

Structural characterization and engineering of Adnectin inclusion bodies

by

Anna O. Schaefer

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Science
in
Chemistry

Waterloo, Ontario, Canada, 2022
© Anna O. Schaefer 2022

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions for manuscripts included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

This thesis consists in part of three manuscripts submitted for publication. Exceptions to sole authorship of material are described at the beginning of each individual chapter. Citations for manuscripts from which data/text are included are provided below:

Schaefer, A.*, Naser, D.*, Siebeneichler, B.*, Tarasca, M. V. and Meiering, E.M. Methodological advances and strategies for high resolution structure determination of cellular protein aggregates. *J. Biol. Chem.* (2022).

Naser, D. *et al.* High Resolution NMR H/D Exchange of Human Superoxide Dismutase Inclusion Bodies Reveals Significant Native Features Despite Structural Heterogeneity. *Angew. Chemie - Int. Ed.* **61**, e202112645 (2022).

Tarasca, M. V., Naser, D., Schaefer, A., Soule, T. G. & Meiering, E.M. Quenched hydrogen-deuterium amide exchange optimization for high-resolution structural analysis of cellular protein aggregates. *Anal. Biochem.* **652**, (2022).

*indicates that these authors contributed equally.

Abstract

Protein aggregation is a fundamental process occurring as a response to aging, disease, protein overexpression and in many areas of biotechnology and medical applications. There has been much recent progress in determining the structural features of protein aggregates that form in cells; yet, owing to prevalent heterogeneity in aggregation, many aspects remain obscure and often experimentally intractable to define. Here we investigate the structural features of unwashed bacterial inclusion bodies (IBs) formed by Adnectins, a family of engineered proteins based on the 10th fibronectin type II domain. These small, stable proteins are actively being designed as biotherapeutics despite being remarkably aggregation-prone, making them an ideal model system for studying aggregation in a biological context. The properties of Adnectin IBs were investigated using a coalition of adapted structural biochemistry techniques including Fourier transform infrared (FTIR) spectroscopy, urea denaturation, Congo red binding and quenched hydrogen deuterium exchange (qHDX) nuclear magnetic resonance (NMR) spectroscopy. It was found that all Adnectin IBs displayed similar structure containing significant native structural features along with a modest contribution from amyloid-like species despite point-mutations affecting their biochemical properties and predicted aggregation behaviour. To further probe aggregate structure, multi- point mutants were rationally designed to investigate how charged residues and monomer stability can influence the ensemble of structures that make up an IB aggregate. As predicted, strongly destabilizing an Adnectin monomer results in an increased contribution from amyloid species. However, these experiments also indicate that charged residues may affect protein stability in ways not anticipated by computational predictors or molecular dynamics simulations, which ultimately demonstrates one of the nuances in understanding the molecular determinants of aggregation as a whole.

These findings are of great significance for understanding the specific molecular determinants of IB formation. Specifically, Adnectin variants may more broadly inform on the determinants of aggregation of proteins with immunoglobulin folds, including antibodies which are a rapidly growing group of novel drug candidates.

Acknowledgements

“Alone we can do so little but together we can do so much”

- Helen Keller.

As is true with many personal journeys, the work within this thesis could not be completed alone. Instead, this thesis is a reflection of the support and guidance of many individuals who have been an integral part of my life over the past two years.

I would like to say a special thank you my supervisor, Dr. Elizabeth Meiering, for providing the opportunity to pursue graduate research and for helping me grow as a scientist.

I also would like to thank all members of the Meiering lab for their continued support, I could not have completed this work without their advice and friendship. I would specifically like to acknowledge Dalia Naser for her mentorship and invaluable insight on the study of aggregation. But most importantly, I would like to thank her for her companionship and never letting me pursue a ‘crazy’ idea on my own.

Lastly, to my partner, my parents, family and friends, your unwavering support and encouragement throughout this journey are what made this possible.

Table of Contents

Author's Declaration	ii
Statement of Contributions	iii
Abstract.....	iv
Acknowledgements	v
List of Figures.....	viii
List of Tables	x
List of Abbreviations.....	xi
Chapter 1 Introduction.....	1
1.1 Protein folding and misfolding.....	1
1.2 Protein aggregation.....	3
1.3 Protein aggregate structure	5
1.4 Inclusion bodies: Structure and applications	7
1.5 Fibronectin type III domain.....	8
1.5.1 Adnectins	10
1.6 Research objectives	11
Chapter 2 Biophysical characterization of Adnectin Inclusion bodies.....	13
2.1 Introduction	13
2.1.1 Context.....	13
2.1.2 Point-mutant selection	14
2.1.3 Fourier Transform Infrared Spectroscopy	15
2.1.4 Methods adapted to investigate aggregate stability	17
2.2 Methods	17
2.2.1 Cell growth and protein expression.....	17
2.2.2 Inclusion body (IB) preparation.....	18
2.2.3 Solubility analysis.....	18
2.2.4 FTIR data acquisition	19
2.2.5 FTIR spectral analysis	19
2.2.6 Proteinase K digestion	19
2.2.8 Congo red binding	20
2.2.7 Chemical solubilization and denaturation	20
2.3 Results and Discussion	21
2.3.1 Adnectin IB formation.....	21
2.3.2 Approaches to FTIR data analysis.....	22
2.3.3 Adnectin IBs contain significant native secondary structure elements	25
2.3.4 Protease digestion reveals similar protected regions despite varying levels of compactness	29
2.3.5 Chemical solubilization of Adnectin IBs expressed at 37°C, overnight.....	32
2.4 Conclusions	36
Chapter 3 High-resolution structural analysis of Adnectin IBs using quenched amide hydrogen-deuterium exchange NMR.....	37
3.1 Introduction	37
3.1.1 Nuclear magnetic resonance spectroscopy	37
3.1.2 Quenched hydrogen deuterium exchange NMR.....	38
2.1.3 Detergent washing	41

3.1.3 Mutant Selection	42
3.2 Methods	42
3.2.1 qHDX-NMR sample preparation.....	42
3.2.3 qHDX NMR experiments	43
3.2.3 qHDX NMR analysis.....	44
3.2.4 Sequence specific resonance assignments	45
3.2.5 Detergent Washing	45
3.3 Results and Discussion	46
3.3.1 Experimental considerations for applying qHDX to study Adnectin IBs	46
3.3.2 qHDX protection for mature Adnectin IBs, expressed at 37°C overnight.....	50
3.3.3 qHDX protection for Adnectin IBs expressed for 6 hours at 37°C	53
3.3.4 Detergent washing of Adnectin IBs.....	55
3.4 Conclusions	60
Chapter 4 Structural engineering of Adnectin inclusion bodies.....	61
4.1 Introduction	61
4.1.1 Context.....	61
4.1.2 Protein engineering and design.....	61
4.1.3 Computational stability predictors.....	62
4.1.4 Aggregation propensity predictors	63
4.2 Methods	64
4.2.1 Adnectin alignment construction and solubility analysis.	64
4.2.3 Aggregation propensity and stability prediction for engineered Adnectin mutants	64
4.2.4 Protein expression and experimental characterization	65
4.2.5 Implicit solvent molecular dynamics simulations and analysis.....	65
4.3 Construct Design	66
4.3.1 Rationale	66
4.3.2 Stability and solubility prediction.....	67
4.3.3 Conformational stability investigation using molecular dynamics	69
4.4 Experimental results and discussion.....	71
4.4.1 Expression levels and aggregation Propensity	71
4.4.2 FTIR and urea solubilization reveal subtle changes in secondary structure and stability of mutant IBs ..	73
4.4.3 Residue-level structure by qHDX.....	75
Chapter 5 Conclusions and future directions.....	78
5.1 Conclusions	78
5.2 Future Experiments	80
References	83
Appendices	92

List of Figures

Figure 1.1: Energetic models of protein folding.....	2
Figure 1.2: General scheme of cellular protein aggregation.	4
Figure 1.3: Human Fibronectin domain structure and sequence.	9
Figure 1.4: Tencon multimers formed by the association of β -strand G domain-swapped dimers.....	10
Figure 2.1 Ribbon diagram depicting location of Adnectin point mutations.	15
Figure 2.2: Comparison of average Adnectin in IB for different growth media types.....	22
Figure 2.3: Data processing workflow for FTIR analysis of protein aggregates.	23
Figure 2.4: Comparison of non-linear curve fitting methods for average SOD1 IB spectrum.	25
Figure 2.5: FTIR of mutant Adnectin IBs expressed overnight exhibit significant native features....	26
Figure 2.6: Point mutations do not have significant effects on the FTIR spectra of Adnectin IBs demonstrating similar aggregate structure.....	27
Figure 2.7: FTIR spectral features are attributed to overexpressed protein and not contaminants.	28
Figure 2.8: Predicted change in consensus amyloid aggregation propensity.	29
Figure 2.9: IBs show greater sensitivity to proteinase K digestion than purified Adnectin.....	30
Figure 2.10 Adnectin IBs and protein fragments released upon proteinase K digestion weakly bind congo red dye.....	31
Figure 2.11 Urea solubilization profiles demonstrating full solubilization of Adnectin IBs at 7M urea despite high thermodynamic stability of pure Adnectin.....	34
Figure 2.12: Correlation between IB denaturation cooperativity and FTIR β -peak position.....	35
Figure 3.1 Quenched hydrogen-deuterium exchange NMR for analysis of inclusion body structure.	39
Figure 3.2: Chemical Structures of mild, non-ionic detergents used to wash bacterial IBs.....	41
Figure 3.3: Protonated ^1H - ^{15}N HSQC spectra for V75R Adnectin IBs expressed for varying amounts of time post-induction with IPTG.....	46
Figure 3.4: ^1H - ^{15}N -2D-HSQC spectrum of Adnectin 6199-CO1 in 95% DMSO/5% H_2O , pH 5.5, with backbone amide assignments.	48
Figure 3.5: Residue-level structure of Adnectin IBs from qHDX analysis.....	51
Figure 3.6 Biphasic Exchange of IB amide protons in D_2O	53
Figure 3.7: Fraction amide protection for Adnectin IBs expressed 6 and 18 hours post induction. ...	55
Figure 3.8: SDS-PAGE gel analysis of washed Adnectin IBs.	56
Figure 3.9: FTIR absorbance spectra of detergent washed IBs.	57
Figure 3.10: Improved ^1H - ^{15}N HSQC spectral resolution upon washing with Triton-X 100.	58
Figure 3.11: Protection ratios for comparing IB structure unwashed vs. washed with Triton-X 100.	59
Figure 4.1: Free energy profile representing altered thermodynamic stability of a protein.	62
Figure 4.2: Computational predictions of (de)stabilizing mutations.....	68
Figure 4.3: Amino acid sequence ‘logos’ (created using WebLogo ¹⁷⁹) for insoluble and soluble FN_3 β -strand sequences.	70
Figure 4.4: Conformational analysis of MD simulations for Adnectin variants.	71
Figure 4.5: Soluble and insoluble expression levels of engineered Adnectin mutants in different media types by SDS-page gel analysis.	72
Figure 4.6: FTIR spectral comparison of parent and engineered Adnectin IBs.	73
Figure 4.7: Urea solubilization profiles of parent Adnectin and engineered Adnectin variants.	74
Figure 4.8: Average qHDX protection of engineered Adnectin IBs compared to parent when expressed at 37 °C overnight.	76

