a conventional analysis (red) dramatically underestimates the fibril length.

the values of R_H (figure 3.5D, 2-12nm) determined from D_T from the Einstein-Stokes equation (section 3.2.3), were approximately three orders of magnitude less than the fibril lengths observed by TEM. In particular, samples containing particularly large fibrils yielded diffusion coefficients comparable with monomeric protein. This observation stimulated the theory described in this chapter, leading to the prediction that D_{eff} should vary with Δ.

When NMR spectra were acquired with an exceptionally long diffusion delay (Δ=1s), the resonances of monomeric $apo-(SH3)_2Cyt$ were clearly observed at 10% G_{max} but were entirely suppressed by 50% G_{max} (figure 3.5B, top). Remarkably, only modest attenuation of the signal intensity was observed between the spectra of fibrillar apo-(SH3)2Cyt at 10% and 50% G_{max} (figure 3.5B, bottom). Fitting these data as above yields an RH≈165 nm, a result still inconsistent with the lengths observed by TEM.

The effects of varying the diffusion delay on the intensities of resonances in the NMR spectra of $apo-(SH3)_2Cyt$ in its fibrillar and monomeric states are shown in figure 3.6A. When plotted as shown, ST behavior predicts all intensity profiles to lie on a single line for all diffusion delays. Such a situation is found for monomeric $apo-(SH3)_2Cyt$ (red), where a diffusion coefficient corresponding to R_H = 2.8 ± 0.8 nm is obtained (figure 3.6B, C and D red), but not for fibrillar apo-(SH3)2Cyt (green). In order to analyse the fibril data, a function with two exponential terms was fitted to the intensity profiles (figure 3.6A green), yielding two effective diffusion coefficients. The larger of these coefficients is closely similar to that of monomeric $apo-(SH3)_2Cyt$, and can be attributed to a small population of monomeric material that exchanges with the fibrils but on a timescale of weeks. The smaller of the effective diffusion coefficients for fibrillar $apo-(SH3)_2Cyt$ (figure 3.6B green) becomes less than that of monomeric $apo-(SH3)_2Cyt$ only when $\Delta >$100ms.

The experimental D_{eff} vs Δ data can be fit to the models described in this chapter. This process has been automated using home built software and can be performed in several different ways, generating $\chi^2 = (D_{\text{eff}} - D_{\text{sim}})^2$ surfaces similar to those shown in figure 3.6C. Assuming that the NMR signal originates from molecules distributed evenly over the length of the fibril, equation 3.16 for 'full length rods' can be used. Here, the best fit (figure 3.6B green line) corresponds to a length of 12μm (figure 3.6D black), a value comparable to the longest fibrils seen in AFM/TEM experiments.

Analysis of AFM/TEM images shows that fibrils studied here are not of uniform length and that the length distributions measured using these two techniques are statistically identical. Heterogeneous length distributions can be taken into account using equation 3.27, using a log-normal function to describe the length distribution (equation 3.30). This distribution function is specified by two independent variables a_0 and a_1, which together describe the width and centre point of the distribution (section 3.6.5). The resulting length distribution (a_0=9μm, a_1=0.6) obtained from fitting this model to the NMR data (figure 3.6D green) has χ^2 values significantly lower than that for the single rod model, and the most probable length (9μm) is closer to that observed by AFM/TEM (2μm).

One possible explanation for the discrepancy between the lengths estimated by AFM and TEM and the length estimated using the rod end fit from the NMR data is that by fixing these structures to a surface for imaging may alter the structure. For example, it has been shown that the length of microtubule fibril preparations can get shorter after vigorous pipetting [133].

The NMR signal intensity per mole of monomer for fibrillar $apo-(SH3)_2Cyt$ samples is much less (ca. 15%) than that of a monomeric sample (section 2.4.7). One explanation for this finding is that the observed NMR signals originate predominantly from the fibril ends, which will have greater freedom of movement than the central regions [104], giving rise to narrower linewidths. In this case, equation 3.10 can be used ('rod end model'). Remarkably the calculated length distribution of the fibrils (figure 3.6D light blue, a_0=2μm, a_1=0.55) is then essentially identical to that measured by AFM/TEM (figure 3.6D histogram). The magnitude of the lowest χ^2 values are similar for both the 'rod end' and the 'full length rod' fits (figure 3.6C).

In conclusion, measurements using a conventional diffusion analysis of NMR diffusion data (figure 3.5D) underestimate fibril lengths by three orders of magnitude, in the system studied here. All fits incorporating rotational diffusion effects (figure 3.6D) results in lengths significantly closer to the distributions observed by AFM/TEM, and with reasonable assumptions are essentially identical to the latter. These results provide a clear example of the effects of rotational diffusion on NMR PFGSE measurements and reveal the potential value of this phenomenon.

The approach described here is applicable to the study of any large molecular structure by solution-state NMR provided that at least one dimension exceeds ca. 500 nm, and that the local molecular dynamics permit resonances to be observed. Experimental D_{eff} vs Δ can be fit to equations 3.11 or 3.17 with suitable models of D_T and D_R such as those present in table 3.1 in order to obtain an estimate of the length scale of the system under study. Several systems are studied in this thesis using these methods, summarised in table 3.2. The technique, therefore, provides a powerful method for probing non-invasively, the distribution of sizes of large molecular assemblies, in solution, and indeed for observing changes in size as a function of time and solution conditions.

System	Section	Rod end fit / μm	Full rod fit / μm
$(SH3)_2Cyt$ fibrils	2.4.8	3.6	15.3
$(SH3_3Cyt$ fibrils	2.4.8	1.8	6.6
αB-crystalin fibrils	4.2.2	1.9	7.2
Human lysozyme fibrils	4.2.2	2.0	7.4
Insulin fibrils	4.2.2	2.4	8.9

Table 3.2: Examples in the thesis where models from this chapter have been used to rationalise experimental data. 'Rod end' and 'full length' best fits are given.

3.4 Discussion

NMR resonances originating from even large macromolecular structures can be observed using conventional solution state experiments when regions of the structure have sufficient mobility to average out local inhomogeneities and residual dipolar interactions. Such 'motional narrowing' leads to narrow solution state NMR linewidths that will normally be observed from these mobile regions at chemical shifts that are approximately those expected for a random coil. One example of such a case is that of amyloid fibrils for which NMR resonances from flexible non-core regions can be observed[51, 50]. While characterising these regions using NMR pulsed field gradients, we observed that measured apparent diffusion coefficients of the fibrils were comparable to those of monomeric protein molecules, rather than to a large molecular assembly.

In light of this observation we have theoretically explored the contribution that rotational diffusion of large assemblies can make to NMR diffusion measurements. Rotational diffusion can lead to displacements that are comparable to, or even larger than those due to translational diffusion. D_{eff} can be seen to be a function of Δ and α. This variance is qualitatively similar to the case of restricted diffusion, where a similar $\text{sinc}^2\alpha R$ function [134] to that in equation 3.11 is observed, although it has a different origin. By observing the variance of D_{eff} with Δ and α, the dimension of the macromolecular species under study can be determined. Application of this approach to calculate length distributions from analysis of the diffusion data has been shown to agree well with those derived from AFM and TEM measurements [94]. In addition, recent work has shown that flexible saccharide groups are present on the surface of starch granules (appendix 3.6.7) and these are the first saccharides in the starch granule to be removed during digestion.

We can therefore conclude that NMR diffusion measurements coupled with the theory described in this work provide a powerful method for probing, non-invasively, the dimensions of large molecular assemblies in solution, and provide a means for observing changes in such systems as a function of time and solution conditions. We therefore believe that this strategy should be of considerable value in the context of macromolecular structure and assembly.

3.5 Methods

3.5.1 Calculations

The model was coded into C++ using functions from the GNU Scientific Library (GSL)[135]. Plots were prepared using GNUplot 4.0 with scripts generated by the program, and Adobe Illustrator 10. The program, including routines for fitting experimental data to the model is available on request. All calculations were performed with η = 0.1cP, T=300K and δ=5.4ms. Using sinusoidal bi-polar gradients. On a Bruker DRX-500, using sinusoidal bi-polar gradients, G_{max}=32 G cm^{-1}.

3.5.2 Monte-Carlo simulations

The triangular rotational distribution functions were simulated using a Monte Carlo method. By drawing angles with a sinusoidal distribution between 0 and 2β, with β defined by $0.5(1 + \cos\beta) = \exp(-2D_R\Delta)$, histograms of $\delta z^2_{\mathrm{rot}}$ were generated using the theoretical models for D_R defined in figure 3.1. Δ was varied from 100ms to 10s with typically 60,000 runs for each (figure 3.2D, inset). The triangular distributions in the freely rotating limit were obtained by drawing start and end angles at random, on the surface of a unit sphere.

3.6 Appendix

3.6.1 Displacement due to translational and rotational diffusion

Rotational D_R and translational D_T diffusion coefficients are related to a friction factor f, by the Einstein relation, $D = \frac{kT}{f}$, where k is the Boltzman constant and T is temperature. For small hard spheres, the translational friction factor is given by Stokes' law as $6\pi\eta R_H$, where η is the viscosity of the solution and R_H is the hydrodynamic radius of the sphere. The Einstein-Stokes equations for translational and rotational diffusion are:

$$\begin{aligned} D_T^{sphere} &= \frac{kT}{6\pi\eta R_H} \\ D_R^{sphere} &= \frac{kT}{8\pi\eta R_H^3} \end{aligned} \tag{3.19}$$

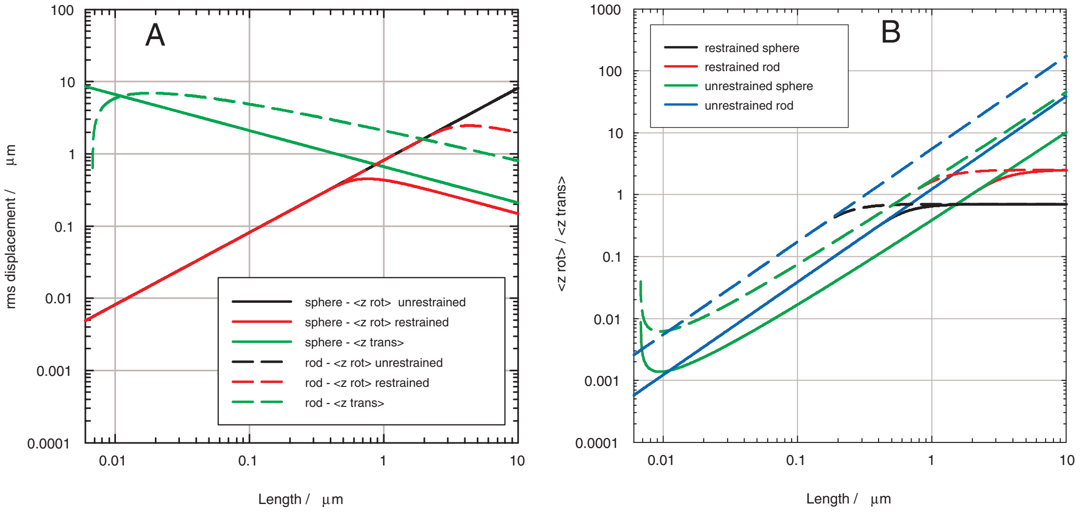


Figure 3.7: Plots of rms displacements due to rotational and translational diffusion for spheres and rods for $T = 300K$ and $\eta = 0.1cP$. **A -** Rms displacement due to rotational and translational diffusion for rods and spheres, where $\Delta = 1s$. **B -** Ratio of rotational to translational diffusion for rods and spheres. Solid lines are for $\Delta = 1s$ and dashed lines are for $\Delta = 50ms$

The microscopic motions of particles due to translation and rotation are directly related to their diffusion coefficients. For translation, the mean squared displacement in one dimension is given by $< z_{\text{trans}} >= \sqrt{2D_T\Delta}$. For rotational diffusion, the mean squared angular displacement is given by $< \theta^2 >= 4D_R\Delta$. As considered in section 3.6.3, the displacement in one dimension due to rotational diffusion, in the limit where rotation is completely free $< z_{\text{rot}} >_{\text{free}}= \sqrt{\frac{2}{3}R^2}$, and for a species restrained by rotational diffusion $< z_{\text{rot}} >= \sqrt{\frac{2}{3}\left(1 - \exp(-2D_R\Delta)\right) R^2}$. Thus the completely free limit applies when $2D_R\Delta$ is large. Displacements due to rotational diffusion can be estimated where the friction coefficient can be estimated. Friction factors for common geometries are given in table 3.1. These displacements are plotted and compared for rods and spheres in figure 3.7.

We note in this analysis that:

- A rod of length L will experience a greater translational rms displacement than a sphere of radius $R = L$.
- The rms displacement due to rotational diffusion becomes comparable to that of translation diffusion for rods of 1μm, and exceeds it for rods longer than 2μm.
- The rms displacement due to rotational diffusion becomes comparable to that of translational diffusion for spheres of radius 1μm, but never exceeds the rms displacement due to translational diffusion.

- Decreasing Δ reduces the length where the rms displacement for translation becomes comparable to the rms displacement for rotation.

For rods, the limit $\lim_{L\to\infty}\frac{<z_{rot}>}{<z_{trans}>} = \sqrt{6}$ is reached where $L = 1\mu m$ for $\Delta = 100ms$ and by $L = 5\mu m$ for $\Delta = 1s$. For spheres, $\lim_{L\to\infty}\frac{<z_{rot}>}{<z_{trans}>} = \sqrt{\frac{1}{2}}$. This limit is reached for spheres of radius 200nm where $\Delta = 100ms$, and for spheres of radius $1\mu m$ where $\Delta = 1s$. In this limit, the formula for restricted rotational diffusion should be applied. The relative contribution of rotational motion to the total rms displacement will be greater in general for a rod than a sphere.

When the characteristic length is of the order 1μm for both rods and spheres, the magnitudes of rms displacement due to rotational and translational diffusion are comparable. Rotational effects will then be manifested in the NMR diffusion experiment. Calculating expressions for the displacement distribution function will allow expressions for PFGSE signal decay to be derived where a particle undergoes both rotational and translational diffusion using equation 3.1.

3.6.2 Explicit displacement functions

Where $b = 2D_T\Delta$, the freely rotating displacement function from equation 3.5 is given by equation 3.20, for a point a distance a from the surface of a sphere.

$$P(z) = \int_{\theta_1=0}^{\pi}\int_{\theta_2=0}^{\pi}\frac{1}{4\sqrt{b\pi}}\sin\theta_1\sin\theta_2 e^{-\frac{(z-a(\cos\theta_2-\cos\theta_1))^2}{b}} \tag{3.20}$$

For a rigid rod, by integrating over a from $-L/2$ to $L/2$, we arrive at the following in terms of error functions, plotted in figure 3.2C:

$$P_{\text{rod}}(z) = \frac{\sqrt{\pi}}{4}\left(\text{Erf}\left(\frac{2r-z}{2D_T\Delta}\right) + \text{Erf}\left(\frac{2r+z}{2D_T\Delta}\right)\right) \tag{3.21}$$

3.6.3 Effects of restricted rotation on a sphere

Proof of equation 3.7. Calculating the displacement due to rotation of a point on a sphere undergoing Brownian diffusion.

The vector $\overline{u}$ is initially at an angle θ to the z axis. After a time Δ, it has rotated through α. The end vector is given by $\overline{u(t)}$ where angle ψ sweeps through the cone defined by $\overline{u(0)}$ and α, as shown in figure 3.8:

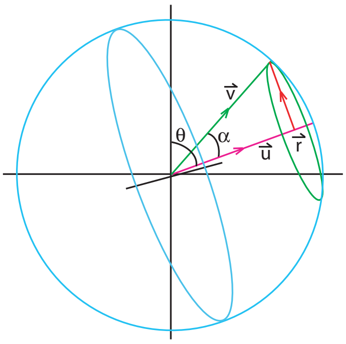


Figure 3.8: The vectors that define the rotation of a point on the surface of a sphere.

$$\begin{aligned}\overrightarrow{u(0)} &= \cos\theta\bar{e}_z + \sin\theta\bar{e}_x \\ \overrightarrow{u(0)}_\perp &= \sin\theta\bar{e}_z - \cos\theta\bar{e}_x \\ \overrightarrow{v} &= \cos(\theta-\alpha)\bar{e}_z + \sin(\theta-\alpha)\bar{e}_x = \cos\alpha\overrightarrow{u(0)} + \sin\alpha\overrightarrow{u(0)}_\perp\end{aligned} \tag{3.22}$$

Rotating $\overrightarrow{v}$ by ψ about $\overrightarrow{u(0)}$ gives $\overrightarrow{u(t)} = \cos\alpha\overrightarrow{u(0)} + \sin\alpha(\cos\phi\overrightarrow{u(0)}_\perp + \sin\phi\bar{e}_y)$, the position on the surface of the sphere at time t. Calculating the square displacement along the z axis, and averaging over the cone gives:

$$\begin{aligned}\frac{1}{r^2}\int_{\psi=0}^{\pi}(z(t)-z(0))^2 d\psi &= \int_{\psi=0}^{\pi}\left(\cos\alpha\cos\theta + \cos\psi\sin\alpha\sin\theta - \cos\theta\right)^2 d\psi \\ &= \cos^2\theta(\cos\alpha-1)^2 + \frac{1}{2}\sin^2\alpha\sin^2\theta\end{aligned} \tag{3.23}$$

Averaging over all start positions

$$\begin{aligned}\frac{<z_{\text{rot}}^2>}{r^2} &= \frac{\int_{\theta=0}^{\pi}\sin\theta\left(\cos^2\theta(\cos\alpha-1)^2 + \frac{1}{2}\sin^2\alpha\sin^2\theta\right)d\theta}{\int_{\theta=0}^{\pi}\sin\theta d\theta} \\ &= \frac{2}{3}(1-\cos\alpha) = \frac{2}{3}\left(1-e^{-2D_R\Delta}\right)\end{aligned} \tag{3.24}$$

This result is equation 3.7. In the limit $t \longrightarrow \infty$, this expression reduces to equation 3.4. By comparing equation 3.7 to equation 3.4 we can make the approximation $r_{\text{eff}} = r\sqrt{1-e^{-2D_R\Delta}}$. At shorter diffusion times, the displacement caused by rotation is reduced as a result of the finite time required for Brownian rotational diffusion.

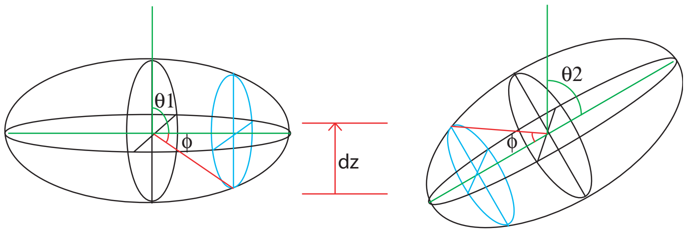


Figure 3.9: The effect of sets of equivalent rotations on δz for ellipsoids.

3.6.4 Displacement function for spheroids

An ellipsoid is defined by the equation $1 = \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2}$, and the spheroid is the case were $b = c = na$. The case where $n < 1$ is therefore prolate, and where $n > 1$, the spheroid is oblate. Considering points along the x, y plane, all points are a distance $r = \sqrt{z^2 + x^2} = \sqrt{n^2 a^2 + z^2(1 - n^2)}$ from the centre of mass. Each distance should be weighted by the circumference of the circle defined by constant ϕ as shown in figure 3.9, and so the degeneracy $2\pi\sqrt{n^2(a^2 - z^2)}$. Using equation 3.5, substituting $r = \sqrt{n^2a^2 + z^2(1 - n^2)}$:

$$P_{\text{total}}^{\text{spheroid}}(\delta z) = \frac{\int_{z=0}^{z} \sqrt{n^2(a^2 - z^2)} P_{\text{total}} dz}{\int_{z=0}^{z} \sqrt{n^2(a^2 - z^2)} dz} \tag{3.25}$$

The mean squared displacement is then given by:

$$< \delta z_{\text{rot}}^2 >_{\text{spheroid}} = \frac{2}{3} \frac{\int_0^a \sqrt{n^2(a^2 - z^2)}(n^2a^2 + z^2(1 - n^2))dz}{\int_0^a \sqrt{n^2(a^2 - z^2)}dz} = \frac{1}{6}\left(3a^2n^2 + a^2\right) \tag{3.26}$$

Which in the limit where $n \to 1$, reduces to equation 3.6, as expected.

3.6.5 Effects of a population distribution on NMR observables

In a heterogeneous sample, we can expect a distribution of the lengths of rods or the diameters of spheres. Certainly in a solution of amyloid fibrils, the lengths are observed to vary over over several orders of magnitude. Account must be taken of this variation if the model is used to analyse experimental data. In all cases, with a prefactor $\zeta(l, G, \delta, \Delta)$, and the Stejskal-Tanner factor $\sigma(G, \delta, \Delta, l)$, where the sample has a distribution of lengths given by the function $C(l)$, the PFGSE signal intensity decay is given by:

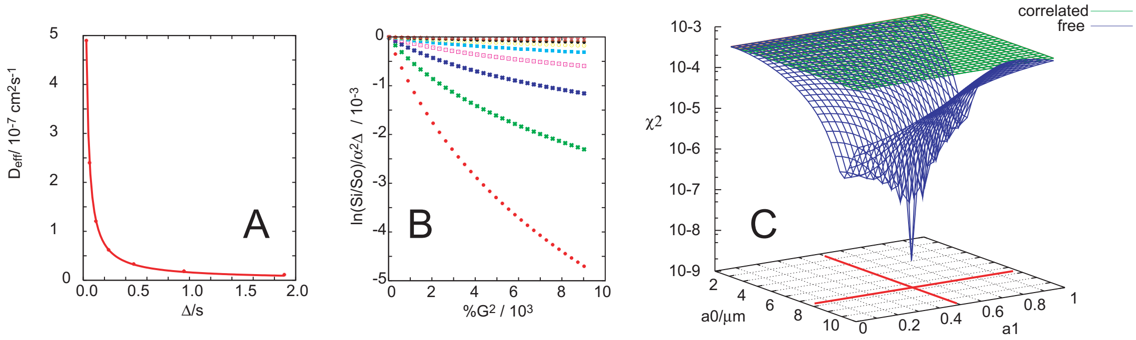


Figure 3.10: NMR diffusion data for rods with a log-normal distribution specified by a_0=6.67μm and a_1=0.5 was calculated, shown A and B. **C -** A χ^2 surface was subsequently calculated, comparing the simulated datasets to calculated data. The minimum on this surface corresponds to a_0=6.67μm and a_1=0.5, demonstrating that data can be quantitatively fit to the model to arrive at a unique length distribution.

$$\ln\frac{S_i}{S_0} = \ln\frac{\sum_{l=0}^{\infty} C(l)\zeta(G,\delta,\Delta,l)\sigma(G,\delta,\Delta,l)}{\sum_{l=0}^{\infty} C(l)\lim_{G\to 0}\zeta(l)} \tag{3.27}$$

In the limit where $C(l)=\delta_l$, this expression reduces to the limits derived in the previous sections. When either ΔD_R or αl is small, the pre-factor is lost and the expression reduces to:

$$\ln\frac{S_i}{S_0} = \ln\frac{\sum_{l=0}^{\infty} C(l)\lim_{\alpha l\to 0}\zeta\sigma}{\sum_{l=0}^{\infty} C(l)\lim_{\alpha l\to 0}\zeta} \tag{3.28}$$

For rigid rods, in this limit we get the following intuitive result:

$$\ln\frac{S_i}{S_0} = \ln\frac{\sum_{L=0}^{\infty} C(l) l e^{-D_T^{\mathrm{rod}}\Delta\alpha^2}}{\sum_{l=0}^{\infty} C(l) l} \tag{3.29}$$

Which states that as each monomer in the rod contributes to the NMR signal, the contribution of each rod of length l is weighted by a the number of monomers in that rod, which, when normalised, leads to the additional factors of l in the numerator and denominator. Where $C(l)=\delta_l$ this equation reduces to equation 3.18. By acquiring NMR data as a function of G, Δ and δ therefore, we can estimate $C(l)$, the distribution of particle sizes of a given species in solution. Empirically, it is found that the length distribution of amyloid fibrils takes the form of a Gaussian function, skewed towards longer rods. This can be approximated well by a log-normal distribution

$$C(L) = \frac{1}{a_1\sqrt{2\pi}L}\exp\left[-\frac{(\ln(L/a_0))^2}{2a_1^2}\right] \tag{3.30}$$

The maximum of this distribution is given by $L_{max} = a_0 e^{-a_1^2}$, and the full width half maximum is $2\sinh(a_1\sqrt{\ln 4})L_{max}$. This distribution has the benefit that as it is specified by two parameters and so the NMR data can be interpreted as a function of these two parameters. Figure 3.10 shows sample NMR from this model for an arbitrary distribution of a_0=6.67E-6μm and a_1=0.5. D_{eff} again is a function of Δ as shown in A, and the intensity decay is well approximated by a single exponential, as shown in B.

A more detailed analysis is shown in figure 3.11. NMR data is calculated for rods over a range of a_1, a_0 and Δ. The decay of the simulated datasets with increasing G are fit to a single exponential to obtain D_{eff}. The curvature is estimated by the coefficient of the x^2 term in a fit to a second order polynomial. Data from the rotating model are shown in blue, and data from the non-rotating model are shown in green.

Figure 3.11A shows how the effective diffusion coefficient varies with the length distribution function. The slice taken along the plane where a_1 is small is the equivalent to the plots seen in figure 3.4, where the $C(l) = \delta(l)$. Where the rods do not exceed 1μm, the results from the rotational model agree with those not considering rotations. When a_1 exceeds 0.1, we see lower diffusion coefficients than would be expected reflecting increasing numbers of larger species in the distribution. D_{eff} is much larger for rods in the rotating model than in the static case.

The curvature of the diffusion plot can be estimated by fitting it to a second order polynomial and taking the second order coefficient. Figure 3.11B shows the variance of the curvature in the data used for figure 3.11. A curvature of 10^{-8} is still linear by eye within typical experimental uncertainty and so the decay of data is well described a single exponential, D_{eff} providing curvature does not exceed this threshold.

Figures 3.11C and D consider the variance of D_{eff} with Δ. The ratio of D_{eff} at Δ=0.05s to Δ=1s is given in C. Where this equals unity, the two are equal, as found where $l < 1\mu m$. Where $a_1 > 0.3$, even in the static case, there is a sufficient variance in D_T to cause the ratio to tend to 0.5. Thus an observed ratio $>$ 0.5 cannot be explained by translational effects alone. In the rotating case, the magnitude of the effect is significantly larger, evident for species where the length exceeds 1μm. Figure 3.11D examples the curvature of the plots of D_{eff} against

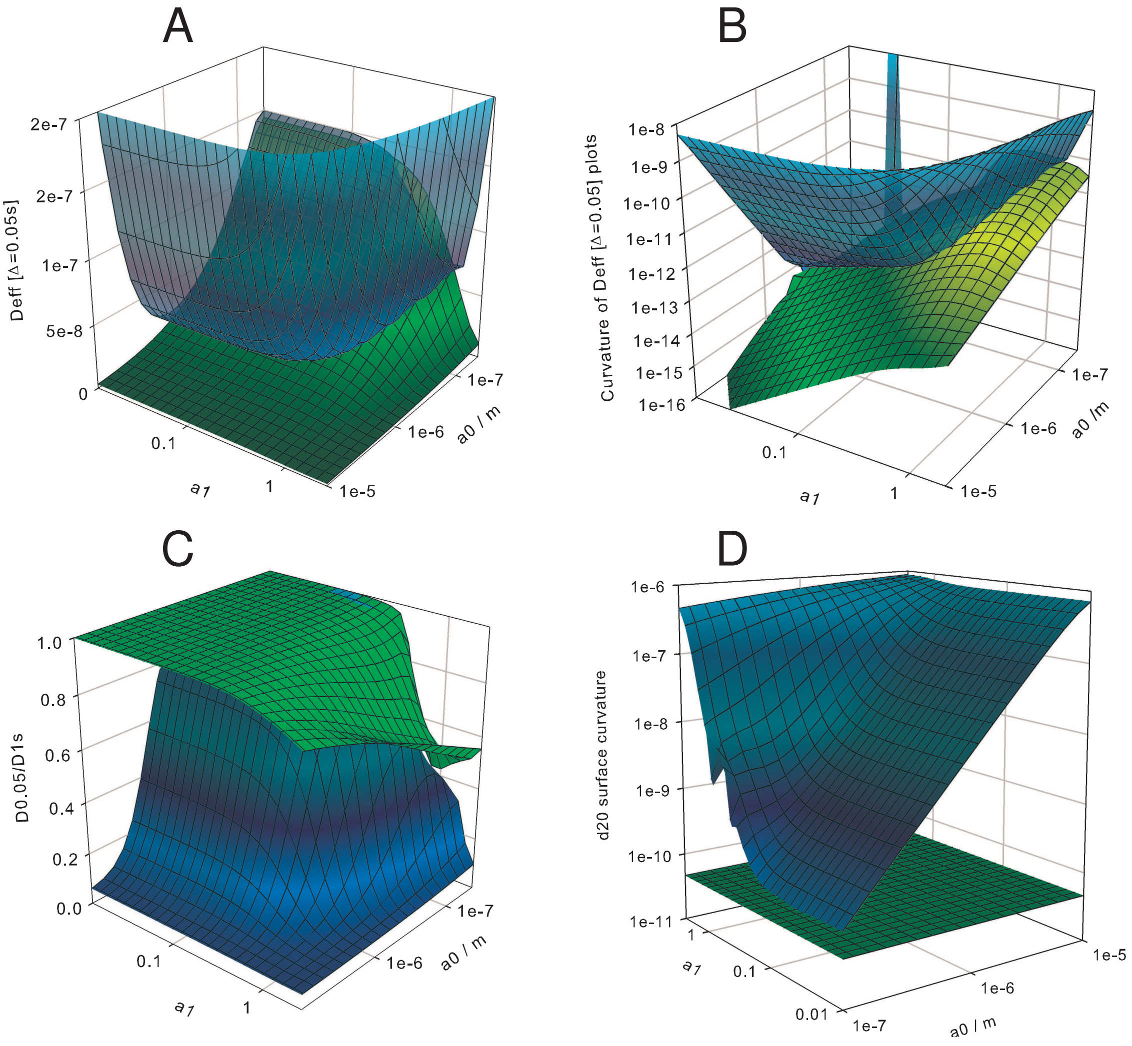


Figure 3.11: The effect of analysing a length distribution by NMR. Calculations using the rotating model are shown in blue, and calculations without rotations are in green. The effect of the correlation term is to move the surfaces between these two limits. **A -** The effective diffusion coefficient with Δ=0.05s. **B -** The curvature of the diffusion plots with Δ=0.05s, estimated as the second order polynomial coefficient. **C -** The ratio of the effective diffusion coefficient measured with Δ=0.05s to the effective diffusion coefficient measured with Δ=1s. **D -** the curvature of a plot of effective diffusion coefficient against Δ.

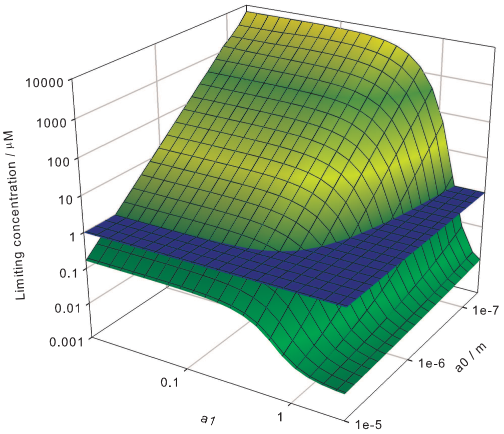


Figure 3.12: The maximum concentration that can be used before the solution can no longer be considered dilute for rods of radius 8nm. The blue surface is fixed at 1μM, the lowest conceivable working concentration for NMR experiments of this type.

Δ. The static plot gives a linear plot whereas where rotational effects, there is pronounced curvature.

The principles that emerge from analysing these effects using a trial function such as the log-normal distribution are general and the same analysis has been performed with several other distribution functions. For experiments performed at sufficient values of gradient strength and diffusion time, the data can be fitted to the models described above, arriving at an approximation of the length distribution present in the sample. A program has been written to simplify this task and is freely downloadable.

Validity of the dilute regime

Where fibrous polymers are under study, the concentration of monomers is usually known, where the total number of fibres is not. Where $C(l)$ is the observed length distribution, the normalised length distribution is $C(l)_N = NC(l)$ and so the total concentration of rods, $C_F = \sum_{l=0}^{\infty} NC(l)$. The total concentration of monomers in rods of a given length is given by $C(l)_P = \frac{l}{l_1} NC(l)$, where L_1 is the number of monomers per unit fibre length. From this we can calculate the total concentration of monomeric protein, $C_P = \sum_{l=0}^{\infty} \frac{l}{l_1} NC(l)$. Rearranging this gives us an expression for N, allowing C_F and $C(l)_N$ to be calculated.

We can use this to estimate whether we are in a dilute or concentrated regime. A single fibre will sweep out a volume of $4/3\pi(l/2)^3$. If in solution there is sufficient volume such that the spheres do not touch, the solution can said to be dilute. If the spheres overlap, then the solution will be in the semi-dilute regime. The frictional coefficient for the diffusion of rod in solution assumes the fibrils to be in the dilute regime. The occupied volume is given by the following, where the concentration unit is mol dm^{-3} and $V(l)_1$ is the volume of a single fibre and N_a is Avogadro's number:

$$\begin{aligned} V_{\%occupied} &= \sum_{L=0}^{\infty} V(L)_1 \cdot N_a \cdot 10^5 \cdot C(L)_N = \frac{\sum_{L=0}^{\infty} \cdot \frac{4}{3}\pi \left(\frac{L}{2}\right)^3 \cdot N_a \cdot 10^5 C(L) C_P L_1}{\sum_{L=0}^{\infty} LC(L)} \\ &= \frac{1}{6}\pi N_a \cdot 10^5 C_P L_1 \frac{\sum_{L=0}^{\infty} L^3 C(L)}{\sum_{L=0}^{\infty} LC(L)} \end{aligned} \tag{3.31}$$

The occupied volume must be less than 100% for the fibrils to be in the dilute regime. Considering again the log-normal distribution, this puts limits on the working concentration in the NMR sample. The limiting concentration where the occupancy equals 100% is shown in figure 3.12 as a function of the distribution parameters a_0 and a_1 for the log-normal distribution. For wide distributions of rods of length greater than one micron, care must be taken to work at the lowest possible concentration. This will have adverse effects on the S/N ratio.

3.6.6 Worm-like chain

The translational diffusion coefficient of the wormlike chain is more complex [136] and is given by the following. The coefficients C_n are functions of the diameter of the chain and its length.

$$f_{worm} = 3\pi\eta L \left(C_1 + \ln\frac{L}{d} + C_2 + C_3 L + C_4 L^2 + C_5 L^3 + C_6 \frac{d}{L} \ln\left(\frac{L}{d} + C_7\frac{d}{L} + C_8\frac{d^2}{L^2} + C_9\frac{d^3}{L^3} + C_{10}\frac{d^4}{L^4}\right) \tag{3.32}$$

The coefficients are as follows.

$$\begin{aligned}
C_1 &= 1 - 0.01412d^2 + 0.00592d^4 \\
C_2 &= 0.3863 - 0.1667d + 0.0016d^2 - 0.0224d^3 - 0.0007d^4 \\
C_3 &= 0.1667 + 0.0222d^2 + 0.0017d^4 \\
C_4 &= 0.01883 - 0.00789d^2 - 0.00038d^4 \\
C_5 &= -0.002039 + 0.000805d^2 + 0.000017d^4 \\
C_6 &= 0.04167d + 0.00567d^3 \\
C_7 &= 0.5 + 0.0786d - 0.0094d^2 + 0.0107d^3 + 0.0039d^4 \\
C_8 &= -0.06250 + 0.00132d^2 - 0.00055d^4 \\
C_9 &= 0.001302d + 0.000181d^3 \\
C_{10} &= 0.001953 - 0.000064d^2 + 0.000027d^4
\end{aligned} \tag{3.33}$$

3.6.7 Starch

Starch granules are semi-crystalline structures composed of the amylose and amylopectin, both of which are polymers of maltose. They are arranged in layers which gives rise to a distinctive 'maltese cross' pattern under cross polarised light. Starch is the main source of carbohydrate in human diet digested primarily by α-amylase [137]. Before heat treatment, starch granules are only partially digested and the structural basis for this is poorly understood. In this study, rice starch granules were studied by solution-state NMR before and after α-amylase addition. The granules are approximately 1 μm in diameter as shown by light microscopy in figure 3.13A, with the characteristic 'maltese cross' pattern obtained when starch granules are viewed under cross polarised light.

^{1}H NMR spectra and integrated PFGSE data from rice starch granules are shown in figures 3.13B, C and D. Maltose resonances are found between 4.2 and 3ppm as no proton is further than two bond lengths from an oxygen atom (figure 3.13B, red). Interestingly, D_{eff} varies as a function of Δ for these resonances (figure 3.13C and D red) suggesting that the residues are attached to a much larger species of sufficient size to undergo rotational diffusion. Of particular interest is the observation that the effective diffusion coefficients obtained where Δ=30 ms is comparable to that of free maltose. Significantly longer diffusion delays where Δ = 500 ms are required before it is apparent that the species is substantially larger than a free saccharide.