Figure 5.1: Adnectin IB aggregation involves multiple pathways including native and non-native interactions.....	79
Figure A.1: Computational aggregation propensity predictions for parent Adnectin.	93
Figure A.2: SDS-page gel analysis of soluble and insoluble expression levels of Adnectin.	94
Figure B1: FTIR curve fitting analysis for pure SOD1 protein.....	96
Figure B.2: FTIR second derivative spectra of mutant SOD1 IBs.	97
Figure B.3: FTIR absorbance spectra of sample replicates for mutant Adnectin IBs from cell cultures grown at 37 °C and harvested at 18 hours post-induction.....	98
Figure B.4: FTIR absorbance spectra of sample replicates for Parent IBs washed with non-ionic detergents.....	99
Figure B.5: FTIR absorbance spectra of sample replicates for V75R IBs washed with non-ionic detergents.....	100
Figure B.6: FTIR absorbance spectra of sample replicates for L78I IBs washed with non-ionic detergents.....	101
Figure C.1: Urea solubilization profiles of sample replicates for mutant Adnectin IBs from cell cultures grown at 37 °C and harvested at 18 hours post-induction.....	100
Figure D.1: Measured intrinsic rate constants of amide exchange for different amino acids.	105
Figure D.2: ¹ H- ¹⁵ N HSQCs showing multiple assignments for N-terminal amino acid sequence 'MGVSD'.	106
Figure D.3:Residue-level structure of Adnectin IBs from qHDX analysis.	107
Figure E.1: Considerations for complementary point-mutations.	112
Figure E.2: ¹ H- ¹⁵ N HSQC assignment transfers using intrinsic exchange rate constants and correlated peak movements.	113
Figure E.3: Measured intrinsic rate constants (K_{int}) of amide exchange for different amino acids in Adnectin monomers unfolded in 95% DMSO/5% DCA.....	114
Figure E.4:Residue-level structure of Adnectin IBs from qHDX analysis.	115
Figure E.5: Extent of overall amide exchange for Adnectin IBs in D2O for 1 hour as determined by 1D ¹ H-NMR spectra.....	116

List of Tables

Table 2.1: Secondary structure of native and aggregated SOD1 as determined by non-linear curve fitting.	24
Table 2.2: Percentage of full-length Adnectin remaining after 30 minutes of PK digestion.	31
Table 2.3: Equilibrium denaturation parameters of Adnectin IBs.....	32
Table 3.1: Comparison of average deviations in chemical shift for each atom type for proteins unfolded in DMSO.	49
Table 3.2: Residues exhibiting deviations from average chemical shift for C α /C β carbons.	49
Table 4.1 Sequence conservation of hydrophobic core residues for FN ₃ domain.....	67
Table 4.2: Equilibrium denaturation parameters of parent and engineered Adnectin IBs.	75
Table A.1 Aggregation Propensity of Adnectins grown at 37 °C in different growth media.	95
Table A.2: Adnectin point-mutant aggregation propensity prediction scores.....	95
Table B.1: Non-linear curve fitting results for SOD1 IB spectrum using variable numbers of components.	102
Table C.1: Change in monomer stability relative to parent Adnectin upon mutation as predicted by FoldX ¹⁵¹	103
Table D.1: Resonance assignments of 6199_CO1 Adnectin (FG Loop: DYRDYC) in 95%.....	107
DMSO/ 5% H ₂ O at pH 5.5.....	107
Table D.2: Rate constants (k _{ex}) of amide exchange in D ₂ O for Adnectin IB aggregates fit to a double exponential decay in OriginPro 2018.	110
Table E.1: Change in Adnectin monomer thermostability ($\Delta\Delta G$) upon point-mutations calculated using computational prediction tools.....	117
Table E.2: Change in Adnectin W22G monomer thermostability ($\Delta\Delta G$) upon a second point-mutation calculated using computational prediction tools.	119

List of Abbreviations

1D	One-Dimensional
2D	Two-Dimensional
3D	Three-Dimensional
A β	Amyloid-Beta
AD	Alzheimer's Disease
AFM	Atomic Force Microscopy
ALS	Amyotrophic Lateral Sclerosis
ANS	1-anilino-8-naphthalene sulfonate
APR	Aggregation-Prone Regions
ATR	Attenuated Total Reflectance
β 2m	β -2 Microglobulin
CD	Circular Dichroism
CDD	Conserved Domain Database
CR	Congo Red
DCA	Dichloroacetic Acid
DTT	Dithiothreitol
DMSO	Dimethyl Sulfoxide
EM	Electron Microscopy
FN	Fibronectin
¹⁰ FN ₃	10th Fibronectin Type 3 Domain
FTIR	Fourier Transform Infrared
GnHCl	Guanidinium Hydrochloride
HD	Huntington's Disease
HPLC	High Performance Liquid Chromatography
HSQC	Heteronuclear Single Quantum Correlation
IB	Inclusion Body
IPTG	Isopropylthio- β -galactoside
LB	Luria Broth/Lysogeny Broth
MD	Molecular Dynamics
MS	Mass Spectrometry
NMR	Nuclear Magnetic Resonance
NOG	N-octyl Glucoside
PD	Parkinson's Disease
PI-SH3	Bovine Phosphatidylinositol-3'-kinase SH3 domain
PMSF	Phenylmethyl Sulfonyl Fluoride
SOD1	Cu, Zn-Superoxide Dismutase
ssNMR	Solid State Nuclear Magnetic Resonance
TEN	Tris, EDTA, NaCl
ThT	Thioflavin T
qHDX	Quenched Hydrogen-Deuterium Amide Exchange
WT	Wild-Type

Chapter 1

Introduction.

Statement of contribution.

Sections 1.2 and 1.3 are part of an invited review article published in the Journal of Biological Chemistry co-written with Bruna Seibeneichler and Dalia Naser ¹. Literature searching and manuscript editing were performed collaboratively, however, Anna Schaefer specifically drafted these introductory sections.

1.1 Protein folding and misfolding.

Proteins are one of the four major classes of biological macromolecules essential for all living species. Proteins are synthesized by sequentially linking specific combinations of amino acid monomers, many of which spontaneously fold to a characteristic three-dimensional (3D) structure ². At a basic level, protein folding is governed by favorable, non-covalent interactions that lead to a stable 3D structure; however, understanding exactly how and why a protein establishes the correct set of interactions is multilayered. In a famous set of experiments Christian Anfinsen and coworkers demonstrated that the small protein ribonuclease is able to fold rapidly and independently ³. This led Anfinsen *et al.* to propose their thermodynamic hypothesis which states: 1) the native structure of a protein is that with the lowest possible Gibbs free energy (Fig. 1.1A) and 2) this structure is ultimately determined by only the primary sequence of amino acids and therefore is pathway-independent ⁴. Also studying protein folding, Levinthal noted that due to the exponential possibilities for protein conformations, it is improbable to assume that the native structure is found so quickly by random sampling ^{5,6}. This led to the proposal that proteins achieve their native state progressing through a directed pathway of distinct intermediate states. Levinthal's model indicated that for such directed pathways, the final structure was dependent on initial conditions and therefore under kinetic control. This meant that the final folded structure was not

necessarily at a global energy minimum but rather could occupy a local energy minimum (Fig. 1.1B) ⁷.