After 1 hour of amylase digestion, the peak intensity in this region is found to increase by a factor of ca. 7.5. In addition, as shown in the inset of figure 3.13B, when the spectrum is normalised for total intensity, the maltose spectrum before amylase treatment is broader than the spectrum after digestion, consis-

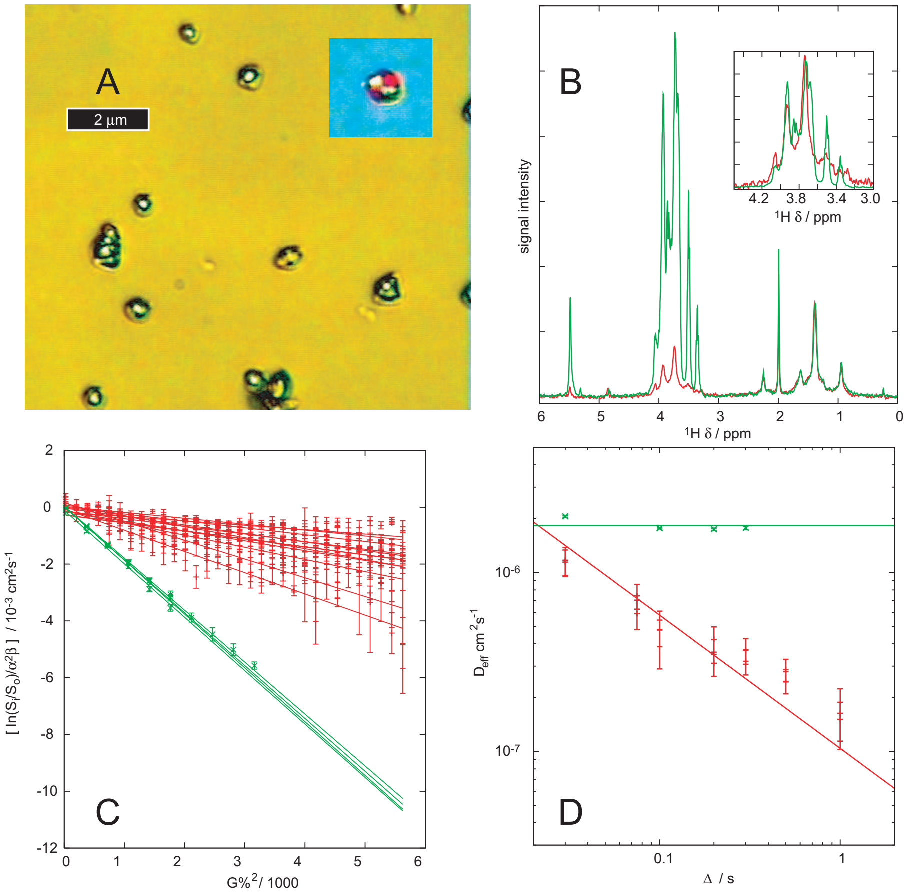


Figure 3.13: NMR behaviour of rice starch granules before (red) and after (green) digestion with α-amylase. **A -** Rice starch granules observed before digestion by light microscopy. When viewed under crossed polarised light, a distinctive maltese cross pattern is observed (inset). **B -** ^{1}H NMR spectra before and after digestion. **C -** Integrated PFGSE data for diffusion delays varied between 30ms and 1s. **D -** The effective diffusion coefficient obtained from the data shown in C as a function of the diffusion delay, Δ. The effective diffusion coefficient from the signal observed after digestion (green) is independent of the diffusion delay and of a value expected for a small molecule. The effective diffusion coefficient after digestion is shown to vary with the diffusion delay, in a manner consistent with the rotational diffusion effects described in this chapter.

tent with the resonances before digestion originating from a much larger species. The NMR diffusion data after digestion show that the maltose resonances after digestion diffuse as expected for free saccharides in solution (figure 3.13C and D, green). Thus the previously observed maltose residues on the surface of the starch granule are cleaved by α-amylase.

The conclusion of this preliminary experiment suggests that α-amylase is preferentially hydrolysing only the more flexible glycosidic linkages bonds on the surface of the starch granule. At higher temperatures it should be possible to determine whether the number of flexible residues increases, and to determine how this correlates with α-amylase activity. This experiment demonstrates the potential utility of the theory described in this chapter in the study of large biological structures.

Chapter 4

Solution-state NMR studies of amyloid fibrils

Abstract

Using PFGSE NMR measurements, it can be determined whether or not flexible regions on the surface of amyloid fibrils can be observed using solution-state NMR spectroscopy. In addition, the concentration of free monomer in a fibril sample can be quantified by centrifuging the fibril sample and measuring the protein concentration in the supernatant. These two tests have been performed for a range of different fibrils and has led to the definition of two apparently mutually exclusive classes of fibrils. Fibrils in the first class have $> 0.1\%$ of their total material as soluble monomer and do not have flexible regions that can be observed using solution-state NMR. Such systems studied here include TTRRGD, SH3 and SH3$_2$. Virtually the entire sequence of such proteins are folded in the β-sheet core of such fibrils. Fibrils in the second class have $< 0.1\%$ of their total material as soluble monomer and have flexible regions on their surface that can be observed using solution-state NMR. Such systems include αB-crystallin, lysozyme, insulin, SH3$_2$Cyt and SH3$_3$Cyt. Remarkably therefore, it is shown that segments not in the β-sheet core of fibrils reduces the tendency of constituent monomers to dissociate from fibrils. The origin of this is most likely to be due to relatively non-specific interactions between non-core regions on the surface of the fibrils. It is shown that these interactions might be expected to either decrease the free energy of elongation, or destabilise the elongation transition state. In both cases, it is estimated that non-core regions must contribute ca. 0.23 kJ mol^{-1} per residue to either the free energy of elongation or the free en-

ergy of the elongation transition state in order to explain the observed trends.

4.1 Introduction

Aggregation of proteins and peptides is a widespread and much-studied problem, with serious implications in contexts ranging from biotechnology to human disease. An understanding of the proliferation of such aggregates under specific conditions requires a quantitative knowledge of the kinetics and thermodynamics of their formation. Aggregation reactions lead to inherently highly heterogeneous products; individual protein molecules have many conformational and aggregated states available to them, the populations of which vary with time. It is a major experimental challenge to distinguish between these states and populations in order to gain an understanding of the underlying aggregation process.

Aggregation reactions are conventionally followed using spectroscopic techniques. Such measurements contain a large quantity of information, averaged over all species in the reaction. For example, fluorescence and CD spectroscopies are often used to follow the average solvent exposure of aromatic residues and secondary structure respectively, and light scattering techniques indiscriminately probe the accumulation of large aggregates in solution. Interpreting these measurements in terms of the individual species involved in an aggregation reaction is difficult to accomplish reliably. Solution-state NMR however, can yield high resolution structural data, and can distinguish small monomeric states from larger oligomers using diffusion measurements [138, 129, 60], the theory of which is discussed extensively in chapter 3.

The current limitation to using solution-state NMR techniques to study biomolecules is the width of the spectral resonance. The theory of Bloch, Wagness and Redfield [139, 140] describes how random fluctuations in the local magnetic fields experienced by nuclear spins will lead to an increase in T_2, and hence broaden the width of the resonance. In solution, the dominant process that leads to fluctuations on an appropriate timescale to influence such relaxation rates is Brownian tumbling, on the timescale of rotational diffusion, $\tau_c = \frac{4\pi\eta r^3}{3kT}$. For large proteins the linewidths are proportional to τ_c, which itself is proportional to the mass of the species under study. Empirically, globular proteins larger than ca. 50 kDa have correlation times that give rise to lines too broad to allow distinction using conventional solution-state NMR experiments, though this limit can be raised to ca. 1 MDa particularly with molecules of high symmetry using techniques based on the TROSY and CRINEPT pulse sequences [141], in conjunction with high field magnets and cryogenically cooled probes. As a consequence, resonances

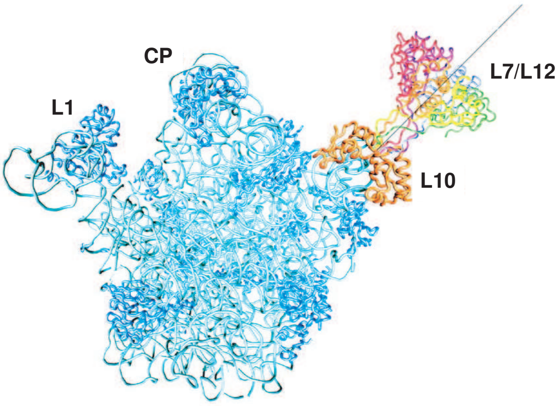


Figure 4.1: Representation of the orientation of L7/L12 on the ribosome. The hinge and CTD regions of L7/L12 (residues 43120), modeled into the structure of the 50S subunit from *T. thermophilus* [PDB 1giy][142] with L10 (in orange indicating the base of the stalk), rRNA (cyan), and the remaining ribosomal proteins (dark blue). L7/L12 is drawn in various colours and positioned in six orientations representing a full rotation around the principal axis (gray line). In this model, the cone representing the possible range of orientations of L7/L12 is placed in an arbitrary position similar to that of the ribosomal stalk, as the precise orientation of the CTD of L7/L12 relative to the ribosomal body is unknown[103].

cannot be observed from the β-sheet core of GDa amyloid fibrils using conventional solution-state NMR techniques.

When a region within a large complex has a high degree of conformational flexibility, the characteristic time of the local fluctuations decreases from that predicted for a rigid body. Consequently, individual subunits in a flexible multi-domain complex will have narrower linewidths than if the complex were rigid [143, 144], and resonances from an unfolded protein are often narrower than those from a folded protein [145]. Interesting extremes of this effect are where flexible groups give rise to narrow NMR resonances, but the species to which the flexible group is tethered does not. One such example is the L7/L12 domain of the bacterial ribosome (figure 4.1). The ribosome is a highly ordered structure of molecular mass approximately 2.5 MDa. When viewed by solution-state NMR, no resonances from the core of the ribosome are observed. However, on the surface of the ribosome, a flexible domain, the L7/L12 stalk, can be observed by solution-state NMR[103]. Similarly, an 11 amino acid stretch is observable on the surface of αB-crystallin complexes, despite the average mass of the complex exceeding 700 kDa [146]. Of particular interest here are the NMR resonances from flexible regions on the surface of amyloid fibrils (chapter 2)[51, 50], despite the mass of the fibril exceeding several GDa, are well resolved. The resonances from the fibril core are too broad to be observed. Thus smaller oligomers and flexible regions on the surface of amyloid fibrils can be reasonably studied by solution-

Size class	Type of species	D_{eff}	Δ variance †
Small molecule	Small molecules	$10^{-5}\mathrm{cm}^2\mathrm{s}^{-1}$	n
Protein sized	Monomer/small oligomer of protein	$10^{-6}\mathrm{cm}^2\mathrm{s}^{-1}$	n
Medium sized	Flexible region on surface of structure <500nm	$10^{-7}\mathrm{cm}^2\mathrm{s}^{-1}$	n
Large	Flexible region on surface of structure >500nm	$10^{-8}\mathrm{cm}^2\mathrm{s}^{-1}$ ‡	y

Table 4.1: Types of species observable to solution-state NMR and how they can be distinguished using NMR diffusion measurements. NMR diffusion measurements and the origin of the variance of D_{eff} with Δ is discussed in chapter 3. † - A Δ variance occurs where D_{eff} is experimentally observed to vary with Δ. ‡ - Approximate D_{eff} is quoted with Δ=1s.

state NMR.

Recently it was observed that the component molecules of SH3 fibrils are in dynamic exchange with a significant population of their constituent monomers [49]; ca. 10% of the total material is found to be in the soluble fraction at equilibrium. As monomers are continually exchanging between fibril and solution this process was termed *recycling*. This treatment describes describes the exchange process between monomers and fibrils as a pseudo-equilibrium avoiding the need to consider longer term rearrangements of the overall length distributions of the fibrils under study. Consistent with this observation, in SH3 fibril preparations purified by ultracentrifugation it was observed that resonances from monomeric protein can be clearly observed (section 2.4). In contrast relatively little monomeric material ($< 0.1\%$ of the total) is observed in preparations of purified fibrils from a derivative system, $SH3_2Cyt$. Interesting, resonances can be detected from flexible regions on the surface of $SH3_2Cyt$ fibrils.

In order to test whether or not the presence of flexible regions on the surface of fibrils correlates with the monomer/fibril association/dissociation rates and the stability of the underlying fibrils, the behaviour of a range of fibril forming systems were studied and compared. This has lead to the remarkable conclusion that the presence of flexible regions on the surface of an amyloid fibril, as confirmed by detectable NMR resonances from the fibril surface, correlates with a large reduction in the apparent concentration of soluble monomeric material in purified fibril preparations.

4.1.1 Methods

Only small oligomers and flexible regions of larger aggregates will have sufficiently sharp resonances to by observed using solution-state NMR spectroscopy. As NMR diffusion measurements can be used to determine the size of the observed species (chapter 3), solution-state NMR can be used to categorise a species as a ‘small molecule’, ‘protein sized’, ‘medium’, or ‘large’ (table 4.1). These meth-

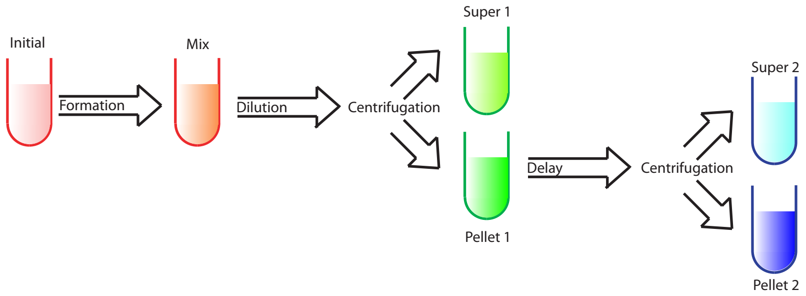


Figure 4.2: Diagram summarising experimentally accessible states for the study of aggregation reactions. The terms defined in this scheme are used extensively both in this chapter, and in chapter 5. Protein molecules can be induced to form aggregates, often by lowering the pH, raising the temperature or by addition of a denaturant. The evolution of the system can be followed by various techniques and at a later time point, usually when the system has reached a steady state, the mixture can be separated by ultracentrifugation yielding a supernatant and pellet. These first two fractions will be referred to here as 'super-1' and 'pellet-1' fractions. Pellet-1 can be resuspended and then can be repurified after a defined delay, yielding a second supernatant and pellet, designated 'super-2' and 'pellet-2'.

	Test for concentration of free monomer	**Test for flexible regions on the surface of a fibril**
NMR signal intensity comparison ‡	pellet-1=super-2	pellet-1$\neq$super-2
Envelope of pellet-1 NMR spectra	monomer-like	not monomer-like
Relative linewidth	sharp	broad
D_{eff} **vs** Δ	D_{eff} independent of Δ, ca. 10^{-6} cm^{2}s^{-1}	D_{eff} varies with Δ

Table 4.2: Two experimental tests are performed on the fibrils studied in this section. The first tests for the quantity of soluble material present in a purified fibril mixture and the second determines whether or not there are flexible regions on the surface of amyloid fibrils. Where there is a large quantity of soluble material, NMR resonances with the properties of monomeric protein will be observed. For the latter, if significant signal is observed after PFGSE experiments with long diffusion delays and high gradient strengths ($\Delta = 1s$, G=25G cm^{-1}) and those resonances show the D_{eff} vs Δ trend, then they can be assigned to originate from the flexible regions of larger aggregates. ‡ the delay between preparing the pellet-1 sample and re-centrifugation to yield super-2 was ca. 1 day.

ods are employed here in conjunction with ultracentrifugation, a method that is able to physically separate species based on their differential sedimentation rates, a factor that correlates with molecular mass [147]. Empirically this method has been found to be well suited to the separation of fibrils from soluble material [80].

An aggregation experiment can be split into stages using the scheme summarised in figure 4.2. During an *in vitro* aggregation experiment a precursor, typically a monomeric protein, is subject to a system-dependent condition that initiates aggregation, often by lowering the pH, raising the temperature or by the addition of denaturant. At a later time point, usually when the system has reached a spectroscopic steady-state, large species can be separated from small by ultracentrifugation, yielding two fractions, termed here *pellet-1* and *super-1* (figure 4.2). Initially, small oligomers and monomers are not expected to be present in a re-suspended pellet-1 sample at an appreciable concentration. If monomers free in solution are in rapid exchange with the fibril, then their concentration in a pellet-1 sample will come to a steady-state. This has been demonstrated with SH3 fibrils, with the equilibrium population of ca. 10% monomer restored within hours [49].

If soluble protein has returned to solution in a pellet-1 sample during a defined delay, as a result of dissociation from a larger aggregate, this material will be transferred to the second supernatant, *super-2*, on re-centrifugation. The NMR spectrum of pellet-1 before re-centrifugation will then be indistinguishable from that of the corresponding super-2. If protein molecules within the fibril do not exchange during the delay period, then the concentration of protein super-2 will be vanishingly small. The total concentration of protein in the samples can be accurately quantified using UV-Vis spectroscopy, observing the aromatic absorption frequency around 280 nm, subtracting baseline scattering effects as described in section 8.2.3. The extinction coefficient in aqueous solution is not influenced by the local electronic environment, and so is independent of the aggregation state of the sample [147] (experimentally verified for $(SH3)_2Cyt$ fibrils in section 2.4.3).

In addition, if resonances are observed in a pellet-1 fraction whose NMR diffusion properties correspond to an amyloid fibril, then the fibril system under study possess regions sufficiently flexible to yield resonances amenable to solution-state NMR study.

Thus by combining UV-Vis spectroscopy, ultracentrifugation and NMR spectroscopy, the dynamic equilibrium between fibrils and their monomeric constituents can be quantitatively studied. The two tests described in table 4.1.1 are performed on a range of fibril systems from different protein and peptide sequences to determine if the presence of flexible regions on the surface of amyloid fibrils correlate with the quantity of monomeric material present in a fibril sample purified by ultracentrifugation.

In this chapter, these tests are used to compare and contrast the fibrils from TTRRGD, αB-crystallin, lysozyme, insulin, SH3, SH3$_2$, SH3$_2$Cyt and SH3$_3$Cyt. Brief descriptions of the systems involved, and the conditions required for fibril formation are given in table 4.3. All pellet fractions were resuspended in a buffer identical to that in which they were formed. UV-Vis spectra, ^{1}H and PFGSE NMR experiments were obtained from pellet-1 and super-2 samples, where the delay between the two was ca. 1 day. Unless specified otherwise, all data was acquired at 300 K. If fibrils were found to have an appreciable concentration of monomeric material ($> 1\%$) in super-2 appear within this time, they were classified as *rapidly recycling*. If flexible regions were observed on the surface of the fibrils, they were classified as having *flexible regions*. The criteria for these classifications are summarised in table 4.1.1.

4.2 Results

4.2.1 Rapidly recycling systems

Fibril samples of TTRRGD, SH3 and SH3$_2$ were found to be rapidly recycling according to the criteria in table 4.1.1. Particularly, ca. 10% of the total material was found to transfer from an equilibrated pellet-1 sample to a super-2 sample. The NMR spectra of pellet-1 samples were identical to those of the super-2 samples (figure 4.3B); the diffusion profiles of all observed resonances correspond to that expected for a monomeric species (figure 4.10). No resonances from larger species were observed in PFGSE experiments with long diffusion times ($\Delta = 1s$) in these fibrils systems.

The nature of these rapidly recycling systems allowed two further types of experiment to be performed.

Recycling After removing smaller species by ultracentrifugation, the rate at which monomeric material accumulates in pellet-1 can be followed with time as the system returns to equilibrium.

Partitioning An equilibrated system can be perturbed by altering the pH, temperature or some other property and the change in concentration of monomer can be followed with time.

In both experiments, the integrated ^{1}H NMR signal intensity can be used as a quantitative measure of the concentration of monomer in a sample once the system has been calibrated using concentrations obtained from UV-Vis spectroscopy. This measure is independent of the conformation provided the integral extends

System	Detail
SH3	*Incubate for 1 week, pH2, 5mg/ml, room temperature* This system was introduced more fully in section 2.1.1. The aggregation properties of an 85 residue SH3 domain from a regulatory subunit of phosphatidylinositol 3-kinase (referred to herein as SH3) have been well studied [60, 30]. The SH3 protein, although not associated with any known disease, forms discrete fibrils at low pH over a period of several days. TEM micrographs and a fibre X-ray diffraction pattern for SH3 fibrils are shown in figures 1.2A and B, and a cryoEM reconstruction of the electron density in figure 1.3. H/D measurements have shown the entire sequence of SH3 folds into the core of SH3 fibrils [49].
SH3 variants	*Incubate for 4 days, pH2, 1mg/ml, room temperature after pH jump* A tandem repeat of SH3, $SH3_2$ forms fibrils significantly faster than does a single SH3 domain and the mechanism of rate enhancement is well studied [80]. $SH3_2$Cyt and $SH3_3$Cyt are described in more detail in chapter 2 [51].
TTRRGD	*Incubate for 1 week, 10% acetonitrile, 10mg/ml, room temperature* Transthyretin (TTR) is a 55 kDa protein involved in the transport of thyroxine and retinol in plasma. The native protein is a homotetramer of 127-residue subunits and has extensive β-sheet structure [148]. WT TTR forms amyloid fibrils *in vivo* in a condition termed senile systemic amyloidosis [149], and a number of naturally occurring TTR variants are associated with familial amyloid polyneuropathy [150]. Full-length TTR [151], TTR variants [152, 153], and two 11-residue peptide fragments derived from the native sequence [151] have been shown readily to form amyloid fibrils *in vitro*. The fragment $TTR_{105-115}$ corresponds to β strand G in the full length protein, has been studied as a model system for amyloid assembly [154]. The conformation of a single peptide within a fibril has been determined [39] (figure 1.6).TTRRGD is a functionalised variant of $TTR_{105-115}$, with an additional 6 residues (Gras et al., unpublished). The sequence is given in table 5.1.
Insulin	*Incubate for 1 hour, pH2, 5 mg/ml, 60^{o}C* A small helical protein hormone consisting of two polypeptides, chain A (21 residues) and chain B (30 residues), linked together by two interchain disulfide bridges [155]. When insulin is heated at high temperature and low pH, a series of structural changes occurs and results in the formation of fibrillar structures [156, 157]. Amyloid fibrils from insulin fibrils have been extensively studied [158, 37, 92] (figure 1.4).
Lysozyme	*Incubate for 10 days, pH 1.5, 10mg/ml, 45^{o}C* Involved in an amyloid related human disorder, hereditary systemic amyloidoses, in which the disease is associated with single-point mutations in the lysozyme gene [24]. Fibrils are deposited widely in tissues in those suffering from this condition. *In vitro* the WT forms fibrils at low pH.
αB-crystallin	*Incubate for 2 hours, 1M GdnHCl, 0.1M phosphate, 2mg/ml, 60^{o}C* A member of the small heat-shock protein (sHSP) family of chaperones and found predominantly in the eye lens where it is thought to play a crucial role in preventing extensive protein aggregation, thus delaying the onset of blindness resulting from cataract formation. The native state of αB-crystallin is oligomeric, with a mass of approximately 600kDa (figure 4.7F shows small spherical particles of radii between 4 to 9nm). Despite its size, the 12 residue segment at the C-terminus is sufficiently flexible to be observable by solution-state NMR [146]. Although αB-crystallin has a proven role as a chaperone, it readily assembles *in vitro* into amyloid fibrils [159] (figure 4.7E) under denaturing conditions and at elevated temperature.

Table 4.3: Summary of the fibril forming peptide and protein systems discussed in this chapter. Fibril forming protocols are shown in italics.

over all aromatic protons and the solvent is D_2O. This is illustrated in figure 4.3A where a super-2 fraction of SH3 is taken from pH 3 to pH 7. The change in spectral envelope reflects the refolding of the protein, but the total integral over the aromatic region does not change. A mathematically rigorous discussion of this method is described in appendix 4.5.1 and only requires that the probe is carefully tuned and matched each time the sample is changed [160]. Thus any observed changes in ^{1}H NMR signal intensity can be related directly to the concentration of monomer in the sample.

Through recycling and partition experiments it should therefore be possible to determine kinetic and thermodynamic parameters for the fibril sytems. In addition this methodology lends itself to screening for the detection of compounds that could induce dissociation of fibrils. For both TTRRGD and SH3, recycling experiments are discussed in section 4.2.1 and partition experiments, where the pH is changes from 3 to 7 are discussed in section 4.2.1. All partitioning and recycling experiments were performed in 100% D_2O unless otherwise specified. Recycling experiments applied to a range of aggregating peptides based on $\mathrm{TTR}_{(105-115)}$ interpreted in terms of polymerisation equilibria are described in chapter 5.

Recycling experiments

Pelleted fibril samples were resuspended and placed into the spectrometer as rapidly as possible. In order to keep the dead time to a minimum (ca. 10 minutes), the NMR spectrometer was pre-calibrated on a similar sample so that adjustments in the tuning, matching and shimming are kept to a minimum. The signal intensity of ^{1}H NMR spectra was followed for several hours (figure 4.4A) and was found to reach a constant level in both TTRRGD and SH3 fibril samples. NMR diffusion experiments verified that the observed species at all times was monomeric protein (figure 4.10). When the NMR signal intensity was found to reach a steady state, samples were re-pelleted via an ultracentrifuge and the concentration of monomeric material present at equilibrium could be quantified by UV-Vis, and thus NMR signal intensity measurements can be converted to concentrations using this calibration. In addition, TEM images were obtained to confirm that fibrils rather than disordered aggregates were present at all time points.

In the case of SH3, the final concentration of monomeric protein in pellet-1 was found to be ca. 10% of the total protein known to be present from UV-Vis measurements (100 μM). Interestingly, when fibril formation is complete, the signal intensity decreases to ca. 10% of its original value [60]. This suggests that the final solution composition of 10% monomer, 90% fibril can be approached from either 100% soluble protein, or where 100% of material is initially in fibrils.

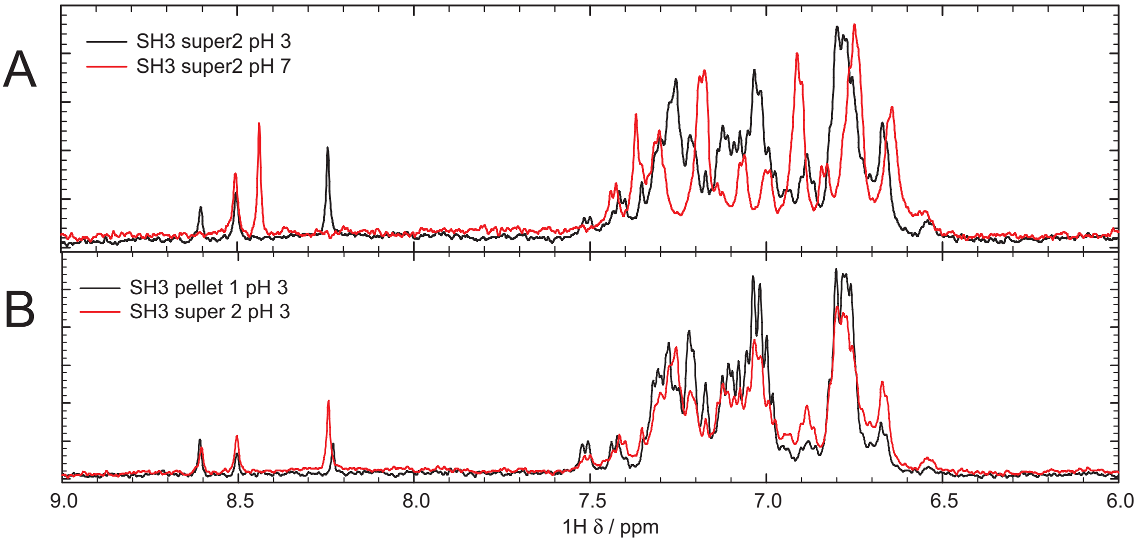


Figure 4.3: ^{1}H NMR spectra from SH3 pellet-1 and super-2 fractions at pH 3 and pH 7 are compared. **A -** NMR spectra of a SH3 super-2 at pH 3 and pH 7. Although the spectral envelope changes, reflected refolding of the protein, the total integral over the aromatic resonances remains the same validating the use of integrated NMR signal intensities to determine concentrations [160]. **B -** The NMR spectra of a super-2 and a pellet-1 of SH3 fibrils are essentially identical. This confirms that these fibrils are rapidly recycling. PFGSE NMR data reveals that the observed species in both cases diffuses like a protein monomer (figure 4.10A).

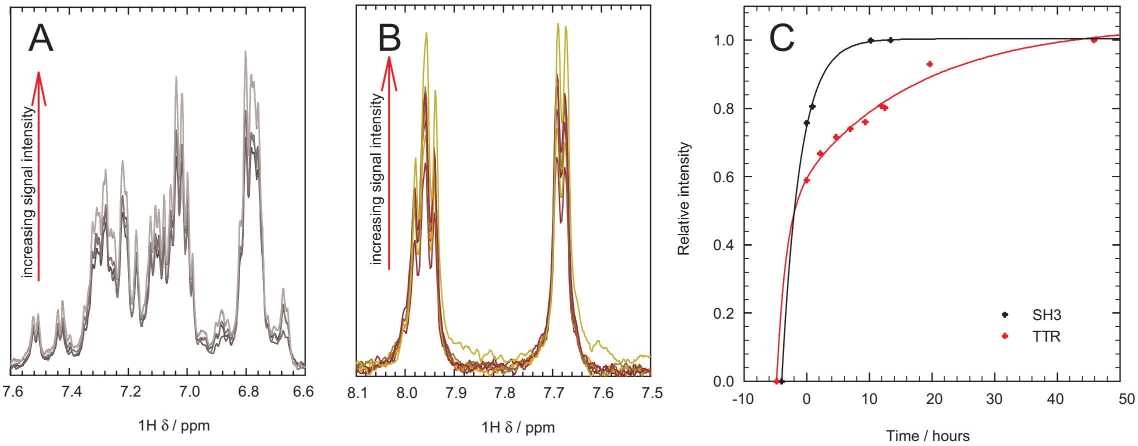


Figure 4.4: Data from recycling experiments. The pellet-1 samples of both SH3 (**A**) and TTRRGD (**B**) were re-suspended, and the first NMR spectrum was recorded at $t = 0$. The dead time between the resuspension and the first NMR spectrum is indicated by the negative time measurement. The ^{1}H NMR signal intensity is shown to increase with time reflecting an increase in the concentration of monomeric material with time (section 4.5.1)[160]. **C -** The integrated intensity of the two samples are compared, normalised to the last datapoint collected. Data are fitted to a bi-exponential function of the form $y = Ae^{-k_1 t} + Be^{-k_2 t}$. In the experimental, through comparing the k_2 values, SH3 is observed to return to equilibrium 6× more rapidly more rapidly than TTRRGD.

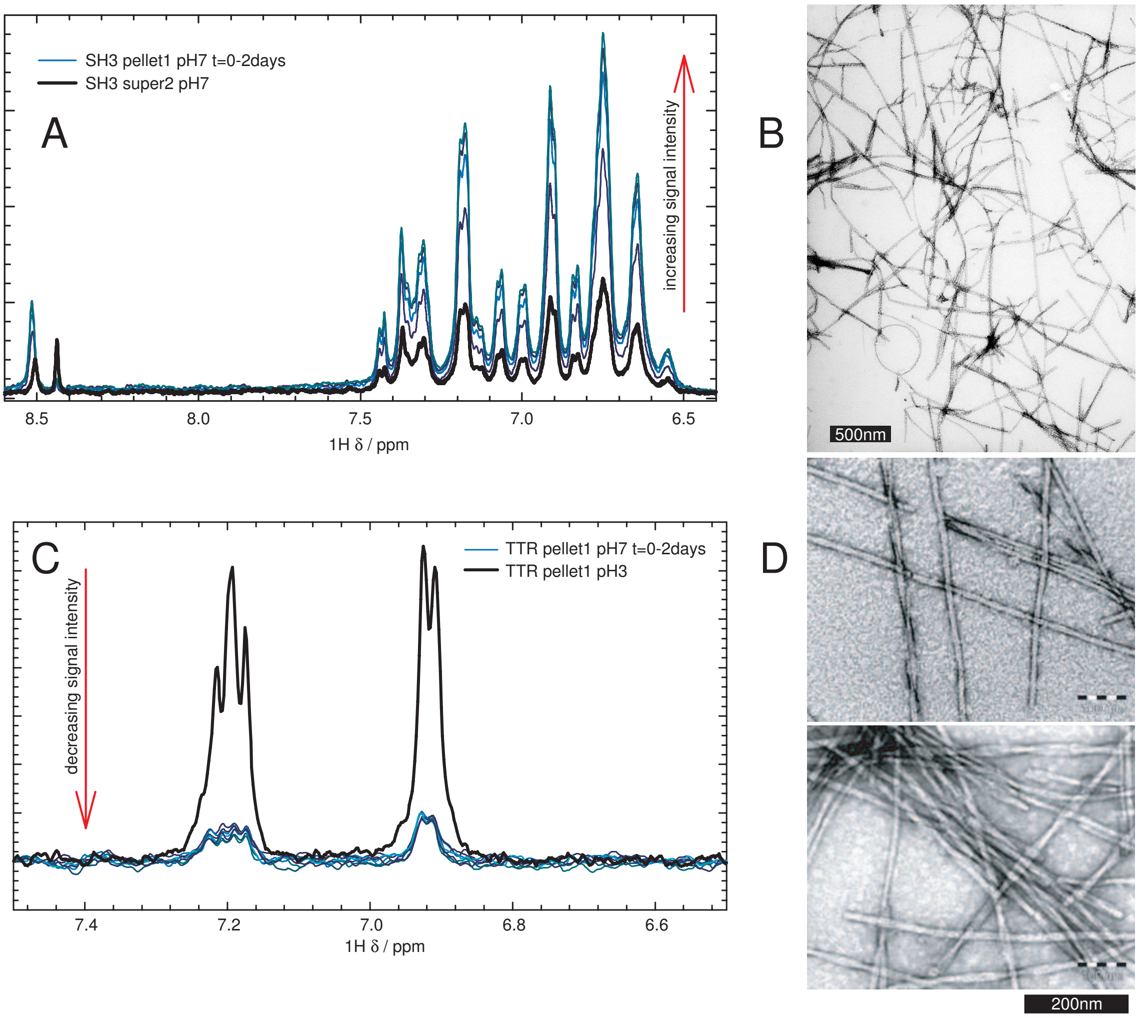


Figure 4.5: Partitioning experiments from both SH3 (**A/B**) and TTRRGD (**C/D**). Equilibrated pellet 1 samples were taken from pH 3 to pH 7 and the change in ^{1}H NMR signal intensity was followed over time. The signal intensity from the SH3 sample was observed to increase, whereas the signal intensity form the TTRRGD sample was observed to decrease. Fibrils were observed in the TEM (**B/D**) at all time points at both pH values.

A more quantitative treatment of these equilibria is discussed in chapter 5.

The rate at which the monomeric protein returns to equilibrium can be seen from this experiment to be system dependent; SH3 returns to equilibrium ca. 6× more rapidly than does TTRRGD (figure 4.4A), under these conditions. This finding is discussed further in section 4.3.2 and in chapter 5.

Partitioning experiments

Data from the recycling experiment indicate that for rapidly recycling fibril systems, the molecular components of amyloid fibrils are apparently in equilibrium with their constituent monomers and recycling experiments follow the concentration of monomer as it increases initially from zero to its equilibrium value.

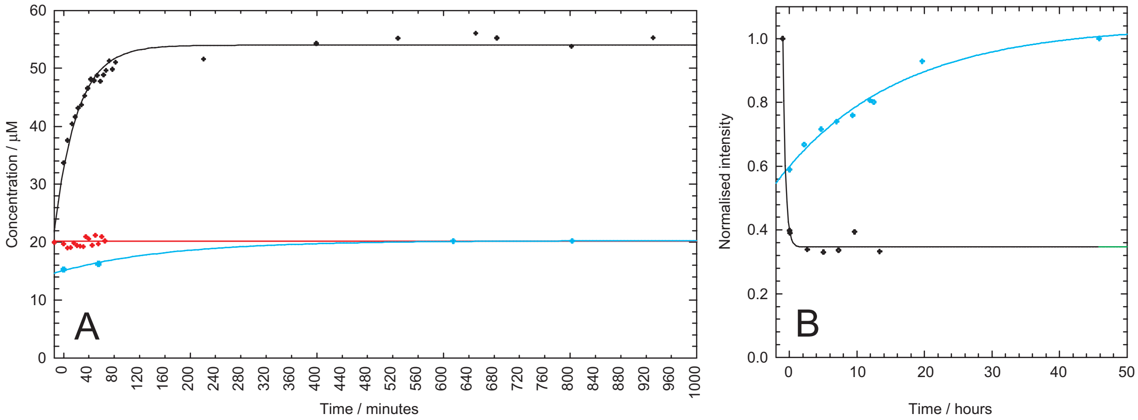


Figure 4.6: A summary of the recycling and partitioning experiments from SH3 (**A**) and TTRRGD (**B**). Data from recycling experiments where the signal intensity is observed to increase over time as monomeric material returns to solution is shown in blue. Data from partitioning experiments where the change in signal intensity is followed on changing the pH of an equilibrated pellet-1 from 3 to 7 is shown in black. Data from a partition experiment on a super-2 fraction of SH3 is shown in red. For SH3, the NMR intensities are converted to absolute concentrations (section 4.5.1). The total concentration of protein in the SH3 sample was 100 μM.

Partition experiments can then follow the change in the equilibrium composition of a rapidly recycling system. Here, the effect of pH on the equilibrium composition of TTRRGD and SH3 fibrils is determined. For both systems, on changing the pH from 3 to 7, large changes in the ^{1}H NMR signal intensity were observed. Fibrils were observed in the TEM at both pH3 and pH7 for both systems, with no detectable variation in morphology (figures 4.5 B and D).

In the case of SH3, the spectra show that the monomers convert from the unfolded state at pH 3, to the folded state that is stable at pH 7 (figure 4.3A) and the overall observable signal intensity increases (figure 4.6A, black). Interestingly, at pH 7, ca. 40% of the material is in the monomeric state at equilibrium (figures 4.5A and 4.6A) compared to the 10% at pH 3. This composition was confirmed by UV-Vis spectroscopy following re-centrifugation. The change in signal intensity can be understood in terms of the equilibrium between fibrillar and monomeric states. The monomeric protein is observed to assume its native conformation in the dead time of the experiment, and the number of observable monomers increases by a factor of 4 over the course of 12 hours. This process is 90% completed within 1 hour (figure 4.6A). By stabilising the native state of monomeric SH3 therefore, the equilibrium is shifted to disfavour the fibrillar state at pH 7.

As the TTRRGD peptide does not have a stable native state, little variation is observed in the chemical shifts between pH 3 and pH 7 (figure 4.5C). The total signal intensity can be observed to decrease rapidly (figure 4.6B, black) to ca. 30% of its initial value, the reverse of the trend observed for SH3. Raising the pH will reduce the net charge on the TTRRGD monomer from +1 to neutral, a factor likely to enhance its stability when bound within a fibril.

In figure 2.7, a similar partition experiment is shown for $SH3_2$. At pH 3, $SH3_2$ forms fibrillar aggregates. Over a time scale of two weeks, the NMR signal intensity decreases to zero. When the pH of the sample of fibrillar aggregates is adjusted from pH 3 to pH 7, the signal is completely restored to its initial value in little over an hour. In contrast, no such variation is observed when $SH3_2$ fibrils are taken from pH 3 to pH 7. It has been noted elsewhere that fibrillar aggregates are less stable than fibrils [161, 162], and so partition experiments of this type could be used to distinguish between fibrils and fibrillar aggregates through observing partitioning rates.

We have shown that for the rapidly recycling fibrils formed from SH3 and TTRRGD, the systems tend to move towards an equilibrium composition. When monomeric material is depleted from the system by ultracentrifugation, the fibrils dissociate to restore the concentration of monomeric material. Experiments of this type provide a powerful method to study the concentration of monomer *in situ*, allowing the determination of thermodynamic and kinetic parameters. These ideas are extended in chapter 5.

4.2.2 Fibrils displaying flexible regions

Pellet-1 fractions of fibrils from all the systems were prepared, and ^{1}H and PFGSE NMR spectra acquired. In pellet-1 samples of αB-crystallin, human lysozyme, insulin, $SH3_2$Cyt and $SH3_3$Cyt, the observed spectral envelope could be distinguished from that of the monomer and the effective diffusion coefficient was found to vary with Δ. From the variance of D_{eff} with Δ, length estimates of the fibril samples could be made using the theory discussed in chapter 3, and were found to be in good accord with the fibril lengths observed by TEM. These fibrils therefore have regions on their surface that are sufficiently flexible to be observed by solution-state NMR (table 4.1.1). In addition, <0.1% of the total material was found to transfer to the second supernatant as determined from UV-Vis spectra, in stark contrast to the rapidly recycling fibrils where ca. 10% of the material was transferred. Furthermore, where this experimented was repeated with a 1 week delay between the pellet-1 and super-2, again only trace material was found in the second supernatant.

The NMR data from αB-crystallin, human lysozyme and insulin are presented below. NMR data from $SH3_2$Cyt and $SH3_3$Cyt are presented in sections 2.4.6 and 2.4.8. In addition, formation of fibrils from native αB-crystallin and their decomposition at low pH were studied. A manuscript including the NMR data from αB-crystallin fibrils has been accepted for publication in J. Mol. Biol. [163].

αB-crystallin

PFGSE NMR data from native αB-crystallin (figure 4.7H, red) shows that the ^{1}H NMR spectrum (figure 4.7A, red) corresponds to a species diffusing as a particle with an average R_H of 5.8 ± 0.4 nm. The R_H of a protein scales with molecular mass according to a Flory scaling law of the form $R_H = aM_w^n$ [164, 165]. By extrapolating the R_H values of proteins studied in this chapter, the effective molecular mass of αB-crystallin was determined, suggesting native αB-crystallin exists as a 10-mer (figure 4.10E). This size is consistent with the dimensions of native αB-crystallin observed using TEM of 4-9nm (average 6.5nm, figure 4.7E) [146].

Native αB-crystallin can be induced to form fibrils (table 4.3). A pellet-1 sample of fibrils gave rise to the ^{1}H NMR spectra shown in figure 4.7A, in green. Once purified, the fibrils were observed to remain in solution, and the NMR spectrum was not observed to vary over a time scale of weeks. Of particular interest is the comparison between the 1D ^{1}H NMR spectrum of native and fibrillar αB-crystallin. The absolute signal intensity, normalised for total concentration of αB-crystallin monomers is ca. 42% that of the native state.

In figure 4.7C, the spectra are normalised such that the total signal intensity observed from each species is equal. Both the fibrillar (green) and native state spectra (red) were processed with a 2Hz exponential window function but when the spectrum from the native state was processed with a 20Hz window function, the (black) spectrum is obtained. This is essentially identical to the spectrum of the fibrillar state, a result that strongly suggests that the same 12 amino acid C-termini observed in the ^{1}H NMR spectrum of the native state is sufficiently flexible to be observed in the fibrillar state, but the effective correlation time τ_c of the resonances in the fibrillar state is ca. $10\times$ greater than that of the native state from the observed line-widths. The dynamic behaviour that could be responsible for the reduction in correlation time are considered more fully in chapter 6.

The PFGSE NMR data show that the D_{eff} of fibrillar αB-crystallin was a function of Δ (figure 4.7H). Fitting to equation 3.17 in chapter 3 yields a length estimate for the fibrils of 7.2μm. This is consistent with the fibril lengths observed by TEM.

Fibril formation. The fibril formation process was followed by solution-state NMR spectroscopy. A sample of native αB-crystallin at pH 2 in 1M GdHCl at 2mg/ml was heated to 60^{o}C and ^{1}H NMR spectra were followed with time. The NMR signal intensity was observed to decay very rapidly, and to come to a steady value in less than 5 minutes (figure 4.7A). A plot of the ^{1}H signal intensity of the aliphatic peaks is shown against time in figure 4.7B. The decaying signal inten-

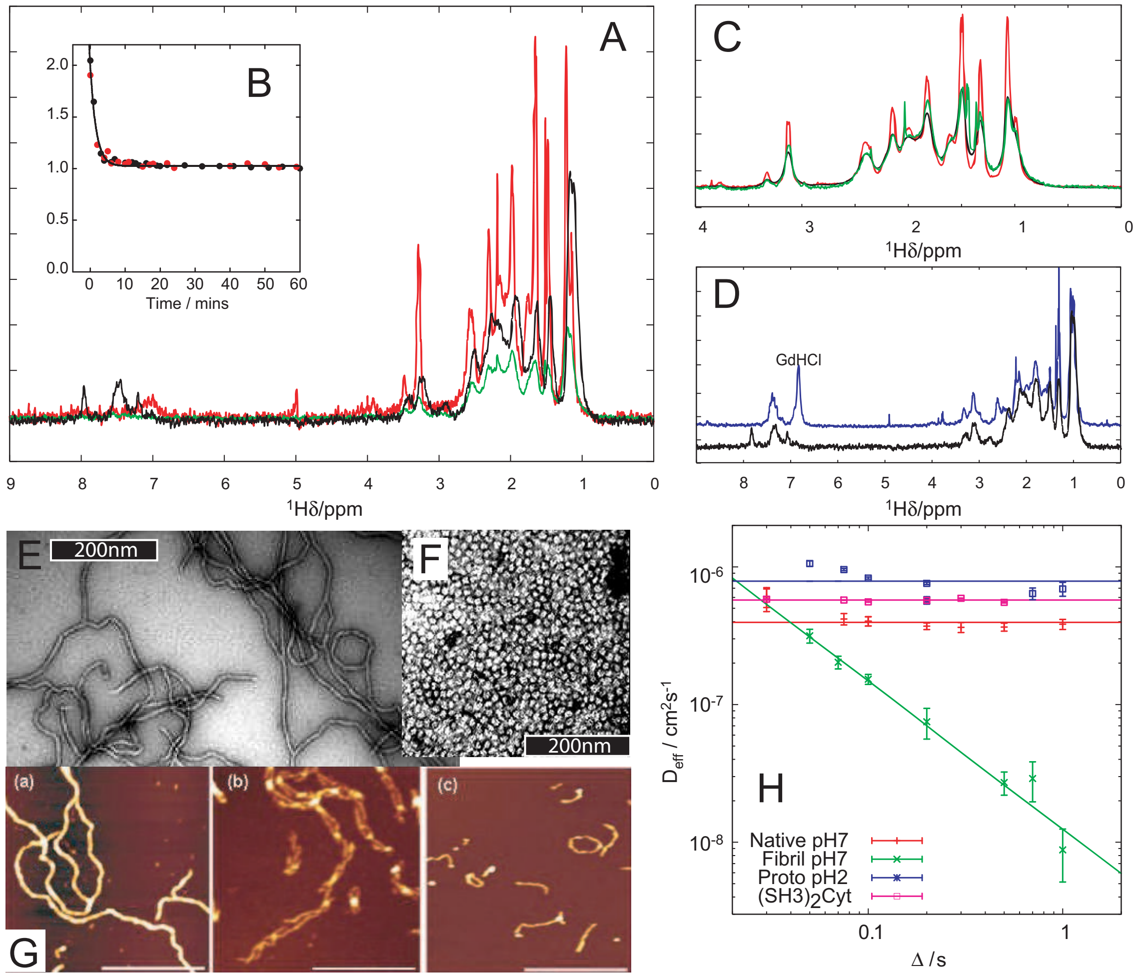


Figure 4.7: A description of the aggregation states of αB-crystallin by NMR (panels A-D and H), TEM (panels E and F) and AFM (panel G). TEM images are courtesy of Dr. Sarah Meehan and AFM images are courtesy of Tuomas Knowles. NMR data from the native states (E) are shown in red, the fibrillar state (F) are shown in green, protofibrils (G(b)) are shown in blue and the formation intermediate is shown in black. **A -** ^{1}H NMR spectra of native and fibrillar states are compared to the formation intermediate. Spectra are normalised to the known concentration of αB-crystallin monomers. **B -** Integrated aliphatic signal intensity shown as a function of time during formation of αB-crystallin fibrils. Two independent runs are shown, normalised to the ending signal intensity. The data are fitted to a function of the form $y = ae^{-kt} + b$. The half life of the decay process is 69±5s. **C -** Comparison of native and fibril NMR spectra, normalised such that the total intensity is the same in both cases. Both are processed with a 2Hz exponential window function. Where the native spectrum is processed with a 20Hz window function, the black spectrum is obtained, which is essentially indistinguishable from the fibrillar spectrum (green). **D -** An NMR spectra of the formation intermediate (black) is compared to a spectrum observed in protofibril samples. The spectra are normalised to ensure the total aliphatic signal intensity of both are equal. The two spectra are similar. Both show resonances in the aromatic region that are not observed in spectra native and fibrillar states. **E/F -** TEM images of fibrillar (E) and native (F) αB-crystallin. **G -** When αB-crystallin fibrils are taken to pH 2, they dissociate into protofilaments. AFM images showing the stages of this transition from fibrils (a, height 10.0±1.7nm) to protofilaments (b and c height 2.2±0.4nm). **H -** NMR Diffusion coefficients obtained as a function of the diffusion delay for the native state, fibrillar state protofilament state of αB-crystallin, and monomeric SH3$_2$Cyt (pink). Native and protofilament states of αB-crystallin and SH3$_2$Cyt show behaviour characteristic of monomeric protein. The data from fibrillar αB-crystallin corresponds to rods of length ca. 7.2 μm, predicted from fitting to equation 3.17 using the theory described in chapter 3.

sity fitted well to a single exponential function of the form $y = ae^{-kt} + b$, decaying with a half life of 69 ± 5s. Interestingly, the spectral envelope observed within the first two minutes (figure 4.7A, black) differs substantially in form from that observed in either the native or fibrillar states of αB-crystallin (figure 4.7A, red and green respectively). This species is termed the *formation intermediate*. In particular, resonances from this formation intermediate are observed in the aromatic region between 7 and 8 ppm in the first two minutes, which are not observed in either the native or the fibrillar state. The native oligomeric state of αB-crystallin must therefore dissociate and unfold substantially before being incorporated into fibrils.

Fibril dissociation. As illustrated in the series of AFM images in figure 4.7G, when the pH of a solution containing fibrils (G(a))is reduced to pH 2, the fibrils dissociate into protofilaments (G(b) and G(c)). The AFM images of the protofilaments indicate considerably less rigid fibrillar structures of height 2.2±0.4nm. This contrasts with a height of 10.0±1.7nm determined also by AFM for the more rigid, mature fibrils. The dissociation of mature fibrils to protofilaments can be followed by solution-state NMR. The total signal intensity is observed to increase rapidly to ca. 10× that observed in the fibrillar state. The form of the protofilament spectrum (figure 4.7D, blue) resembles neither that of native or fibrillar αB-crystallin. However, it does closely resemble that of the formation intermediate (figure 4.7D, black). For example, both the protofilaments and the formation intermediate have comparable resonances in the aromatic region. PFGSE NMR data however show that these resonances arise form a species whose D_{eff} does not vary with Δ (figure 4.7H, blue) and whose diffusion coefficient is larger than that of native αB-crystallin (figure 4.7H, red). This suggests that these resonances arise from a monomeric species, and resonances are not observed from the surface of the dissociated protofilaments. These data suggest therefore, that the protofilaments of αB-crystallin are in rapid exchange with their constituent monomers.

In conclusion the native, fibrillar, protofibrillar and formation intermediate states of αB-crystallin have been studied using solution-state NMR, AFM and TEM under a variety of conditions. 11 residues of the C-terminus are observed in both the native and fibrillar states, though the decrease in signal intensity in the fibrillar state suggests that more C-termini are buried in the fibrillar state than the native state. From the dependence of D_{eff} on Δ, it can be concluded that the resonances from fibrillar αB-crystallin correspond to amyloid fibrils of length 7.2μm, using equation 3.17. The NMR spectra of the species observed during fibril formation, and the species observed when fibrils are partially dissociated at low pH show similarities particularly in the aromatic region, and are both very

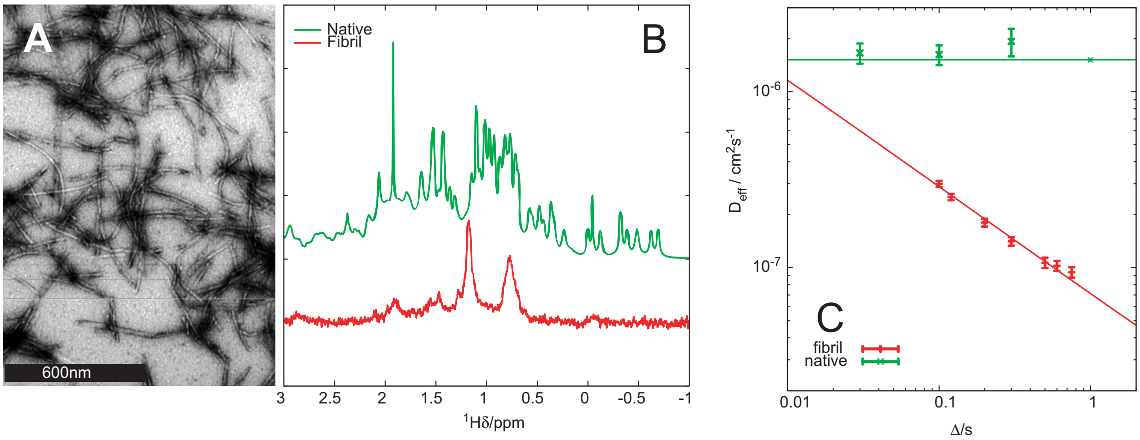


Figure 4.8: NMR Data from human lysozyme in both monomeric and fibrillar states. **A - ** TEM image of human lysozyme fibrils courtesy of Anne Dhelusia. **B -** ^{1}H NMR spectrum of fibrils (red) is quite different from that of the native state (green). No resonances were observed in the aromatic region in the fibrillar sample. **C -** D_{eff} varies with Δ for the fibrils (red), but does not for the native state (green).

Observed peak	BMRB suggested assignment
0.72 ppm	Ile (γ2 or δ) Val (γ)
1.20 ppm	Lys (γ) Leu (γ) Ile (γ1) Ala (β)

Table 4.4: Possible assignments for the two resonances observed from lysozyme fibrils, from the BioMagResBank [102]

different from the spectra of native and fibrillar states. Interestingly, whereas the concentration of monomeric material in a purified fibrillar αB-crystallin is small (<0.1%), under conditions where the fibrils are partially dissociated into protofilaments at low pH, the concentration of dissociated monomeric material is significantly greater.

Lysozyme

A ^{1}H NMR spectrum of the native wild type human lysozyme at pH 2 (figure 4.8A, green) is compared to that obtained from a purified fibrillar sample (figure 4.8A, red), showing clear differences between the two. The majority of the resonances of the native state are not observed in the fibrillar state. Observed resonances in the fibrillar spectrum are significantly broader than those observed in the native spectrum.

The two largest observed resonances from the fibrillar human lysozyme spectrum are at 0.72 and 1.20 ppm are likely to originate from the methyl groups of aliphatic side chains. Comparing these shifts to random coil shifts found in the BioMagResBank [102, 101] reduces the set of amino acids from which they can be assigned (table 4.4). All of the possible assignments are from the side chains of hydrophobic amino acids. Less than 0.1% of the total protein was found in the super-2 of a fibril sample and the ^{1}H NMR spectrum was not observed to change over a time scale of weeks.

NMR diffusion data for both fibrillar and native states are shown in figure 4.8B. The effective diffusion coefficient is a function of the diffusion delay for the fibrils suggesting they originate from flexible regions on the surface of lysozyme fibrils. These data can be fitted to equation 3.17 yielding a rod length estimate of 7.4 μm. It is surprising that despite ca. 50 residues having been previously identified as being flexible and proteolytically accessible [48] that so few are observed by solution-state NMR, whereas ca. 180 resonances are observed from SH3$_2$Cyt fibrils. It is likely that the residues from lysozyme fibrils that are not in the core, nor sufficiently flexible to be observed by solution-state NMR are in molten globular state.

Insulin

NMR diffusion data recorded on monomeric insulin at pH 2 indicates that it diffuses as a monomer with $D_T = 2 \times 10^{-6}$cm^{2}s^{-1} (figure 4.9 blue). A ^{1}H NMR spectrum of fibrillar insulin is shown in figure 4.9C, black. PFGSE NMR data showed the large peaks at 1.1 ppm and 2.0 ppm to originate from small molecules. The broad peak centred on 1.4 ppm however appeared to diffuse significantly more slowly. Using the methodology described in section 9.2.1 it is possible to separate the two species using their differing diffusion behaviour and reconstruct the spectra of the two. Each discrete recorded chemical shift i is fitted to a bi-exponential function $y_i = A_i e^{-D_1 c\%G^2} + B_i e^{-D_2 c\%G^2}$, where $c = \alpha^2\beta$ (see section 9.2 for a more detailed discussion on the analysis of NMR diffusion measurements). The data are then sorted to obtain the spectrum A_i of a species of diffusion coefficient D_1 and the spectrum B_i of a species of diffusion coefficient D_2. The spectrum corresponding to small molecules is shown in figure 4.9C, green, with a diffusion coefficient histogram shown in figure 4.9D, green. The spectrum corresponding to a significantly larger species is shown in figure 4.9C, red with corresponding diffusion coefficient histogram shown in figure 4.9D, red.

The diffusion coefficient of the larger component of the fibrillar insulin spectra has the profile shown in figure 4.9B. The effective diffusion coefficients vary as a function of Δ. Fitting to equation 3.17 shows these data correspond to rods of mean length 8.9μm, consistent with that observed in TEM images.

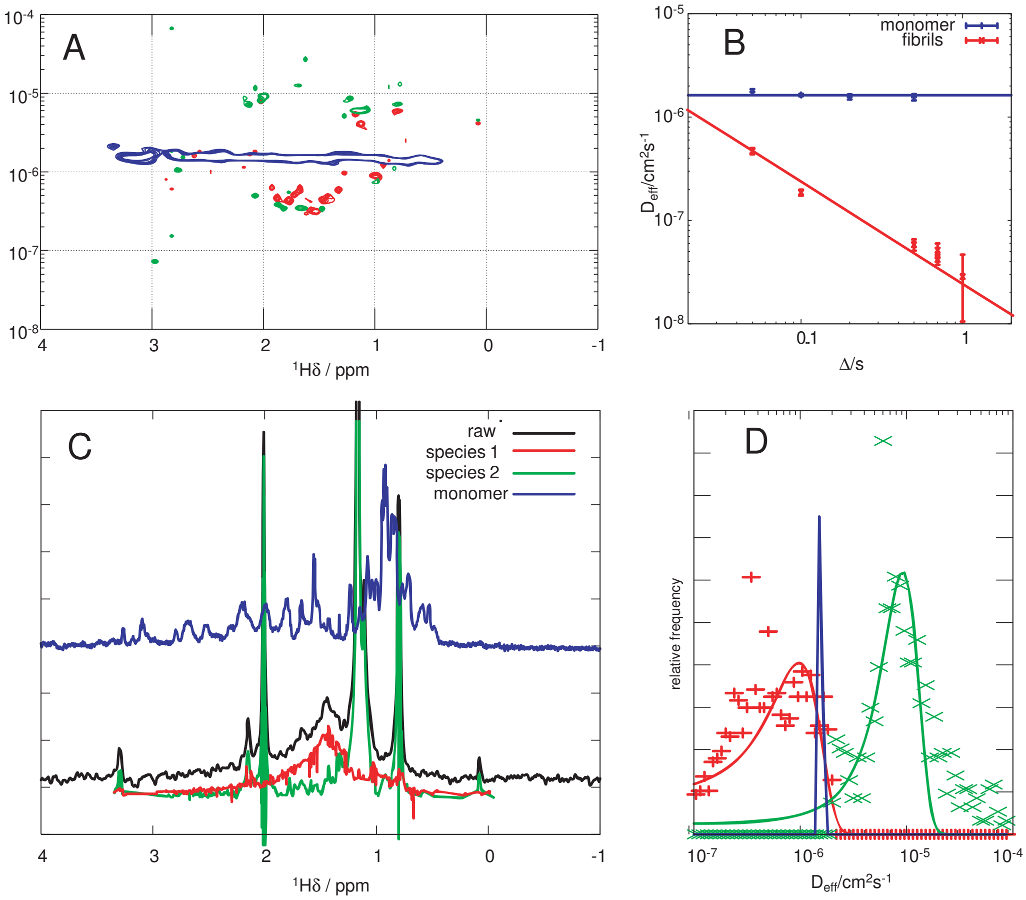


Figure 4.9: Comparison of NMR properties of monomeric and fibrillar insulin, and the separation of the spectra of small molecules from that of fibrillar insulin. NMR Data from monomeric insulin are compared to that from fibrillar insulin. **A -** Raw D_{eff} coefficients obtained from a single exponential fit to monomeric insulin (blue), a monomeric fit to fibrillar insulin (red) and a biexponential fit to fibrillar insulin (green), showing the need to treat the fibrillar sample as a mixture of two species (methodology described in more detail in section 9.2.1). **B -** D_{eff} as a function of Δ for monomeric (blue) and deconvolved fibrillar (red) insulin. The fit shown for the fibrillar data corresponds to a rod length of 8.9μm. **C -** A spectral deconvolution of insulin fibril PFGSE NMR data with Δ=50 ms. The raw spectrum is shown in black. The smaller species with $D_{\text{eff}} > 2 \times 10^{-6} cm^2 s^{-1}$ (green) is distinguished from the larger species (red), whose $D_{\text{eff}} > 2 \times 10^{-6} cm^2 s^{-1}$ using the methodology described in section 9.2.1. Both spectra are compared to that of monomeric insulin (blue). **D -** D_{eff} histograms from the deconvolution shown in C, showing that the smaller species has a diffusion coefficient typical of a small molecule (ca. 10^{-5} cm^{2}s^{-1} green), whereas the second species is considerably larger (red).

	Fibril NMR[i]	Rapid recycling[ii]	Residues	Length (end) / μm[iii]	Length (full) / μm[iii]	Monomer R_H/nm[iv]	Core % [v]	50% rule[vi]
SH3	n	y	86	-	-	2.10 ± 0.04	100% [49]	n/a
$SH3_2$	n	y	168	-	-	2.6 ± 0.1	n/a	n/a
$SH3_2$Cyt	y	n	308	3.6	15.3	3.02 ± 0.08	n/a	y
$SH3_3$Cyt	y	n	386	1.8	6.6	-	n/a	y
TTRRGD	n	y	10	-	-	0.62 ± 0.02	n/a	y
Insulin	y	n	53	2.4	8.9	1.35 ± 0.04	n/a	y
αB-crystallin	y	n	175	1.9	7.2	5.9 ± 0.1	n/a	y
Lysozyme	y	n	130	2.0	7.4	1.67 ± 0.03	50% [48]	y

Table 4.5: Summary of data from amyloid fibril forming systems. **i** - Whether or not resonances were observed from flexible regions on the surface of amyloid fibrils of this system. **ii** - Whether or not a concentration $> 0.1\%$ of the total protein present was found to be monomeric protein in a purified fibril sample. **iii** - The length of the fibrils estimated from the variation of D_{eff} with Δ, and fitting to the 'full rod' and 'rod end' models discussed in section 3.3. The difference in the predicted lengths does not effect the conclusions of this chapter. **iv** - The R_H values were obtained from monomeric samples where D_{eff} did not vary with Δ. **v** - Independent estimates of the proportion of the sequence in the fibril core, as measured by H/D exchange [49] and limited proteolysis [48]. **vi** - The 50% rule is described more fully in section 2.6.3, and describes the ratio between the concentration of solvent exposed monomers to the total monomer concentration. It is argued in section 2.6.3 that this ratio of ca. 50% measured in many amyloid fibril systems is indicative of a general protofilament packing effects.

4.3 Discussion

The NMR properties of the species described in this chapter: SH3, $SH3_2$, $SH3_2$Cyt, $SH3_3$Cyt, TTRRGD, insulin, crystallin and lysozyme are compared in both their monomeric and fibrillar forms in figures 4.10 and table 4.5. The tests described in table 4.1.1 were performed. Remarkably, all systems studied fall into two mutually exclusive classes.

Class 1 Approximately 10% of the total material is observed to be in super-2 after delays of hours after the initial purification. The NMR signal observed in pellet-1 and super-2 fractions indicates the species present diffuse as monomeric protein. This situation has been found for TTRRGD, SH3, $SH3_2$.

Class 2 Less than 0.1% material of the total material was found in super-2, after delays up to 7 days after the initial purification. Interestingly, NMR signals observed in pellet-1 fractions correspond to large fibrils with no detectable monomeric protein. This situation has been found for $SH3_2$Cyt, $SH3_3$Cyt, insulin, αB-crystallin and lysozyme.

In figure 4.10A, D_{eff} vs Δ profiles for both monomeric and fibrillar forms of the proteins are shown. Where D_{eff} does not vary with Δ, the system under

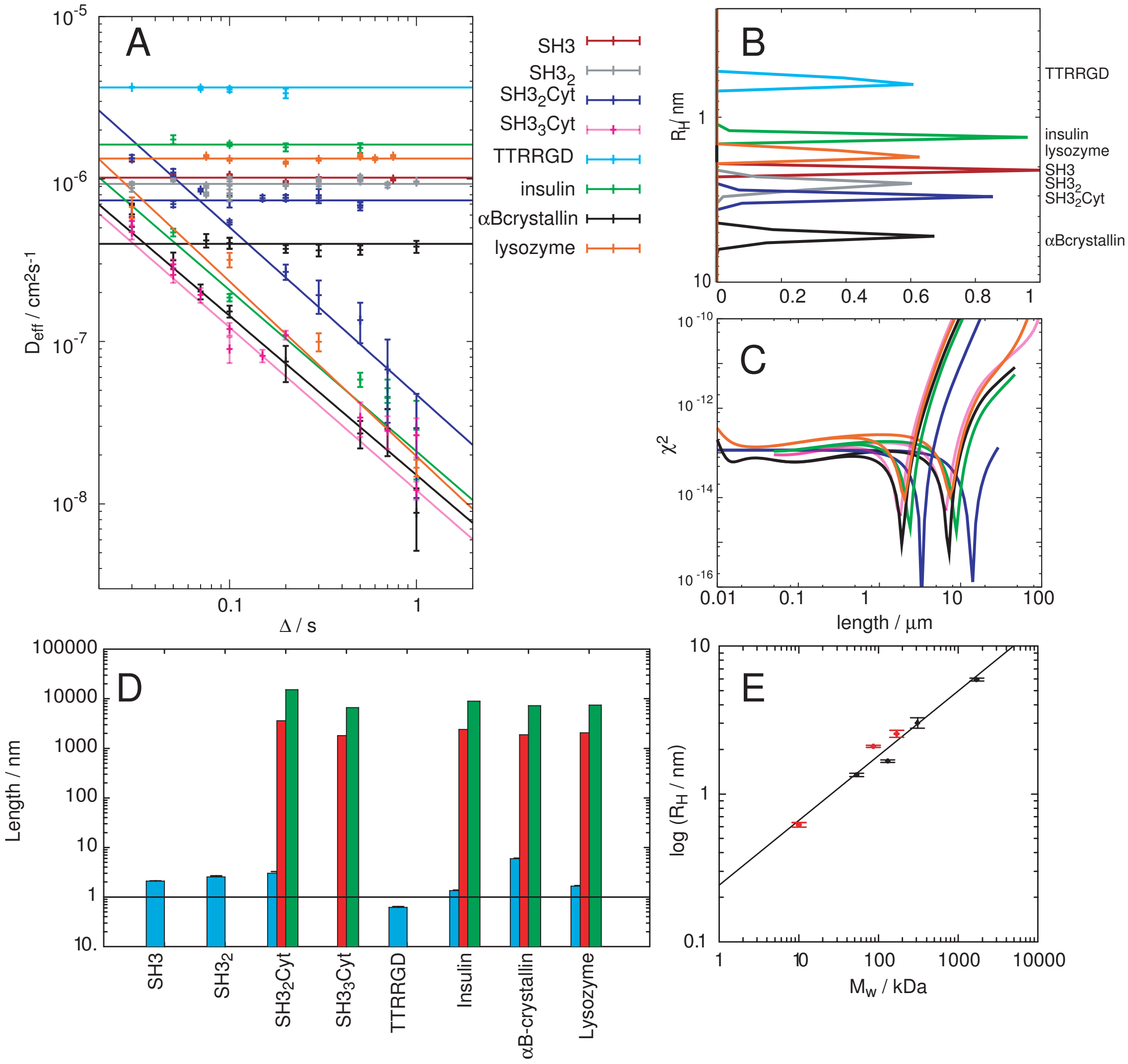


Figure 4.10: NMR Diffusion properties of amyloid forming proteins. Monomeric and fibrillar diffusion data from SH3, SH3$_2$, SH3$_2$Cyt, SH3$_3$Cyt, TTRRGD, insulin, αB-crystallin and lysozyme are compared, revealing two classes of fibril. **A -** $D_{\rm eff}$ vs Δ for both monomeric, and fibrillar forms of the studied proteins. **B -** R_H estimations for samples that show monomeric, ST behaviour in A. The analysis method is described in more detail in section 9.2. **C -** $\chi^2 = \sum(y_{exp} - y_{theory})^2$ results from fitting the fibrillar data show non-ST behaviour to equation 3.17 derived using the theory described in chapter 3. **D -** R_H values of the monomeric protein (blue) are compared with the lengths of fibril, where applicable, as estimated from NMR measurements. The 'rod end' fits are shown in red, and the 'full rod' fits are shown in green. The difference between these numerical estimates is discussed in more detail in section 3.3 and does not affect the conclusions drawn in this chapter. **E -** The molecular mass of a monomer plotted against its determined radius of hydration. Recycling systems are shown in red, and non-recycling systems are shown in black. The data are fitted to $R_H = aM_w^b$, excluding αB-crystallin, with $a = 0.22 \pm 0.05$ and $b = 0.46 \pm 0.03$. The molecular mass of native αB-crystallin is then ca. 1700 kDa, and so can be estimated to be a 10mer in its native state.

study behaves as a monomer (ST behaviour, see chapter 3). Under these conditions, using the Einstein-Stokes equation, a radius of hydration, R_H can be estimated (section 9.2). This is possible for all non-fibrillar states of the proteins considered, and for the fibrillar states of TTRRGD, SH3 and $SH3_2$. The values of R_H measured with Δ = 100 ms are summarised in figure 4.10B, where the width of the plots reflects the experimental uncertainty in the measurement (section 9.2). As expected, this scales with the number of amino acid residues with SH3<$SH3_2$ <$SH3_2$Cyt<$SH3_3$Cyt. This relation is shown more quantitatively in figure 4.10E, with values scaling as $R_H = aM_W^b$ with $b = 0.46 \pm 0.03$. This value is close to 0.5 observed by Jens et al. [165] for unfolded states, and that of 0.57 by Wilkins et al. [164] for highly denatured states, consistent with a description of the unfolded state as an incompletely swollen Gaussian chain [166].

The fibrillar states of $SH3_2$Cyt, $SH3_3$Cyt, insulin, αB-crystallin and lysozyme all show non-ST behaviour, in that D_{eff} varies with Δ. This can be understood in terms of the significance of rotational diffusion of large species (chapter 3), and data can be fitted to the diffusion theory to estimate fibril lengths. Data are fitted to the 'full rod' (figure 4.10, D, green) and 'rod end' (figure 4.10, D, red) models, described in more detail in section 3.3. The χ^2 values for the two models are shown in figure 4.10C. While there is slight discrepancy between the two values, both are in reasonable accord with the lengths observed using TEM and AFM. The relative sizes of both monomeric and fibrillar species identified through NMR diffusion measurements are summarised in figure 4.10D (where monomeric species are shown in light blue, 'rod end' fits are shown in red and 'full rod' fits are shown in green), and in table 4.5.

Interestingly, where regions on the surface of a fibril are sufficiently flexible to give rise to solution-state NMR resonances, very little material is found to be present as soluble monomer in purified fibril samples. Where no resonances are observed from the fibril, a large concentration of monomeric material is found to be present as soluble monomer is a purified fibril sample. Intermediate classes of fibrils were not observed; no sample was observed with both fibril and monomer resonances, nor was a pellet-1 fraction obtained with zero NMR intensity. Both of these cases would have been readily detected using the methodology described here.

In order to observe solution-state NMR resonances from amyloid fibrils, regions of sufficient flexibility to average the dipolar interactions must exist (chapter 6). These resonances therefore fall at random coil chemical shifts but are shown to arise from fibrils because of their measured NMR diffusion behaviour. These species that give rise to these resonances therefore, cannot be in the β sheet core of the fibril, nor is it likely that can they arise from the loops linking two stretches of sheet within a single molecule [103].

The proportion of a sequence within the core of the fibril does not correlate with the molecular mass of the monomer, as shown in figure 4.10E (recycling systems are shown in red, and non-recycling systems are shown in black); for example, insulin forms non-recycling fibrils, but is of lower molecular mass than SH3 and $SH3_2$ which are recycling proteins. The relative proportions of sequence incorporated into the fibril core is therefore sequence dependent. Algorithms designed to estimate aggregation propensities from a primary sequence have been developed [62, 63], and efforts are underway to determine if the incorporation proportion can be estimated in a similar fashion.

This surprising correlation would suggest that where fibrils have flexible regions on their surface, the tendency of the constituent molecules of a fibril to dissociate is greatly reduced. This would suggest that interactions between non-core regions in the fibril are responsible for this dramatic change in behaviour.

It does however seem intuitively unlikely that flexible residues will interact with each other to the extent required to bring about the large effects on the dissociation rate. In the case of lysozyme, it is known through limited proteolysis experiments that ca. 50 residues are not in the fibril core [48]. Only a small fraction of these residues are observed in the solution-state NMR spectrum. Thus, the residues that are not observed by solution-state NMR, and also are not in the fibril core are those likely to be those responsible for the observed change in recycling behaviour. While flexible regions are observed by solution-state NMR, it is possible that they could be diagnostic of a population of molten globular set of non-core residues whose interactions dominate the effects discussed in this chapter.

Therefore it is important to distinguish between non-core regions, and flexible regions. Flexible regions can be observed by solution-state NMR and are a sub-set of the non-core regions. Their observation therefore is highly indicative of a second type of non-core region, in a molten globular state likely to be interacting with adjacent non-core regions more strongly than the observable flexible regions. Experimental methods such as limited proteolysis and H/D exchange however cannot make this distinction. It is interesting therefore to compare SH3 fibrils and lysozyme fibrils. Both systems have a similar number of residues in the fibril core (ca. 80). Assuming to first approximation that each residue contributes equally to thermodynamic parameters, the differences between the two systems can be discussed in terms of the additional 50 residues of lysozyme, found not to be in the core.

Two possible interpretations are presented to understand the role of the non-core regions in the fibril assembly mechanism. The first is a thermodynamic model that assumes non-core interactions stabilise the fibrillar state using a linear polymerisation model. The second assumes the effect is kinetic and considers the effects of the interactions on the free energy of activation for dissociation.

4.3.1 Thermodynamic rationale

One possible explanation is that the flexible regions on the surface of the fibrils are stabilising the fibrils. As described in more detail in chapter 5, the recycling behaviour of TTR based fibrils is well described by the linear polymerisation model of Oosawa [167].

$$\begin{aligned} M + f_1 &\rightleftarrows f_2 \\ M + f_2 &\rightleftarrows f_3 \\ M + f_3 &\rightleftarrows f_4 \\ M + f_{i-1} &\rightleftarrows f_i \end{aligned} \tag{4.1}$$

In this model, the association constant of each step is given by K_a, and the partition coefficient K_P, the fraction of protein in a soluble state (the concentration of protein in super-2 vs the total protein concentration C_{tot} estimated from the pellet-1 concentration, in the limit $[pellet - 1] >> [super - 2]$, is given by;

$$K_P = \frac{[S2]}{[P1]} = \frac{1}{C_{tot} K_a}$$

If flexible regions on the surface of fibrils acted to increase the association constant to favour higher order aggregates, the concentration of monomer at equilibrium will be smaller. For TTR and SH3 where ca. 10% of the material is in the soluble fraction at equilibrium, the corresponding free energy for elongation is -27 kJ mol^{-1}. Using the electronic spectroscopy it should be possible to detect protein concentrations ca. 0.1 μM, suggesting that using this approach, the free energy of elongation of the class two fibrils will be less than -41 kJ mol^{-1}. In this case, comparing SH3 and lysozyme, non-core residues would be expected to contribute $\geq$ 0.25 kJ mol^{-1} per residue.

4.3.2 Kinetic rationale

The flexible regions act to kinetically trap monomeric material onto the fibril. The correlation between slow recycling rate and flexible regions can be considered using the Eyring equation, $k_{\text{rate}} = \frac{k_B T}{h} e^{-\Delta G \ddagger / k_B T}$. Here, the equation is applied to the dissociation step of the following equilibrium, relating the dissociation rate to the free energy of activation is that between the bound state, and the dissociation transition state (figure 4.11).

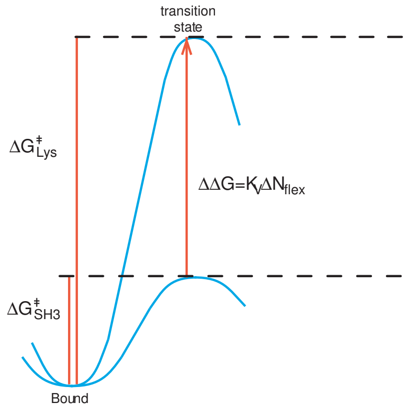


Figure 4.11: Schematic free energy surface comparing the free energy profiles of the off rate for SH3 and lysozyme. As both systems have a comparable number of residues in the fibril core using the coarse grained model described in this section the increase activation free energy in lysozyme fibrils is due to the increase number of residues outside the fibril core.

$$F_{i+1} \rightleftarrows F_i + M \tag{4.2}$$

Two dissociation rates are then related to their respective activation enthalpy and entropy by;

$$\begin{aligned} k_{\text{fast}} &> k_{\text{slow}} \\ \Delta G^{\ddagger}_{\text{fast}} &< \Delta G^{\ddagger}_{\text{slow}} \\ \Delta H^{\ddagger}_{\text{fast}} - \Delta H^{\ddagger}_{\text{slow}} &< T\left(\Delta S^{\ddagger}_{\text{fast}} - \Delta S^{\ddagger}_{\text{slow}}\right) \end{aligned} \tag{4.3}$$

And so, in order for the inequality to be satisfied, $\Delta\Delta H^{\ddagger} < T\Delta\Delta S^{\ddagger}$. Expressions for the activation entropy and enthalpy will now be estimated in terms of the number of residues in the sequence N. To first approximation in this model, solvent effects are neglected and all residues will be considered equivalent. Let ζ be the fraction of the sequence in the core, such that ζN is the number of residues in the core, and $(1-\zeta)N$ is the number of non-core residues.

The enthalpy of activation $\Delta H^{\ddagger}$ will be positive, as favourable interactions will be broken moving between the fibril and the transition state. Let the core residues contribute K_C per residue, related to the breaking of hydrogen bonding in the core. In addition, less specific sticky interactions between flexible domains will add an additional attraction K_V that will scale linearly with number of residues outside the core. This term will lower $\Delta H^{\ddagger}$. With these assumptions, the enthalpy of activation is then $\Delta H^{\ddagger} = N(\zeta K_C + (1-\zeta)K_V)$.