From further experimental work, it became evident that the thermodynamic and kinetic views of protein folding are not mutually exclusive; therefore, a new model was proposed that aimed to integrate these observations ⁸. The newer and widely accepted model of protein folding describes the energy landscape as a funnel formed by converging kinetic pathways (Fig. 1.1C). These pathways progress through different intermediates but can lead to the same pathway-independent final structure. At each point along the pathway the protein can be viewed as an ensemble of partially folded states. A smooth funnel with no kinetic traps, at left in Figure 1.1C, is effective for describing small proteins that exhibit two-state folding kinetics. More complicated funnels with a rugged landscape (right in Fig. 1.1C) can describe the folding pathway of a protein with multi-state kinetics and distinct folding intermediates. The funnel model aligns with results produced by atomistic molecular dynamics simulations as well as experimental folding

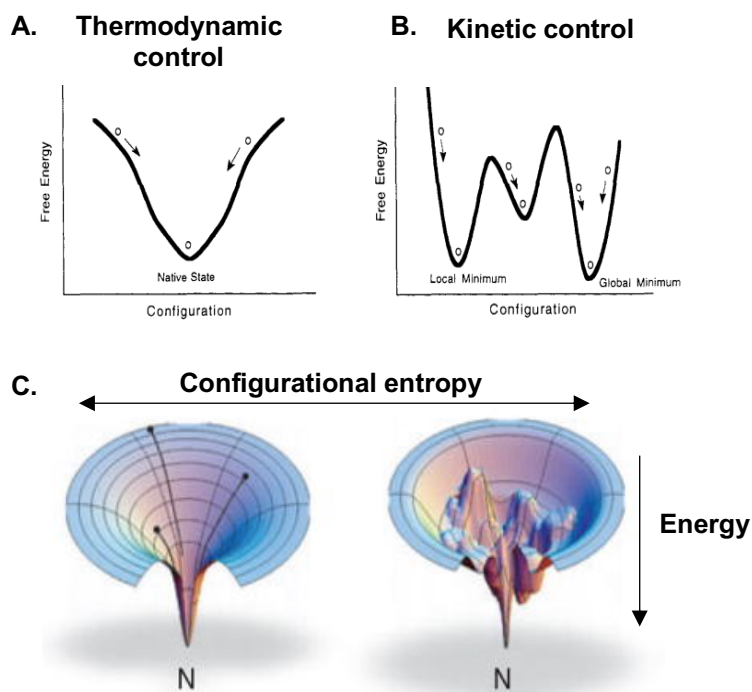


Figure 1.1: Energetic models of protein folding. A) Energy diagram of protein folding under thermodynamic control where native state is global energy minimum. B) Energy Diagram of protein folding under kinetic control where local energy minimum may be the native-state. C) Folding funnel for proteins exhibiting two-state folding (left) and for proteins exhibiting multi-state kinetics (right). Adapted from Baker and Agard, 1994 ⁷ and Dill and Chan, 1997 ⁸.

observations ^{9,10}. However, in addition to illustrating the validity of past theories, the energy landscape model effectively represents the complexity of the protein folding problem. Energy funnels have a vast number of possible shapes and a physical model to explain folding for a broad range of proteins remains elusive.

Misfolded proteins are a product of errors in the folding process. Long-lived misfolded protein species can be viewed as off-pathway folding intermediates, which are stabilized by the same non-covalent forces that control protein folding. The probability of protein misfolding can be increased by intrinsic properties such as primary sequence mutations or the complexity of the folding pathway ^{11,12}. Misfolding can also be influenced by external factors including failure of cellular chaperones, cellular crowding, lack of proper post-translational modifications, temperature, or other environmental conditions ¹¹. Cells have several interconnected systems to control the concentration of misfolded proteins; misfolded proteins can be dealt with through chaperone assisted refolding, sequestration, or proteasome-mediated degradation. However, if misfolded species are not properly dealt with, their major fate is aggregation ¹¹. Misfolded and other non-native protein species are particularly susceptible to aggregation due to solvent-exposed hydrophobic residues. This is a problem as it is more thermodynamically stable for hydrophobic amino acids to be buried away from polar solvent molecules ². Therefore, exposed residues in misfolded proteins seek to form intermolecular interactions when native intramolecular interactions are absent. The population of alternative protein conformations at equilibrium is related to the protein's thermodynamic stability and consequently, improving stability has long been associated with preventing aggregation ¹³. However, in recent years, several research groups have described that increasing global stability does not necessarily suffice to prevent aggregation, thereby indicating the possibility of native-state aggregation and opening up a new avenue of research ¹⁴.

1.2 Protein aggregation

Protein aggregation is a widespread phenomenon that is of great importance to many branches of life, medicine, and biotechnology. Numerous pathological conditions such as Alzheimer's Disease (AD), Parkinson's Disease (PD), Huntington's Disease (HD), amyotrophic

lateral sclerosis (ALS), diabetes, and cataracts are associated with the deposition of amyloid fibrils and other protein aggregates ^{15,16}. Moreover, aggregation may occur at several points during the expression and purification of recombinant proteins for biotechnological purposes. Due to their relevance, great effort has been put into determining the structural and biochemical properties of these aggregates under native conditions ¹⁷. Many of these studies have been performed with hope of selectively preventing or controlling aggregation, while others investigate how protein folding, maturation and environment can influence toxic aggregates. However, limitations such as structural heterogeneity, sample impurity and size constraints of structure determination techniques have obstructed a high-resolution understanding of protein aggregates.

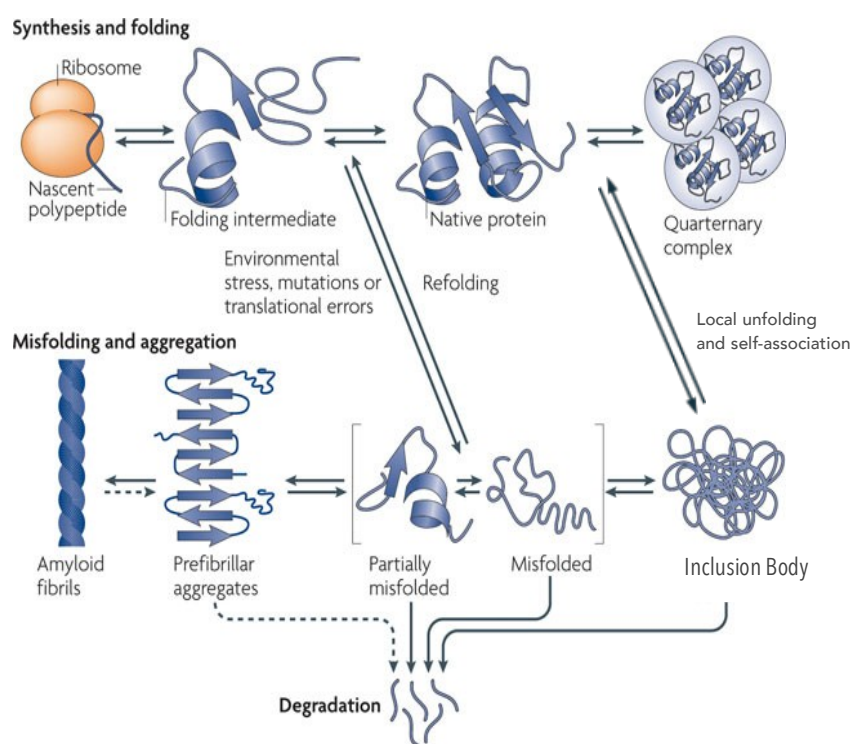


Figure 1.2: General scheme of cellular protein aggregation. Schematic diagram illustrating the multiple fates of translated protein. Figure illustrates how internal and environmental factors can trigger destabilization leading to misfolding and aggregation. Adapted from Tyedmers, Mogk, and Bukau, ¹⁸

In general, the aggregation process is thought to proceed by the association of two or more proteins subunits, followed by formation of small, soluble aggregates, then further assembly into larger insoluble structures (Fig. 1.2). However, the final structures and stabilities of aggregates are

variable. Historically, aggregate structures were classified into two structural categories: amyloid or amorphous aggregates. By definition, amyloid describes fibrillar aggregates with ordered cross β -sheet structure as identified using electron microscopy and X-ray diffraction ¹⁹. Mature amyloid fibers are often found in the characteristic plaques of protein deposition diseases and thus have been widely studied in terms of their structure, formation, and toxicity. This has contributed to an accepted macroscopic aggregation mechanism by which amyloid is produced via a nucleation-dependent polymerization. However, many theories exist for the exact molecular mechanism ²⁰. In contrast, large, insoluble deposits known as inclusion bodies (IBs) were categorized as amorphous when first discovered ^{21,22}. IBs most commonly occur when recombinant proteins are overexpressed in bacteria but have also been compared with inclusions found in neurodegenerative diseases such as ALS and HD. For a long time, they were considered deposits of misfolded and inactive proteins that posed an obstacle in recombinant protein production. However, recent studies have revealed that this is not strictly the case and IBs have gained much attention in the fields of biotechnology and medicine as self-immobilized catalysts and drug delivery systems ²³.

To date, descriptions of protein aggregates have become much more detailed and include several other distinct types such as oligomers, protofibrils, and secretory granules. Briefly, protofibrils are defined as initial, transiently formed aggregates that can lead to the formation of mature amyloid; most often they are reported to have antiparallel in-register β -sheet character. The common description of oligomer aggregates is quite similar to that of protofibrils ²⁴; however alternative structures such as β -helices have been reported ²⁵. Finally, secretory granules, also derived of β -sheets, are amyloid ensembles produced by the human endocrine system to sequester hormones for induced release.

1.3 Protein aggregate structure

Despite the importance of protein aggregates, current knowledge of their structure and mechanisms of formation is limited. Thus, control of their formation remains an inefficient, trial-and-error endeavour. Classical techniques for obtaining structural information about protein aggregates typically provide low-resolution data on whether a protein is aggregating and must be coupled with other methods to obtain a more detailed view. One of the simplest techniques to detect the presence of amyloid is using UV-visible absorbance or fluorescence spectroscopy to monitor

optical changes of the sample upon sequestration of dye by the amyloid fibril. Commonly used dyes include Congo red (CR), thioflavin T (ThT) and 1-anilino-8-naphthalene sulfonate (ANS). Upon binding amyloid aggregates, CR exhibits characteristic green–yellow birefringence under polarized light and a shift of the absorption maximum from ~490 nm to ~540 nm. Similarly, ThT and ANS exhibit increases in fluorescence when they bind to amyloid species ^{26,27}. While these dyes have proved useful as diagnostic tools, they are non-specific and as a result can bind to other cellular components; for example, CR binding to certain lipids and ThT to other protein hydrophobic pockets has been described ^{28,29}. ANS is even more limited because its binding to amyloid structure can induce protein conformational changes ²⁷.

Other characterization techniques, including protease digestion, atomic force microscopy (AFM) and electron microscopy (EM), report on global structure or morphology of aggregates. The degree of digestion with proteases can provide insight on the compactness of the aggregate structure ^{30,31} and taking this one step further, fragment sizes can be obtained using subsequent detection by a variety of methods, each providing different resolution. Approximate fragment sizes can be identified using SDS-PAGE or Western blotting while high performance liquid chromatography (HPLC) or mass spectrometry (MS) provide much more specific information on what regions of the protein are exposed in the aggregate ^{32–34}.