To estimate the enthalpy of activation we will assume the difference in energy between the dissociated state and the fibrillar state is sufficiently large such that the transition state resembles the dissociated state. With such an assumption, using Hammond's postulate, we can associate the free entropy of activation with the entropy of the free chain. Considering the free chain to behave as a swollen Gaussian chain, $\Delta S^{\ddagger} = k_B \zeta N \log \gamma + c$, where the entropy scales linearly with N. This result is consistent with theoretical estimates for the entropy of a random coil (section 4.5.2), and has been demonstrated experimentally to be a good description of poly-alanine chains [168]. γ is related to the conformational freedom of an individual amino acid. Empirically, from studies on poly alanine chains $\gamma = 2.17$ [168] and from simulations, $\gamma = 4.68$ [169].

The ratio of the two rate constants can then be calculated using the Eyring equation, and the expressions for $\Delta\Delta H^{\ddagger}$ and $\Delta\Delta S^{\ddagger}$;

$$\begin{aligned}
\Delta G^{\ddagger} &= \Delta H^{\ddagger} - T\Delta S^{\ddagger} \\
&= (K_C)N\zeta + K_V N(1-\zeta) - T(\zeta N \log \gamma + c) \\
&= N((K_C - K_V - k_B T \log \gamma)\zeta + K_V) + c k_B T \\
&= N_{\text{core}}(K_C - k_B T \log \gamma) + N_{\text{flex}} K_V + c k_B T
\end{aligned} \tag{4.4}$$

The dissociation rates can now be compared between two systems labelled 1 and 2, where $\Delta N_{\text{core}} = N^1_{\text{core}} - N^2_{\text{core}}$ and $\Delta N_{\text{flex}} = N^1_{\text{flex}} - N^2_{\text{flex}}$, and $\Delta\Delta G^{\ddagger} = \Delta G_1 - \Delta G_2$;

$$\begin{aligned}
\frac{k_1}{k_2} &= \exp\left[\frac{-\Delta\Delta G^{\ddagger}}{k_B T}\right] \\
&= \exp\left[-\Delta N_{\text{core}}\left(\frac{K_C}{k_B T} - \log\gamma\right) - \Delta N_{\text{flex}}\frac{K_V}{k_B T}\right]
\end{aligned} \tag{4.5}$$

Thus increasing the number of residues in the flexible chain will be expected to reduce the dissociation rate. The number of residues in the core of SH3 and lysozyme fibrils are comparable, approximately 80 residues, and for lysozyme, 50 of the 130 residues have been found not to be in the core[48]. From the generic hypothesis, interactions within the core would be expected to be similar in both cases, and so any difference in dissociation rate should be attributed to the non-core regions. No recycling rate was measured for lysozyme fibrils.

If the origin for this is that the free energy of activation for dissociation is too large to permit dissociation on these timescales, then the recycling rate for lysozyme can be estimated to be $\geq 1000 \times$ slower than SH3. This is shown by the lack of monomeric material as detected by solution-state NMR months after the sample was prepared. With these assumptions;

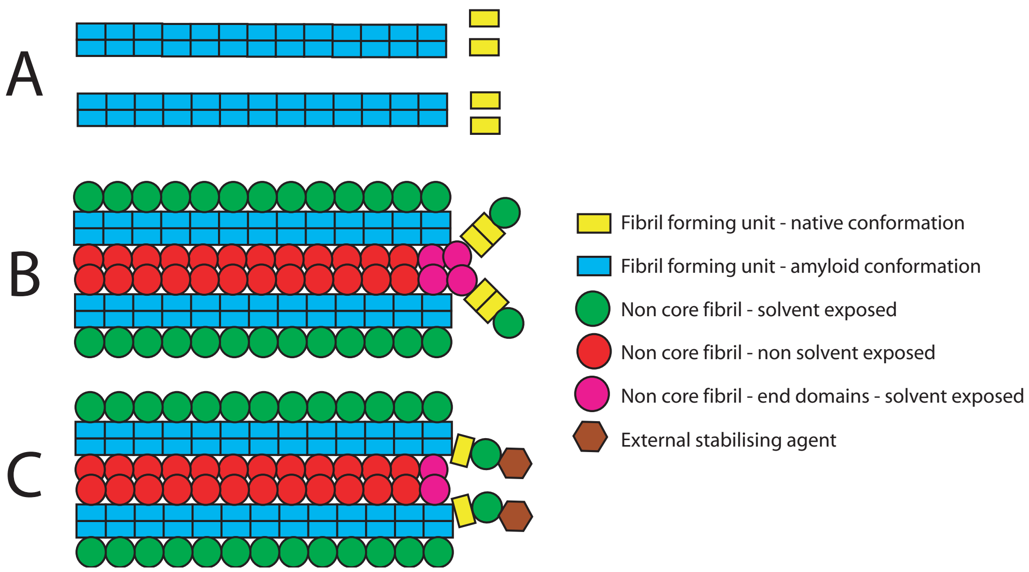


Figure 4.12: Importance of the non-core fibril groups. **A -** Where 100% of a given monomer is in the core, the rate decreases with increasing length as the entropy gain increases with length. **B -** Where there are non-core fibril regions, additional non-specific interactions lead to an increase in the enthalpy of activation that acts to increase the off rate. **C -** Where an independent agent that encourages refolding of the non-core groups is present, degradation of the fibrils from the ends would be expected.

$$\frac{k_{\mathrm{SH3}}}{k_{\mathrm{Lys}}} = \exp\left[N_{\mathrm{flex}}^{\mathrm{Lys}} \frac{K_V}{k_B T}\right] \tag{4.6}$$

K_V can then be estimated to be ca. 0.21 kJ mol^{-1} per residue for lysozyme fibrils. This situation is summarised in figure 4.11. Thus many weak hydrophobic interactions can be expected to give rise to a large increase in the activation free energy, and can readily explain the trends observed in this chapter. These interactions therefore would be expected to contribute very little to the overall stability of the fibril.

If the activation barrier for dissociation increases, then the activation barrier for association must also increase. Lysozyme fibrils form rapidly at 45^{o}C. At this temperature, the association barrier must be relatively small, otherwise fibrils would not form. When a fibril sample is held at this temperature, soluble monomers are not observed to re-appear. Therefore, in the case of lysozyme, it is likely that the thermodynamic rationale is more appropriate than the kinetic rationale.

4.4 Conclusion

The properties of a range of amyloid fibrils from different polypeptides have been examined, and two classes of fibrils are revealed according to their behaviour measured by NMR spectroscopy. One class has a significant quantity of monomeric material ($> 10\,\%$) that dissociates from purified amyloid fibrils within hours (SH3, $SH3_2$ and TTRRGD). Virtually the entire sequence of the monomers in this class of fibril are folded into the β-sheet core of the fibril. The second class is of fibrils have sufficiently flexible regions that can be observed using solution-state NMR, and $< 0.1\%$ monomeric material is found to dissociate from purified fibrils (insulin, αB-crystallin, lysozyme, $SH3_2Cyt$ and $SH3_3Cyt$).

The origin of these two classes must lie in the interactions between non-core regions within the fibrils. These interactions must reduce the tendency of molecular components of fibrils to dissociate. These interactions could either be expected to stabilise the fibrillar state, or increase the dissociation free energy. Non-specific interactions might between flexible regions on the surface of a fibril might be expected to stabilise the associated state, and destabilise the elongation transition state by ca. 0.23 kJ mol^{-1} per residue.

It is likely however, that the stabilisation of the associated state cannot increase in an unconstrained fashion. As described in chapter 2 that the presence of non-core regions perturb the protofilament packing within the fibril. Thus where the packing of the non-core regions begins to significantly distort the fibril core, the dissociated state will again be favoured. This would appear to be the situation for both TTR+10 and TTR+20 systems, introduced and discussed in chapter 5.

In chapter 2 fibrils from SH3, $SH3_2$, $SH3_2Cyt$ and $SH3_3Cyt$ are studied. A significant proportion of the total material from SH3 and $SH3_2$ fibrils are observed to rapidly recycle, whereas negligible material is found to recycle in fibrillar $SH3_2Cyt$ and $SH3_3Cyt$. This transition in behaviour is induced entirely through the addition of sequence, with a low propensity to aggregate, to a core fibril based on that formed by SH3. In addition, the morphology of SH3 and $SH3_2$ fibrils are readily distinguished from the morphology of $SH3_2Cyt$ and $SH3_3Cyt$ fibrils. Thus non-core regions not only effect the energetics of fibril assembly, they also influence protofilament packing.

The fact that non-core regions can destabilise the transition state is demonstrated in chapter 5. Fibrils of TTRRAD and TTRRGD are both shown to have a similar elongation free energy as fibrils from TTR. However, the association and dissociation rates are both increased by an order of magnitude.

These factors are summarised in the 'glue' model (figure 4.12). Where there are flexible regions, additional non-specific interactions increase the enthalpy of activation, decreasing the tendency of a molecular component of a fibril to dissociate (B). Reagents that stabilise the non-core regions, and prevent inter-molecular interactions, or the removal of the flexible regions will therefore be expected to drastically increase the tendency of a monomer to dissociate (C).

Interestingly, the protein sequences implicated in amyloidoses are members of the second class of fibrils (αB-crystallin, lysozyme and insulin); whereas the systems chosen primarily for the property of forming fibrils amenable to biophysical characterisation, with the majority of their sequence in the fibril core are members of the rapidly recycling class (SH3, TTRRGD). If the fibrils that form *in vivo* sequester toxic pre-fibrillar aggregates, then fibrils with a reduced tendency to dissociate would be desirable and might be expected to persist in tissues for longer.

It should be possible, where a significant proportion of a fibril recycles rapidly, processes such as dialysis would be expected to remove the fibrils. Selectively removing the flexible regions of a slowly recycling fibril would be expected to encourage dissociation of monomeric material from fibrils perhaps to a point where dialysis strategies can be employed. In addition, the equilibrium formed between monomers and fibrils would be expected to be sensitive to conditions such as pH, chaperones and other external conditions. In both classes of fibril, stabilising the native state has been shown to disfavour the fibrillar state (TTR and SH3 section 4.2.1, and SH3$_2$Cyt section 2.4.14).

In conclusion therefore, non-core regions appear to reduce the tendency of monomeric material to dissociate from amyloid fibrils. Although the precise details of these interactions on the complex mechanism of amyloid formation are not fully understood, plausible thermodynamic and kinetic arguments have been presented. These results emphasise the unexpected importance of non-core regions in the fibril assembly process.

4.5 Appendix

4.5.1 The measurement of concentrations using NMR spectroscopy

The following arguments are based on those of Wider et al. [160]. Using the principle of reciprocity, the signal strength S measured by a given NMR coil can be expressed by [170]:

$$S \propto B_{90} \int \frac{\partial}{\partial t} \left(\frac{\overrightarrow{B_{1u}} \cdot \overrightarrow{M}}{|\overrightarrow{B_{1u}}|} \right) dV \tag{4.7}$$

where B_{90} is the rf field amplitude of a 90^o pulse measured with the sample in the tuned and matched coil, B_{1u} is the field created by unit current in the volume element dV M is the magnetisation perpendicular to the main magnetic field in dV, and the integration runs over the sample volume. For measurements in aqueous solutions the high temperature approximation is fulfilled and the magnetisation M follows the Curie law $M \propto c/T$ where c is the concentration and T is the temperature in Kelvin. Combining these gives the following proportionality of the signal strength:

$$S \propto \frac{cB_{90}}{T} \int \frac{\partial}{\partial t} \left(\frac{\overrightarrow{B_{1u}} \cdot \overrightarrow{M}}{|\overrightarrow{B_{1u}}| \cdot |\overrightarrow{M}|} \right) dV \tag{4.8}$$

For a given coil, this integral is a constant if the sample in the NMR tube fills the active volume of the coil in the z direction. The rf magnetic strength field strength B_{90} is inversely proportional to the duration of the 90^o. The concentration of an unknown sample c_U can be calculated from he known concentration c_R of a reference sample with:

$$c_U = c_R \frac{S_U}{S_R} \cdot \frac{\theta_{90U} T_U f_U n_U}{\theta_{90R} T_R f_R n_R} \tag{4.9}$$

Where S_U and S_R are the signal strengths of the two samples as an integral over measured resonance intensities, θ is the duration of the 90^o pulse used in the two experiments, n is the number of transients acquired and f is an experiment dependent factor that accounts for different experimental schemes used in measuring S_R and S_U that is, for example, linearly proportional to the receiver gain. In this chapter, where two identical spectral regions are quantitatively compared using identical receiver gains, carrier frequencies and pulse times, n, f, T and θ were identical in both S_R and S_U. Therefore simply;

$$c_U = c_R \frac{S_U}{S_R} \tag{4.10}$$

As discussed by Wider et al. [160], these results are in perfect accord with those obtained independently using the 280 nm absorption in UV-Vis spectra.

The overall error of concentration measurement using this method critically depends on the quality and reproducibility of the integration. For separated resonance lines very reproducible integrals can be obtained. As soon as the lines are no longer separated and close to the baseline, the error increases since the integration range is not clearly defined. But even in this situation the error stays within the limits of corresponding UV measurements unless there is severe overlap of resonances. Integrating over the aromatic region where the total number of protons is constant, relieves this problem. Furthermore, where the relaxation properties of R and U differ, the ratio of the two integrals no longer reflects sample concentration.

The reliability of this method depends on the rf power delivered to the coil in the NMR probe, perfect tuning/matching of the coil, and identical rf cabling, rf filter, and other components in the path from the amplifier to the probe, as is the case where NMR spectra are compared on a single NMR spectrometer. If proper matching cannot be obtained (e.g., due to high salt content), a reliable concentration measurement is not possible and only lower limits are obtained representing the NMR-active concentration. In addition, the ability to use only one reference measurement for all concentration determinations relies on identical filling of the receiving coil. Thus, the NMR tube has to contain a minimal amount of solution which fills the active volume of the coil; any additional sample volume then does not influence the concentration measurement.

4.5.2 Entropy of activation

The entropy of activation can then be estimated from considering free protein to behave as a Gaussian chain. The probability of an end-to-end distance r occurring within a Gaussian chain of N subunits of length l and total length $L = Nl$ can be determined, using central limit theorem, to be $P(r) = \left(\frac{3}{2\pi <r^2>}\right)^{3/2} \exp\left(-\frac{3r^2}{2<r^2>}\right)$, where $< r^2 >= l^2 N = Ll$ and r is the modulus of the end-to-end distance. The entropy is given by $S(r) = \log n_N(r)$. The quantity $n_N(r)$ is the number of chains made of N links having end-to-end distance r, which depends on the probability $p(r)$ in the form $n_N(r) = n_N^{\mathrm{tot}} L^3 \cdot p(r)$ where n_N^{tot} is the total number of chains one can build out of N monomers and L the length scale used to distinguish among conformations. Consequently the Gaussian chain entropy is $S(r) = \log n_N^{\text{tot}} L^3 (2\pi N l^2/3)^{-3/2} - \frac{3r^2}{2<r^2>}$. The total number of chains scales as $n_N^{\text{tot}} = \sum_r n_N(r) = \gamma^{N-1}$ and so the equilibrium entropy, where $r = 0$ is given by:

$$\begin{aligned} S(0) &= \log \gamma^{N-1} L^3 (2\pi N a^2/3)^{-3/2} \\ &= N \log(\gamma) - 3/2 \log(N) + \log(\frac{L^3}{\gamma}(2\pi a^2/3)^{-3/2}) \end{aligned} \tag{4.11}$$

Experiments on poly alanine [168] show that that the entropy does indeed scale as $N \log \gamma + c$. The entropy of the non-core region of the fibril will not change moving from the bound to the dissociated state and noting $S(0)_{\text{fibril}} << S(0)_{\text{flexible}}$, the entropy of activation is reflects on the change in number of core residues moving from the bound state, to the dissociated state. The term in $\log(N)$ will be negligible $N \gtrsim 10$ so:

$$\Delta\Delta S^{\ddagger} = S(0)_{\zeta_1} - S(0)_{\zeta_2} = (N_1\zeta_1 - N_2\zeta_2)\log(\gamma) \tag{4.12}$$

Chapter 5

The effects of flexible regions on TTR fibrils

Abstract

In chapter 4, it was proposed that the presence of non-core regions in amyloid fibrils reduce the tendency of the monomeric constituents of amyloid fibrils to dissociate. To test this, a series of peptides were synthesised based on the amyloidogenic peptide $\mathrm{TTR}_{(105-115)}$. Peptides starting with this sequence were synthesised together with additional sequence designed to be non-aggregating. The length of this additional sequence was varied between 6 and 22 residues. All sequences formed class 1 (section 4.3) amyloid fibrils; $>$ 10% of the total material was found to be present as monomer at equilibrium and no resonances originating from flexible regions on the surface of fibrils were observed using solution-state NMR. The addition of 6 residues changes the morphology of the fibrils without significantly effecting the elongation free energy (ca. -23.5 ± 0.8 kJ mol^{-1} for both TTR+0 and TTR+6). The association and dissociation rates however, were both decreased by over one order of magnitude by the addition of 6 additional residues, corresponding to an increase in the activation free energy of elongation by 8.6 ± 1 kJ mol^{-1}. The addition of a larger number (10 - 22) of residues dramatically reduces the stability of the fibril, whose free energy of elongation rises to ca. -11 kJ mol^{-1}. These results suggest that the non-core residues of amyloid fibrils play an unexpected but important role in fibril assembly.

5.1 Introduction

The studies described in chapter 4 suggest that the presence of residues not in the β-sheet core of amyloid fibrils reduces the tendency of monomeric material within fibrils to dissociate. In order to rationalise this observation, it was proposed that the non-core regions interact with adjacent non-core regions on the surface of a fibril. These interactions might be expected to stabilise the fibril and destabilise transition state of elongation as discussed in section 4.3.

In order to investigate this further, a series of peptides were synthesised based on the highly amyloidogenic $\mathrm{TTR}_{(105-115)}$ fragment [151, 154]. This sequence was elongated with additional sequence chosen to have a low β sheet propensity and hydrophobicity, factors that have been shown to correlate well with the propensity of a sequence to aggregate [62, 63]. The TTR peptide is derived from human transthyretin and is described more fully in appendix 5.4.1. The peptides were synthesised as described elsewhere [154]. The peptides investigated here are summarised in table 5.1, and have 0, 6, 10 and 22 additional residues. Fibrils were formed from these systems and purified according to the scheme described in figure 4.2. The terms pellet-1 and super-2 are defined in this figure and are used extensively in this chapter.

For each fibril, the two tests described in table 4.1.1 were performed, to determine the concentration of soluble monomer present in a purified and equilibrated fibril sample and whether resonances can be observed from flexible residues on the surface of the fibrils using solution-state NMR. In addition, recycling experiments as described in section 4.2.1 were performed. In these experiments, the increase in the concentration of soluble material can be followed by solution-state NMR when pellet-1 samples are resuspended.

The ratio of soluble to total protein in an equilibrated pellet-1 can be used to infer thermodynamic quantities such as the elongation free energy. The rate at which soluble material is observed to increase with time in a recycling experiment combined with the elongation free energy, can be used to obtain association and dissociation rates using Oosawa's linear polymerisation model [167]. In this scheme, monomeric material is free to associate and dissociate from aggregates of a known size, f_i, increasing the size of the aggregate by one:

$$\begin{aligned} M + M &\rightleftarrows f_2 \\ M + f_2 &\rightleftarrows f_3 \\ M + f_3 &\rightleftarrows f_4 \\ M + f_{i-1} &\rightleftarrows f_i \end{aligned} \tag{5.1}$$

Each step is assumed to have the same equilibrium constant, K_{eq}, corresponding to the free energy of elongation, $\Delta G = -RT \ln K_{eq}$. The partition coefficient, reflecting the ratio of soluble material to total material is defined as $K_P = \frac{[S2]}{[P1]}$. In section 5.4.2 *vide infra* that the equilibrium concentration of soluble monomer, given by the concentration of a super-2 ($S2$) sample, and the total concentration of protein, equal to that of a pellet-1 ($P1$) sample are related to K_{eq} by:

$$K_{eq} = \frac{1}{[S2]} - \frac{1}{\sqrt{[S2][P1]}}$$

In the limit where $[S2] << [P1]$, the model reduces to the simpler case where soluble monomers are considered to be in equilibrium with fibril ends E according to the equilibrium scheme:

$$E + M \rightleftarrows E$$

According to this scheme, $K_{eq} = \frac{1}{[S2]} = \frac{k_{-1}}{k_1}$, where k_1 and $k_{-1}$are respectively the forward and backward rates. In addition, in this limit $K_P = \frac{1}{[P1]K_{eq}}$. Furthermore, again in this limit, the differential equation that governs the rate of change of monomer is:

$$\frac{dM}{dt} + k_1 M \cdot E = k_{-1} E$$

This can be solved for $M(t)$ with the boundary condition $M(0) = 0$ as would be the case in a recycling experiment to show that;

$$M(t) = M_{eq} \left(1 - e^{-k_1 E t}\right) \tag{5.2}$$

Where M_{eq}, the concentration of soluble material at equilibrium is equal to the super-2 concentration. This function is of the form of an exponential rise to maximum, $M(t) = M_{eq}\left(1 - e^{-k_{exp} t}\right)$, of characteristic decay $k_{exp} = k_1 E$. This kinetic scheme can be employed proving that the change in monomer concentration is investigated on a timescale where the total number of ends does not change. Provided that the equilibrium concentration of monomer is small and the length distribution of the fibrils is not observed to significantly change, this scheme can be used to analyse experimental data.

In this study, K_{eq} and K_P are estimated for the fibril systems formed from the elongated TTR peptides. In addition, fibrils from TTR and the two TTR+6 peptides were found to be in the correct regime (i.e. $[S2] << [P1]$) for analysis using equation 5.2 allowing k_1 and k_{-1} to be estimated.

Peptide	Residues	M_W / g mol^{-1}	Sequence
TTR	11	1198.3	YTIAALLSPYS
TTRRGD	11+6	1727.8	TTR-GGRGDS
TTRRAD	11+6	1741.9	TTR-GGRADS
TTR10c	11+10	2029.3	TTR-GGKGGEGGKC
TTR22c	11+22	3000.2	TTR-GGKGGEGGKGGEGGKGGEGGKC

Table 5.1: Sequences of modified TTR peptides used, the number of additional residues, molecular mass, and the conditions required for fibril formation.

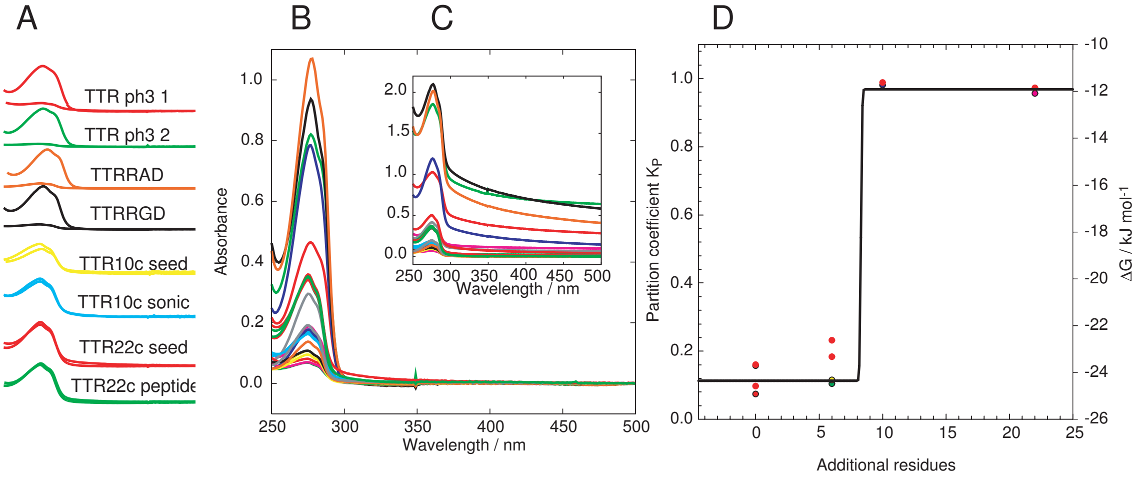


Figure 5.1: Sedimentation analysis of TTR fibrils. The UV-Vis spectra of an equilibrated pellet-1 were compared to that of a super-2, allowing ΔG_{eq} to be estimated. **A -** Direct comparisons of the UV-vis spectra of pellet-1 and super-2 fractions for the TTR species investigated showing the relative absorbance for both fractions for each fibril species studied. The range shown is 250nm to 350nm. **B -** The UV-Vis spectra obtained after removing scattering from the signal by fitting non-absorbing regions to a third order polynomial function. This method of baselining electronic spectra is described in more detail in section 8.2.3. **C -** The raw UV-Vis spectra show the large contribution of light scattering to the measurements. **D -** ΔG_{eq} (red) and K_P (black) estimates for the fibrils as a function of additional groups. The partition constants of TTRRAD/RGD are similar to those of TTR ca. 0.11, whereas those for TTR10c and 22c are significantly larger ca. 0.95.

5.2 Results

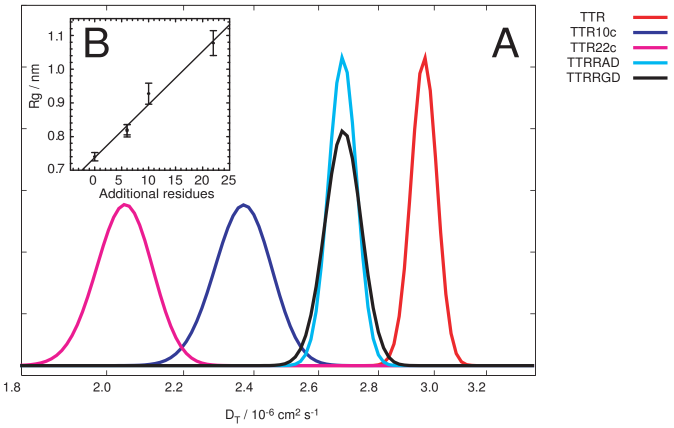
All the peptides studied successfully formed fibrils. Limited proteolysis experiments coupled with SDS and mass spectrometry have shown that the additional residues are solvent exposed and proteolytically accessible, unlike the TTR$_{(105-115)}$ core. The morphologies of the fibrils were however different in each case, with the fibrils becoming more twisted with as the number of additional resonances was increased (data not shown). However, the fibre XRD profiles show that the structure of the underlying protofilament is similar in all cases (data not shown) suggesting that the protofilament packing is perturbed by the addition of non-core regions. This type of perturbation of the protofilament packing has been observed elsewhere [46, 51].

Fibril samples were purified by ultracentrifugation, and pellet-1 samples were prepared. Subsequent centrifugation of equilibrated pellet-1 samples determined that significant quantities of soluble monomeric material was present in the super-2 fractions. Comparison of the protein concentration in pellet-1 and super-2 samples allowed the partition coefficients and elongation free energies of various fibrils to be estimated. These are shown as a function of additional residues in figure 5.1.

Remarkably, fibrils from the two systems with +6 additional residues were found to have similar partition constants and elongation free energies as the fibrils formed from TTR alone (0.11 ± 0.06 and 0.11 ± 0.02 respectively), corresponding to 11% of the total material present as soluble material at equilibrium. In contrast, fibrils formed from sequences with significantly longer sequences (+10 and +22) were found to be considerably less stable (partition coefficient 0.97 ± 0.02). As controls, the same recycling experiments were performed on monomeric peptide only and the spectra were found to be identical in both fractions (K_P=1), and no fibrils were observed by TEM. These data are summarised in table 5.2.

Pellet-1 samples were studied by solution-state NMR. At all times, diffusion measurements showed that monomeric protein was observed, whose translational diffusion coefficients decrease with the number of additional residues (figure 5.2). As was discussed in chapter 4, flexible regions on the surface of amyloid fibrils would be readily detected using NMR diffusion measurements. However, these were not observed in any of the systems studied. All the TTR based systems therefore, are members of the class 1 fibrils, as defined in section 4.3.

Recycling experiments as defined in section 4.2.1 were performed on resuspended pellet-1 samples, and the ^{1}H NMR signal intensity was observed to increase over time, eventually reaching a saturation point. Integrated ^{1}H NMR signal intensities are shown as a function of time, normalised to the signal end point in figure 5.3. NMR diffusion measurements confirmed that these resonances orig-



	D_T / 10^{-6}cm^{2}s^{-1}
TTR	2.97 ± 0.05
TTRRAD	2.68 ± 0.05
TTRRGD	2.69 ± 0.06
TTR10c	2.37 ± 0.08
TTR22c	2.04 ± 0.07

Figure 5.2: Diffusion data from equilibrated pellet-1 fractions of the fibrils studied in this chapter. **A -** The width of the trace indicated the experimental uncertainty in the measurement. The diffusion data indicates that monomeric protein is observed in all cases, and that the diffusion coefficient scales well with the mass of the observed peptide. **B -** R_G as a function of additional residues.

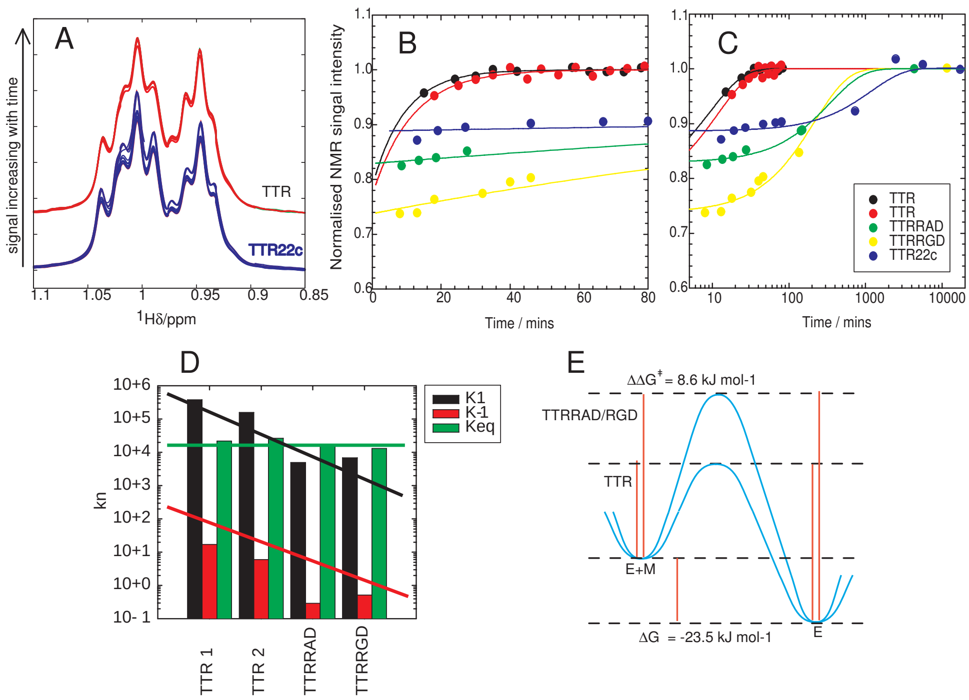


Figure 5.3: Rate of re-appearance of monomer after resuspension of pellet for TTR fibrils. **A -** ^{1}H NMR spectra showing the increase in signal with time in resuspended pellets of TTR and TTR22c. **B/C -** Change in signal intensity of resuspended pellets with time for TTR, TTRRAD, TTRRGD and TTR22c fibrils. The intensity change was fit to an exponentially increasing function. The first 80 minutes are shown in B, and the time course over two weeks is shown with a logarithmic scale in C. **D -** The rate constants k_1 and k_{-1} for two repeats of TTR, TTRRAD and TTRRGD. **E -** A comparison of the elongation free energy surfaces of TTR+0 and TTR+6. The free energy of elongation was found to be comparable in both cases (-23.6 ± 0.8 kJ mol^{-1}). However the free energy barrier is significantly higher in the case of TTR+6 ($\Delta\Delta G$ = 8.6 ± 1 kJ mol^{-1}).

inate from monomeric protein (figure 5.2). As the only source of monomeric material in the samples are the fibrils, it is reasonable to conclude that this increase in signal intensity reflects molecular components dissociating from fibrils, restoring the equilibrium composition. The increase in signal intensity was found to be well described by equation 5.2 (figure 5.3B and C). The experimental rate constant was found to decrease by 3 orders of magnitude (table 5.2), progressing to TTR variants with increasing added residues, from TTR to TTR+6 to TTR+22. In particular, the experimental rate constants of TTRRAD and TTRRGD were identical within experimental uncertainty, and are two orders of magnitude lower than that found for TTR. An apparent burst phase was observed in all cases as the NMR signal intensity extrapolated to the point of re-suspension is not zero. The origin of this is unclear. Some monomeric material could co-sediment with fibrils or the mechanical action of pipetteing fibrils on re-suspension could be expected to shear fibrils [133]. Such effects would not effect the later time points from which the rate constants have been derived.

With a suitable model for the number of fibril ends, E, it is possible to estimate both k_1 and k_{-1}. The coarse grained model described here is identical to that described in section 2.4.12. The internal density of globular proteins has been shown to be $\rho = 1.37 \times 10^3$kg m^{-3} from analytical ultracentrifugation measurements [147]. Amyloid fibrils are modeled here as being cylinders of this density. The mass of a fibril of length L and radius r will this model is given by $M = \rho\pi r^2 L$. The number of moles of monomer N, within a fibril is given by $N = \frac{M}{M_w}$ where M_w is the molar mass of the monomer. This leads to $N = \frac{\rho\pi r^2 l}{M_w}$. The mean length due to a mole of monomers is then $L_1 = \frac{M_w}{\rho\pi r^2}$, and so $N = \frac{L}{L_1}$. As the concentration of protein within the fibrils, C_P, is known from the difference between the total protein concentration $[P1]$ and the concentration of soluble protein $[S2]$, the concentration of fibrils, C_F can be estimated to be $C_F = \frac{C_P}{N_A N} = \frac{C_P L_1}{N_A L}$ where N_A is Avogadro's constant. The concentration of fibril E ends is therefore $2 \cdot C_F$. For fibrils of length 1μm and radius 2nm [53], E is ca. four orders of magnitude lower than C_P, with ca. 8,600 molecules per fibril. In order for the mean spacing of monomers to be 4.68Å, the fibril would be expected to be formed from 4 protofilaments, consistent with AFM and CryoEM measurements. k_1 and k_{-1} can then be distinguished, where the partition constant and the mean length of the fibrils is known.

$$\begin{aligned} k_1 &= \frac{k_{exp}}{E} = \frac{k_{exp} N_A L}{2L_1([P1] - [S2])} \\ k_{-1} &= k_1[S2] \end{aligned} \tag{5.3}$$

System	Residues	K_P	ΔG_{eq}/ kJ mol^{-1}	$k_{\exp}$/ min^{-1}	k_1/ mol dm^{-3} min^{-1}	k_{-1}/ min^{-1}
TTR	+0	0.12 ± 0.05	-24.1 ± 0.6	$(1.1 \pm 0.2) \times 10^{-1}$	$(1.34 \pm 0.8) \times 10^{6}$	57.9 ± 20
TTR+6	+6	0.11 ± 0.08	-23.0 ± 0.5	$(3.8 \pm 1) \times 10^{-3}$	$(2.94 \pm 0.7) \times 10^{4}$	2.02 ± 0.8
TTRRAD	+6	0.10	-23.3	$(2.9 \pm 0.4) \times 10^{-3}$	4.96×10^{3}	2.91×10^{-1}
TTRRGD	+6	0.12	-22.6	$(4.6 \pm 0.8) \times 10^{-3}$	6.82×10^{3}	5.17×10^{-1}
TTR10c	+10	0.97	-11.6	-	-	-
TTR22c	+22	0.95	-11.9	$(8.5 \pm 4) \times 10^{-4}$	-	-

Table 5.2: Partition constants and recycling rates obtained for the TTR systems studied in this chapter. The uncertainty estimates for TTR come from 2 repeats. The uncertainty estimates for TTR+6 come from averaging the TTRRAD and TTRRGD values. The partition coefficients and ΔG_{eq} are plotted against addition residues in figure 5.1D.

For both TTR and TTR+6 fibril preparations, the mean fibril length was observed to be ca. 1 μm by TEM. This estimate allows the the rate constants k_1 and k_{-1} to be estimated for these systems (figure 5.3D). As previously described, these systems have identical equilibrium and partition coefficients within experimental uncertainty. However, both rate constants are reduced by a factor of 0.028 ± 0.01 in the TTR+6 peptide when compared to the TTR+0 peptide, corresponding to a $\Delta\Delta G$ of 8.6 ± 1 kJ mol^{-1}. The calculation of a $\Delta\Delta G$ only assumes that the mean length of the two fibril preparations are comparable. The corresponding free energy surface describing the elongation barrier of TTR+0 and TTR+6 is shown in figure 5.3E.

This value could not be determined for the TTR+22 system as ca. 96% of the material returns to solution and the $[S2] << [P1]$ required for application of equation 5.2 does not apply. These parameters are summarised in table 5.2.

A further experiment to test the approach of equilibrium from the opposite direction is to add a quantity of monomers M_1. In this case, we should expect to observed the signal to decrease with time according to $M(t) = M_{eq} + M_1 e^{-k_1 E t}$.

5.3 Conclusion

The effects of additional sequence of low aggregation propensity onto a sequence with an intrinsically high propensity to form amyloid fibrils has been studied, and several effects have been identified. All fibrils studied in this chapter are class 1 fibrils as defined in section 4.3 as a significant proportion ($>$ 10%) of the total material is monomeric at equilibrium and no resonances were observed from the fibrils using solution-state NMR.

The β sheet core of all fibrils was shown to be highly similar in each case

(XRD), though the higher order structure imposed by the protofilament packing was found to vary. The stability of the fibrils does not necessarily correlate with the morphology observed in the TEM. Particularly, TTR, TTRRAD and TTRRGD fibrils were found to have similar elongation free energies (table 5.2), but the observed morphologies were found to vary distinctly, with TTRRAD/TGD fibrils being more twisted that those of TTR. It has been noted elsewhere [53] that the differences in the angular correlations [53] between fibrils of TTR are greater than the differences in the angular correlations within a single fibril, suggesting that several protofilament arrangements can have similar energies, and it has been shown theoretically that twisting twisted ribbon structures could relieve steric pressure with little energetic cost [112, 116]. Thus it would appear that the additional non-core regions in the TTR+6 fibrils apply a steric pressure that is relieved through the favouring of a more twisted conformation. It is remarkable that this twisting does not effect the elongation free energy.

Adding a larger number of residues significantly destabilises the fibrils; for both TTR+10 and TTR+22 96 ± 1% of the total material was found to be monomeric in purified equilibrated fibril samples, corresponding to a decrease in the elongation free energy of 11 kJ mol−1. This would indicate that the steric constraints imposed by the further addition of residues is no longer relieved by twisting the fibril, and the protofilament packing is significantly destabilised as a consequence. This observation is in accord with the design principles outlined in chapter 2, where it was noted that a large fibril forming system such as $SH3_2$ would be required to support a large displayed functional group, such as an electron transport protein in order to successfully produce fibrils.