On a microscopic level, secondary structure determination methods such as circular dichroism (CD) spectroscopy and Fourier transform infrared (FTIR) spectroscopy can report on aggregates in many physical states ³⁵. While both are able to detect increased β -sheet content, a common feature of aggregating proteins, FTIR provides more detailed observations due to its higher extinction coefficient for β -structures and better fitness for turbid samples; while CD gives proportionally higher absorbance for α -helical structures, FTIR provides a more equal output allowing easier visual assessment of structural changes ^{36,37}. At the atomic level, methods such as quenched hydrogen deuterium exchange (qHDX) and solid state nuclear magnetic resonance (ssNMR) provide high resolution information on which residues are participating in the aggregate structure, allowing for a mechanistic interpretation of aggregate formation ^{38–43}. FTIR and qHDX will be emphasized further in Chapter 2 and 3 respectively as they provide a wealth of complementary and high-resolution information for Adnectin IBs.

There is clearly an abundance of techniques to study protein aggregation. Unfortunately, due to limitations caused by sample heterogeneity or impurity, many investigations have largely been limited to studying aggregation *in vitro* - under conditions that may be far different from physiological or pathological conditions. Furthermore, most of the methods applied *in vitro* for the biophysical and structural characterization of these species usually involve purification processes that can perturb the equilibrium of the aggregation process and ultimately may change the aggregate structure. Thus, it is critical that structure determination methods be adapted to study cell-derived aggregates both *ex vivo* and in whole cells or tissue.

1.4 Inclusion bodies: Structure and applications

Inclusion bodies (IBs) as described here refer to large, insoluble deposits of protein that result from the aggregation of protein in bacterial cells. The formation of these IBs occurs during protein overexpression, particularly when large amounts of recombinant protein are produced for biotechnological or pharmaceutical applications. Upon their discovery, bacterial IBs were thought to be large, unorganized structures formed by misfolded or unfolded proteins²³. This was challenged by FTIR results demonstrating native secondary structure elements in IBs formed by interleukin- β and corroborated by studies demonstrating that some IBs retain enzymatic activity and therefore native conformation^{44,45}. Further work revealed that bacterial IBs formed by β -galactosidase share several interesting properties with amyloid fibrils. These properties include binding to diagnostic dyes, the ability to seed self-aggregation, and an FTIR spectrum characteristic of an ordered β -structure different from the native structure of β -galactosidase⁴⁶. A seminal study by Wang *et al.* reported additional structural information for amyloid formation in IBs⁴⁰. Four small proteins were chosen based on known folding deficiencies and/or high aggregation propensities in *E. coli*. Purified IBs of each protein resembled amyloid proteins in low-resolution diagnostic tests including ThT binding and CR staining. Furthermore, with amide hydrogen deuterium exchange and complementary X-ray diffraction data, it was possible to determine structural patterns and therefore previously unattainable high-resolution structural information. While large segments of each protein appeared to be conformationally disordered, several distinct segments exhibited non-native cross- β structure⁴⁰. Additional research on IBs formed by superoxide dismutase 1 (SOD1) using experimental techniques such as ATR-FTIR, CR binding and NMR spectroscopy provided strong evidence that IB structure is much more complex;

bacterial IBs are composed of an ensemble of protein conformations including native-like, partially folded and amyloid-like fibrils ²³.

Protein aggregation occurs as a conserved response to overexpression and various types of cellular stress and therefore is not unique to bacterial cells ¹⁸. The sequestration of protein into inclusion-like aggregates also occurs in eukaryotic cells ⁴⁷. There is increasing desire to identify similarities and differences between different types of cellular aggregates, particularly to identify if bacterial IBs can serve as a model system for studying amyloid formation and other aggregation mechanisms that occur in neurodegenerative diseases. Currently there is some evidence that these aggregates are similar with several groups having used bacterial IBs to identify effective amyloid aggregation inhibitors ^{48–50} and other groups identifying that aggregates of yeast prion proteins Sup35 and Ure2 produced in *E. coli* were able to induce similar levels of prion conversion as homologous amyloid fibrils ⁵¹. Bacterial IBs are also of growing interest as biotherapeutics and in biotechnology. There is evidence that native protein is released from IBs simply during their incubation in buffer at physiological conditions ^{52,53}. Protein release can also be assisted or accelerated by biological chaperone proteins. This release property may be exploited for slow-release protein drugs and targeted drug-delivery systems ^{7,53}. Consequently, gaining an advanced knowledge of IB aggregate structures and aggregation mechanisms are of great interest to specifically tailor IB properties for many diverse applications.

1.5 Fibronectin type III domain

This thesis uses engineered, single-domain, immunoglobulin proteins based on a fibronectin type III scaffold as a model system for investigating protein aggregation. Fibronectin (FN) is a ubiquitous, multi-subunit, glycoprotein that serves several functions in the extracellular matrix. FN's overarching functions are in cellular communication and mechanoregulation; these broad definitions encompass specific functions in tissue repair, regulating cell attachment and motility, and in embryogenesis ⁵⁴. FN molecules are composed of three types of repeating domains, FN₁, FN₂, and FN₃ with variable β -structures and binding partners (Fig. 1.3A). FN₁ and FN₂ domains form β -pleated sheet structures stabilized by disulfide bonds. In contrast, FN₃ domains form an antiparallel β -barrel structure without stabilizing disulfide linkages making them more susceptible to structural changes caused by chemical or mechanical stimuli (Fig. 1.3B).

Fibrillogenesis involving FN-FN interactions is a main contributor to FN function and thus has been extensively studied. Fibrillogenesis is mediated by bound integrins which contribute to contractile forces that expose binding regions. Although, N-terminal FN₁ domains 1-5 are considered to be the main site of self-association in fibril assembly, many other binding locations in FN₃ domains have been identified⁵⁴⁻⁵⁶. For example, mechanical unfolding simulations of the 10th FN₃ domain containing the RGD cell binding sequence have identified an unfolded intermediate with exposed “SLLISWD” sequence that can mediate FN assembly. Knowledge of such self-association mechanisms is extremely advantageous for engineered extracellular matrices and biotherapeutics based on a fibronectin scaffold^{57,58}. Additional association mechanisms have been identified upon examination of the ordered units of crystals formed by Tencon, a consensus designed protein with the same fold as ¹⁰FN₃⁵⁹. These multimers involve β -strand G domain-

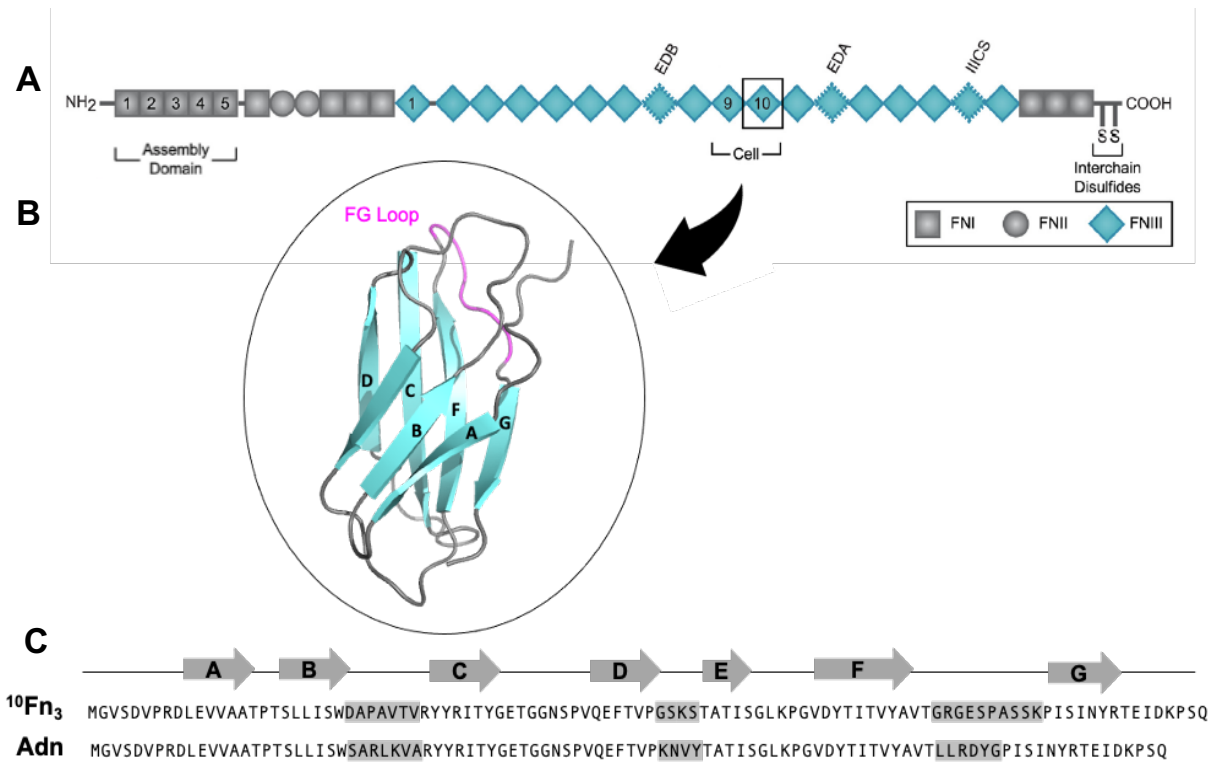


Figure 1.3: Human Fibronectin domain structure and sequence. A) Human fibronectin multidomain organization with domain types denoted by different shapes. The ninth and tenth type III domains are responsible for cell binding. B) Homology model of an engineered 10th fibronectin type III (¹⁰FN₃) domain created by Julia Steckner using SwissModel⁶⁰. Structures rendered using PyMol⁶¹. C) Sequence comparison of wild-type ¹⁰FN₃ and parent Adnectin with engineered binding loops highlighted in gray. Figure adapted from Gee *et al.*⁵⁵

swapping resulting in hexamers and higher order multimeric structures (Fig 1.4). In particular, the large, left-handed, helical structure provides an interesting potential self-association mechanism as it is open-ended with the possibility of forming very long strand-swapped chains (Fig 1.4B). Multimeric structures have also been observed experimentally through size-exclusion chromatography providing further support for this plausible aggregation mechanism for proteins with a fibronectin scaffold ⁵⁹.

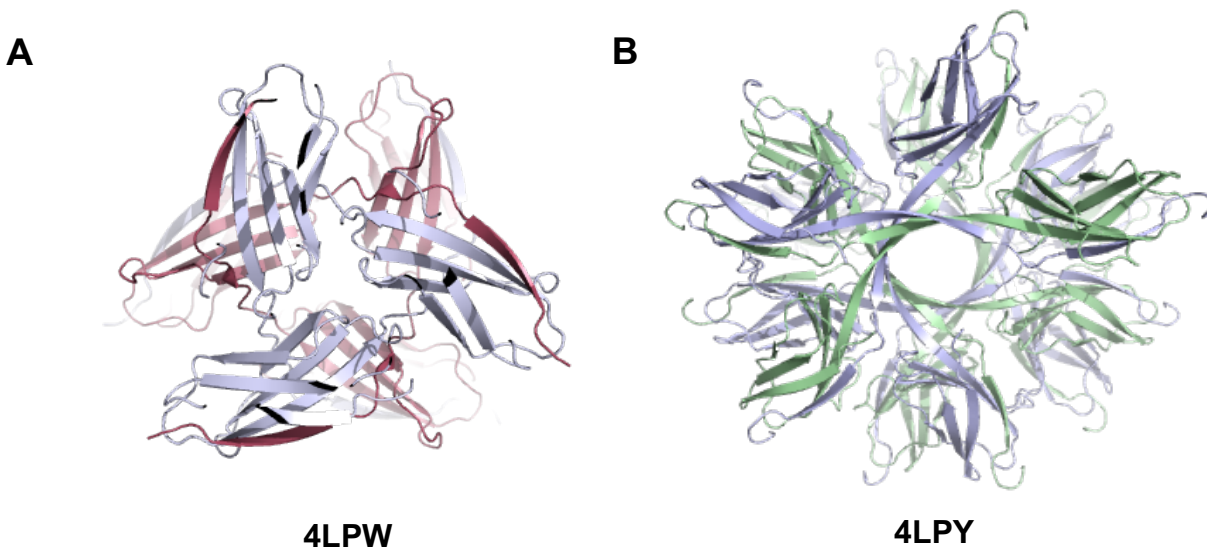


Figure 1.4: Tencon multimers formed by the association of β -strand G domain-swapped dimers. Hexamers (A) and helical multimer (B) shown with PDB identifier underneath. Figure reproduced from Teplyakov *et al* ⁵⁹.

1.5.1 Adnectins

Adnectins, also known as monobodies, are a family of human ¹⁰FN₃-derived proteins that bind with high affinity and specificity to a wide variety of targets making them prime candidates for use as protein drugs ^{62,63}. ¹⁰FN₃ proteins have β -sandwich immunoglobulin-like structure with three loops analogous to complementarity determining regions 1-3 of antibody variable domains (Fig. 1.3B). Adnectin proteins have been engineered for various biotherapeutic applications; examples include binders of Vascular Endothelial Growth Factor Receptor (VEGFR) and Proprotein convertase subtilisin kexin-9 (PCSK9) produced in *Escherichia coli* shown to decrease tumour formation and low-density lipoprotein concentrations respectively ^{64,65}. As these proteins

are being developed as protein drugs to be produced in bacteria, there is concern about their aggregation during production, storage and human administration. Therefore, studies of their stability and solubility are of great interest.

As described above, there is mounting evidence that local unfolding of beta strands in the Fibronectin fold plays a critical role in their self-association and aggregate formation. In particular, instability of the G-strand has been implicated in causing a non-native aggregation prone conformation^{59,66}. Recently, an Adnectin library was used by the Meiering group as a model system to quantitate and model primary sequence and structural effects on IB formation^{67,68}. During these studies, Adnectins were subject to several similar techniques as the ones discussed in this chapter including computational solubility prediction, mutational studies, HDX NMR, and extensive molecular modelling simulations. Combined results support the predictions that local conformational changes in and around the FG loop promote aggregation. Specifically, HDX NMR on monomeric Adnectins reveals that beta-strand G is significantly less protected than its hydrogen-bonding partner, strand F, in aggregation-prone variants. This indicates that a non-native conformation where strand G is not protected may be formed during Adnectin aggregation similar to what has been shown for another immunoglobulin-fold protein β -2 microglobulin (β 2M)⁶⁹.

Existing knowledge of Adnectin variant aggregation propensity, amenability of Adnectins to NMR analysis owing to their small size, and the availability of previously proposed determinants of aggregation make Adnectins a powerful model system for unravelling specific molecular determinants of IB formation. Specifically, Adnectin variants may more broadly inform on the determinants of aggregation of proteins with immunoglobulin folds, including antibodies which are a rapidly growing group of novel drug candidates^{63,70,71}.

1.6 Research objectives

The focus of this thesis will be an in-depth structural analysis of cell-derived Adnectin inclusion bodies using a subset of the techniques described above. The Adnectin variants studied herein have a very similar sequence to wild-type $^{10}\text{FN}_3$ excluding the loop regions (Figure 1.3C) Chapter 2 will explore low-resolution methods such as chemical denaturation, protease digestion and ATR-FTIR to investigate relative stability, morphology and secondary structure respectively.

A set of point mutants for analysis were selected based on previously obtained solubility data and aggregation propensity predictions ^{67,68}. Chapter 3 will investigate the structure of a subset of these mutants at a single amino acid resolution using qHDX and address important considerations when applying this method to a new protein system. Lastly, using results from the previous chapters and a variety of computational prediction software, new Adnectin mutants were designed to test specific aggregation mechanism hypotheses. The rationale behind their design and a detailed structural analysis are provided in Chapter 4. By comparing exchange data for IBs with the biophysical properties of the expressed variant proteins, aggregate processing conditions, and aggregate maturation time, important molecular features of the aggregation process have been identified. Ultimately this provides critical new insights into the molecular mechanisms governing aggregate formation and could lead to a better understanding of how to control protein aggregation, especially in the case of inducing or preventing IB formation for biotechnological applications.

Chapter 2

Biophysical characterization of Adnectin Inclusion bodies.

Statement of contribution

Some FTIR data in this chapter (Table 2.1, Figure 2.4) have been published previously ⁴¹. These data were acquired in collaboration with Bruna Siebeneichler. Data analysis and interpretation were performed jointly with Dalia Naser and Bruna Siebeneichler. SDS-page gels of SOD1 IB samples were performed by Dalia Naser and Gyana Gourab Mishra. Subsequent SDS-page gel densitometry was performed by Anna Schaefer. Descriptions of FTIR included in sections 2.1.3 and 2.3.2 were modified from Schaefer *et al.* ¹. Initial selection and characterization of the Adnectin mutants analyzed in this chapter were performed by Kyle Trainor. FTIR data from section 2.3.2 was acquired in collaboration with Xiaoyue Liu as part of an undergraduate research project. Purified Adn included for comparison was prepared by Julia Steckner and Kennedy Weir.

2.1 Introduction

2.1.1 Context

The current view is that IBs are composed of an ensemble of protein conformations including partially folded, unfolded and near-native. This chapter will first explore the effects of mutation on predicted aggregation behavior. Then, these predictions will be evaluated through an investigation of the ensemble of structures within IBs formed by Adnectin mutants of interest. Through analysis with low resolution methods including FTIR, chemical denaturation, protease digestion, and CR binding, key structural elements and relative stabilities have been identified. Further, comparison to purified protein, and known amyloid species allows an estimation of the relative amount of native structure present in the IB, which is a parameter critical for biotechnological relevance.

2.1.2 Point-mutant selection

Aggregation propensities of proteins with similar sequences can vary greatly ⁴¹. Therefore a structural study of Adnectin point mutants is valuable for understanding their aggregation mechanism. A subset of mutants for study were selected based on previously determined aggregation and structural data ⁶⁷ and were further examined by SDS-PAGE to determine their aggregation propensity under a set of defined conditions (Table A.1). Regions of interest for Adnectin aggregation include β -strands F and G as well as the intervening loop (FG loop) (Fig 2.1); this region is computationally predicted to very highly aggregation-prone (Fig. A.1), and demonstrates evidence of conformational rearrangement through hydrogen deuterium exchange performed on purified protein ⁷². Wild-type (WT), referred to henceforth as parent Adnectin, displays asymmetric amide protection different than what would be seen for an interaction between native β -strands. The proposed biological explanation for this observation was that β -strand F participates in protective hydrogen bonds while β -strand G is able to access an exchange competent state. In contrast, a solubility enhanced point-mutant, V75R, displays symmetric protection and associated H-bonding of the F and G strands, much like what would be expected for the native ¹⁰FN₃ structure. Therefore for parent Adnectin, reorganization of hydrogen-bonding may be promoted by self-association. Both parent and V75R Adnectin will be analyzed herein. As a point of comparison, L78I, a more aggregation-prone mutant ⁶⁷, will also be analyzed.

To further investigate the relevance of β -strand G dissociation to Adnectin IB formation and structure, additional point-mutants were designed. The mutation Y88F was originally designed in an attempt to anchor strand G by inserting the side chain into the core of the protein. While this did not have the predicted effect, when combined with a L18V mutation, the double mutant L18V/Y88F was successful in increasing solubility. It was hypothesized that L18V provided additional space for the phenylalanine residue to pack, however, L18A did not have the same effect. These mutants (L18V/Y88F, L18A/Y88F) as well as the single mutant L18V will also be analyzed in this chapter. Locations of all previously discussed mutations are shown on a ribbon structure for convenience (Fig. 2.1).