The addition of additional residues to TTR has been shown to have a significant effect on the association and dissociation rates. By modelling the recycling process as a two-state equilibrium, both rates are seen to decrease by more than one order of magnitude through the addition of 6 residues (figure 5.3E). This corresponds to the raising of the barrier height by 8.6 ± 1 kJ mol^{-1}. A physical rationale for this is described in section 4.3.2. Additional non-core regions in a fibril forming peptide would be expected to decrease the likelihood of a monomer successfully colliding with a fibril end in a conformation that will lead to elongation, and additional non-specific interactions between adjacent non-core regions on the surface of a fibril would be expected to decrease the dissociation rate. It is intriguing however that these interactions have little effect on the overall stability of the fibril (figure 5.1). It is possible that the destabilisation of the fibril core and the stabilisation of non-core residues exactly cancel.

If the disordered regions within fibrils were entirely flexible, it is likely that it would be possible to observe solution-state NMR resonances from these residues.

This was not found to be the case. This observation is in accord with the kinetic data, which imply that the flexible regions are interacting on the surface of the fibril, destabilising the transition ensemble. These interactions are most likely responsible for the reduced mobility of the non-core regions thus preventing the motional narrowing required for flexible regions to be observed by solution-state NMR when attached to larger bodies.

This conclusion further refines the model discussed in section 4.4, suggesting that there are two types of non-core region. The first behaves like a free chain and results in sufficiently flexibility to be observable by solution-state NMR. The second, between the fibril core and the truly flexible regions are engaged in transient, non-specific interactions that have large effects on the kinetics of fibril formation. As a consequence are insufficiently mobile to be observed by solution-state NMR.

This observation is supported by the experiments described in chapter 4. Limited proteolysis experiments have shown that ca. 80 residues are in the core of fibrils from human lysozyme, and ca. 50 non-core residues that are proteolytically accessible [48]. Despite this, very few residues are observable to solution-state NMR (section 4.2.2) and so the majority of the non-core residues are not observed and act to reduce the tendency of constituent monomers of lysozyme fibrils to dissociate.

The non-core regions of amyloid fibrils are not observed by solid-state NMR experiments. Thus the non-core regions that are unobservable to both solid and solution-state NMR techniques most likely have properties that are molten globular, where many conformations in slow or intermediate exchange are populated at equilibrium.

The work described in this chapter demonstrates that non-core regions of amyloid fibrils can have large effects on the kinetic pathway of fibril elongation. If the interactions between flexible residues are sufficiently strong, then the dissociation rate could be decreased to the point where the all monomers are kinetically trapped within the fibril, as observed in class two fibrils (section 4.4), particularly in the transition from class one, to class two behaviour in the SH3, $SH3_2$, $SH3_2Cyt$, $SH3_3Cyt$ series. The interactions between the flexible regions of TTR10c and TTR22c fibrils are insufficiently strong to cause this change in behaviour. The sequences were selected for their low propensity to aggregate, and are composed primarily of glycine residues. Thus the strength of the interactions between flexible regions on the surface of amyloid fibrils are partially mediated by the nature of the amino acids involved. These matters are of particular importance in understanding the role of amyloid fibrils in amyloidoses, and to those who would build functionalised nanomaterials from amyloid fibrils.

5.4 Appendix

5.4.1 Transthyretin (TTR)

Full length transthyretin (TTR) is a 55 kDa protein involved in the transport of thyroxine and retinol in plasma. WT TTR forms amyloid fibrils *in vivo* in a condition termed senile systemic amyloidosis [149], and a number of naturally occuring TTR variants are associated with familial amyloid polyneuropathy [150]. Full-length TTR [151], TTR variants [152, 153], and two 11 residue peptide fragments derived from the native sequence [151] have been shown readily to form amyloid fibrils *in vitro*, one of which is $\mathrm{TTR}_{105-115}$. This fragment corresponds to β strand G in the full length protein, and has been studied as a model system for amyloid assembly [154]. Fibrils are formed from this peptide by incubating the fragment at room temperature, at 10 mg ml^{-1} in 10% acetonitrile for one week. The conformation of a single peptide within a fibril has been determined [39] (figure 1.6).

5.4.2 Equilibrium schemes for polymerisation

The following schemes were considered to understand the data presented in this chapter. The 'Two State B' model more commonly termed the liner polymerisation model of Oosawa, and is the model chosen to best describe the data presented in this chapter.

Two species

Considering an equilibrium between two chemical species, A and E such that $A \rightleftarrows E$ where $E_0 = A + E$, $K_{eq} = \frac{E_{eq}}{A_{eq}} = \frac{k_1}{k_{-1}}$. This leads to the following differential equation $\frac{dA}{dt} + (k_1 + k_{-1})A = k_{-1}A_0$, which can be solved to show $A = A_{eq}\left(1 - e^{-(k_1+k_{-1})t}\right)$, with the boundary condition of (0,0). This model is un-physical as it assumes the aggregation reaction is uni-molecular. In addition, there is no distinction between the bulk of the material in a fibrils, and the ends of a fibril.

Two state A

If the number of ends on the fibril is constant throughout the reaction, then the addition of an extra monomer yields another fibril end. The equilibrium under consideration is $A + E \rightleftarrows E$ where $K_{eq} = \frac{1}{A_{eq}} = \frac{k_1}{k_{-1}}$. This leads to the following differential equation, $\frac{dA}{dt} + k_1 A \cdot E = k_{-1}E$ which can be solved to show that $A(t) = A_{eq}\left(1 - e^{-k_1 E t}\right)$ under the boundary condition $A(0) = 0$.

This model assumes the number of fibril ends is a constant during the recycling process and so negates scission processes, and that the concentration of monomer at equilibrium reflects k_1 and k_{-1}. The quantity of monomer at equilibrium reflects the position of the association equilibrium. As k_{-1} falls, or k_1 increases, the concentration of monomer expected at equilibrium will increase. This model is a limit of the linear polymerisation model.

Two state B

Oosawa's linear polymerisation model [167]. The following scheme is considered where successively increasing aggregates f_i are in equilibrium with monomeric material, M.

$$\begin{aligned} M + f_1 &\rightleftarrows f_2 \\ M + f_2 &\rightleftarrows f_3 \\ M + f_3 &\rightleftarrows f_4 \end{aligned} \tag{5.4}$$

This is more efficiently represented by $M + F_{i-i} \rightleftarrows F_i$ and so $k_a = \frac{[F_i]}{[M][F_{i-1}]}$. The total concentration of monomer is therefore $M_0 = \sum_{i=0}^{\infty} i[F_i] = \sum_{i=0}^{\infty} i k_a^{i-1}[M]^i = \frac{[M]}{(1-k_a[M])^2}$,[167] where the last step requires use of the series for $(1-x)^{-2}$. Rearranging gives $k_a = \frac{1}{[M]} - \frac{1}{\sqrt{[M]M_0}}$. At equilibrium therefore $[M] = \frac{1}{k_a} - \frac{1-\sqrt{1+4k_aM_0}}{2M_0k_a^2}$.

$$K_P = \frac{1}{C_{tot}k_a}\left(1 - \frac{1}{2M_0k_a} + \sqrt{\frac{1}{4M_0^2k_a^2} + 1}\right)$$

In the limit $M_0 > M$, the equilibrium expression reduces to that of the TWO STATE A model.

Three state A

If a 'docked' protein is represented by ES, and once 'locked' this protein cannot dissociate, we can represent the equilibrium by $A + E \rightleftarrows ES \rightarrow E$.

$$\begin{aligned} \frac{dA}{dt} &= -k_1A \cdot E + k_{-1}ES \\ \frac{dE}{dt} &= k_2ES \\ \frac{dES}{dt} &= +k_1A \cdot \textbf{E -}k_{-1}ES - k_2ES = -\frac{dA}{dt} - \frac{dE}{dt} \end{aligned} \tag{5.5}$$

Placing the docked protein in the steady state yields the following expressions:

$$\begin{aligned}\frac{dES}{dt} &= k_1AE - k_{-1}ES - k_2ES = k_1A(E^0 - ES) - k_{-1}ES - k_2ES = 0\\ ES &= \frac{k_1AE^0}{k_1A + k_{-1} + k_2} = \frac{A\cdot E^0}{A + \frac{k_{-1}+k_2}{k_1}}\\ k_M &= \frac{k_{-1}+k_2}{k_1}\end{aligned} \tag{5.6}$$

This allows the derivation for an expression giving the rate of change of concentration of monomer:

$$\begin{aligned}\frac{dA}{dt} &= k_{-1}ES - k_1A\cdot E = (k_{-1} + k_1A)ES - k_1A\cdot E^o\\ \frac{dA}{dt} &= (k_{-1} + k_1A)\frac{A\cdot E^0}{A + k_M} - k_1A\cdot E^o\\ \frac{dA}{dt} &= \frac{k_2A\cdot E^o}{A + k_M}\end{aligned} \tag{5.7}$$

Which applies only when the steady state criterion of $\frac{dES}{dt} = 0$ is satisfied, and is analogous to the Michaelis-Menten expression for enzyme kinetics. The model will result in 100% of monomer eventually being converted into fibrils. Where $k_2 > k_{-1}$, this model becomes kinetically indistinguishable from the two species model. This model is used to analysed the aggregation of A$\beta_{(1-40)}$ by Esler et al. [171]

Three state B

A more complex situation is revealed when the protein is allowed to 'unlock', as represented by the equilibrium $A + E \rightleftarrows ES \rightleftarrows E$, where $E_0 = ES + E$. Let $K_0 = K_DK_L$ and $K_D = \frac{f_i^*}{f_{i-1}M}$ and $K_L = \frac{f_i}{f_i^*}$. $f_i = K_D^{i-1}K_L^{i-1}[M]^i$, $f_i^* = K_D^{i-1}K_L^{i-2}[M]^i$, and so $M_0 = \sum_{i=1}^{\infty} i(f_i + f_i^*) - \frac{[M]}{K_L} = \left(1 + \frac{1}{K_L}\right)\frac{[M]}{(1-K_DK_L[M])^2} - \frac{[M]}{K_L}$.

Three coupled differential equations result from kinetic considerations of this treatment:

$$\begin{aligned}\frac{dA}{dt} &= -k_1E\cdot A + k_{-1}ES\\ \frac{dE}{dt} &= -k_1AE + k_{-1}ES + k_2ES - k_{-2}E = \frac{dA}{dt} + k_2ES - k_{-2}E\\ \frac{dES}{dt} &= k_1AE - k_{-1}ES - k_2ES + k_{-2}E = -\frac{dA}{dt} - k_2ES + k_{-2}E = -\frac{dES}{dt}\end{aligned} \tag{5.8}$$

There exist three equilibrium constants $K_D = \frac{ES_{eq}}{A_{eq}E_{eq}} = \frac{k_1}{k_{-1}}$, $K_L = \frac{E_{eq}}{ES_{eq}} = \frac{k_{-2}}{k_2}$

and $K_O = \frac{1}{A_{eq}}$, where $K_3 = K_2 \cdot K_1$. The equilibrium concentrations are then $A_{eq} = \frac{k_{-1}k_{-2}}{k_1 k_2}$, $E_{eq} = \frac{k_2}{k_2+k_{-2}}E^o$ and $ES_{eq} = \frac{k_{-2}}{k_2+k_{-2}}E^o$. The equilibrium concentrations of the relevant species are specified by combinations of the fundamental rate constants. Putting ES and E in the steady state:

$$\begin{aligned} \frac{dES}{dt} &= k_1 A(E_0 - ES) - k_{-1}ES - k_2 ES + k_{-2}(E_0 - ES) = 0 \\ ES &= E_0 \frac{k_1 A + k_{-2}}{k_1 A + k_{-1} + k_2 + k_{-2}} = E_0 \frac{A+\alpha}{A+\beta} \end{aligned} \tag{5.9}$$

with $\alpha = \frac{k_{-2}}{k_1}$ and $\beta = \frac{k_{-1}+k_2+k_{-2}}{k_1}$. These expressions can be used to obtain a differential equation for the rate of change of A.

$$\begin{aligned} \frac{dA}{dt} &= -k_1 E \cdot A + k_{-1}ES = -k_1 A \cdot E^o + (k_{-1} + k_1 A)ES \\ &= -k_1 A E^o \left(1 - \left(\frac{k_{-1}}{k_1 A} + 1\right) \frac{k_1 A + k_{-2}}{k_1 A + k_{-1} + k_2 + k_{-2}}\right) \end{aligned} \tag{5.10}$$

Alternatively, it can be expressed as:

$$\begin{aligned} \frac{dA}{dt} &= -k_2 ES + k_{-2}E = -k_2 ES + k_{-2}(E_0 - ES) \\ \frac{dA}{dt} &= E_0 k_{-2} - (k_2 + k_{-2})E_0 \frac{A+\alpha}{A+\beta} \\ &= k_{-2}E_0 \left(1 - \left(\frac{k_2}{k_{-2}} + 1\right) \frac{k_1 A + k_{-2}}{k_1 A + k_{-1} + k_2 + k_{-2}}\right) \end{aligned} \tag{5.11}$$

In order for an appreciable concentration of monomer to appear at equilibrium, $k_{-1}k_{-2} > k_1 k_2$. If $k_1 A >> k_{-1} + k_2 + k_{-2}$ the situation becomes kinetically indistinguishable from the TWO STATE A model and the initial rate becomes $k_{-1}E^o$. Where $k_2 ES >> k_{-2}E$, the model becomes kinetically indistinguishable from the THREE STATE A model.

5.4.3 Summary

The schemes considered in this chapter are summarised below.

Model	A_{eq}	rate	initial rate
Two species	$\frac{k_1}{k_{-1}} = K_{eq}$	$A = A_{eq}(1 - e^{-(k_1+k_{-1})t})$	$k_1 + k_{-1}$
Two state A	$\frac{k_{-1}}{k_1} = \frac{1}{K_{eq}}$	$A = A_{eq}(1 - e^{-k_1 E t})$	$k_{-1}E$
Three state A	-	-	$\frac{k_2 A E^o}{A+k_M}$
Three state B	$\frac{k_{-1}k_{-2}}{k_1 k_2}$	-	$k_{-2}E_0 \left(1 - \left(\frac{k_2}{k_{-2}} + 1\right) \frac{k_1 A + k_{-2}}{k_1 A + k_{-1} + k_2 + k_{-2}}\right)$

Chapter 6

On the effective τ_C of flexible regions

Abstract

The internal dynamics of flexible groups covalently attached to larger structures, narrows the observed resonances in a solution-state NMR experiment. This effect is particularly remarkable where resonances from the bulk of the biological structure are too broad to be interpreted, whereas due to the motional narrowing, the flexible domain can be readily observed and studied. It is demonstrated here, by explicitly calculating the effective correlation times of simulated trajectories of flexible regions attached to slowly tumbling species that the correlation time can be reduced by up to six orders of magnitude, in line with experimental observations.

6.1 Introduction

In order for high resolution solution-state NMR techniques to yield structural and dynamic information, the width of the observed resonances must be sufficiently narrow such that individual resonances can be uniquely identified. The width of the resonance is therefore of crucial importance in determining the information content of a solution-state NMR spectrum. The theory of Bloch, Wagness and Redfield [139, 172, 140] describes how random fluctuations in the local field at characteristic timescales decreases the transverse correlation time, T_2, and hence broadens observed resonances. For spin 1/2 nuclei in biological systems, the primary mechanism for broadening is the dipolar interaction between protons. For like spins, it is given by [173]:

$$\frac{1}{T_2} = \frac{3}{8}\gamma^4\hbar^2 I(I+1)\left(J^2(2\omega) + 5J^1(\omega) + J^0(\omega)\right) \tag{6.1}$$

For rigid molecules, the process that gives rise to randomly fluctuating fields is rotational tumbling. In this case, the spectral density functions J, are functions of the rotational correlation time $\tau_c = \frac{4\pi\eta r^3}{3kT} = \frac{V\eta}{kT}$. This timescale is of great importance in solution-state NMR as it characterises the timescale over which local magnetic fields fluctuate, and so describes the timescale and hence sets the line-width of observable resonances. In the limit of $\omega\tau_c < 1$, as is the case of large bio-molecules such as proteins, T_2 becomes inversely proportional to τ_c, and so the line-width is proportional to τ_c.

Empirically it has been observed that local dynamics decrease the width of observed NMR linewidth. When a region within a larger molecule has a high degree of conformational flexibility, the characteristic time of the local fluctuations decreases from that expected for a rigid body. For example, unstructured states of protein generally have smaller linewidths than corresponding structured states[145] and individual domains within multi-domain complexes have sharper lines than a single domain of equal mass[143, 144].

A particularly interesting manifestation of this effect is where the tumbling time of the molecule of interest is so short that it leads to unresolvably broad resonances, but the motion of flexible regions attached the molecule is sufficient to yield resonances whose linewidth is comparable to a significantly smaller protein, free in solution. This situation is observed in large 800KDa chaperone complexes of αB crystallin [146], in the L7/L12 stalk of the ribosome [103], and in flexible regions on amyloid fibrils [51, 50, 52]. NMR diffusion measurements can be used to unambiguously determine the effective size of the underlying species from which the resonances originate [174].

In this chapter the extent to which internal dynamics can be expected to narrow NMR resonances is investigated. By explicitly simulating trajectories of flexible chains attached to larger domains undergoing rotational diffusion, the effective correlation time of the local field fluctuations can be estimated. It is demonstrated that short flexible regions can be expected to lead to a reduction reduction in the correlation time by several orders of magnitude, in line with the experimental observations.

6.1.1 NMR Theory

The local magnetic field experienced by the ith nuclear spin due to the dipolar field of the jth at time t is given by the following, where $r_(t_{ij})$ is the distance between i and j at time t, $\theta(t_{ij})$ is the angle between the $\overline{ij}$ vector and the external field:

$$B_i(t) = \sum_{j}^{N} \frac{1}{r(t_{ij})^3} (3\cos^2\theta(t_{ij}) - 1) \tag{6.2}$$

The autocorrelation function of the fluctuating field around residue i, for n time points is given by

$$G_i(\tau) =< B_i(t)B_i(t+\tau) >= \frac{1}{(n-1)\overline{B_i^2}} \sum_{j}^{n-\tau-1} \left(B_i(t) - \overline{B_i(t)}\right)\left(B_i(t+\tau) - \overline{B_i(t+\tau)}\right) \tag{6.3}$$

The autocorrelation function will decay from unity at $\tau = 0$, to zero. Fitting the decay curve to a single exponential function yields a correlation time defined by $G_i(\tau) = e^{-\tau/\tau_c}$. By simulating the dynamic trajectories of a polymer chain will allow $B_i(t)$ to be explicitly calculated, and hence the calculation of an effective correlation time.

6.1.2 Polymer Theory

Two suitable models for a flexible chain are the Gaussian chain, and the worm-like chain. In both, the chain is modelled as an N-unit chain of bond length r_b. The geometry of the peptide bonds means that each peptide unit in the chain can be modeled as a hard sphere of diameter 5Å. Excluding bond vibration, the energy of the chain in a given conformation is given by:

$$E_{\text{chain}} = \frac{1}{2} k_s \sum_{i}^{N} (\theta_i - \theta_0)^2 \tag{6.4}$$

where k_s is the spring constant of the chain, θ_i is the bond angle between two adjacent segments and θ_0 corresponds to the conformation of lowest energy. In this work, θ_0 is set to be π, and so the chain has its lowest energy when it is fully extended. For a Gaussian chain, k_s=0 and all conformations have equal energy. The spring constant k_s is related to the persistence length l_P, where $\overline{u}$ is the vector pointing from the ith position, to the $(i + 1)$th position:

$$
\begin{aligned}
E &= \frac{1}{2} l_P k_B T \int_0^L \left(\frac{\partial \overline{u}}{\partial s}\right)^2 ds \\
\int_0^L \left(\frac{\partial \overline{u}}{\partial s}\right)^2 ds &= \sum_i^N \left(\frac{\overline{u}(s_i) - \overline{u}(s_{i-1})}{\Delta s}\right)^2 \Delta s \\
&= \sum_i^N \frac{1}{\Delta s}\left(\overline{u}(s_i)^2 + \overline{u}(s_{i-1})^2 - 2\overline{u}(s_i)\cdot\overline{u}(s_{i-1})\right) \\
&= \sum_i^N \frac{2}{r_b}(1-\cos\alpha) \approx \sum_i^N \frac{\alpha^2}{r_b}
\end{aligned}
\tag{6.5}
$$

And so where θ is small, setting $\theta_0 = \pi$ and noting $\alpha = \theta - \pi$:

$$
k_s = \frac{l_P}{r_b} k_B T \tag{6.6}
$$

An ensemble of conformations that represent a trajectory can be built up using a time approximated Monte-Carlo technique [175]. For each step, a random position within the chain is taken, and a new solid angle is randomly selected. The acceptance of each step is subject to the Metropolis condition $e^{-\gamma} > \xi$ with ξ is a random number between 0-1 and $\gamma = E_{\text{chain}}/k_B T$ where k_B is the Boltzman constant and T is the temperature [176]. To make the walk self-avoiding, if a step leads to two hard spheres to overlap, a large energy penalty is added.

6.1.3 Monte Carlo justification

The most important requirement for achieving quantitative agreement with experimental data is that the free energy landscape must be sufficiently realistic. To sample all the events to obtain a correlation time, kinetic trajectories ca. 10 ms are required, sampling conformations on the ps timescale (10^{10}) structures) in order to understand the relaxation phenomena of interest. Simulation times of this length are difficult using conventional molecular dynamics techniques; a Monte Carlo approach is has been employed. The use of a Monte Carlo approach to estimate the relaxation rate of simulated systems [177, 178, 179, 175] is justified heuristically. If kinetic events are separated by a sufficiently large number of MC steps, and if they are observed in a large ensemble of trajectories, they represent a population shift, and reflect the properties of the free energy landscape such as the depth of the local minima and the barrier heights. Examining an ensemble of trajectories therefore highlights the major kinetic events during the simulated process by averaging out the short MC events.

In this case the quantity of interest, $(B(t))$, decays as $\frac{1}{r^3}$ and its autocorrelation function with $\frac{1}{r^6}$. Only the nearest neighbours along the chain will influence these measurements. The chosen move-set consists of a random update in a randomly chosen torsional angle. In order to be able to make quantitative predictions and arrive at reliable estimates for the time, the MC trajectories have to be calibrated against experimental and theoretical results. The computational efficiency of this model allows computation of a large number of trajectories for each case considered. In this chapter there are two simulated dynamics. The first is the rotational diffusion of large spheres, where calculated relaxation times are compared to theoretical values derived from statistical mechanics. The second is the mean rate of fluctuations within flexible polymers, which are compared to experimental results from long-lived triplet state femtosecond fluorescence measurements.

6.2 Results

6.2.1 Flexible chain dynamics

The mean angular fluctuations in simulations of flexible chains were measured as a function of γ, where $\gamma = k_s/k_BT$ (figure 6.1B). Using $< \cos\alpha >= e^{-Nr_b/l_P}$, these can be converted into persistence lengths (figure 6.1C). Where k_s is smaller than k_BT, the chain behaves as a Gaussian chain and the mean angular fluctuation is $\pi/2$. As k_s is increased, the chain behaves as a worm-like chain and the persistence length increases with increasing spring constant. To further test the validity of the simulated trajectories, ensemble averages were taken from the pseudo time trajectories. The mean end-to-end distance was compared to that expected for a random coil model. The probability for an end-to-end distance r in an ensemble of Gaussian chain polymers, is given by:

$$P(r) = \frac{4\pi r^2}{(\frac{2}{3}\pi < r^2 >)^{3/2}} e^{-\frac{3r^2}{2<r^2>}} \tag{6.7}$$

where $< r^2 >= r_b^2 N$. Comparing simulated ensembles to this model yield the end to end distance distributions shown in figure 6.1A, red. For small values of γ, the simulated conformations closely resemble that expected for the Gaussian chain. For the worm-like chain, the root mean squared end to end displacement is;

$$< r^2 >= 2Nr_b l_P \left(1 - \frac{l_P}{Nr_b}(1 - e^{-Nr_b/l_P})\right) \tag{6.8}$$

The trajectories obtained from the pseudo time Monte Carlo simulations have

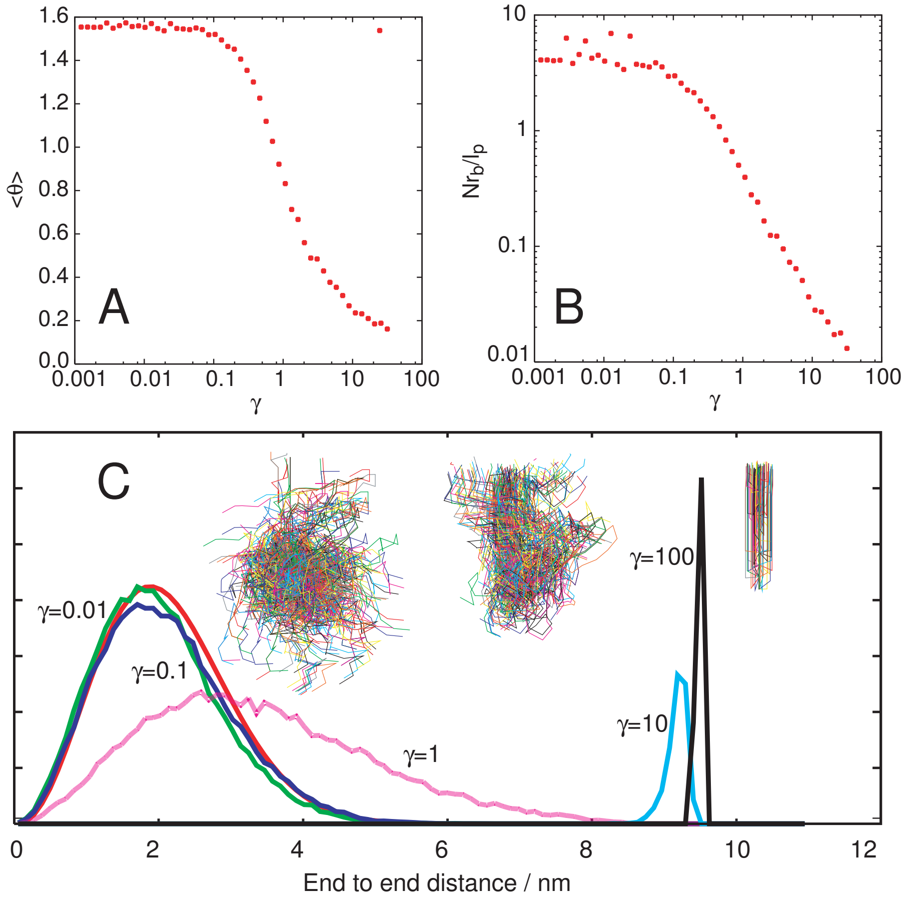


Figure 6.1: Calibration plots, with N=10. **A -** End-to-end distance probability distributions for varying chain flexibility constants for k_s=10^{-2*}(blue),10^{-1}(green),1*(pink),10(light blue) and 10^{2*}(black), where $*$ indicates that 1k structures from ensemble are shown (inset). These distributions are compared to the that of equation 6.7 (red). The simulated distributions where $\gamma < 1$ therefore, precisely agree with the distributions expected for a Gaussian chain. **B -** the mean angle as a function of γ. **C -** The ratio of the segment length to the persistence length as calculated from the mean angle, against γ. For γ¿0.1, the persistence length scales linearly with k_s, as expected for a worm-like chain (equation 6.6). For γ¡0.1, the chain behaves as a Gaussian chain.

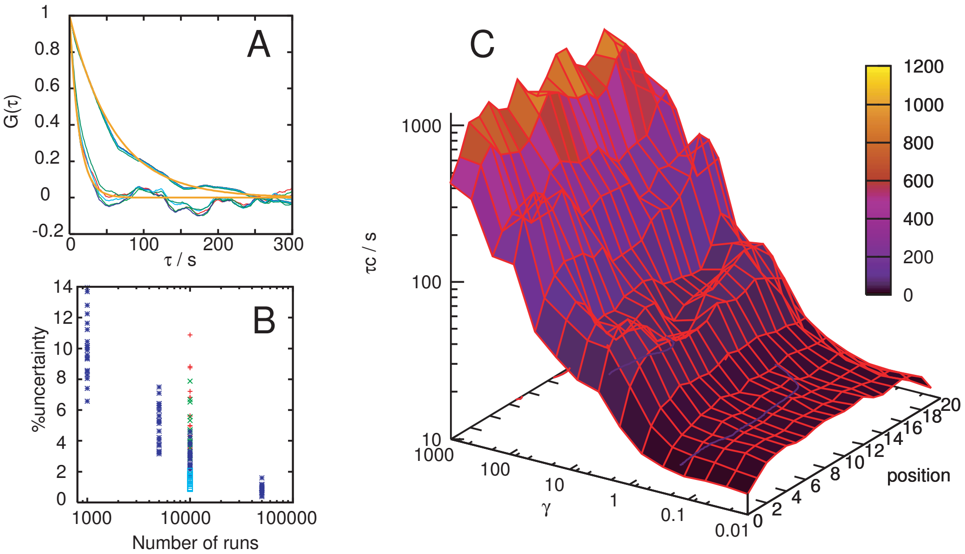


Figure 6.2: **A -** Auto correlation functions for 10k structures with $\gamma = 10^{-2}$ and 10^2. These are highly local functions - red includes all interactions, green includes the nearest 10 neighbours , dark blue 5, light blue 3 and dark green 2. All are essentially indistinguishable. These functions are fitted to a single exponential decay (orange), whose decay rate yields the effective correlation time. **B -** How τ_c uncertainties estimates with replicates and step number, as estimated by the variance between replicas for a 20 member chain (3 replicas are shown in red, 5 are shown in green, 10 in dark blue and 30 in light blue). 10^5 steps are required before replicas converge. **C -** variance of simulated τ_c values, with the chain position and γ.

the expected ensemble averages, and so can be taken to be good representations of the local internal dynamics these types of systems undergo.

Autocorrelation functions and correlation times were calculated for flexible chains as a function of γ. Calculated autocorrelation functions (figure 6.2) for $\gamma = 10^{-2}$ and 10^2 fit well to a function with a single exponential decay (orange). The autocorrelation function for the stiffer chain decays much more slowly than that for the flexible chain, reflecting the reduced internal mobility. Unsurprisingly, the autocorrelation functions of all simulated chains are dominated by local interactions - functions including the 2 adjacent residues (dark green) are essentially indistinguishable from those including the adjacent 3 residues (light blue), the adjacent 5 (dark blue) and the adjacent 10 (green). This is a consequence of the $\frac{1}{r^3}$ averaging, and contributions of fields from residues that are further away than 3 residues are much smaller than the contributions from the nearest neighbours.

The effective correlation time, τ_c, can be estimated from the exponential decay. Values of τ_c require 10^5 replicates before the standard deviation between independent calculations falls below 2%, as shown in figure 6.2B. τ_c is shown as a function of γ, and position within the chain in figure 6.2C. τ_c is observed to increase in two orders of magnitude as increases in γ lead the chain to move from a Gaussian chain ($\gamma = 10^{-2}$)to a stiff semi-flexible chain ($\gamma = 10^3$). Interestingly, τ_c is lower at the ends of the chain that it is in the middle, for all values of γ. This reflects the observation that as the chain ends have fewer nearest neighbours, the fluctuations in the local field is reduced and so the correlation time decreases. This is similar to the finding of Schwalbe et al[104], where R_2 for a given position within a random coil is assumed to have a contribution from adjacent groups that decays exponentially with distance along the chain.

Thus the time dependent Monte-Carlo simulations give time dependent chain trajectories that behave as their ensemble averages dictate they should, providing sufficient steps have been incorporated such that the chain can explore the free energy surface. Confidence can therefore be placed in the calculated correlation times. Internal dynamics are shown to have a large effect on τ_c, varying by two orders of magnitude moving from a Gaussian chain to a rigid semi-flexible rod. The 'effective' mass of a flexible region can therefore be expected to decrease by two orders of magnitude than that of a rigid body; a flexible region in a 1MDa protein would give similar relaxation properties to that expected for a 10 kDa protein. This is exactly the situation observed in the Ribosome, where the L7/L12 stalk region has line-widths comparable to a 10 kDa protein, whereas the bulk of the resonances are too broad to be detected. These simulation results rationalise why one can observe resonances from large complexes, given they have sufficient flexibility.

6.2.2 Sphere dynamics

Rotational diffusion of a rigid body can be described by a random walk of constant step size on the surface of a sphere. The direction of each step is determined here by drawing a random vector on the unit sphere. For rotational diffusion, $< \cos\alpha >= e^{-2D_R\Delta}$ where α is the angle between any two vectors on the walk, Δ is the total time allowed for the walk, and D_R, the rotational diffusion of a sphere of radius r_H in a solvent of viscosity η will be given by $D_R = \frac{k_BT}{8\pi\eta r_H^3}$. The rotational correlation time is related to the diffusion coefficient by $D_r = \frac{1}{6\tau_c}$. Thus, at long times, the mean angle α tends to $\frac{\pi}{2}$ and for short times, the mean angle is 0. This can be used to estimate the mean step size dr in the simulated random walk. For small steps, $dr = \alpha r_H$, and so $\alpha = \arccos e^{-2D_R\Delta}$. dr was set to 10% of the bond length of the flexible chain, r_b. With these parameters, the walk can be simulated for a defined time period. The total period is linearly proportional to the number of steps. Two views of a random walk on a 50nm sphere are shown for 20k (blue), 50k (green) and 100k (red) steps (figure 6.3A). The time between steps is 100μs so the walks correspond to 2,5 and 10s, respectively.

Calculated autocorrelation functions (figure 6.3B) converge when the random

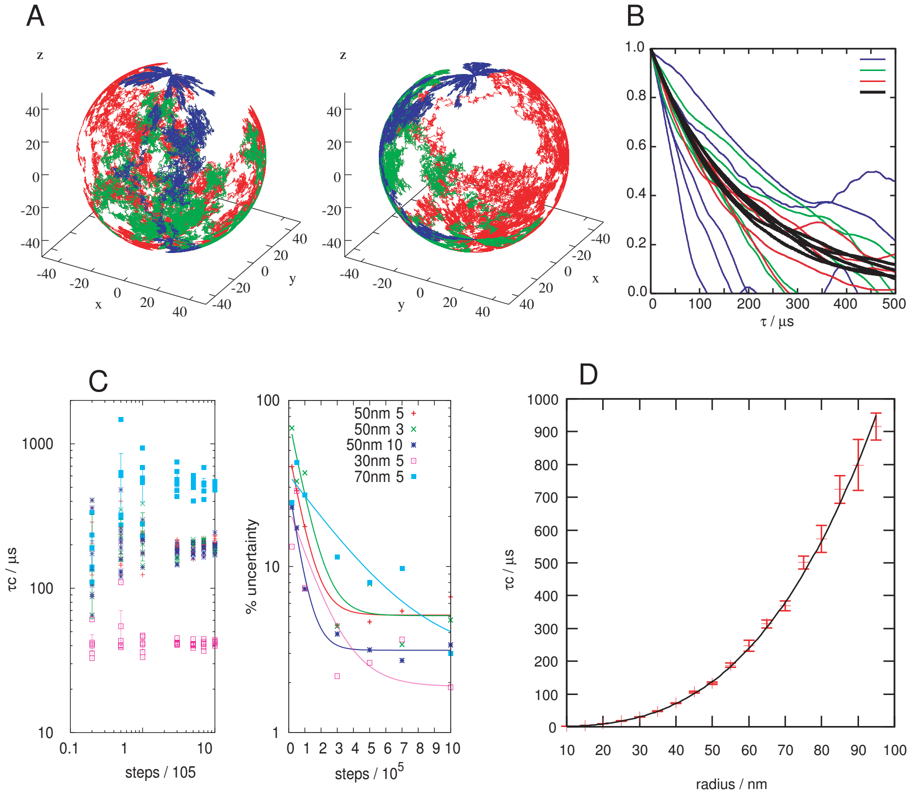


Figure 6.3: **A -** Two views of random walks on the surface of a 50nm sphere with 20k (blue), 50k (green) and 100k (red) steps. The time between steps is 100μs so the walks correspond to 2, 5 and 10s. **B -** 5 replicates of autocorrelation functions together with that for 10^6 steps (100s, black). The correlation plots converge when the number of steps is sufficiently large. **C -** A description of how quickly simulations converge on τ_c. The random walk has to cover the sphere before the uncertainty in τ_c, estimated from the variance in replicas, falls below 5%. As a consequence, more steps are required for larger spheres. 3 repetitions of 300k steps are required for convergence. The total number of simulation steps determines the uncertainty - for 50nm spheres, 10x50k gives a similar uncertainty to 5x100k. **D -** Plot of simulated τ_c with 10^6 steps as a function of sphere radius is shown, scaling with the cube of the radius $\tau_c = kr^3$. The constant k is equal to $\frac{4\pi\eta}{3kT}$, as expected for the Debye rotational correlation time.

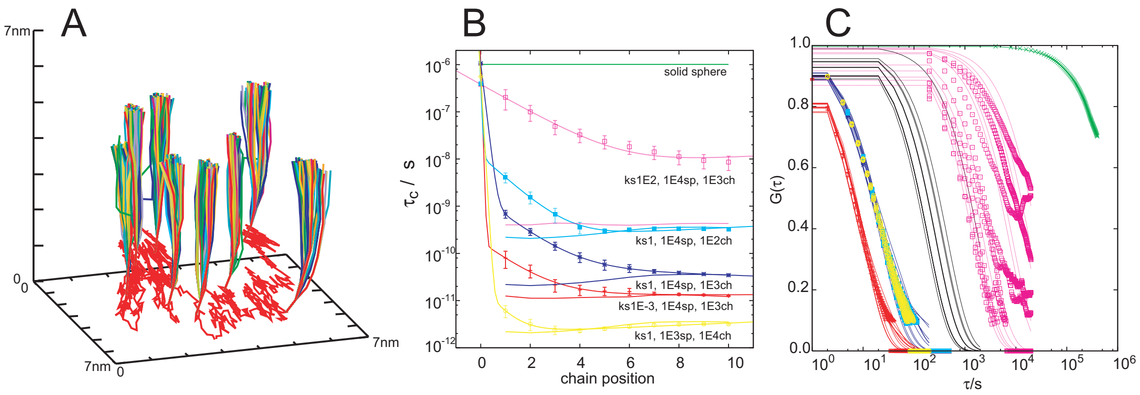


Figure 6.4: A - 1μs simulation of a 10 residue flexible domain on a 10nm sphere, with γ=100. The simulation here has 1k chains steps between each sphere step. Every 10th structure is shown (100 structures total) for every 100 sphere steps. B and C show the correlation functions (for the 10th position along the chain) and characteristic times for a 10 member chain, with 1k chain steps for $\gamma=10^2$ (pink) $\gamma=10^{-3}$ (red) γ=1 (dark blue). Data for γ=1 and 100 steps (light blue) and 10k steps (yellow) are also shown.

walk has sufficient time for the walk to cover the surface of the sphere. With 10^6 steps, independent runs agree to within 2% (figure 6.3C). Correlation times are shown as a function of sphere radius in figure 6.3D where Debye's law is recovered from the simulations. This further demonstrates that the ensembles of simulated Monte Carlo trajectories reflect kinetic events.