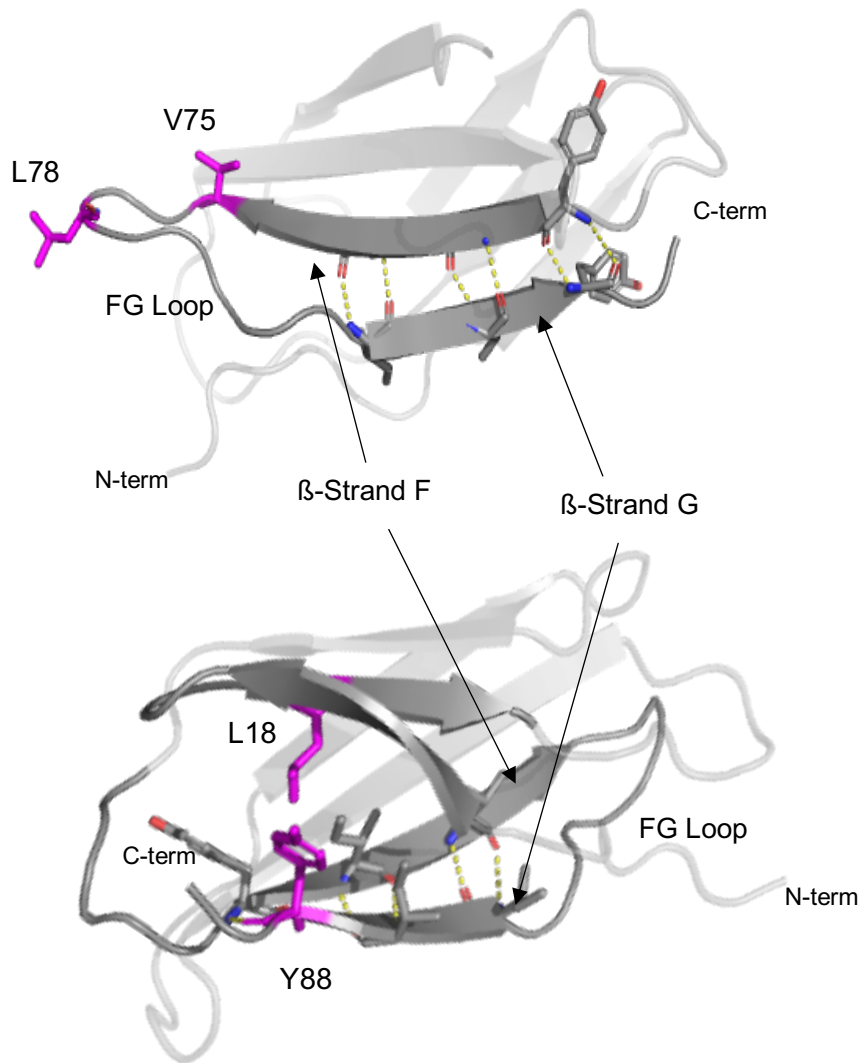


Figure 2.1 Ribbon diagram depicting location of Adnectin point mutations. Adnectin homology model showing backbone-backbone hydrogen bonds joining β -strands F and G in native $^{10}\text{FN}_3$ and residues of interest colored magenta (Top: V75 and L78 and bottom: L18 and Y88). Hydrogen bonds between β -strands F and G are indicated by dashed yellow lines. All atoms are shown for residues Y68, I70, V72, V75, I88, I90, and Y92. Carbon atoms are colored to match the cartoon backbone, nitrogens are blue, oxygens are red, and hydrogens are white.

2.1.3 Fourier Transform Infrared Spectroscopy

FTIR is a rapid and sensitive means of monitoring protein secondary structure and performing compositional analysis for biological samples. FTIR spectra report on covalent bonds which absorb in the infrared region and result in a unique fingerprint for any molecular mixture. Scientists studying protein conformational changes in the context of protein dynamics, misfolding

and aggregation rely specifically on the carbonyl (C=O) stretching vibration of amide bonds which appears in the 1600-1700 cm^{-1} range, termed the amide I region. This range has characteristic frequency ranges for α -helical, β -strands, random coil, loop and turn structures due to unique hydrogen-bonding patterns of the protein backbone, and is largely independent of primary sequence allowing facile structure comparison between proteins ^{35,73}. As extensive FTIR studies have been performed on protein aggregates, a wealth of knowledge is available about their spectral characteristics which can assist current researchers with their conclusions. Additional regions, termed amide II and III, exist at 1500-1600 cm^{-1} region and 1175-1310 cm^{-1} regions respectively. These regions are infrequently used for structure determination due to their lower intensity and overlap with absorbance bands of amino acid side chains in the case of amide II.

Protein aggregates have been extensively studied with FTIR as this method circumvents the typical limitations of structural analysis techniques: it has short sample measurement times allowing the observation of time dependent changes, it requires relatively low sample amount (<100 ng), and is applicable to insoluble and impure samples ⁷³. Although both transmission and attenuated total reflectance (ATR) FTIR have been used to study aggregates, ATR-FTIR has become the preferred method because there is little to no sample preparation required. IR spectra of proteins associated with neurodegenerative disease such as A β (Alzheimer's), α -synuclein (Parkinson's), and many others have been acquired for fibrils produced *in vitro* ⁷⁴. These studies have provided a strong foundation for studies of cellular aggregates. Key spectral characteristics include a shift in the β band towards lower wavenumber (1611 cm^{-1} to 1630 cm^{-1}) in fibrils, allowing differentiation of native and amyloid-like intermolecular β -sheets ⁴⁵. Additionally, meaning has been extracted from the broad range of wavenumbers observed for intermolecular β -sheets; an FTIR study on synthetic amyloid beta revealed the size of oligomers correlated with the FTIR absorbance with larger oligomers having lower wavenumbers ⁴⁵. Others have distinguished between parallel and antiparallel- β -sheet shift assignments based on the appearance of a higher order β peak above 1685 cm^{-1} in antiparallel aggregates ⁷⁵⁻⁷⁷. All the above pieces of information are unique to analysis by FTIR and are very valuable for the interpretation of Adnectin IB secondary structure.

2.1.4 Methods adapted to investigate aggregate stability

Chaotropic agents such as urea or guanidium hydrochloride (GnHCl) have been used to denature or solubilize IBs ^{34,78}. The exact molecular mechanism of these agents is still unclear, however, proposed modes of action include screening of intramolecular hydrogen bonds and by making the exposure of hydrophobic residues more favorable ⁷⁹. Urea denaturation analysis can be useful for determining relative stability of aggregates formed by different proteins, different mechanisms (ie. thermal aggregates or cell-derived) or assessing the effect of a point-mutation. Through comparison with purified protein and other studies on amyloid-forming proteins, these studies can help characterize the ensemble of structures participating in an IB. Protease digestion has been used similarly to report on the stability as well as the global structure or morphology of aggregates. As mentioned previously, the degree of digestion with proteases can provide insight on the compactness of the aggregate structure as well as locate regions that are protected from digestion ^{30–33}. Furthermore, the concept of a proteinase-resistant core is often attributed to amyloid structures with digestion-resistant fragments often having fibril-like morphology and the ability to bind amyloid-specific dyes such as ThT ³³. Therefore, a combination of denaturation and digestion analyses were conducted to support conclusions made using FTIR data.

2.2 Methods

2.2.1 Cell growth and protein expression.

Adnectin variants were expressed in BL21(DE3) pLysS cells containing a pET-9d vector and grown overnight at 37 °C with shaking at 225 RPM in 50 mL of Luria broth (LB) containing 50 µg/mL Kanamycin and 34 µg/mL chloramphenicol. Cells from 30 mL of the overnight culture were collected by centrifugation at 4000g at 4 °C for 10 minutes. The pellet was resuspended in 10 mL of the supernatant and used to inoculate 1 L of sterile M9 minimal media (6 g/L Na₂HPO₄, 3 g/L KH₂PO₄, 0.5g/L NaCl, 2 mM MgSO₄, 0.1 mM CaCl₂, 4 g/L glucose, and 0.5 g/L ¹⁵NH₄Cl) containing the same concentration of antibiotics. Flasks were incubated at 37 °C with shaking at 225 RPM to an A600 of 0.6-0.8. Expression was induced using 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). To harvest, cells were pelleted by centrifugation at 4000g at 4 °C

for 15 minutes, resuspended in 30 mL of TEN buffer (20 mM Tris pH 8.0, 1 mM EDTA, 100 mM NaCl) and flash frozen in liquid nitrogen. Cell pellets were stored at -80 °C until further use.

In addition to Adnectins, FTIR was performed to analyze mutants of SOD1 and determine appropriate secondary structure determination methods. Protein growth and IB preparation was performed by Dalia Naser and Michael Tarasca. FTIR data acquisition was performed by Buna Siebeneichler. SOD1 variants were expressed in E. coli BL21(DE3) pLysS cells containing a pET-21 vector. Cells were grown as described for Adnectin with the following changes: selection was achieved using 100 µg mL⁻¹ Ampicillin and 30 µg mL⁻¹ chloramphenicol, expression was induced using 0.1 mM IPTG and cells were grown for 4 hours following induction ⁴¹.

2.2.2 Inclusion body (IB) preparation

For cell lysis, cell pellets in TEN buffer were subjected to four freeze-thaw cycles using liquid nitrogen and a 37°C water bath. After the first freeze-thaw cycle cell lysates were incubated with 0.1 mg/mL DNase at room temperature for 25 minutes. After the final cycle, insoluble material was collected by centrifugation at 16,300g for 15 minutes (room temperature). Insoluble material from 1L of cell culture was resuspended in 8 mL of TEN buffer to obtain a homogenous solution. Aliquots of this solution were stored at -80°C.