6.2.3 Combined dynamics

Combining the flexible chain model with the rigid body, it is possible to estimate how τ_c will vary along a flexible chain attached to a rigid body as a function of the flexibility of the chain. To simulate this, the first point of the flexible chain (section 6.2.1) is modelled to undergo a random walk, as described in section 6.2.2. The bond vector between the first and second residues were fixed in the direction of the vector from the origin to the first residue. A single random walk step then explicitly moves both the first and second residues. The new angle between the second and third residues is set equal to the angle formed by the second residue in its new position, to the position of the 3rd residue in its previous position. This strains the chain, as the angle is increased away from equilibrium. The chain then undergoes a series of MC moves that bring the chain back into equilibrium before the next sphere step. The time between the sphere steps is known exactly (section 6.2.2). The number of chain steps, m, is allowed before the next sphere rotation is a parameter that reflects the chain fluctuation rate, $T_f = \Delta/m$. τ_c can therefore be estimated at each position along the chain, as a function of γ and m.

In order to simulate the L7/L12 domain on the surface of the ribosome, the radius of the sphere was set to 10nm, with a rotational correlation time of $1\mu s$. With a step time of $1ns$, the mean angle is 0.05^o, and a step distance of 0.1Å, 10% of the bond length, r_b. 10k sphere steps therefore corresponds to a total simulation time of $10ms$. Structures from a short simulation of chain length 10 are shown in figure 6.4A, with 1k chain steps between each sphere steps. 1 in every 10 of these is shown for 1 in every 100 sphere steps, for 1k total sphere steps (10 families of 100 structures are shown, with a total simulation time of 1 μs), showing the mechanical coupling between the flexible chain, and the larger, rotating sphere. Corresponding correlation functions for varying γ and m are shown in figure 6.4C, and fit well to functions with a single exponential function. The corresponding correlation times shown in figure 6.4B.

The first residue in the chain, fixed to the surface of the sphere (green) has a correlation time identical to that of the sphere. The correlation time then decreases up to six orders of magnitude, depending on γ and m. Decreasing γ from a rigid chain ($\gamma = 10^2$, pink) to a semi-flexible chain ($\gamma = 1$, dark blue) to a Gaussian chain ($\gamma = 10^{-3}$, red), decreases τ_c. Increasing m for a chain of $\gamma = 1$, from $m = 10^2$ (light blue) through $m = 10^3$ (dark blue) to $m = 10^4$ (yellow) decreases the correlation time. Moving along the chain, τ_c becomes tends to that calculated for a free, flexible chain after 5 residues where $\gamma < 1$. For the rigid rod with $\gamma = 10^2$, the free chain model does not converge on its free chain limit. However, the correlation time of a point on the end of the chain is two orders of magnitude lower than that of a point on the surface of the sphere.

The effective correlation time of a group within a flexible chain attached to a larger, rigid body, decreases significantly along the chain. A decrease in two orders of magnitude with the correlation time will lead to a resonance from a 1MDa complex appearing as a resonance from a 10KDa protein. Results here show that within 5 residues along a flexible chain, the relaxation properties are comparable to a free chain. This firmly demonstrates the possibility of observing solution-state NMR resonances from flexible groups attached to larger bodies.

6.3 Further calculations

Four further calculations will compliment these results.

- Running some simulations of longer sequences with 'realistic' choices of γ and m. The persistence length of a protein random coil is approximately 3 bond lengths, corresponding to $\gamma \approx 3$ (fig 6.1A), and so the average change in bond length per step is 45^o (fig 6.1B). The time required for this displacement has been experimentally estimated to be on the order of ps.

- Simulating a chain attached to a rod. It is likely that the rapid axial rotation of the rod will substantially decrease the effective correlation time expected for end-over-end tumbling. The expected result is that flexible regions would be expected to be readily observable on the surface of amyloid fibrils, supporting the observations of chapter 4.

- Final calculation is for a modular protein. The correlation time of a sphere of mass M, of known radius will relax at a well described rate (section 6.2.2). If the sphere is then split into two smaller spheres of mass $M/2$ connected by a flexible linker, then the effective correlation time of the two spheres can be calculated. By further breaking down the mass into smaller units, it should be possible to calculate the reduction in correlation times obtained for modular proteins.

- Finally, running the early chain simulations with a molecular dynamics script to ensure the methods produce similar local dynamic trajectories.

6.4 Conclusion

These simulations demonstrate that the correlation time of flexible regions attached to larger bodies can be significantly lower than that of the larger body. The dipolar interaction that leads to line-broadening is a highly local effect, dominated by the fields generated by only the nearest neighbours. For a chain with a stiffness comparable to an unfolded polypeptide, the field experienced by a given residue five residues will relax like a free chain. Physical factors that act to reduce the flexibility such as interactions between side chains will act to increase the correlation time. However, this study predicts that short, flexible regions on the surface of larger structures are predicted to be readily observable by solution-state NMR. NMR diffusion experiments on the flexible regions, using the theory described in chapter 3, will allow the study of the underlying larger structure as a function of external variables such as temperature, time or pH.

These results combined with the results of chapter 3 potentially open up a large range of large biological structures such as starch (data not shown) to solution-state NMR studies. The results from this chapter show that τ_c can be expected to be sufficiently low for short flexible regions on large bodies to give sharp resonances, and the results from chapter 3 explain that due to rotational diffusion, one might expect relatively small diffusion coefficients at conventional diffusion delays for these resonances, an observation that may have curtailed their study in the past.

Chapter 7

Thesis conclusions

By harnessing the remarkable stability of amyloid fibrils and their propensity to self assemble, there exists great potential to utilise amyloid fibrils as a scaffold for the display of catalytic or other functions. As such, they could provide the basis for novel materials for technological applications. As fibrils rapidly self-assemble from a starting polypeptide into well defined, stable, nanoscale architectures, they are of particular use in an application where functional groups require alignment with sub-nm precision. It is demonstrated in chapter 2 that fibrils displaying a high density of electron transport proteins can be produced, providing that care is taken with the design of the starting monomeric precursor. These fibrils can reversibly exchange electrons with external molecules with an appropriate electrode potential.

It was interesting to observe the effects that the displayed functional group has upon the structure of the underlying fibril scaffold. Fibre XRD measurements determined that while the addition of functional groups does not significantly alter the underlying core protofilament structure, the protofilament packing is greatly perturbed. The underlying helicoid fibril structure was observed twist into a helical ribbon conformation when the large cytochrome domains were added to the fibril. In addition, through measuring the concentration of surface exposed groups to total groups, we demonstrate that 50% of the groups are not exposed to solvent. This figure reflects how the protofilaments are packed and has been observed in many other amyloid fibril systems.

It would appear that greater understanding of the structure of these fibrils is required if we are to control the molecular interactions between adjacent functional groups with the necessary precision to make subsequent generation fibrils for useful electronic applications.

Remarkably, well resolved solution-state NMR resonances were observed to originate from the amyloid fibrils. As the chemical shifts of these resonances fall at random coil chemical shifts and rapid local motion is required to narrow the resonances to a point where they can be detected by solution-state NMR, these resonances have been assigned to flexible regions on the surface of the fibril, regions that are not folded into the β-sheet core of the fibril.

This surprising observation led to the simulations in chapter 6, where the dynamics of flexible regions on the surface of very large structures are modelled. It is shown that short flexible regions are expected to have significantly sharper linewidths than those in the core of a large structure and that relatively little motion is required in order to narrow the resonances. This observation suggests that it should be possible to study NMR resonances from flexible regions of large biological structures usually considered to be unsuitable for study by solution-state NMR.

The NMR diffusion properties of the resonances from large structures were considered in chapter 3. Though the local motion of the flexible regions is rapid, their diffusion properties reflect that of the underlying structure to which they are attached. The displacements of individual nuclear spins on the surface of a large structure due to translational diffusion of large structures will decrease with size. In contrast, on a millisecond timescale the displacements that result from rotational diffusion will increase with size.

The conclusion of this analysis is that apparent diffusion coefficients measured using PFGSE NMR experiments will vary as a function of the diffusion delay used. This variance can then be related to the size of the large structure under study. In particular, the observed diffusion coefficients of such large structures at commonly used diffusion delays (30-100ms) are predicted, remarkably, to yield values comparable to monomeric protein. NMR resonances from flexible regions of large structures may have not received further study as a consequence. A particularly interesting case in point is that of rice starch granules where PFGSE experiments with longer diffusion times (Δ = 500ms -1s) are required to determine whether or not resonances originate from a large species in solution. The methods developed in this thesis are therefore readily applicable to the study of any large molecular assemblies.

In order determine the generality of these effects, a range of amyloid fibrils were studied. Remarkably, a correlation between the presence of flexible regions and a reduction in the quantity of soluble material found in a fibril preparation was observed. This is particularly apparent in the series SH3, SH3$_2$, SH3$_2$Cyt, SH3$_3$Cyt where ca. 10% of the material is found as soluble monomer in preparations of SH3 and SH3$_2$ fibrils, but $>$ 0.1% material was found as soluble monomer in preparations of SH3$_2$Cyt and SH3$_3$Cyt fibrils. NMR resonances were only observed from fibril preparations of the latter two.

The origin of this effect is likely to be due to non-specific interactions between the non-core regions on the surface of the fibrils. These interactions could in principle manifest as a change in either the free energy of elongation or in the free energy of the elongation transition state. The comparison of fibrils from SH3 and human lysozyme shows that these effects are expected to contribute ca. 0.23 kJ mol^{-1} per residue.

The addition of six non-core residues to $TTR_{(105-115)}$ fibrils has been shown to destabilise the elongation transition state by 9 kJ mol^{-1}, but leave the free energy of elongation unchanged. However, adding 4 addition residues to form TTR+10 increases the free energy of elongation by 11 kJ mol^{-1}, destabilising the fibrils.

The role of non-core regions in amyloid formation has not previously been considered in studies of amyloid fibrils. In this thesis it has been shown that non-core regions can have significant effects on the assembly of fibrils. The presence of non-core regions decreases the tendency of a monomeric component of a fibril to dissociate, and perturb the protofilament packing leading to large changes in the observed morphology of the fibrils. It has been possible to draw these conclusions through the development of novel NMR methodology developed during the course of thesis.

7.1 Publications

A biophysical characterisation of $(SH3)_2Cyt$ fibrils as described in chapter 2, has been published in J. Am. Chem. Soc. [51]. The diffusion theory described in chapter 3 has been published in J. Chem. Phys. [174], and an experimental application of the theory where the lengths of $SH3_2Cyt$ fibrils are determined has been accepted for publication by Angewandte Chemie [94]. A manuscript including the NMR data from αB-crystallin fibrils described in section 4.2.2 has been published in J. Mol. Biol [163].

A manuscript is in preparation comparing protofilament packing between $SH3$, $SH3_2$, $SH3_2Cyt$ and $SH3_3Cyt$ as described in chapter 2. In addition, a manuscript describing the relationship between flexible regions and the dissociation rates of amyloid fibrils as discussed in chapters 4 and 5 is in preparation. Finally, a manuscript describing the dynamics behind the motional narrowing that gives rise to solution state NMR resonances on the surface of large structures, discussed in chapter 6, is in preparation.

7.2 NMR analysis software

The NMR processing software has been developed in order to assist quantitative analysis of NMR data, and is particularly adept at the batch analysis of diffusion data (section 8.2.1). The software reads 1, 2 and 3D NMR datasets and rapidly performs batch integrations, and subsequent curve fitting as required by the user. The analysis of large datasets particularly in chapters 4 and 5 would not have been possible within the period of PhD without the software. The C++ source code and the version of the program compiled for Linux, are available on request.

Chapter 8

Methods and materials

8.1 Protein sequence and expression

The GST fusion protein was expressed in *E.coli* strain HMS174 in Luria Bretani (LB) medium [80, 51]. After cell lysis, the soluble GST fusion was passed down a Glutathione-sepharose column, washed and cleaved on-column with thrombin using an AKTA FPLC. The fusion protein released was passed down a mono-Q column with a KCl gradient of 0-200 mM KCl (to separate holo protein due to endogenous heme) on a Waters HPLC, achieving separation of apo and holo constructs typically in the ratio of 5:1. A typical elutant trace is shown in figure 8.1. The pooled fractions of each of these species were passed down a gel filtration column in ammonium bicarbonate. Fractions were then lyophilised giving final yields of 3-7mg per litre of culture. A native state mass spectrum of a holo sample is shown in figure 8.1 showing holo protein and a small population of apo protein.^{15}N Labeled protein was prepared in the same way substituting LB for M9 taking yields down to 2-5mg per litre M9, though under these conditions, the ratio of apo to holo is increase to 50:1.

Abbreviation	ϵ_{280}/cm^{-1}mol^{-1}	Residues	M_w / g mol^{-1}	PI
Cyt	2560	107	11,837.2	5.38
$(SH3)$	15,930	86	9,638.6	4.66
$(SH3)_2$	31,860	182	19,889.7	4.66
$(SH3)_2Cyt$	33,000	294	32,024.34	4.85
$(SH3)_3Cyt$	46,510	388		
$(SH3)_2B1$	38,830	244	26,466.9	4.63

Table 8.1: Constructs produced for this investigation with some useful physical parameters.

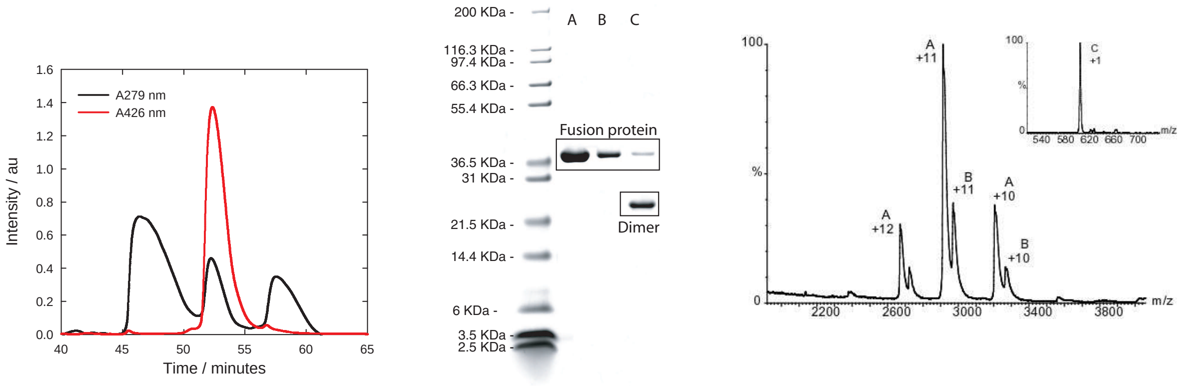


Figure 8.1: Separation of apo and holo $(SH3)_2Cyt$ from residual $(SH3)_2$ followed by ion exchange. Three fractions are separated by a gradient of 0-200 mM KCl. apo fusion protein is the first off the column followed by holo fusion protein (large absorbance at 420nm) and an linker cleaved pseudo SH3 dimer species. SDS/PAGE and electrospray mass spectrometry of apo fusion protein from band A are shown

8.1.1 Full sequences of designed constructs

$(SH3)$

GSMSAEGYQYRALYDYKKEREEDIDLHLGDILTVNKGSL
VALGFSDGQEAKPEEIGWLNGYNETTGERGDFPGTYVE
YIGRKKISP

Cyt

ADLEDNMETLNDNLKVIEKADNAAQVKDALTKMRAA
ALDAQKATPPKLEDKSPDSPEMKDFRHGFDILVGQIDD
ALKLANEGKVKEAQAAAEQLKTTRNAYHQKYR

$(SH3)_2$

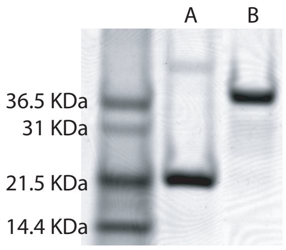
$SH3 - GGGGSGGGGS - Cyt$

$(SH3)_2Cyt$

$SH3 - GGGGSGGGGS - SH3 - GSGGGG - Cyt$

$(SH3)_3Cyt$

$SH3 - GGGGSGGGGS - SH3 - GSGGGG - SH3 - GSGGGG - Cyt$



8.1.2 Fibril formation and purification

The rate enhanced 'pH jump' protocol was used for fibril formation [80]. Lyophilised protein was resuspended in 3mM Tris-HCl and taken down to pH 2 with 1M HCl. Sodium acetate (5.6M) buffered at pH 3.6 is added and the mixture was kept at room temperature for 15 minutes. The mixture was then taken back down to pH 2 and kept at room temperature for 4-6 days. The mixture was purified by ultracentrifugation at 90,000rpm in a TLA 120 rotor (sedimenting material 10S and greater) at 20°C to remove remaining smaller aggregated states. The pelleted material was resuspended in 3mM HCl and its morphology confirmed by TEM. Total protein concentration quantified using extinction coefficients calculated initially using protparam[91] shown in table 8.1, that were verified to be accurate using amino acid analysis. Yields of pelleted material vary between 70-90%, with balance of the protein present as monomer or small aggregates.

8.2 Biophysical analysis techniques

Systems as complex as biological systems require multiple techniques to begin to understand the complexity. No one technique can give full information, but the pictures built up using a combination of these techniques can begin to tweak information out of biological systems. In order to optimise the analysis of spectral data, several programs were written in c++. The programs and the biophysical techniques are described below.

8.2.1 Analysis Programs

Four programs have been written for the purposes of efficient and accurate spectral analysis with uncertainty estimates. They read in source data from text files and perform various mathematical operations and fittings using the routines found in the GNU Scientific Library (GSL)[135], an excellent resource. Output is in the form of text files with tailor made Gnuplot scripts. Running the Gnuplot scripts automatically produces postscript outputs for visualising the data. The programs are compiled with GCC run on a Linux platform.

diffprocess.exe

General purpose NMR analysis software. Processed NMR spectra are read in as ascii text files and diffusion analysis, rate analysis and principle components analysis (PCA) analysis can be performed. The function is useful for the batch analysis of data, and can be easily tuned to fit any user specified regions from any dimensionality of data to any arbitary function. Ascii outputs are produced, together with tailored gnuplot scripts for ready visualisation. Sets of spectra of any dimensionality can be analysed in batch, very rapidly. Should one reprocess the data, any analysis is re-run with a single command. A manual for this software is found in section 9.1.

diffmodel.exe

Performs calculations and fittings for the model described in chapter 3. For fitting data, the program reads in a file, dataset.txt produced by diffprocess.exe that holds signal intensity data as a function of gradient strength. Decay constants are calculated, and compared to model parameters to find a minimum $\chi^2 = (Y_{exp} - Y_{model})^2$ corresponding to the best fit. More details are given in section 9.4.

assigner.com

An automatic assignment program. It reads in peak lists, sequence and random coil shifts, and performs a Bayesian assignment process. This is discussed in more detail in section 9.3 and was found to function efficiently when the spectra are highly overlapped, as in the case for $(SH3)_2Cyt$ fibrils.

uvvis.com

A program analysing UV-Vis data. A series of raw spectral data are read in, with the details of the given series required for analysis (concentrations, volumes etc). Baseline subtraction and deconvolution algorithms are run to separate the spectra into basic components. A manual for the program is found in section 9.5.

8.2.2 Spectroscopy

The theory of the types of spectroscopy particularly amenable to biophysical analysis is described elegantly by Cantor et al.[147]. In all cases, where spectra are acquired as a function of an external variable such as time or pH, deconvolution techniques such as principle components analysis have been used to separate the data into its principle components. This is a far more informative method of spectral analysis than inferring changes from measurements a single wavelength.

8.2.3 Nuclear Magnetic Resonance (NMR) spectroscopy

NMR data was acquired on Bruker DRX500 and 750 spectrometers equipped with cryoprobes. 3mm, 5mm and Shigemi tubes were used depending on the sample conditions. For ease of sample transfer 3mm tubes were used for titrations, and Shigemi tubes used for higher dimensional experiments. Sample concentration was limited by the preparative methods and varied between 20-100 μM for fibril samples.

^{1}H 1D spectra using bruker sequence zgpw5 (1D with w5 water suppression) were acquired generally with 128 scans for all samples 2D ^{15}N HSQC experiments were acquired using selective pulses, or with watergate water suppression.

For NMR diffusion work, the pulse field gradient stimulated echo (PFGSE) pulse sequence STEBPFG was used, with the length of each individual gradient pulses, p30 typically 2.7 ms. As bipolar gradients are used, δ=2×p30. The diffusion delay, Δ=d20 was varied as described. For a 'conventional' diffusion experiment, d20 was set to 100ms. For the majority of protein work, the pulse sequence was coded into a two dimensional form with constant δ and Δ, using the xau script DOSY to specify a gradient list between user specified limits. Typically a quadratically spaced gradient list was used, spaced between 5 and 95% of the machine's calibrated maximum gradient strength. The processing and error analysis of noisy protein spectra is a more complex problem than it might appear. Analysis was performed using the program diffprocess.exe, coded over the course of this thesis, described in section 9.2. NMR diffusion experiments were always ran in conjunction with a PFGSE sequence with no water suppression to moniter the diffusion of water molecules during experiments to provide an internal viscosity control.

2D diffusion measurements acquired used an PFG/HMQC pulse sequence [96]. E-burp and re-burp shaped pulses was used on the proton channel for the 90^o and 180^o pulses respectively. Large diffusion coefficients cannot be measured with this technique as the pulse sequence is based on a spin echo and so signal is degraded by T_2. While it is possible to use pulse sequences based on a stimulated echo such as the PFGSE [96], these were not use during this thesis. A more complex expression is required for the diffusion delay using the HMQC/PFG pulse sequence:

$$\Delta = 2 \times d5 + p12 + 1m + p30 + 4u$$

.

NMR titrations were acquired with the 1D sequence described above. Batch processing and analysis of the titration data was performed with the program described in section 9.1. Where the NMR titration was linked to UV-Vis data acquisition, 3mm tubes were used to minimise volume loss in passing between the NMR tube and cuvettes for UV-Vis.

The theory behind NMR spectrsocopy is exceptionally elegant. For an introduction, Malcolm Levitt's book [180] teaches spin dynamics in an exceptionally clear way. Canvanagh [101] takes a more rigorous approach and considers particularly the implications of applying advanced NMR techniques to biological molecules. The principles of nuclear magnetism by Abragam is a classic text that describes the theory more rigorously [173]. Pulse field gradient methods are discussed concisely in [181].

Processing of NMR data to produce fourier transformed spectra was performed using NIH's NMRPipe[182], before the data was passed onto the home written analysis software. Analysis of processed NMR data was performed using home written software (section 9.1).

Electronic absorption (UV-Vis) spectroscopy

UV-Vis data typically in the range 250-700nm was acquired on a Cary 400 using 1cm path length 100μl volume Hellma cuvettes. Once the measured signal is subtracted from baseline species, the extinction coefficient ϵ is related to the absorbance of a species A by Beer's law, $A = \epsilon c l$ where c is the concentration of the absorbing species and l is the optical path length. Extinction coefficients of relevant species are shown in figure 2.4, and summarised in table 2.1.

Large species (50 nm - 1 μm) in solution give rise to scattering effects that will lead to additional apparent absorption. Phenomenologically, these can be fitted to 3rd order polynomials and removed form the spectrum to separate scattering effects from true absorption. In order for this to be effective, the spectral width must be extended to best characterise the curvature effects due to scattering. Mie scattering, numerical solutions to Maxwell's equations show that globular protein, small molecules and thin rods do not give rise to large scattering effects, whereas larger aggregates will give rise will.

Once scattering effects have been removed the absorption spectra A_i^j as a function of frequency i and a second variable such as time or pH j. Difference spectra, D_i^j or obtained from subtracting a reference spectrum, typically the first spectrum in the series or the spectrum in the absence of titrant from each subsequent spectra. Following the difference spectra as a function of titrant yields the quantitative information required for the various fits required of the data.

$$D_i^j = A_i^j - A_i^{j=1}$$

	P1/nm	P2/nm	—P2—/—P1—
Free ZnPPIX	590	642	0.38
Bound ZnPPIX	596	653	0.50

Figure 8.2: Left - Excitation spectra of free and bound ZnPPIX, measuring emission at 600nm. Middle - emission profile of free and bound ZnPPIX. The profile is independent of B or Q band excitation. The peak frequencies and ratios are diagnostic of free and bound states, shown in the table on the right

Difference normalised spectra N_i^j are useful for comparing the spectral envelope of a set of data in a dataset. As they are normalised for at a single intensity $i = p$, a comparison of relative lineshapes shows the emergence of other spectral species. A wavelength, usually the maximum of a peak or an isosbestic point is taken as the normalising frequency.

$$N_i^j = \frac{D_i^j}{D_{i=p}^j}$$

The program uvvis.com performs these functions automatically and is discussed in section 9.5.

Fluorescence emission spectroscopy

Fluorescence data was acquired on a Cary Eclipse. Hellma cuvettes of 2 mm × 1 cm path with a minimum volume of 200 μl were used with the longest path aimed at the detector. Two peaks are observed in the emission spectrum of ZnPPIX, corresponding to the stokes reflection of the Q band. This emission spectrum was acquired from exciting either the B band or the Q band, resulting in the fluorescence intensity being weighted by the extinction coefficient of the absorbing frequency. Thus for maximum intensity yield, the B band of the ZnPPIX was excited at 430nm. Bound and free states of ZnPPIX can be distinguished as described in figure 8.2.

FarUV circular dichroism spectroscopy

Far-UV circular dichroism spectra were obtained on a Jasco-820. Concentrations are typically constrained to $< 10\mu$M to ensure the solution scattering remained under the 700 mV machine limits. 1 mm path length Hellma cuvettes were used typically with volumes > 200 μl. CD spectra F_i^j give ellipticity θ of a sample as a function of wavelength i in units of milli-degrees. This is converted to molar ellipticity, Θ in a manner analogous to Beer's law, using $\theta = \Theta cl$ where c is the concentration of the sample and l is the path length. This can be scaled further by the number of amino acids to give a *mean residue weight ellipticity*.

The ellipticity of the carbonyl π-π^* transition at 200 nm gives information on local chirality of each individual carbonyl. These overlay to give an overall far-UV spectrum characteristic of the secondary structure of a protein. Following CD spectra as a function of an external variable j shows how the secondary structure of a biomolecule changes with respected to the variable. Taking a single wavelength, normalised transitions t_j between the initial spectrum $j = 1$ ($t_1 = 0$) and a final value $j = k$ ($t_k = 1$) are calculated using the following.

$$t_j = \frac{F_i^j - F_i^1}{F_i^k - F_i^1}$$

Normalised transitions are readily fit to sigmoidal functions to determine the mid-points of such transitions.

8.2.4 Microscopies

Approximately one monolayer is achieved by depositing 25 pico moles (5μl of a 5μM) of protein onto a 1cm^2 substrate. The nature of the surface and the method of preparation will affect structures seen.

Transmission electron microscopy (TEM)

Bright field TEM images were acquired on a Philips CM100 microscope operating at 80kV. Contrast between sample and substrate was achieved with uranyl acetate negative staining. Typically 5μl of 5μM (25 p moles) of fibril were added to UV irradiated formvar and carbon coated 400 mesh nickel grids. Three repeats of 5μl 2% uranyl acetate were added, left to sediment for 1 minute and blotted with filter paper. Finally, grids were washed with H_2O, blotted then dried.

Atomic force microscopy (AFM)

AFM microscopy data were acquired using a Molecular Imaging Pico Plus microscope operating in AC-AM mode (tapping mode) in air. Ultrasharp Micromasch

NCS36 silicon cantilevers were used at resonance frequencies between 150 kHz and 225 kHz. Samples for AFM analysis were prepared by diluting the fibril solutions by a factor of 30200 in dilute HCl at pH 3.0, and 20 μl were subsequently deposited onto a freshly cleaved mica surface. The samples were then left to dry in air for 60 min whilst shielded from dust particles.

Scanning electron microscopy (SEM)

Sample is deposited on conducting substrate (gold / graphite / silicon) and the solvent is left to evaporate. This is followed by a wash step consisting of 2 x volume H_2O being added then blotted.

8.2.5 Mass spectrometry

Data was collected on an LCT and a tandem mass spectrometer (QToF) modified for high pressure and high mass operation equipped with a nanoflow Z-spray source. The pressures and accelerating potentials in the mass spectrometer were adjusted to preserve non-covalent interactions.

8.3 Thesis preparation

Written in WinEdt5 and typeset using LaTeX2e. Gnuplot 4.0 and Sigmaplot 9.0 were used to produce graphical representations of data. Sparky was used for visualising and overlaying 2D NMR data. All figures were optimised and compiled in Adobe Illustrator 10 as encapsulated postscript files.

Chapter 9

Appendix

9.1 NMR processing software manual

In Fourier Transform NMR, a free induction decay (FID) is the observable NMR signal generated by non-equilibrium nuclear spin magnetisation precessing about the magnetic field. This non-equilibrium magnetisation is generally created by applying a pulse of resonant radio-frequency close to the Larmor frequency of the nuclear spins. If the magnetisation vector has a non-zero component in the xy plane, then the precessing magnetisation will induce a corresponding oscillating voltage in a detection coil surrounding the sample. This time-domain signal is typically digitised and then Fourier transformed in order to obtain a frequency spectrum of the NMR signal i.e. the NMR spectrum.

If the NMR spectra need to be compared quantitatively, resonances in the spectrum are integrated and the integrated intensity compared between spectra. The integrated will then be fitted to a function to yield a physical parameter such as diffusion coefficients or hydrogen exchange protection factors. There are many NMR processing programs, many of which are freely available that perform these functions. However, due to insufficient documentation, and inaccessibility of source code, the operations of the programs are largely unclear and documentation is far from unclear. Furthermore, few functions are available for batch processing of data.

This is the role of the software described in this section; large numbers of NMR spectra can be read in, peaks are integrated according to flexible user specifications and the resulting intensity trends are then fit to a user specified function. The program is written in C++, a powerful straightforward programming language. Combined with libraries from the GNU Scientific Library (GSL), advanced mathematical functions can be readily employed to perform fitting and statistical analysis. Text files are then produced with the summary of the analysis and tailored gnuplot scripts are printed that generate plots of the data as postscript files.

This section is written so that others can use the program and to allow users to be able to read the code and to make amendments where appropriate.

9.1.1 Processing

Processing of data was performed using the excellent based NMRPipe[182]. This is can be run either on Linux systems, or under windows using Microsoft's Services For Unix (SFU) unix shell program. Many NMR spectra can be processed with a single command where experiment numbers are referred to using a loop. Using the C-shell, the syntax to batch process data is as follows;

```
set x = i while ($x < j)
echo "$x"              /*print the current experiment number */
SCRIPT                 /*with $x inserted as variable        */
set x = ‘expr $x +1‘   /*increment the experiment number     */
end
```

In the script, directory and file information is obtained from the loop counter variable. Adding a similar line with an scp command allows batch downloading of the files from the institution's file server. A typical processing script for a 2D series of 1D diffusion data would read as follows;

```
nmrpipe -in ./$x/test.fid                       \
| nmrPipe  -fn POLY -time                       \
| nmrPipe -fn EM -lb 3.0 -c 0.5                 \
| nmrPipe  -fn SP -off 0.5 -pow 2 -end 0.98     \
| nmrPipe  -fn ZF -auto                         \
| nmrPipe  -fn FT -auto -verb                   \
| nmrPipe -fn PS -p0 0 -p1 0 -di                \
| nmrPipe  -fn EXT -x1 10ppm -xn -2ppm -sw      \
| nmrPipe -fn BASE -nw 2 -nl 0% 4% 90% 100%     \
| nmrPipe -out ./$x.ft1 -ov -verb
pipe2txt.tcl ./$x.ft1 -index ppm > ./$x.out
```

The final pipe2txt.tcl command outputs the real part of the processed spectrum in a text format. The outputted text file is read in by the NMR analysis program. Thus the batch processing of data can be performed with a single command.

9.1.2 Analysis

Once ASCII files of the processed data are generated, the analysis becomes a numerical problem. The files have to be read into memory, operations are performed then output files are generated, with gnuplot scripts for visualising the data. For 1D titrations were the aim is to rapidly overlay 1D spectra, running the program then compiling the gnuplot script `1dplot.gp` will give such a plot. The raw data is read into arrays within C, which are then sorted, ready for further analysis. Two types of analysis have been coded, a titration analysis or a diffusion analysis for 1 or 2 dimensional data. The primary output of the program is sorted spectra and a gnuplot script to plot them.

For two dimensional data a peaklist is supplied showing the centre of each peak. There are several methods to obtain an integral. Either the peak can be fitted to a two dimensional Gaussian to obtain a volume, or an integral can be obtained by summing over the spectrum, with either an ellipsoidal or rectangular area, with a size defined in terms of the full width half maximum of the peak. An additional output from 2D spectra are two projections of the spectra and a 2D mask that can overlay with 2D spectra that explicitly show the integrated areas. For one dimensional data individual peaks can be fitted to a Gaussian function, or the intensity between a user specified range can be summed.

The intensity is then analysed as a function of a variable associated with each spectrum such as gradient strength for when analysing diffusion data, or time. Error analysis of such fitting are employed using several methods. Through fitting each individual chemical shift separately, a range of fitting parameters are obtained which can be assembled into a histogram. Fitting this histogram to a Gaussian yields a mean and standard deviation of the parameter. A more detailed discussion of processing and analysing diffusion data is found in section 9.2.

The following section will discuss how files should be arranged for the program and the files that are produced as output. In addition, there is a section discussing the various functions within the source code that would only be of interest to those who want to develop the program further.

9.1.3 File inputs

Pragmatically the difficulty faced by the user of this program is to ensure the files required by the program are entered in the correct format. Limited error checking has been added, the proof of a correct calculation coming mainly in the form of the output files. If a mistake has been made, there will be obviously spurious output tables.

The form of the input files are summarised in figure 9.1.4. There are three input modes. The first is for a string of 1D spectra, use `input.com`, the second is for a 2D array of 1D spectra (a DOSY spectrum for example) use `data.com`and the third is for 3D spectra (HMQC/PFG data for example), use `input.com`. The

*.temp	temporary files
*.inp	input files with no manipulation
*.sol	analysis files
*.int	integration files
*.gp	dumped gnuplot script files

Table 9.1: File classes produced from diffprocess.com

columns should be tab delimited. Having the correct number of columns for a given input mode is crucial, with >> representing a tab, as column delimiter. 2D spectral datasets should be submitted with a peaklist file, `peaklist.com`. Sparky provides excellent peak picking algorithms. One has yet to be included in this program. The final file that has to be specified is `control.com`, of column format shown at the top of figure 9.1.4, depending on the data input mode. Getting the files setup correctly is the only barrier to successful operation of the program.

Three other files can be optionally read in. `rg.com`, used for T1/T2 analysis, and is the set of receiver gain values of the successive experiments if they are different. Where solvent or small molecules were used for internal calibration and a separate set of spectra exist for their comparison, `soldata.com` should be submitted, with form identical to that of `data.com`. For diffusion analysis of both 1D and 2D datasets, roi.com can be supplied to deconvolve different spectra from the data based on their diffusion coefficient. With a separate line for each species, The highest ppm, followed by the lowest ppm, then the highest diffusion coefficient followed by the lowest diffusion coefficient should be submitted, defining a rectangle in diffusion/chemical shift space.

For 1D datasets and DOSY datasets, the file integration.com specifies the centre of an integration summation in ppm and absolute width, also in ppm. Different integrations should be submitted on different lines of the file. A more detailed discussion of processing and analysing diffusion data is found in section 9.2.

9.1.4 File Outputs

All outputs are printed to the bottom of the specified files. Old files must either be deleted or moved away from the running directory. Different output files appear depending on the menu choices and are summarised in table 9.2. The required input files and output types for a given choice of running mode are summarised in figure 9.1.4 and the gnuplot output scripts are summarised in table 9.1.4. The five file types are;

INP_raw.inp	Raw data as submitted from pipe2txt.tcl as a check for reading in data
INP_rawi.inp	Where i is an integer. When 2D spectra data is inputted, these files are generated as each 2D dataset in gnuplot format
INP_xyproj.inp	For 2D spectra, these are the x and y projections obtained by summing across the slices at a given value of x and y.
INP_sort.inp	For 2D diffusion datasets, this will be a collection of the raw diffusion data. Each block of text is for a given value of Δ, with the gradient points increasing with column number. If this file is in the correct format, the data has been read in correctly.
INT_full.int	Will sequentially output integration data from the experiment.
OUT_hist.sol	Generates histograms according to the user specified region of interest of diffusion data
OUT_hist4.sol	Generates 2D histograms of chemical shift versus diffusion coefficient
OUT_histanal.sol	Gives the calculated diffusion coefficients and uncertainties obtained by several methods
OUT_gaus.sol	Fitted Gaussians to the ROI histograms
OUT_gplot.sol	The semi log diffusion plots for the region of interest, with an uncertainty measure from averaging across the region of interest
OUT_solhist.sol	Same as hist.sol but for the solvent files
OUT_solhist4.sol	Same as hist4.sol but for the solvent files
OUT_solhistanal.sol	Same as gplot.sol but for the solvent files
OUT_2dpeak.sol	For 2D spectral data, this will regenerate the peaklist, together with columns for the types of analysis required to be performed.
OUT_svd.sol	The results from an SVD analysis of the data
OUT_ppmfit.sol	The results from fitting mon, bi and tri exponentials to each chemical shift
OUT_specfit.sol	The results from an roi.com spectral deconvolution
monplot.gp	DOSY analysis plot script for mono-exponential plotting
biplot.gp	DOSY analysis plot script for bi-exponential plotting
decon.gp	Deconvolution plot
1dplot.gp	1d titration plot script
2dplot.gp	2d plot with region of interest integrals marked

Table 9.2: Summary of file outputs from diffprocess.com

DIFFPROCESS.COM

modes

control.com
1D diffusion - #datafiles >> Noisemax >> Noisemin >> Nwmax >> Nwmin >> Gmax >> Gmin >> Grad
2D diffusion - #gradient points >> Noisemax >> Noisemin >> Nwmax >> Nwmin >> Gmax >> Gmin >> Grad
2D spectral diffusion - #datafiles >> xNoisemax >> xNoisemin >> yNoisemax >> yNoisemin >> Gmax >> Gmin >> Grad
1D titration - #datafiles >> Noisemax >> Noisemin >> Nwmax >> Nwmin >> Gmax >> Gmin >> Grad
2D titration - #datafiles >> Noisemax >> Noisemin >> Nwmax >> Nwmin >> Gmax >> Gmin >> Grad

1.diffusion

Linear or quartic?
Use for 2D diffusion where linear or quartic gradient spacings are used

1D for 1D diffusion data

data.com (2D)
file >> d20 >> d9

input.com (1D)
file >> g% >> d20 >> d9

REGION OF INTEREST?
enter ppm range for a gplot and user specified diffusion histograms

INPUT SOLVENT FILES?
enter ppm range for gplot and diffusion histograms for small molecule internal control

RUN SVD?
perform a singular valued decomposition in the data

RUN TRI EXPONENTIALS?
Run tri-exponential fits. Computationally expensive

RUN SIMPLEX?
Perform a global analysis on the dataset

RUN T1/T2 ANALYSIS?
where several diffusion datasets of several values of Delta have been used, T1 and T2 can be estimated

OUTPUTS
INP_raw.inp
INP_sort.inp
INT_full.sol
OUT_ppmfit.sol
OUT_hist.sol
OUT_hist4.sol
OUT_histanal.sol
OUT_gaus.sol
OUT_gplot.sol
OUT_svd.sol
OUT_solhist.sol
OUT_solhist4.sol
OUT_solgplot.sol
OUT_simplex.sol
OUT_spec.sol
OUT_specfit.sol

monplot.gp
biplot.gp
decon.gp

2D data?
For analysis of 2D spectral data

input.com
file >.> g% >> d20 >> d9 >> +y >> +x

peaklist.com
1Hmax >> 1Hmin >> 15Nmax >> 15Nmin

INTEGRATION TYPE?
Using the supplied peak information, rectangular or ellipsoidal integration can be specified

INTEGRATION FACTOR?
The peak list will have an estimate for peak width. The first factor multiplies the x by this number and the second multiples y.