2.2.3 Solubility analysis

Samples were grown as described above in a total culture volume of 50 mL. Duplicate 1-mL samples were taken at 4 and 24 hours post-induction (T₄, and T_{ON}, respectively). The 1 mL samples were centrifuged at 8000 rpm for 2 minutes and the pellet was resuspended in 30 µL of TEN buffer. The cell lysis procedure was performed as described above. The soluble and insoluble fractions were analyzed using 15% SDS-PAGE gel. The percentage (%) of protein in IB was the ratio of the pellet band over the sum of both the pellet and supernatant band was calculated using Eq 2.1.

$$\text{insoluble}/(\text{soluble} + \text{insoluble}) \times 100\%$$

Eq. 2.1

2.2.4 FTIR data acquisition

Attenuated total reflectance Fourier transform infrared spectroscopy (ATR-FTIR) was performed following established recommendations^{80,81} using a Tensor 37 FTIR spectrometer (Bruker Optics, Germany) continuously purged with N₂ gas, equipped with a LN-MCT (mercury-cadmium-telluride) detector cooled with liquid N₂. The sample cell is a Bio-ATR II cell with a ZnSe crystal. For all measurements, four spectra were acquired consecutively with 256 scans each with spectral range of 4000-1200 cm⁻¹ and 4 cm⁻¹. For each protein sample, the following series was acquired. A measurement of the background (air) and buffer (TEN) were collected. Next, the Adnectin IB sample (prepared as above) was measured. The crystal was then flushed thoroughly with buffer before collecting post-sample buffer spectra. As recommended, due to clear evidence of sample adsorbed on the crystal despite the extensive flushing, the post-sample buffer was used for liquid water subtraction from the sample spectrum. For each mutant, six independent spectra were measured, comprising two biological replicates with three technical replicates each. Measurements were made at 25°C controlled with a Thermo HAAKDC water bath (Haake).

2.2.5 FTIR spectral analysis

All data processing and analyses for IB and pure protein samples were performed using the fourth spectrum to ensure thermal equilibrium. Initial processing and analyses of FTIR spectra used the OPUS 6.5 software (Bruker) to perform atmospheric compensation and buffer subtraction. Next, spectra were baseline-corrected by subtracting a straight line fit between 1700 cm⁻¹-1600 cm⁻¹, followed by normalizing the area of the amide I band to 1, using OriginPro 2018. The final average spectrum of each Adnectin mutant was obtained by averaging the six replicates.

Component peaks in the absorbance spectra were determined by multiple peak fitting in OriginPro 2018. Peaks were fit to a convolution of Lorentzian and Gaussian curve profiles. For a given spectrum, initial peak position estimates for the fitting were obtained based on the corresponding second derivative spectrum. Quality of fit was evaluated based on Chi-squared (χ^2).

2.2.6 Proteinase K digestion

Adnectin inclusion bodies obtained after different time points of induction were subjected to proteolytic digestion with 0.1 mg/mL proteinase K. IBs were diluted 1/10 with 20 mM Tris-HCl,

pH 8.1 (final volume 200 μ L). Digestion was inhibited with 1 mM phenylmethylsulfonyl fluoride (PMSF) at specified time points. Fragment sizes were identified using Tris-Tricine SDS-PAGE gels with 4% stacking and 16 % resolving gel containing 6M urea.

2.2.8 Congo red binding

A 300 μ M stock solution of Congo red (CR) was prepared in TEN buffer with 10% ethanol and filtered 3 times through a 0.2 μ m Acrodisc filter. CR-IB assays were conducted using a protocol modified from Klunk et al.⁸². All measurements were made using a Cary 4000 spectrophotometer (300 nm - 700 nm) blanked with TEN buffer. For each sample, a dilution of the IB was made to obtain an absorbance spectrum below 1, bringing IB concentrations to an average of $\sim$ 2 mg/mL. As controls, 20 μ M CR and diluted IB were measured separately before each binding experiment. Each diluted IB was combined with 300 μ M CR for a final concentration of 20 μ M CR and was incubated for 10 mins before being scanned 3 times. Absorbance was unchanged when remeasured after 15 minutes. Differences plotted correspond to the IB+CR solution with diluted IB and blank CR subtracted.

$$\text{CR bound protein} = \text{IB+CR} - (\text{diluted IB}) - (\text{CR}) \quad \text{Eq. 2.1}$$

2.2.7 Chemical solubilization and denaturation

Adnectin inclusion body samples were diluted 10 times in 20 mM tris, pH 8.1 with urea concentrations ranging from 0.5 – 8 M. Dissolving of the IB was measured via turbidity of the solution by monitoring optical density at 350 nm, OD₃₅₀, on a SpectraMax Plus 384 microplate reader (Molecular Devices). IBs were monitored for 3 hours, which was considered to be the point where no further changes were observed. Urea solubility profiles for each type of IBs were obtained by plotting fractional change in OD₃₅₀ at 3 hours post-dissolution against urea concentration. Curves were fit using a two-state denaturation model described by Eq 2.2, where C_{mid} is the midpoint of denaturation, m is the [urea]-dependence of denaturation, Y is the optical signal of the native (N) or unfolded (U) states, while S is the [urea]-dependence of each of those states respectively. Denaturation data were obtained as a point of comparison for purified parent Adnectin in 40 mM sodium citrate at pH 4.0. The protein stocks were diluted to a concentration of 20 μ M with denaturant in 40 mM sodium citrate to reach the final urea concentration spanning

from 0 M to ~8M. These samples were equilibrated at 25 °C for at least 1 day. Unfolding was quantified by measuring tyrosine fluorescence at 307 nm with excitation at 277 nm on a Fluorolog3-11 spectrofluorometer (Horiba, Kyoto, Japan).

$$y = \frac{(Y_N + S_N[urea]) - ((Y_N + S_N[urea]) - (Y_U + S_U[urea])) * (e^{\frac{-(C_{mid} * m) + m[urea]}{RT}})}{1 + e^{\frac{-(C_{mid} * m) + m[urea]}{RT}}} \quad \text{Eq. 2.2}$$

2.3 Results and Discussion

2.3.1 Adnectin IB formation

Previous aggregation propensity data for Adnectin mutants was obtained from cells grown in lysogeny broth (LB) media⁶⁷. It has been shown that growth media can largely affect the kinetics and extend of inclusion body formation in *E. coli* due to changes in cell size and cellular processes such as protein translation rate, availability of chaperones and other stress response proteins^{83–86}. However, in order to study IBs by NMR, they must be isotopically labelled and would need to be grown in minimal media. Therefore, for all experiments within this thesis, IB formation and structure are being evaluated for protein expressed in M9 minimal media so that a direct comparison can be made between all experiment types. Furthermore, aggregation propensities of the mutants studied herein were redetermined for cells grown in M9 minimal media (Fig. A.2, Table A.1) (See methods section 2.2.1-2.2.3). The correlation constant between aggregation propensity in minimal vs. rich media was determined to be 0.362 with a p-value of 0.206 indicating that this correlation is not statistically significant (Fig. 2.2). Therefore, while some general trends in aggregation propensity may be conserved, the apparent measured effect of a mutation relative to parent is not always consistent between growth conditions.

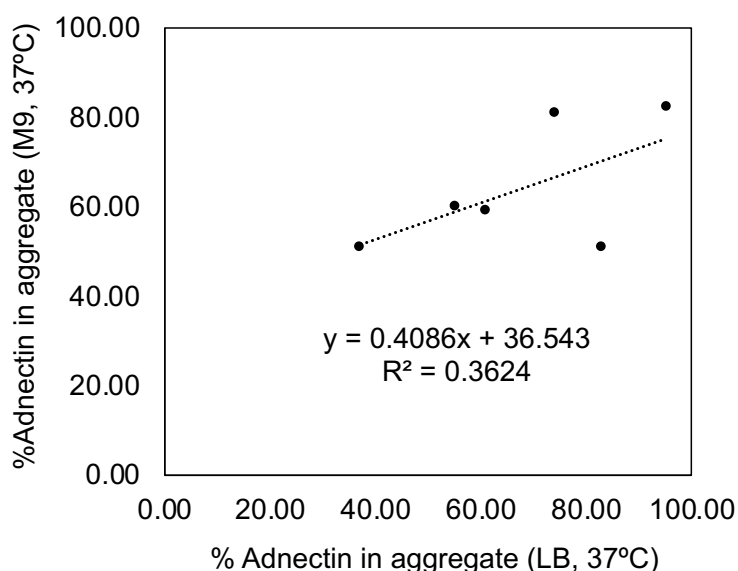


Figure 2.2: Comparison of average Adnectin in IB for different growth media types. Values determined by densitometry of SDS-PAGE analysis. % insoluble=insoluble/(soluble + insoluble) x 100%.

2.3.2 Approaches to FTIR data analysis

Despite unique maxima for different secondary structures in the amide regions, the wide bands and small spectral range cause significant overlap and result in a broad absorbance spectrum. Therefore, for a complete interpretation, additional information must be extracted; common ways to extract secondary structure information include using second derivative analysis, Fourier self-deconvolution, multivariate peak fitting, and cluster analysis⁸¹. While the methods noted above can be combined to provide absolute quantitation, it should be cautioned that peak fitting and Fourier self-deconvolution can be easily biased without proper instruction or expertise⁸¹. Furthermore, the Savitzky-Golay algorithm used throughout the analysis to remove artifacts caused by noise can also cause the loss of finer details. Therefore, the field generally relies on comparative analyses to avoid incorrect interpretation⁸⁷. A graphical description of FTIR data processing methods used in this chapter, is outlined in Figure 2.3.