OUTPUTS
INP_rawi.inp
INP_xyproj.inp
OUT_2dpeak.sol

2dplot.gp

2.titration

1D data?
For analysis of 1D spectral data

input.com
file >> # >> 1 >> 1

No options
Relatively self explanitory

OUTPITS
INP_raw.inp
INP_sort.inp
INT_full.sol

1dplot.gp

2D data?
For analysis of 2D spectral data

input.com
file >.> +y >> +x

peaklist.com
a b c d e f g h i j k l m n o p

INTEGRATION TYPE?
Using the supplied peak information, rectangular or ellipsoidal integration can be specified

INTEGRATION FACTOR?
The peak list will have an estimate for peak width. The first factor multiplies the x by this number and the second multiples y.

OUTPUTS
INP_rawi.inp
INP_xyproj.inp
OUT_2dpeak.sol

2dplot.gp

1dplot.gp	1d	Outputs all 1D spectra in a single plot for rapid inspection of the data
monplot.gp	monplot	A master plot for diffusion data. 1D spectra is shown, with the 2D histogram of probability density of diffusion coefficient against ppm. A histogram of the user specified region of interest is shown on the right, with gradient plots shown in the top right. Diffusion coefficient against diffusion delay is shown bottom right.
biplot.gp	biplot	The same as monplot.gp but with bi exponential fits.
grplot.gp	grplot	The gradient plots of the user specified region of interesting. Plots of $\ln Si/S0/N$ against $\%G^2$ are given with mono and bi-exponential fits if requested.
decon.gp	decon	Where regions of interest has been specified, this file will output the deconvolution data
ppmmon.gp	ppmmon	Will output a 1D spectra and raw fits of diffusion coefficient against ppm. Erroneous fits to low signal to noise data will dominate.
ppmbi.gp	ppmbi	Same as above but with bi exponential fits
ppmtri.gp	ppmtri	Same as above but with tri exponential fits

Table 9.3: Summary of the gnuplot files, the postscripts they produce after they are compiled and a brief description of the data the plots contain.

9.1.5 Details

The following is a description of the main functions defined in the program, for those who might wish to further develop the source code, or adapt it for another purpose. Original source code is available on request.

Reading in files - 'input' Files are read in as streams and so having the correct structure for the files is crucial. These are saved internally as arrays and used for subsequent analysis.

Determination of diffusion data structure For 2D diffusion data for example, the pipe2txt.tcl output is in a three column format. The first task for processing is to separate the data into individual 1D spectra each associated with appropriate Δ, δ and G, coming from `control.com` and `data.com`. Similar functions are performed for 1D input spectra and 2D input spectra. These functions are performed by the function 'collator'.

Determination of noise The file searches through a regions specified in `control.com` to determine the maximum absolute magnitude of noise signals and saves it for future reference. Any signal less than 1.5 × the noise will not be counted in any subsequent analysis.

1D integration - function 'sorter' All intensity between the integration limits specified in `integration.com` will be summed to give an integral associated with each spectrum for future analysis. These integrals are printed in `INT_full.sol`. Where multiple integrals are defined, the function 'joiner' sums them for the diffusion analysis.

Mono/bi/tri exponential fittings - function 'ppmfit' The GSL Levenberg-Marquardt algorithm for non-linear regression is used to fit functions of the form $f(x) = \sum_i A_i e^{-k_i \% G^2}$ to intensity to each recorded chemical shift, where i is either one, two or three (functions 'fitmon', 'fitbi' and 'fittri'). For diffusion analysis, $k = \alpha^2\beta$. Each chemical shift is only included in the count if there are more than 5 signals above the noise for a mono or bi exponential fitting, or more than 7 for a tri exponential fit.

SVD A principle component analysis based on an SVD algorithm is performed on the data using a GSL SVD algorithm.

Simplex The program fits $f(x) = \sum_i A_i e^{-kD_i \% G^2}$ to the dataset. It contrasts from the above fitting in that the user supplies the number of species to expect, and it holds the diffusion coefficient constant for that species. The output will be the set of A_i as a function of chemical shift, which can be interpreted as the deconvolved spectrum of that species.

Diffusion analysis and error analysis Described in more detail in section 9.2.1. Analysing diffusion data to gain an accurate estimate of experimental uncertainty is a tricky problem and several strategies have been included. Using the decay rates from fitting each chemical shift, histograms (function 'hist') are constructed that count the frequency of each decay rate. These are then fitted to Gaussian functions (function 'fithist', output `OUT_gaus.sol`)to yield a mean value x_0, and an uncertainty σ from the Gaussian function $y = Aexp\left(-\frac{(x-x_0)^2}{2\sigma^2}\right)$. The user can specify a chemical shift range for this process if required. The program will generate a two dimensional histograms (function 'hisogram2') by looking between regions of chemical shift and observed diffusion coefficients to overview the spectrum for rapid assimilation of data. The number of frequency buckets is by default set to 100, and the diffusion coefficients are bucketed in 100 logarithmically spaced groups and printed in `OUT_hist4.sol`. A summary of the diffusion measures and estimates can be found in `OUT_histanal.sol`. A traditional semi log plot is constructed from data in this region (`OUT_gplot.sol`) from the function 'globalsort'. Here, the uncertainty in each integrated intensity is estimated from the integrated range.

Inclusion of small molecules Where small molecule data have been recorded in a separate dataset, an independent routine resets the program and re-runs the program looking at the solvent datafiles, directly after processing the experimental files. For the analysis of water, as the signal is so large, an additional threshold constraint is applied; signal is only incorporated into the diffusion analysis if the signal has lost less than 4 orders of magnitude of its original value.

Spectral deconvolution - function 'seeker' Decribed in more detail in section 9.2.1. Where a biexponential fit has been run, this function searches for bi-exponential intensities whose diffusion coefficient and chemical shift fall into the range defined in `roi.com`. Individual spectra in a mixture can then be reconstructed and printed in `OUT_spec.sol`. Diffusion hisograms are created, and Gaussian functions are fitted to them (function 'fithist2'). Both are printed in `OUT_specfit.sol`.

Relaxation estimate - function 'relax' The variance of the signal intensity in the absence of an applied gradient, S_0 can be used to estimate the relaxation parameters T_1 and T_2. Where S_0 is known for varying values of δ and Δ, $S_0 = \exp\left(2\frac{\tau_2}{T_2} + \frac{\tau_1}{T_1}\right)$, where τ_1 and τ_2 are the periods where magnetisation is stored on the z axis and the xy axis respectively.

2D data analysis - 1D projections - function 'proj' The 2D dataset is first sorted for re-occurring values of x and y to determine the structure of the dataset, giving tdx and tdy for the 2d dataset. The x projection is obtained by holding each value of x constant and summing over all y, and the y projection is obtained by holding each value of y constant and summing over all x.

2d data analysis - function 'dd' and 'dd2' Using a peaklist obtained from a peak picking program such as sparky, a 2d peak analysis can be performed. Function 'dd' sums all signal intensity within either a rectangular or ellipsoidal range whose width is determined as a fraction of the full width half maximum of the peak. These intensities are used in diffusion and rates analysis. Integrals are printed in `OUT_2dpeak.sol`.

Outputs - function 'dump', 'soldump' and 'gnuplot' The calculation arrays are outputted at the end of the program after calling these two functions. The function 'gnuplot' writes the relevant gnuplot scripts, described in table 9.1.4 with an appropriate number of 'plots' as determined by the number of files that were analysed.

9.2 Analysis of NMR diffusion data

The translational self-diffusion coefficient D_T in PFGSE experiments is obtained using the Stekjal-Tanner equation [95] (see chapter 3 for derivation), which gives the PFGSE NMR signal intensity S_i in the presence of pulsed field gradients of strength G of duration δ separated by a delay Δ relative to the signal with zero gradient S_0:

$$S_i = S_0 e^{-\gamma^2 G^2 \delta^2 (\Delta - \delta/3) D_T} \tag{9.1}$$

The signal intensity in the absence of gradients S_0 varies with is reduced by both T_1 and T_2 relaxation and so it is common to acquire a series of PFGSE experiments by varying G at constant δ and Δ, ensuring S_0 is a constant. The experimental parameter varied is generally the ratio $x = G/G_{max}$ where G_{max} is the maximum gradient the machine can generated. With $\alpha = \gamma\delta G_{max}$ and $\beta = \Delta - \delta/3$ the gradient of a plot of $\ln \frac{S_i}{S_0}/\alpha^2\beta$ against x^2 will be $-D_T$.

The observed translational self-diffusion coefficient is a complex function. The pure self diffusion coefficient was shown by Einstein to be $D_0 = kT/f$ where k is the Boltzman constant, T is temperature and f is a geometry dependent friction factor, generally proportional to viscosity η. Temperature in all experiments in this these was set to be 300 K in all experiments in this thesis unless explicitly stated otherwise.

The observed translational diffusion coefficient is a complex function of the

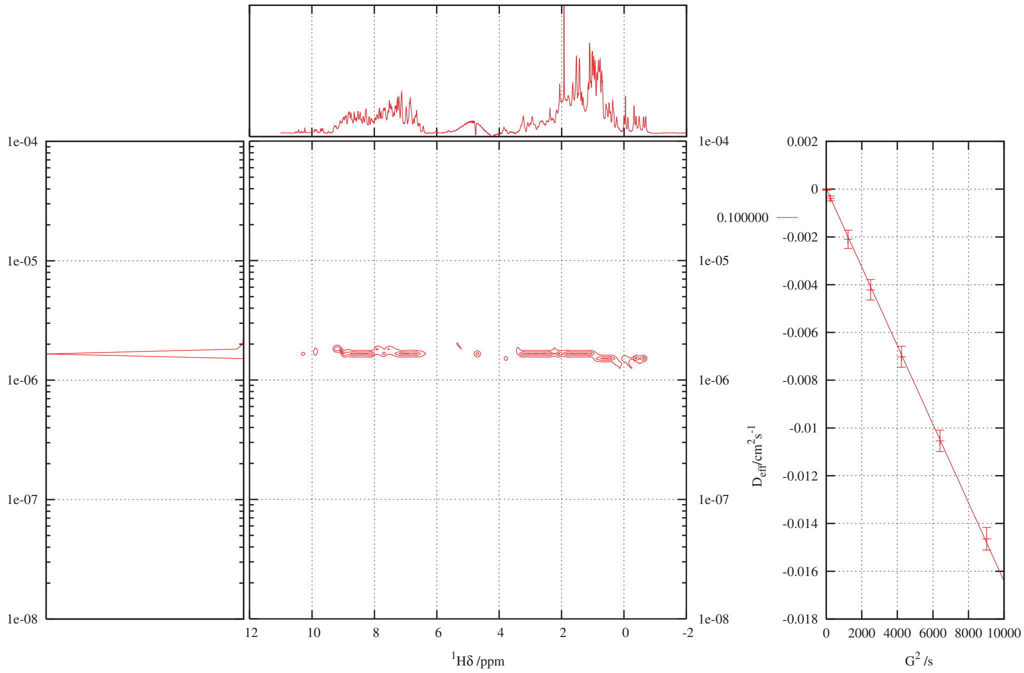


Figure 9.1: A sample processing from Human Lysozyme. On running the program and compiling the gnuplot script monplot.gp this figure was produced as monplot. These plots has four sections; 1 - a sample 1D spectrum; 2 - a conventional plot of $\ln(S_i/S_0)/\alpha^2\beta$ against x^2 on the right; 3 - a 2D histogram of diffusion coefficient vs chemical shift; 4 - a 1D projection of this histogram. Where the user inputs a region of interesting, the 1D histogram and the semi-log plot will be based on this specified region of interest.

solution conditions. The viscosity of the solution can be estimated by measuring the self-diffusion of a small molecule whose diffusion properties are well studied. Proteins are studied here in an aqueous environment, and so the diffusion properties of of water which serves as a convenient internal standard [183]. At 300 K, water is expected to diffuse at 2.38 × 10^{-5} cm^{2}s^{-1}. At the studied total protein concentrations of 10-100 μM at 300K this was not found to deviate significantly, even in fibril samples. This will be due to the huge excess of water molecules relative to solute molecules.

The diffusion of a solute is concentration c dependent, reflecting the loss of mobility as the concentration of larger molecules is increased. The diffusion coefficient is expected to vary as $D = D_0(1+c\phi)$ [184, 123]. The protein concentrations studied in this these were in the range 10-100 μM unless otherwise stated. Within this range, the observed diffusion coefficient were found not to vary significantly with concentration.

9.2.1 Analysing diffusion data with the program

The primary motivation behind coding the NMR analysis software was to perform analysis of diffusion data. Empirically fitting the Stejskal-Tanner relation

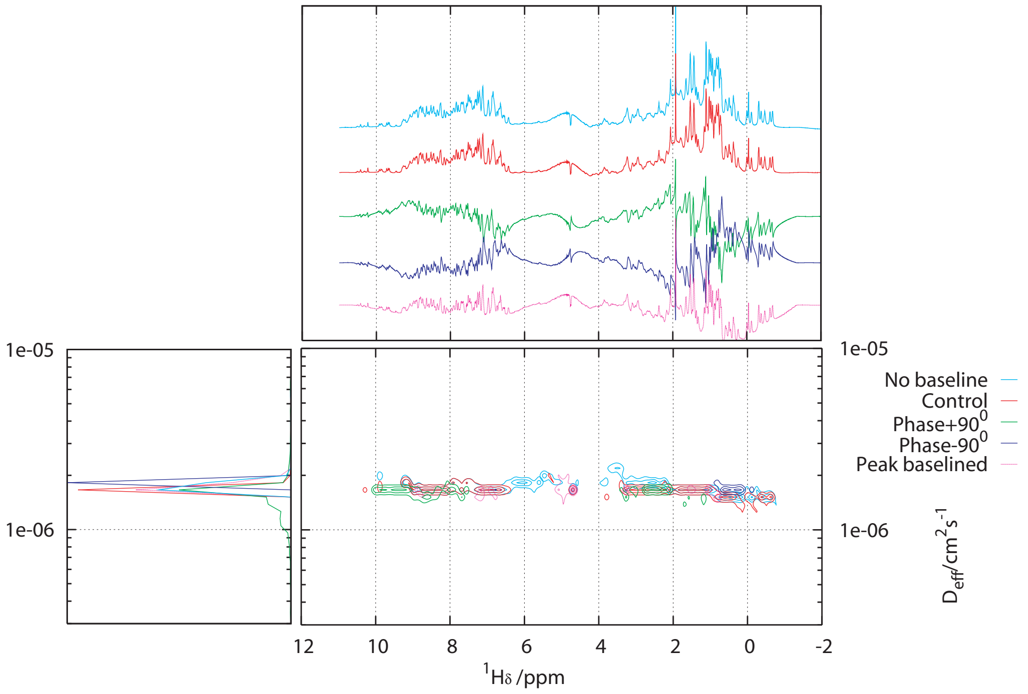


Figure 9.2: The effects of different processing schemes on a Human Lysozyme diffusion dataset. Diffusion data is remarkably resilient to poor processing. These data are of a single Human Lysozyme dataset processed with different schemes, whose corresponding 1D spectrum is given in the top panel. The diffusion histogram for each processing scheme with its projection. Sending the phasing off by as much as $\pm 90^o$ does not effect the results of the diffusion analysis. Also, remarkably, using a baselining function that sets the tops of various randomly selected peaks to zero despite visibly distorting the spectra, does not effect the diffusion analysis. The only effect of poor processing is that fewer regions where signal intensity is greater than the noise are found. Thus the uncertainty in the diffusion measure tends to increase. The diffusion analysis is not perturbed significantly by major offsets in phase ($\pm 90^0$) and arbitrarily misplaced baselines.

(equation 9.1) to taking a single recorded chemical shift results in a near perfect fit. Estimating the uncertainty in the measurement using fitting residuals leads to exceedingly low uncertainties. Summing a set of intensities within a range was found to have a similar property leads to decay curves with very low uncertainty in measured parameters estimated from fitting residuals. Crucially, the uncertainties obtained through averaging signal in this method led to estimates significantly less than the estimates obtained through averaging repeats of the experiment on independent samples.

Interestingly, the diffusion coefficient is found to fluctuate in a random fashion across a spectrum. The measured diffusion coefficients can then be grouped together in histograms whose form is Gaussian. Fitting a Gaussian function to this distribution yields an excellent estimate of the mean value of D_T and its uncertainty. The uncertainties measured in this way were found to agree with the uncertainties obtained through multiple repeats of independently prepared samples. This is the analysis the program performs. To further minimise noise in the spectrum, the number of data points above the noise is first counted. This number must be greater than 5 before a D_T fit is added to the analysis. Any single data point whose intensity is lower than the maximum noise is removed from the analysis.

The primary output of the program is a two dimensional histogram showing the 1D NMR spectrum with histograms constructed from counting observed diffusion coefficients within a chemical shift range and a one dimensional histogram that surveys counts over the entire dataset, or a user specified range. A sample output is shown in figure 9.1, though compiling the `monplot.gp` gnuplot script to produce `monplot`. Individual diffusion coefficients against chemical shift are also obtained through compiling `ppmmon.gp`. Interestingly, diffusion datasets are surprisingly resilient to poor processing (figure 9.2). Similarly, each individual chemical shift can be fitted to a bi-exponential function;

$$S_i = A_0 e^{-\alpha^2 \beta x^2 D_1} + B_0 e^{-\alpha^2 \beta x^2 D_2}$$

This situation is expected for when two species are contributing to a given chemical shift. Analysing the two diffusion coefficients yields an analysis similar to that described above yields the output generated by the gnuplot file `biplot.gp`. Where bands around a single diffusion coefficient are observed it is likely that the assumption of a second species is reasonable. Such a situation is shown in figure 9.3 where a sample of calmodulin was shown to be contaminated by a small quantity of a peptide.

Where a mixture exists, it is possible to separate the spectra to reconstruct the spectra of the two species in the absence of gradients, given by A_0 and B_0.

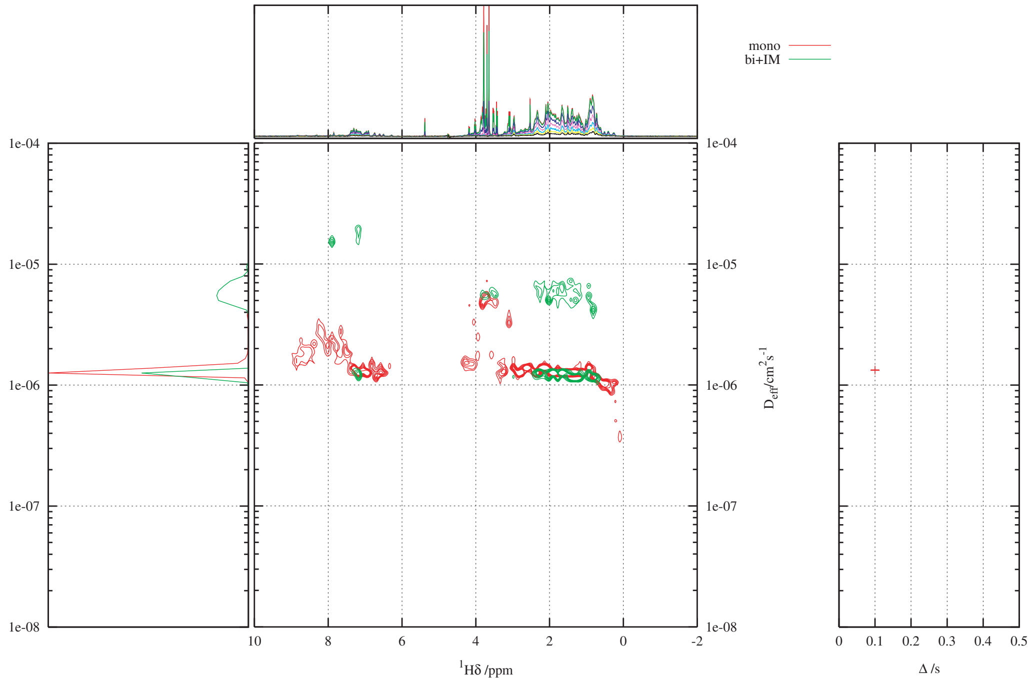


Figure 9.3: Diffusion data acquired from Calmodulin sample before (red), and after imidazole addition. Before, the data fits well to a single exponential. After imidazole addition the data require a bi-exponential fit (green). In addition to the imidazole, a third species is observed diffusing significantly faster than calmodulin. The spectrum of this species can be deconvolved (figure 9.4).

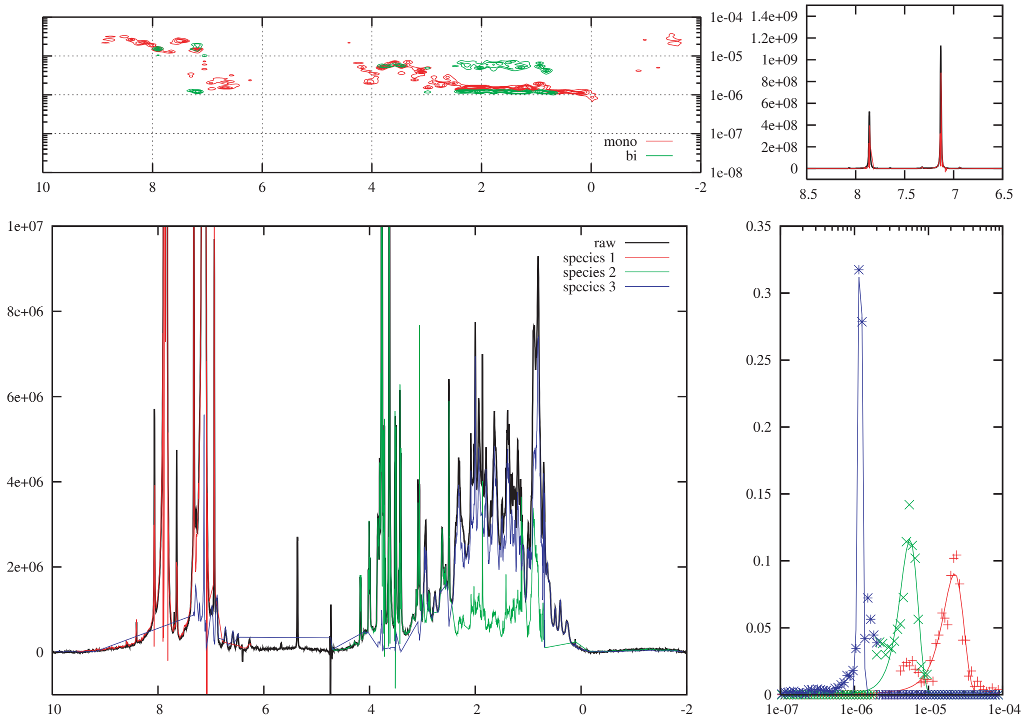


Figure 9.4: The spectra of the smaller species was deconvolved from that of calmodulin, yielding the output gnuplot script decon.gp which when compiled produces this figure decon. The dispersion of the spectrum combined with its rapid diffusion suggest a polypeptide is cleaved from calmodulin on imidazole addition. This was confirmed by subsequent mass-spec analysis.

Protein	D$_{lit}$ / $cm^2 s^{-1}$	D$_{exp}$ / $cm^2 s^{-1}$	D$_{lit}$/D$_{exp}$
Human lysozyme	1.19×10^{-6}	$(1.16 \pm 0.02) \times 10^{-6}$	9.97
BSA	6.8×10^{-7}	$(6.80 \pm 0.06) \times 10^{-7}$	9.99
Cytochrome C	1.00×10^{-6}	$(1.05 \pm 0.07) \times 10^{-6}$	1.05
Myoglobin	9.57×10^{-7}	$(9.8 \pm 0.3) \times 10^{-7}$	1.02

Table 9.4: Experimental values for the diffusion coefficient compared to literature values for selected systems. All test systems are in aqueous buffer.

The currently implementation of this is semi-automatic. The user specified a file `roi.com` which contains information specifying the program to obtain A_0 and B_0 values within a chemical shift range and a diffusion coefficient range. These extract coefficients are taken together to give a spectrum of the two observed components. Such a deconvolution is described in figure 9.4 where the spectrum of the contaminating peptide is separated from calmodulin, together with a histogram of the diffusion coefficients D_1 associated with the individual intensities used in the analysis. This type of information is lost when chemical shifts are integrated by summing intensity information within a single spectrum.

Thus through iteratively processing NMR data and adjusting the regions of interest one can rapidly obtain an understanding of the diffusion behaviour of a sample.

In this thesis several samples were understood to have varying diffusion coefficients with the diffusion delay Δ. The program can read in up to 60 datasets at one time and superimpose the outputs. Using this immensely useful function allows one to rapidly ascertain whether a sample is shown anomalous or conventional diffusive behaviour. Without this analysis tool, the analysis described in this thesis would not have been possible.

9.2.2 Calibration of the magnet

Diffusion data was recorded from a range of systems and compared to known values to calibrate absolute diffusion measurements from PFGSE measurements as shown in table 9.4 and figure 9.5. A summary of the diffusion data from the full dataset is shown in figure 9.6. The trend is that the larger the system, the smaller the diffusion coefficient. However, unfolded proteins (SH3 and Cytb562) are shown to diffuse slightly slower than would be expected as expected [164]. Hen egg white lysozyme (HEWL) is also found to diffuse slower than would be expected. Thus the experimental arrangement is in good working order and can reproducibly reproduce trends found in the literature. Datasets were acquired on both the 500 TCI and 500 ATM spectrometers in the Chemistry Department and the measured diffusion coefficients were identical within experimental un-

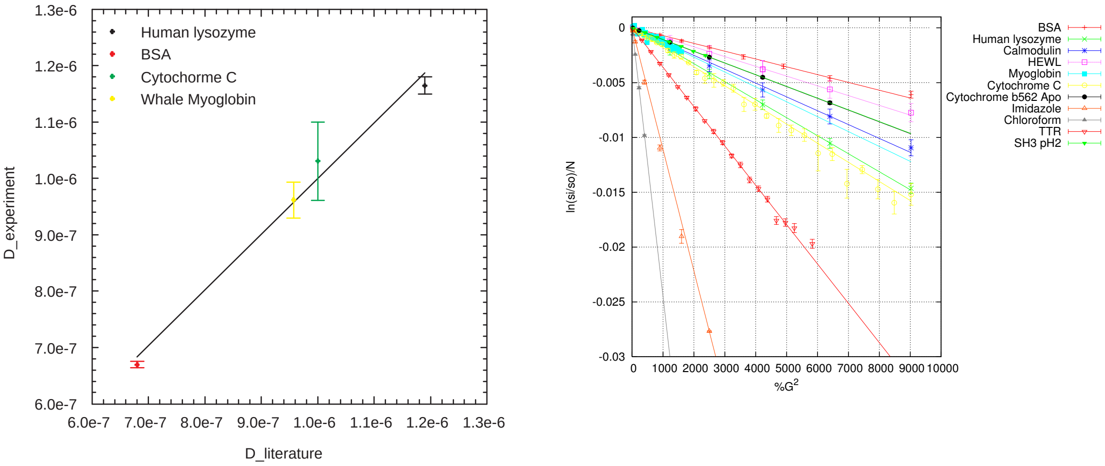


Figure 9.5: Left - A calibration plot of test systems. Experimentally determined diffusion coefficients compared with literature values. Diffusion coefficients determined on a range of test systems. Right - a more complete data from a range of systems. For all, Δ=0.1s and δ=5.4ms. To compare datasets, the y axis has $ln(si/so)/(\gamma^2 G_{max}^2 \delta^2 (\Delta - \delta/3))$ and the x axis has user supplied % of maximum gradient strength. The temperature was set to 300K.

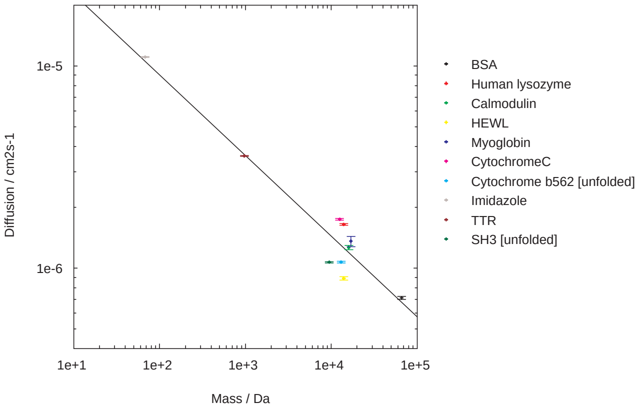


Figure 9.6: A plot of the measured diffusion coefficients against mass.

certainty (data not shown).

HNCACO	$H_{i-1}N_{i-1}C^{CO}_{i-1}$, $H_iN_iC^{CO}_i$	2
HNCO	$H_{i-1}N_{i-1}C^{CO}_{i-1}$	1
HNCACB	$+veH_{i-1}N_{i-1}C^{\alpha}_{i-1}$, $-veH_{i-1}N_{i-1}C^{\beta}_{i-1}$, $+veH_iN_iC^{\alpha}_i$, $-veH_iN_iC^{\beta}_i$	4
CBCACONH	$+veH_iN_iC^{\beta}_i$, $-veH_iN_iC^{\beta}_i$	2

Table 9.5: A summary of the type of information revealed from selected 3D spectra, the chemical shifts resonances will appear with relative phase, and the number of peaks expected per amino acid residue

9.3 Auto assignment of residues

The assignment of backbone resonances in an NMR spectrum is a logic puzzle and as such, can largely be automated. 3-dimensional spectra assignment provide connectivity information of a given residue through selecting for specific coupling evolutions with carefully crafted pulse sequences. Examples are shown in figure 9.5

Once processed, the difficulty is generating a peaklist. Once a peaklist has been generated, the problem fits into an automatable form. In the case of the $SH3_2Cyt$ fibrils, we are attempting to assign an unfolded protein with 304 amino acids. Due to overlap this is a difficult assignment. Nearest neighbours are not obvious. To deal with this difficulty, a Bayesian approach can be taken. The first assumption is that we can estimate the carbonyl chemical shifts from random coil shifts, narrowing the possibilities for a given assignment resonance.

Once there, using the relation between i and i-1 spectra, we can begin to make assignments. Of the possible i+1 and i-1 resonances, the possible sequences can be compared to the known sequence and either accepted or eliminated. Where there are still several options, a further step along the chain can be performed, and the combinations again compared to the sequence. This can be performed up and down the chain, until the most probable sequence is found.

The scale of the problem makes this impossible to attempt by hand. Several autoassignment programs are available, but none are tailored for this purpose. With this assignment program it is trivial to add other constraints. In the assignment of $SH3_2Cyt$, additional information from overlap between $SH3_2Cyt$, $SH3_2$ spectrum and apo $Cytb_{562}$ spectra narrow down the possibilities for some peaks. This information can be used to further narrow down the sequence possibilities. Other autoassignment programs cannot be constrained in this fashion.

9.4 Diffusion model calculations

The model in chapter 3 has been coded into c++ using mathematical functions available in the Gnu Scientific Library (GSL). Fitting algorithms have recently been implemented such that experimental data can be fit to the model. At the time of writing, these functions were yet to prove to be functional.

9.5 Spectroscopic analysis program

A spectral analysis program was written to perform functions similar to those performed by the NMR analysis software. This was used primarily to analyse series of UV-Vis spectroscopic data. Two files need to be submitted to the program. The raw data be entered where x is the wavelength y_i is the intensity either as $x_1 y_1, x_2 y_2, x_n y_n$ or as x, y_1, y_2, y_n in file `input.com`. The second and third columns vary on whether a sequential or static titration is to be analysed. In the former case, the absolute volume is in the second column and the third column is the concentration of titrant. In the latter case, the second column shows the volume added. The file `control.com` has the following information.

The program pulls all spectra to zero at the highest wavelength point. Discontinuities that occur where the lamp changes are also fixed. The Raw date, the difference spectra, scattering is removed with the three specified wavelengths in the control file, and the difference normalised spectra are calculated, and printed as text files. Gnuplot files for their viewing are also created.

The protein concentration is calculated from the 280 extinction coefficient supplied, and absorbance at a user specified frequency is quantified. A singular value decomposition algorithm derives the principle spectral components. Finally, a table of hyperbolic and job outputs are generated with gnuplot scripts for their visualisation. The data shown in section 2.20 were analysed using this program.

Bibliography

[1] L Pauling and M Delbruck. The nature of the intermolecular forces operative in biology. *Science*, 92:77–79, 1940. 1

[2] M. Karplus and J.A. McCammon. Dynamics of proteins: Elements and function. *Annual Review of Biochemistry*, 52(1):263–300, 1983. 1

[3] L.U. Linderstrom-Lang. *Proteins and Enzymes*, volume 6 of *Lane Medical Lectures*. Stanford University Press, Stanford, 1952. 1

[4] C. B. Anfinsen. Principles that govern the folding of protein chains. *Science*, 181(96):223–30, 1973. 1

[5] D. Leckband and J. Israelachvili. Intermolecular forces in biology. *Q Rev Biophys*, 34(2):105–267, 2001. 1

[6] A.J. Stone. *The theory of intermolecular forces*. The international series of monographs on chemistry. Oxford University Press, 2000. 1

[7] F London. The general theory of molecular forces. *Trans. Faraday Soc.*, 33(8b):26, 1937. 1

[8] A.J. Dingley and S. Grzesiek. Direct observation of hydrogen bonds in nucleic acid base pairs by internucleotide $^2\mathrm{j}_{NN}$ couplings. *J. Am. Chem. Soc.*, 120(33):8293–8297, 1998. 1

[9] M. Karplus and J. A. McCammon. Molecular dynamics simulations of biomolecules. *Nat Struct Biol*, 9(9):646–52, 2002. 1

[10] M. Vendruscolo. Determination of conformationally heterogeneous states of proteins. *Current Opinion in Structural Biology Folding and binding / Protein-nucleic interactions*, 17(1):15–20, 2007. 1

[11] A. E. Cass, G. Davis, G. D. Francis, H. A. Hill, W. J. Aston, I. J. Higgins, E. V. Plotkin, L. D. Scott, and A. P. Turner. Ferrocene-mediated enzyme electrode

for amperometric determination of glucose. *Anal Chem*, 56(4):667–71, 1984. 1.1

[12] R. Feynman. There's plenty of room at the bottom. *Caltech's Engineering and Science*, December 29th 1959. 1.2

[13] M. Jormakka, S. Tornroth, B. Byrne, and S. Iwata. Molecular basis of proton motive force generation: Structure of formate dehydrogenase-n. *Science*, 295(5561):1863–1868, 2002. 1.3, 1.1

[14] M. Czjzek, L. ElAntak, V. Zamboni, X. Morelli, A. Dolla, F. Guerlesquin, and M. Bruschi. The crystal structure of the hexadeca-heme cytochrome hmc and a structural model of its complex with cytochrome c(3). *Structure*, 10(12):1677–86, 2002. 1.3, 1.1

[15] O. Einsle, A. Messerschmidt, P. Stach, G. P. Bourenkov, H. D. Bartunik, R. Huber, and P. M. Kroneck. Structure of cytochrome c nitrite reductase. *Nature*, 400(6743):476–80, 1999. 1.3, 1.1

[16] M. L. Rodrigues, T. F. Oliveira, I. A. Pereira, and M. Archer. X-ray structure of the membrane-bound cytochrome c quinol dehydrogenase nrfh reveals novel haem coordination. *Embo J*, 25(24):5951–60, 2006. 1.3, 1.1, 1.3

[17] D. Leys and N. S. Scrutton. Electrical circuitry in biology: emerging principles from protein structure. *Curr Opin Struct Biol*, 14(6):642–7, 2004. 1.3

[18] H. B. Gray and J. R. Winkler. Long-range electron transfer. *Proc Natl Acad Sci USA*, 102(10):3534–9, 2005. 1.3

[19] L. Michaelis. The Enzymes, Chemistry and Mechanism of Action. Academic, New York, 1951. 1.3

[20] C. G. Mowat and S. K. Chapman. Multi-heme cytochromes–new structures, new chemistry. *Dalton Trans*, (21):3381–9, 2005. 1.3

[21] B. Loll, J. Kern, W. Saenger, A. Zouni, and J. Biesiadka. Towards complete cofactor arrangement in the 3.0[thinsp]a resolution structure of photosystem ii. 438(7070):1040–1044, 2005. 1.3

[22] R. A. Marcus. On the theory of electron-transfer reactions. vi. unified treatment for homogeneous and electrode reactions. *J. Chem. Phys.*, 43(2):679–701, 1965. 2

[23] C.C. Page, C.C. Moser, X. Chen, and L.P. Dutton. Natural engineering principles of electron tunnelling in biological oxidation-reduction. *Biochemistry*, 402(6757):47–52, 1999. 2, 2.4.12

[24] S. Y. Tan and M. B. Pepys. Amyloidosis. *Histopathology*, 25(5):403–14, 1994. 1.4.1, 4.1.1

[25] M. R. Krebs, L. A. Morozova-Roche, K. Daniel, C. V. Robinson, and C. M. Dobson. Observation of sequence specificity in the seeding of protein amyloid fibrils. *Protein Sci*, 13(7):1933–8, 2004. 1.4.1, 1.4.3, 2.6.3

[26] C. F. Wright, S. A. Teichmann, J. Clarke, and C. M. Dobson. The importance of sequence diversity in the aggregation and evolution of proteins. *Nature*, 438(7069):878–81, 2005. 1.4.1, 1.4.3, 2.6.3

[27] F. Chiti, P. Webster, N. Taddei, A. Clark, M. Stefani, G. Ramponi, and C. M. Dobson. Designing conditions for *In-vitro* formation of amyloid protofilaments and fibrils. *Proc Natl Acad Sci USA*, 96(7):3590–4, 1999. 1.4.1, 1.4.3

[28] J. I. Guijarro, M. Sunde, J. A. Jones, I. D. Campbell, and C. M. Dobson. Amyloid fibril formation by an SH3 domain. *Proc Natl Acad Sci USA*, 95(8):4224–8, 1998. 1.4.1, 2.1.1

[29] C. M. Dobson. The structural basis of protein folding and its links with human disease. *Philos Trans R Soc Lond B Biol Sci*, 356(1406):133–45, 2001. 1.4.1, 2.1.1

[30] J. Zurdo, J. I. Guijarro, and C. M. Dobson. Preparation and characterization of purified amyloid fibrils. *J Am Chem Soc*, 123(33):8141–2, 2001. 1.4.2, 4.1.1

[31] P. S. Vassar and C. F. Culling. Fluorescent stains, with special reference to amyloid and connective tissues. *Arch Pathol*, 68:487–98, 1959. 1.4.2, 1.4.3

[32] Cohen, Calkins, and Levene. *Am. J. Path.*, 35:971, 1959. 1.4.2, 1.4.3

[33] M. R. Krebs, E. H. Bromley, and A. M. Donald. The binding of thioflavin-t to amyloid fibrils: localisation and implications. *J Struct Biol*, 149(1):30–7, 2005. 1.4.2, 1.4.3

[34] R. Khurana, C. Coleman, C. Ionescu-Zanetti, S. A. Carter, V. Krishna, R. K. Grover, R. Roy, and S. Singh. Mechanism of thioflavin t binding to amyloid fibrils. *J Struct Biol*, 151(3):229–38, 2005. 1.4.2, 1.4.3

[35] J. L. Jimenez, J. I. Guijarro, E. Orlova, J. Zurdo, C. M. Dobson, M. Sunde, and H. R. Saibil. Cryo-electron microscopy structure of an SH3 amyloid fibril and model of the molecular packing. *Embo J*, 18(4):815–21, 1999. 1.2, 1.3, 1.4.2, 2.6, 2.4.4, 2.29, 2.6.3, 2.6.4

[36] S. B. Padrick and A. D. Miranker. Islet amyloid: phase partitioning and secondary nucleation are central to the mechanism of fibrillogenesis. *Biochemistry*, 41(14):4694–703, 2002. 1.2

[37] J. L. Jimenez, E. J. Nettleton, M. Bouchard, C. V. Robinson, C. M. Dobson, and H. R. Saibil. The protofilament structure of insulin amyloid fibrils. *Proc Natl Acad Sci USA*, 99(14):9196–201, 2002. 1.4, 1.4.2, 2.6.3, 4.1.1

[38] R. Nelson, M. R. Sawaya, M. Balbirnie, A. O. Madsen, C. Riekel, R. Grothe, and D. Eisenberg. Structure of the cross-beta spine of amyloid-like fibrils. *Nature*, 435(7043):773–8, 2005. 1.5, 1.4.2

[39] C. P. Jaroniec, C. E. MacPhee, V. S. Bajaj, M. T. McMahon, C. M. Dobson, and R. G. Griffin. High-resolution molecular structure of a peptide in an amyloid fibril determined by magic angle spinning NMR spectroscopy. *Proc Natl Acad Sci USA*, 101(3):711–6, 2004. 1.6, 1.4.2, 4.1.1, 5.4.1

[40] C. Ritter, M. L. Maddelein, A. B. Siemer, T. Luhrs, M. Ernst, B. H. Meier, S. J. Saupe, and R. Riek. Correlation of structural elements and infectivity of the HET-s prion. *Nature*, 435(7043):844–8, 2005. 1.7, 1.4.2, 2.6.3

[41] A. T. Petkova, G. Buntkowsky, F. Dyda, R. D. Leapman, W. M. Yau, and R. Tycko. Solid state NMR reveals a ph-dependent antiparallel beta-sheet registry in fibrils formed by a beta-amyloid peptide. *J Mol Biol*, 335(1):247–60, 2004. 1.4.2

[42] N. Ferguson, J. Becker, H. Tidow, S. Tremmel, T.D. Sharpe, G. Krause, J. Flinders, M. Petrovich, J. Berriman, H. Oschkinat, and A. R. Fersht. General structural motifs of amyloid protofilaments. *PNAS*, 103(44):16248–16253, 2006. 1.4.2, 2.6.3

[43] O. S. Makin, E. Atkins, P. Sikorski, J. Johansson, and L. C. Serpell. Molecular basis for amyloid fibril formation and stability. *Proc Natl Acad Sci USA*, 102(2):315–20, 2005. 1.4.2

[44] M.R. Sawaya, S. Sambashivan, R. Nelson, M.I. Ivanova, S.A. Sievers, M.I. Apostol, M.J. Thompson, M. Balbirnie, J.J.W. Wiltzius, H.T. McFarlane, A.O. Madsen, C. Riekel, and D. Eisenberg. Atomic structures of amyloid cross-β spines reveal varied steric zippers. 447(7143):453–457, 2007. 1.4.2

[45] I. Kheterpal, A. Williams, C. Murphy, B. Bledsoe, and R. Wetzel. Structural features of the abeta amyloid fibril elucidated by limited proteolysis. *Biochemistry*, 40(39):11757–67, 2001. 1.4.2

[46] U. Baxa, K. L. Taylor, J. S. Wall, M. N. Simon, N. Cheng, R. B. Wickner, and A. C. Steven. Architecture of Ure2p prion filaments: the N-terminal domains form a central core fiber. *J Biol Chem*, 278(44):43717–27, 2003. 1.4.2, 2.6.4, 5.2

[47] M. Monti, A. Amoresano, S. Giorgetti, V. Bellotti, and P. Pucci. Limited proteolysis in the investigation of beta2-microglobulin amyloidogenic and fibrillar states. *Biochim Biophys Acta*, 1753(1):44–50, 2005. 1.4.2

[48] E. Frare, M. F. Mossuto, P. Polverino de Laureto, M. Dumoulin, C. M. Dobson, and A. Fontana. Identification of the core structure of lysozyme amyloid fibrils by proteolysis. *J Mol Biol*, 361(3):551–61, 2006. 1.4.2, 2.5.2, 2.6.3, 4.2.2, 4.3, 4.5, 4.3, 4.3.2, 5.3

[49] N. Carulla, G. L. Caddy, D. R. Hall, J. Zurdo, M. Gairi, M. Feliz, E. Giralt, C. V. Robinson, and C. M. Dobson. Molecular recycling within amyloid fibrils. *Nature*, 436(7050):554–8, 2005. 1.4.2, 2.4.3, 2.4.6, 4.1, 4.1.1, 4.3, 4.5

[50] A. Sillen, A. Leroy, J. M. Wieruszeski, A. Loyens, J. C. Beauvillain, L. Buee, I. Landrieu, and G. Lippens. Regions of tau implicated in the paired helical fragment core as defined by NMR. *Chembiochem*, 6(10):1849–56, 2005. 1.4.2, 3.4, 4.1, 6.1

[51] A. J. Baldwin, R. Bader, J. Christodoulou, C. E. MacPhee, C. M. Dobson, and P. D. Barker. Cytochrome display on amyloid fibrils. *J. Am. Chem. Soc.*, 128(7):2162–3, 2006. 1.4.2, 3.4, 4.1, 4.1.1, 5.2, 6.1, 7.1, 8.1

[52] A.B. Siemer, A.A. Arnold, C. Ritter, T. Westfeld, M. Ernst, R. Riek, and B.H. Meier. Observation of highly flexible residues in amyloid fibrils of the HET-s prion. *J. Am. Chem. Soc.*, 128(40):13224–13228, 2006. 1.4.2, 6.1

[53] T.P.J. Knowles, J.F. Smith, A. Craig, C.M. Dobson, and M.E. Welland. Spatial persistence of angular correlations in amyloid fibrils. *Physical Review Letters*, 96(23):238301, 2006. 1.4.2, 2.4.4, 2.6.4, 2.8.3, 5.2, 5.3

[54] C. Chothia. Conformation of twisted beta-pleated sheets in proteins. *J Mol Biol*, 75(2):295–302, 1973. 1.4.2

[55] D. Hall and H. Edskes. Silent prions lying in wait: a two-hit model of prion/amyloid formation and infection. *J Mol Biol*, 336(3):775–86, 2004. 1.8, 1.4.3

[56] F. Oosawa and S. Asakura. *Thermodynamics of the polymerisation of protein*. Academic, New York, 1975. 1.4.3

[57] M. Eigen. Prionics or the kinetic basis of prion diseases. *Biophys Chem*, 63(1):A1–18, 1996. 1.4.3

[58] M. Bucciantini, E. Giannoni, F. Chiti, F. Baroni, L. Formigli, J. Zurdo, N. Taddei, G. Ramponi, C. M. Dobson, and M. Stefani. Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases. *Nature*, 416(6880):507–11, 2002. 1.4.3, 2.4.15

[59] M. Groenning, L. Olsen, M. van de Weert, J. M. Flink, S. Frokjaer, and F. S. Jorgensen. Study on the binding of thioflavin t to beta-sheet-rich and non-beta-sheet cavities. *J Struct Biol*, 158(3):358–69, 2007. 1.4.3

[60] J. Zurdo, J. I. Guijarro, J. L. Jimenez, H. R. Saibil, and C. M. Dobson. Dependence on solution conditions of aggregation and amyloid formation by an SH3 domain. *J Mol Biol*, 311(2):325–40, 2001. 1.4.3, 2.1.1, 2.4.1, 2.4.15, 4.1, 4.1.1, 4.2.1

[61] N. Ferguson, J. Berriman, M. Petrovich, T. D. Sharpe, J. T. Finch, and A. R. Fersht. Rapid amyloid fiber formation from the fast-folding WW domain FBP28. *Proc Natl Acad Sci USA*, 100(17):9814–9, 2003. 1.4.3

[62] K. F. DuBay, A. P. Pawar, F. Chiti, J. Zurdo, C. M. Dobson, and M. Vendruscolo. Prediction of the absolute aggregation rates of amyloidogenic polypeptide chains. *J Mol Biol*, 341(5):1317–26, 2004. 1.4.3, 4.3, 5.1

[63] A. P. Pawar, K. F. Dubay, J. Zurdo, F. Chiti, M. Vendruscolo, and C. M. Dobson. Prediction of ”aggregation-prone” and ”aggregation-susceptible” regions in proteins associated with neurodegenerative diseases. *J Mol Biol*, 350(2):379–92, 2005. 1.4.3, 4.3, 5.1

[64] W. P. Esler, E. R. Stimson, J. B. Fishman, J. R. Ghilardi, H. V. Vinters, P. W. Mantyh, and J. E. Maggio. Stereochemical specificity of Alzheimer’s disease beta-peptide assembly. *Biopolymers*, 49(6):505–14, 1999. 1.4.3

[65] J. S. Griffith. Self-replication and scrapie. *Nature*, 215(5105):1043–4, 1967. 1.4.3

[66] J. S. Griffith. A theory of the nature of memory. *Nature*, 211(5054):1160–3, 1966. 1.4.3

[67] J. Shorter and S. Lindquist. Prions as adaptive conduits of memory and inheritance. *Nat Rev Genet*, 6(6):435–50, 2005. 1.4.3

[68] S. B. Prusiner. Prion diseases and the bse crisis. *Science*, 278(5336):245–51, 1997. 1.4.3

[69] D.C. Gajdusek. *Unconventional Viruses and the Origin and Dissappearance of Kuru*. Nobel Lectures, Physiology or Medicine 1971-1980. World Scientific Publishing Co., Singapore, 1976. 1.4.3

[70] S. Lesne, M. T. Koh, L. Kotilinek, R. Kayed, C. G. Glabe, A. Yang, M. Gallagher, and K. H. Ashe. A specific amyloid-beta protein assembly in the brain impairs memory. *Nature*, 440(7082):352–7, 2006. 1.4.3

[71] M. Reches and E. Gazit. Casting metal nanowires within discrete self-assembled peptide nanotubes. *Science*, 300(5619):625–627, 2003. 1.10, 1.5

[72] T. Scheibel, R. Parthasarathy, G. Sawicki, X.M. Lin, H. Jaeger, and S.L. Lindquist. Conducting nanowires built by controlled self-assembly of amyloid fibers and selective metal deposition. *PNAS*, 100(8):4527–4532, 2003. 1.10, 1.5

[73] C. E. MacPhee and C. M. Dobson. Formation of mixed fibrils demonstrates the generic nature and potential utility of amyloid nanostructures. *J. Am. Chem. Soc.*, 122(51):12707–12713, 2000. 1.5, 2.6.3

[74] N. Sondheimer and S. Lindquist. Rnq1: an epigenetic modifier of protein function in yeast. *Mol Cell*, 5(1):163–72, 2000. 1.5

[75] U. Baxa, V. V. Speransky, A.C. Steven, and R.B. Wickner. Inaugural article: Mechanism of inactivation on prion conversion of the saccharomyces cerevisiae ure2 protein. *PNAS*, 99(8):5253–5260, 2002. 1.5

[76] M. R. Chapman, L. S. Robinson, J. S. Pinkner, R. Roth, J. Heuser, M. Hammar, S. Normark, and S. J. Hultgren. Role of *E. coli* curli operons in directing amyloid fiber formation. *Science*, 295(5556):851–5, 2002. 1.5

[77] D. M. Fowler, A. V. Koulov, C. Alory-Jost, M. S. Marks, W. E. Balch, and J. W. Kelly. Functional amyloid formation within mammalian tissue. *PLoS Biol*, 4(1):e6, 2006. 1.5

[78] G. W. Booker, I. Gout, A. K. Downing, P. C. Driscoll, J. Boyd, M. D. Waterfield, and I. D. Campbell. Solution structure and ligand-binding site of the SH3 domain of the p85 alpha subunit of phosphatidylinositol 3-kinase. *Cell*, 73(4):813–22, 1993. 2.2

[79] C. M. Dobson and M. Karplus. The fundamentals of protein folding: bringing together theory and experiment. *Curr Opin Struct Biol*, 9(1):92–101, 1999. 2.1.1

[80] R. Bader, R. Bamford, J. Zurdo, B. F. Luisi, and C. M. Dobson. Probing the mechanism of amyloidogenesis through a tandem repeat of the PI3-SH3 domain suggests a generic model for protein aggregation and fibril formation. *J Mol Biol*, 356(1):189–208, 2006. 2.1.1, 2.4.1, 2.4.3, 2.4.15, 4.1.1, 8.1, 8.1.2

[81] M. T. Fisher. Differences in thermal stability between reduced and oxidized cytochrome b_{562} from *E. coli*. *Biochemistry*, 30(41):10012–8, 1991. 2.1.2

[82] P. D. Barker, J. L. Butler, P. de Oliveira, H. A. O. Hill, and N. I. Hunt. Direct electrochemical studies of cytochromes b_{562}. *Inorganica Chimica Acta*, 252(1-2):71–77, 1996. 2.1.2, 2.1.2, 2.4.11

[83] F. Arnesano, L. Banci, I. Bertini, J. Faraone-Mennella, A. Rosato, P. D. Barker, and A. R. Fersht. The solution structure of oxidized *E. coli* cytochrome b_{562}. *Biochemistry*, 38(27):8657–70, 1999. 2.1.2, 2.1.2, 2.6

[84] M. Gouterman and D. Dolphin. The porphyrins. In *The Porphyrins*, volume 3, pages 1–165. Academic Press, New York, 1978. 2.1.2

[85] P. A. Bullock and Y. P. Myer. Circular dichroism and resonance raman studies of cytochrome b_{562} from *E. coli*. *Biochemistry*, 17(15):3084–91, 1978. 2.1

[86] G. R. Moore, R. J. Williams, J. Peterson, J. A. Thomson, and F. S. Mathews. A spectroscopic investigation of the structure and redox properties of *E. coli* cytochrome b_{562}. *Biochim Biophys Acta*, 829(1):83–96, 1985. 2.1.2

[87] K. Hamada, P. H. Bethge, and F. S. Mathews. Refined structure of cytochrome b_{562} from *E. coli* at 1.4 a resolution. *J Mol Biol*, 247(5):947–62, 1995. 2.1.2

[88] J. Z. Wu, G. N. La Mar, L. P. Yu, K. B. Lee, F. A. Walker, M. L. Chiu, and S. G. Sligar. 1H NMR study of the solution molecular and electronic structure of *E. coli* ferricytochrome b_{562}: evidence for s = 1/2 in equilibrium s = 5/2 spin equilibrium for intact his/met ligation. *Biochemistry*, 30(8):2156–65, 1991. 2.1.2

[89] C. R. Robinson and R. T. Sauer. Optimizing the stability of single-chain proteins by linker length and composition mutagenesis. *PNAS*, 95(11):5929–5934, 1998. 2.2

[90] R. A. George and J. Heringa. An analysis of protein domain linkers: their classification and role in protein folding. *Protein Eng.*, 15(11):871–879, 2002. 2.2

[91] Expasy protparam tool. *http://www.expasy.ch/tools/protparam.html*. 2.4.3, 8.1.2

[92] J.F. Smith, T.P.J. Knowles, C.M. Dobson, C.E. MacPhee, and M.E. Welland. Characterization of the nanoscale properties of individual amyloid fibrils. *PNAS*, 103(43):15806–15811, 2006. 2.4.4, 4.1.1

[93] J. L. Jimenez, G. Tennent, M. Pepys, and H. R. Saibil. Structural diversity of *ex-vivo* amyloid fibrils studied by cryo-electron microscopy. *J Mol Biol*, 311(2):241–7, 2001. 2.4.4

[94] A. Baldwin, S.J Anthony-Cahill, J. Christodoulou, G. Lippens, P. D. Barker, and C. M. Dobson. Measurement of amyloid fibril length distributions by inclusion of rotational motion in solution-state NMR diffusion measurements. *Ange. Chem. Int. Ed.*, accepted, 2007. 2.4.4, 3.4, 7.1

[95] E. O. Stejskal and J. E. Tanner. Spin diffusion measurements: Spin echoes in the presence of a time-dependent field gradient. *J. Chem. Phys.*, 42(1):288–292, 1965. 2.4.8, 3.1, 9.2

[96] A.J. Dingley, J.P. Mackay, G.L. Shaw, B.D. Hambly, and G.F. King. Measuring macromolecular diffusion using heteronuclear multiple-quantum pulsed-field-gradient NMR. *Journal of Biomolecular NMR*, 10(1):1–8, 1997. 2.4.9, 8.2.3

[97] P. D. Barker and S. M. Freund. Bis-methionine ligation to heme iron in mutants of cytochrome b_{562}. 2. characterization by NMR of heme-ligand interactions. *Biochemistry*, 35(42):13627–35, 1996. 2.4.11, 2.5.4

[98] F. A. Armstrong, J. N. Butt, and A. Sucheta. Voltammetric studies of redox-active centers in metalloproteins adsorbed on electrodes. *Methods Enzymol*, 227:479–500, 1993. 2.4.11

[99] K. M. Lundberg, C. J. Stenland, F. E. Cohen, S. B. Prusiner, and G. L. Millhauser. Kinetics and mechanism of amyloid formation by the prion protein H1 peptide as determined by time-dependent ESR. *Chem Biol*, 4(5):345–55, 1997. 2.4.15

[100] M.-H.J. Lee and R.J. Collier. Sheet resistance measurement of thin metallic films and stripes at both 130 GHz and DC. *Instrumentation and Measurement, IEEE Transactions on*, 54(6):2412–2415, 2005. 2.5.4

[101] J. Cavanagh, W. Fairbrother, Palmer, and N. A.G. Skelton. *Protein NMR Spectroscopy: Principles and Practice*. Academic Press, 1995. 2.6.1, 4.2.2, 8.2.3

[102] Biological Magnetic Resonance data Bank. www.bmrb.wisc.edu. 2.6.1, 4.4, 4.2.2

[103] J. Christodoulou, G. Larsson, P. Fucini, S. R. Connell, T. A. Pertinhez, C. L. Hanson, C. Redfield, K. H. Nierhaus, C. V. Robinson, J. Schleucher, and C. M. Dobson. Heteronuclear NMR investigations of dynamic regions of intact *E. coli* ribosomes. *Proc Natl Acad Sci USA*, 101(30):10949–54, 2004. 2.6.1, 4.1, 4.3, 6.1

[104] H. Schwalbe, K. M. Fiebig, M. Buck, J. A. Jones, S. B. Grimshaw, A. Spencer, S. J. Glaser, L. J. Smith, and C. M. Dobson. Structural and dynamical properties of a denatured protein. heteronuclear 3D NMR experiments and theoretical simulations of lysozyme in 8 M urea. *Biochemistry*, 36(29):8977–91, 1997. 2.6.2, 3.3, 6.2.1

[105] T. Scheibel, A. S. Kowal, J. D. Bloom, and S. L. Lindquist. Bidirectional amyloid fiber growth for a yeast prion determinant. *Curr Biol*, 11(5):366–9, 2001. 2.6.3

[106] M. R. Nilsson and C. M. Dobson. Chemical modification of insulin in amyloid fibrils. *Protein Sci*, 12(11):2637–41, 2003. 2.6.3

[107] I. Kheterpal, S. Zhou, K. D. Cook, and R. Wetzel. Aβ amyloid fibrils possess a core structure highly resistant to hydrogen exchange. *PNAS*, 97(25):13597–13601, 2000. 2.6.3

[108] C. S. Goldsbury, G. J. Cooper, K. N. Goldie, S. A. Muller, E. L. Saafi, W. T. Gruijters, M. P. Misur, A. Engel, U. Aebi, and J. Kistler. Polymorphic fibrillar assembly of human amylin. *J Struct Biol*, 119(1):17–27, 1997. 2.6.3, 2.6.3

[109] I. Cardoso, C. S. Goldsbury, S. A. Muller, V. Olivieri, S. Wirtz, A. M. Damas, U. Aebi, and M. J. Saraiva. Transthyretin fibrillogenesis entails the assembly of monomers: a molecular model for *In-vitro*assembled transthyretin amyloid-like fibrils. *J Mol Biol*, 317(5):683–95, 2002. 2.6.3, 2.6.3

[110] H. H. Bauer, U. Aebi, M. Haner, R. Hermann, M. Muller, and H. P. Merkle. Architecture and polymorphism of fibrillar supramolecular assemblies produced by *in-vitro* aggregation of human calcitonin. *J Struct Biol*, 115(1):1–15, 1995. 2.6.3, 2.6.3

[111] O. N. Antzutkin, R. D. Leapman, J. J. Balbach, and R. Tycko. Supramolecular structural constraints on Alzheimer's beta-amyloid fibrils from electron microscopy and solid-state nuclear magnetic resonance. *Biochemistry*, 41(51):15436–50, 2002. 2.6.3, 2.6.3

[112] Rouzbeh Ghafouri and Robijn Bruinsma. Helicoid to spiral ribbon transition. *Physical Review Letters*, 94(13):138101, 2005. 2.30, 2.6.4, 5.3

[113] A. Bigi, L. Dovigo, M. H. Koch, M. Morocutti, A. Ripamonti, and N. Roveri. Collagen structural organization in uncalcified and calcified human anterior longitudinal ligament. *Connect Tissue Res*, 25(3-4):171–9, 1991. 2.6.4

[114] A. M. Emons and B. M. Mulder. The making of the architecture of the plant cell wall: how cells exploit geometry. *PNAS*, 95(12):7215–9, 1998. 2.6.4

[115] S. Pakhomov, R. P. Hammer, B. K. Mishra, and B. N. Thomas. Chiral tubule self-assembly from an achiral diynoic lipid. *Proc Natl Acad Sci USA*, 100(6):3040–2, 2003. 2.6.4

[116] S. Zhao, S. Zhang, Z. Yao, and L. Zhang. Equilibrium conformation of polymer chains with noncircular cross section. *Physical Review E*, 74(3):032801, 2006. 2.6.4, 5.3

[117] H. C. Torrey. Bloch equations with diffusion terms. *Physical Review*, 104(3):563 LP – 565, 1956. 3.1

[118] A. Dehner and H. Kessler. Diffusion NMR spectroscopy: folding and aggregation of domains in p53. *Chembiochem*, 6(9):1550–65, 2005. 3.1, 3.1

[119] X. Qin, Y. Ishizuka, J. S. Lomas, T. Tezuka, and H. Nakanishi. 17o, 13c and 1h NMR and IR spectral study of crowded ketones: possible intramolecular C-HO interactions. *Magnetic Resonance in Chemistry*, 40(9):595–598, 2002. 3.1

[120] H. Ponstingl and G. Otting. Detection of protein-ligand NOEs with small, weakly binding ligands by combined relaxation and diffusion filtering. *J. Biol. NMR*, 9(4):441–444, 1997. 3.1

[121] O. Mayzel and Y. Cohen. Diffusion coefficients of macrocyclic complexes using the PGSE NMR technique: Determination of association constants. *J. Chem. Soc. Chem. Commun.*, page 1901, 1994. 3.1

[122] K. Bleicher, M. Lin, M. J. Shapiro, and J. R. Wareing. Diffusion edited NMR: Screening compound mixtures by affinity NMR to detect binding ligands to vancomycin. *J. Org. Chem.*, 63(23):8486–8490, 1998. 3.1

[123] J. J. Chou, J. L. Baber, and A. Bax. Characterization of phospholipid mixed micelles by translational diffusion. *J Biomol NMR*, 29(3):299–308, 2004. 3.1, 9.2

[124] A. Dehner, E. Planker, G. Gemmecker, Q.B. Broxterman, W. Bisson, F. Formaggio, M. Crisma, C. Toniolo, and H. Kessler. Solution structure, dimerization, and dynamics of a lipophilic α3 10-helical, ca;-methylated peptide. implications for folding of membrane proteins. *J. Am. Chem. Soc.*, 123(27):6678–6686, 2001. 3.1

[125] X. Chang, D. Keller, S. I. O'Donoghue, and J. J. Led. NMR studies of the aggregation of glucagon-like peptide-1: formation of a symmetric helical dimer. *FEBS Letters*, 515(1-3):165–170, 2002. 3.1

[126] J. A. Jones, D.K. Wilkins, L. J. Smith, and C. M. Dobson. Characterisation of protein unfolding by NMR diffusion measurements. *J. Biol. NMR*, 10(2):199–203, 1997. 3.1

[127] J. Balbach. Compaction during protein folding studied by real-time NMR diffusion experiments. *J. Am. Chem. Soc.*, 122(24):5887–5888, 2000. 3.1

[128] J. E. Tanner. Use of the stimulated echo in NMR diffusion studies. *J. Chem. Phys.*, 52(5):2523–2526, 1970. 3.1

[129] D. K. Wilkins, C. M. Dobson, and M. Gross. Biophysical studies of the development of amyloid fibrils from a peptide fragment of cold shock protein B. *Eur J Biochem*, 267(9):2609–16, 2000. 3.1, 4.1

[130] Y. Tao, W. K. den Otter, J. T. Padding, J. K. G. Dhont, and W. J. Briels. Brownian dynamics simulations of the self- and collective rotational diffusion coefficients of rigid long thin rods. *J. Chem. Phys.*, 122(24):244903, 2005. 3.2.1

[131] H. Yamakawa and G. Tanaka. Translational diffusion coefficients of rod-like polymers: Application of the modified Oseen tensor. *J. Chem. Phys.*, 57(4):1537–1542, 1972. 3.2.3

[132] S. Broersma. Rotational diffusion constant of a cylindrical particle. *J. Chem. Phys.*, 32(6):1626–1631, 1960. 3.2.3

[133] F. R. Hallett and R. Keates. Rapid estimation of length distributions of microtubule preparations by quasi-elastic light scattering. *Biopolymers*, 24(12):2403–2415, 1985. 3.3, 5.2

[134] J. E. Tanner and E. O. Stejskal. Restricted self-diffusion of protons in colloidal systems by the pulsed-gradient, spin-echo method. *J. Chem. Phys.*, 49(4):1768–1777, 1968. 3.4

[135] http://www.gnu.org/software/gsl/, 2007. 3.5.1, 8.2.1

[136] H. Yamakawa and M. Fujii. Translational friction coefficient of wormlike chains. *Macromolecules*, 6(3):407–415, 1973. 3.6.6

[137] N Roder, P Ellis, and P.J. Butterworth. Starch molecular and nutritional properties; a review. *Advances in molecular medicine*, 1:5–14, 2005. 3.6.7

[138] M. Gross, D. K. Wilkins, M. C. Pitkeathly, E. W. Chung, C. Higham, A. Clark, and C. M. Dobson. Formation of amyloid fibrils by peptides derived from the bacterial cold shock protein CspB. *Protein Sci*, 8(6):1350–7, 1999. 4.1

[139] R. K. Wangsness and F. Bloch. The dynamical theory of nuclear induction. *Physical Review*, 89(4):728 LP – 739, 1953. 4.1, 6.1

[140] K. T. Dayie, G. Wagner, and J. Lefevre. Theory and practise of nuclear spin relaxation in proteins. *Annual Review of Physical Chemistry*, 47(1):243–282, 1996. 4.1, 6.1

[141] I. J. Griswold and F. W. Dahlquist. Bigger is better: megadalton protein NMR in solution. *Nat Struct Biol*, 9(8):567–8, 2002. 4.1

[142] M. M. Yusupov, G. Z. Yusupova, A. Baucom, K. Lieberman, T. N. Earnest, J. H. Cate, and H. F. Noller. Crystal structure of the ribosome at 5.5 a resolution. *Science*, 292(5518):883–96, 2001. 4.1

[143] S. E. Radford, E. D. Laue, R. N. Perham, J. S. Miles, and J. R. Guest. Segmental structure and protein domains in the pyruvate dehydrogenase multienzyme complex of *E. coli*. genetic reconstruction *In-vitro*and ^{1}H-NMR spectroscopy. *Biochem J*, 247(3):641–9, 1987. 4.1, 6.1

[144] R. E. Oswald, M. J. Bogusky, M. Bamberger, R. A. Smith, and C. M. Dobson. Dynamics of the multidomain fibrinolytic protein urokinase from two-dimensional NMR. *Nature*, 337(6207):579–82, 1989. 4.1, 6.1

[145] S. Schwarzinger, G. J. Kroon, T. R. Foss, J. Chung, P. E. Wright, and H. J. Dyson. Sequence-dependent correction of random coil NMR chemical shifts. *J Am Chem Soc*, 123(13):2970–8, 2001. 4.1, 6.1

[146] J. A. Carver, J. A. Aquilina, R. J. Truscott, and G. B. Ralston. Identification by ^{1}H NMR spectroscopy of flexible C-terminal extensions in bovine lens alpha-crystallin. *FEBS Lett*, 311(2):143–9, 1992. 4.1, 4.1.1, 4.2.2, 6.1

[147] C.R. Cantor and P.R. Schimmel. *Biophysical Chemistry, Vols I, II, III*. W H Freeman, 1980. 4.1.1, 5.2, 8.2.2

[148] C. C. Blake, M. J. Geisow, S. J. Oatley, B. Rerat, and C. Rerat. Structure of prealbumin: secondary, tertiary and quaternary interactions determined by fourier refinement at 1.8 Ä. *J Mol Biol*, 121(3):339–56, 1978. 4.1.1

[149] P. Westermark, K. Sletten, B. Johansson, and 3rd Cornwell, G. G. Fibril in senile systemic amyloidosis is derived from normal transthyretin. *Proc Natl Acad Sci USA*, 87(7):2843–5, 1990. 4.1.1, 5.4.1

[150] M. J. Saraiva. Transthyretin mutations in hyperthyroxinemia and amyloid diseases. *Hum Mutat*, 17(6):493–503, 2001. 4.1.1, 5.4.1

[151] A. Gustavsson, U. Engstrom, and P. Westermark. Normal transthyretin and synthetic transthyretin fragments form amyloid-like fibrils *In-vitro*. *Biochem Biophys Res Commun*, 175(3):1159–64, 1991. 4.1.1, 5.1, 5.4.1

[152] W. Colon and J. W. Kelly. Partial denaturation of transthyretin is sufficient for amyloid fibril formation *In-vitro*. *Biochemistry*, 31(36):8654–60, 1992. 4.1.1, 5.4.1

[153] M. J. Bonifacio, Y. Sakaki, and M. J. Saraiva. '*In-vitro*' amyloid fibril formation from transthyretin: the influence of ions and the amyloidogenicity of TTR variants. *Biochim Biophys Acta*, 1316(1):35–42, 1996. 4.1.1, 5.4.1

[154] C. E. MacPhee and C. M. Dobson. Chemical dissection and reassembly of amyloid fibrils formed by a peptide fragment of transthyretin. *J Mol Biol*, 297(5):1203–15, 2000. 4.1.1, 5.1, 5.4.1

[155] T. L. Blundell, J. F. Cutfield, G. G. Dodson, E. Dodson, D. C. Hodgkin, and D. Mercola. The structure and biology of insulin. *Biochem J*, 125(3):50P–51P, 1971. 4.1.1

[156] D. F. Waugh. A fibrous modification of insulin. i. the heat precipitate of insulin. *J. Am. Chem. Soc.*, 68(2):247–250, 1946. 4.1.1

[157] J. Brange, L. Andersen, E. D. Laursen, G. Meyn, and E. Rasmussen. Toward understanding insulin fibrillation. *J Pharm Sci*, 86(5):517–25, 1997. 4.1.1

[158] M. Bouchard, J. Zurdo, E. J. Nettleton, C. M. Dobson, and C. V. Robinson. Formation of insulin amyloid fibrils followed by ftir simultaneously with cd and electron microscopy. *Protein Sci*, 9(10):1960–7, 2000. 4.1.1

[159] S. Meehan, Y. Berry, B. Luisi, C. M. Dobson, J. A. Carver, and C. E. MacPhee. Amyloid fibril formation by lens crystallin proteins and its implications for cataract formation. *J Biol Chem*, 279(5):3413–9, 2004. 4.1.1

[160] G. Wider and L. Dreier. Measuring protein concentrations by NMR spectroscopy. *J. Am. Chem. Soc.*, 128(8):2571–2576, 2006. 4.2.1, 4.3, 4.4, 4.5.1, 4.5.1

[161] M. Calamai, C. Canale, A. Relini, M. Stefani, F. Chiti, and C. M. Dobson. Reversal of protein aggregation provides evidence for multiple aggregated states. *J Mol Biol*, 346(2):603–16, 2005. 4.2.1

[162] C. Dirix, F. Meersman, C. E. MacPhee, C. M. Dobson, and K. Heremans. High hydrostatic pressure dissociates early aggregates of ttr105-115, but not the mature amyloid fibrils. *J Mol Biol*, 347(5):903–9, 2005. 4.2.1

[163] S. Meehan, T.P.J. Knowles, A.J. Baldwin, J.F. Smith, A.M. Squires, P. Clements, T.M. Treweek, H. Ecroyd, G.G. Tartaglia, M. Vendruscolo, C.E. MacPhee, C.M. Dobson, and J.A. Carver. Characterisation of amyloid fibril formation by small heat-shock chaperone proteins human αA-, αB- and R120G αB-crystallins. *J. Mol. Biol.*, 372(2):470–484, 2007. 4.2.2, 7.1

[164] D.K. Wilkins, S.B. Grimshaw, V. Receveur, C.M. Dobson, J.A. Jones, and L.J. Smith. Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. *Biochemistry*, 38(50):16424–16431, 1999. 4.2.2, 4.3, 9.2.2

[165] J. Danielsson, J. Jarvet, P. Damberg, and A. Grslund. Translational diffusion measured by PFG-NMR on full length and fragments of the Alzheimer Aβ(1-40) peptide. determination of hydrodynamic radii of random coil peptides of varying length. *Magnetic Resonance in Chemistry*, 40(13):S89–S97, 2002. 4.2.2, 4.3

[166] P.G. de Gennes. *Scaling concepts in polymer physics*. Cornell University Press, 1979. 4.3

[167] F. Oosawa and M. Kasai. A theory of linear and helical aggregations of macromolecules. *J Mol Biol*, 4:10–21, 1962. 4.3.1, 5.1, 5.4.2, 5.4.2

[168] Y. Ohkubo, Z. Brooks, and L. Charles. Exploring flory's isolated-pair hypothesis: Statistical mechanics of helix-coil transitions in polyalanine and the C-peptide from RNase A. *PNAS*, 100(24):13916–13921, 2003. 4.3.2, 4.5.2

[169] D. S. McKenzie. Polymers and scaling. *Physics Reports*, 27(2):35–88, 1976. 4.3.2

[170] J. J. van der Klink. The NMR reciprocity theorem for arbitrary probe geometry. *Journal of Magnetic Resonance*, 148(1):147–154, 2001. 4.5.1

[171] W.P. Esler, E.R. Stimson, J.M. Jennings, H.V. Vinters, J.R. Ghilardi, J.P. Lee, P.W. Mantyh, and J.E. Maggio. Alzheimer's disease amyloid propagation by a template-dependent dock-lock mechanism. *Biochemistry*, 39(21):6288–6295, 2000. 5.4.2

[172] Redfield AG. Theory of nuclear relaxation processes. *Adv. Magn. Reson.*, 1:1–32, 1965. 6.1

[173] A Abragam. *The principles of nuclear magnetism*. The international series of monographs on physics. Clarendon press, Oxford, 1961. 6.1, 8.2.3

[174] A.J. Baldwin, J. Christodoulou, P.D. Barker, C.M. Dobson, and G. Lippens. Contribution of rotational diffusion to pulsed field gradient diffusion measurements. *J. Chem. Phys.*, 127(11):114505, 2007. 6.1, 7.1

[175] J. Shimada, E. L. Kussell, and E. I. Shakhnovich. The folding thermodynamics and kinetics of crambin using an all-atom monte carlo simulation. *J Mol Biol*, 308(1):79–95, 2001. 6.1.2, 6.1.3

[176] Nicholas Metropolis and S. Ulam. The monte carlo method. *Journal of the American Statistical Association*, 44(247):335–341, 1949. 6.1.2

[177] Antonio Rey and Jeffrey Skolnick. Comparison of lattice monte carlo dynamics and brownian dynamics folding pathways of [alpha]-helical hairpins. *Chemical Physics*, 158(2-3):199–219, 1991. 6.1.3

[178] V. I. Abkevich, A. M. Gutin, and E. I. Shakhnovich. Free energy landscape for protein folding kinetics: Intermediates, traps, and multiple pathways in theory and lattice model simulations. *J. Chem. Phys.*, 101(7):6052–6062, 1994. 6.1.3

[179] V. I. Abkevich, A. M. Gutin, and E. I. Shakhnovich. Specific nucleus as the transition state for protein folding: Evidence from the lattice model. *Biochemistry*, 33(33):10026–10036, 1994. 6.1.3

[180] Malcolm H. Levitt. Spin dynamics. *Wiley*, 2001. 8.2.3

[181] J Keeler. http://www-keeler.ch.cam.ac.uk/, 1999. 8.2.3

[182] F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, and A. Bax. NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biol. NMR*, V6(3):277–293, 1995. 8.2.3, 9.1.1

[183] K. Krynicki, C.D. Green, and D.W. Sawyer. Pressure and temperature dependence of self-diffusion in water. *Faraday Discuss. Chem. Soc.*, 66:199–208, 1978. 9.2

[184] G. Petekidis, D. Vlassopoulos, G. Fytas, R. Rulkens, G. Wegner, and G. Fleischer. Diffusion dynamics of hairy-rod polymers in concentrated solutions. *Macromolecules*, 31(18):6139–6147, 1998. 9.2