As there is not an available crystal structure of the Adnectins studied here, previously acquired SOD1 data were used to determine an appropriate fitting procedure (Fig B.1)⁴¹. Following previously published protocols, the number of component bands and their positions in

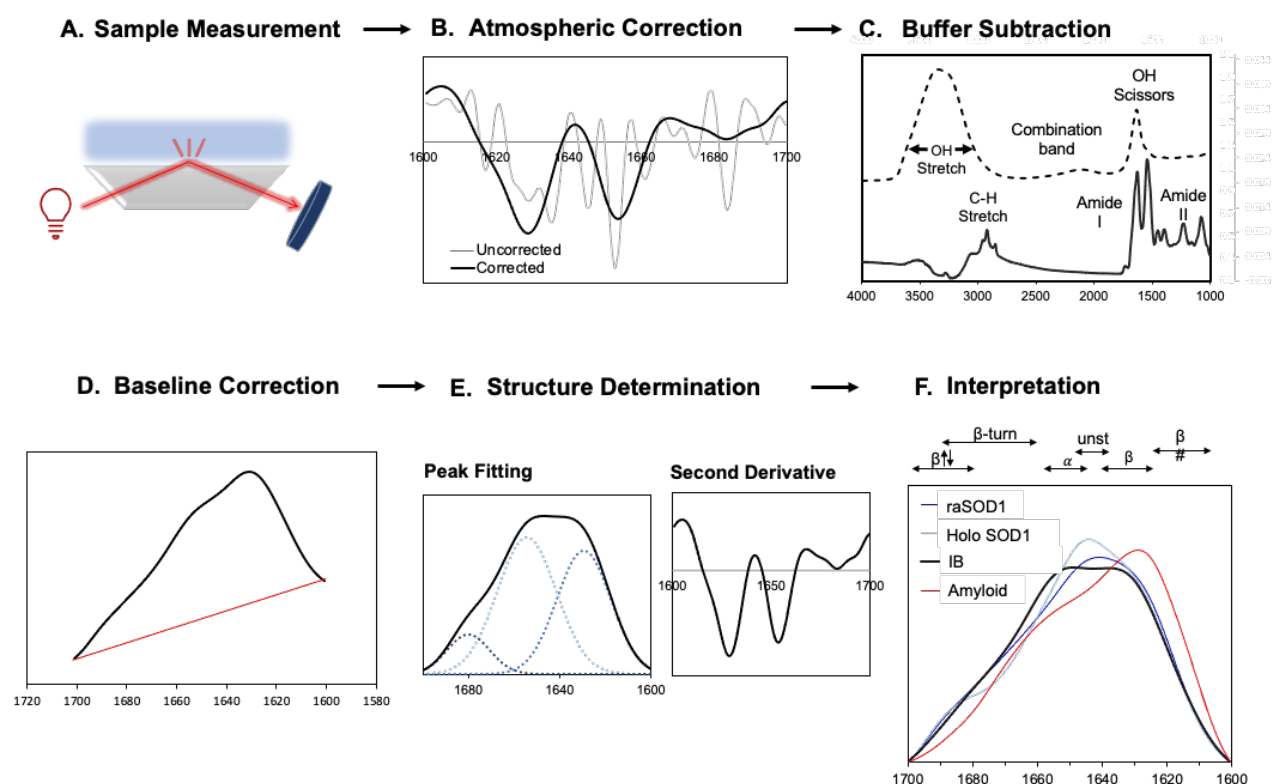


Figure 2.3: Data processing workflow for FTIR analysis of protein aggregates. Following sample measurement on ATR-FTIR, data is transformed to give an absorbance spectrum. This absorbance spectrum is sensitive to water and carbon dioxide in the atmosphere and therefore can be corrected using patented algorithms from acquisition and analysis software such as Bruker Opus. Following atmospheric correction, a post-buffer scan must be subtracted from the sample to remove contributions from water. A flat baseline can be applied to the amide I region (1600-1700 cm^{-1}) and then the resulting spectrum can be analyzed using structure determination methods including multivariate peak fitting, second derivative analysis, Fourier self-deconvolution and more⁸¹. Interpretation is often aided by comparison to purified or native FTIR spectra.¹

the spectra were based on the second derivatives of the corresponding absorbance spectra⁸¹. The secondary structures obtained from fitting the spectrum for holo pWT (Table 2.1) agree well with the X-ray structure (comprised of 34 % beta, 41 % disorder, 21 % turn and 4 % helix)⁸⁸, as was also observed previously by Dong et al.⁸⁹. Also, pure sonicated and heat-denatured SOD1 fits are consistent with expected structural differences compared to native SOD1: sonicated SOD1 shows increased beta structure with a shift to lower wavenumber, characteristic of strong intermolecular hydrogen bonding in aggregates and amyloid^{45,90}, and heat denatured SOD1 shows a large increase in disordered signal (Table 2.1).

Table 2.1: Secondary structure of native and aggregated SOD1 as determined by non-linear curve fitting.

SOD1 variant	Secondary structures				
	# β -sheet	β -sheet	unstructured/ helix	turn	Antiparallel β -sheet
average IB	-	40.5%	49.2%	-	10.3%
ra	-	36.1%	31.3%	28.8%	3.8%
holo	-	33.7%	35.4 %	24.6%	10.3%
sonicated	40.2%	-	50.6%	9.7%	-
unfolded	3.6%	22.8%	65.9%	7.7%	-

Values in the table correspond to integrated areas for fitted bands with absorbance maxima within the following ranges which were assigned to: Intermolecular (#) β -sheet: 1620-1625 cm^{-1} ; β -sheet: 1628-1630 cm^{-1} ; unstructured/helix: 1640-1660 cm^{-1} ; turns: 1660-1680 cm^{-1} ; antiparallel β -sheet, 1680-1690 cm^{-1} ^{36,73}.

IB curve fitting results are consistent with the conclusions made from comparing absorbance spectra that the IBs are composed of both native-like and disordered components. However, the presence of a small shoulder at 1615 cm^{-1} and a small band at ~ 1690 cm^{-1} in the second derivative spectra (Fig. B.2) beg the question of whether the IBs should be fit with additional peaks. To test this hypothesis, the average IB spectrum was also fit with 4 or 5 separate components using Origin Pro 2018 (Fig. 2.3). It should be noted that a curve may be fit to infinite components and yield a result, and that it is up to the user to establish what is a reasonable and unbiased fit. It was determined that these additional minor components do not affect the overall interpretation of secondary structure data. Furthermore, despite a small decrease in the adjusted χ^2 (Table B.1) additional components drastically increase the fitting error generated. Therefore, Adnectin IB data were fit using three components.

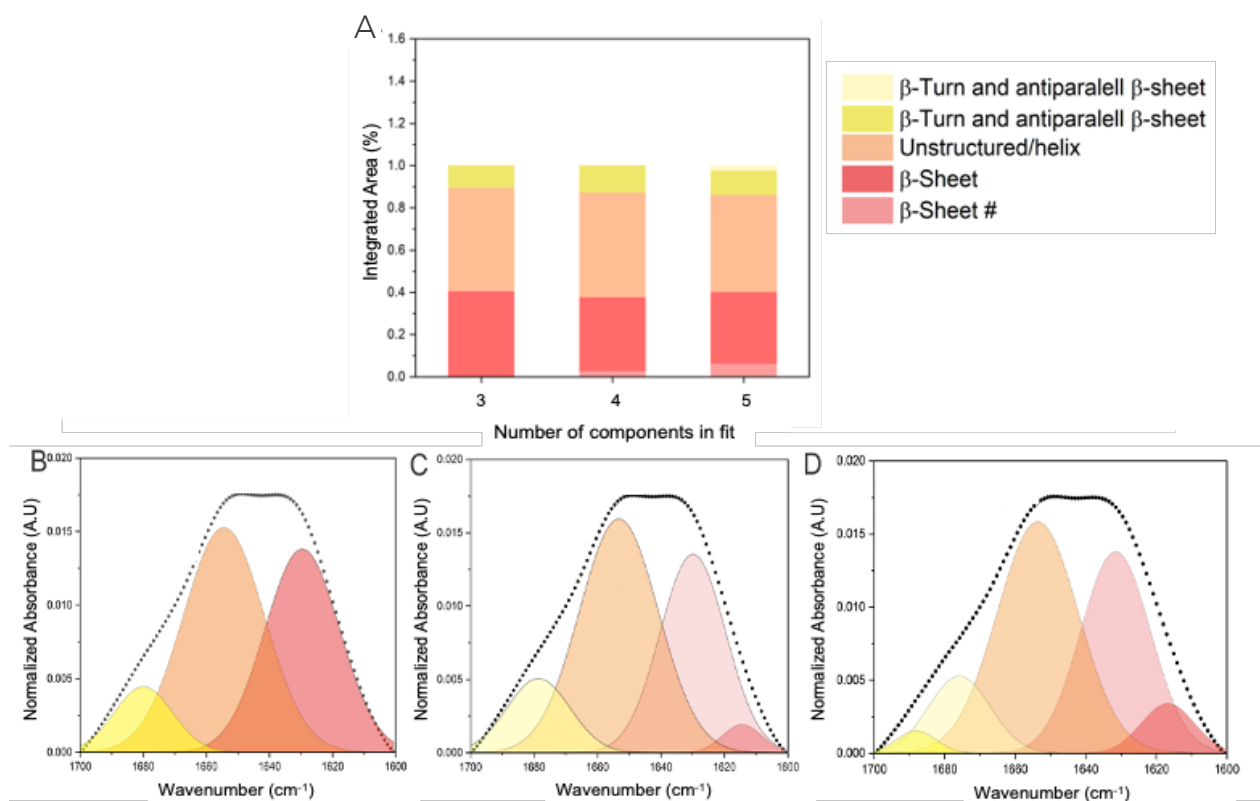


Figure 2.4: Comparison of non-linear curve fitting methods for average SOD1 IB spectrum. Upon comparing the results of fitting with 3, 4 or 5 components, shown in panels B, C and D, respectively. The fitting results show that additional components do not have a large effect on secondary structure interpretation.

2.3.3 Adnectin IBs contain significant native secondary structure elements

Given that FTIR measures total signal from the entire sample, it is important that the sample contains enough Adnectin to outweigh contributions from other contaminating proteins. Therefore, FTIR spectra were obtained for Adnectin IBs expressed overnight to analyze their secondary structure. Amide I absorbance profiles and integrated areas from curve fitting analyses reveal extensive similarities between Adnectin IBs and purified protein indicating that a high level of native-structure is retained in the aggregate (Fig 2.5). Although β -sheets exhibit a main absorbance at low wavenumbers, an additional, high wavenumber component, above 1680 cm⁻¹, can be attributed to anti-parallel β -sheets and turns^{75–77}. This is consistent with the large shoulder seen for purified parent Adnectin, a significant amount of which remains in the IB spectra (Fig 2.5). In contrast, a lack of signal in this region is seen for mature amyloid fibrils with known parallel

amyloid structures^{76,91,92}. Amyloid forming proteins also exhibit a right-shifted β -sheet signal to lower wavenumbers (1611 cm^{-1} - 1630 cm^{-1}) derived from short intermolecular hydrogen-bonds⁴⁵.

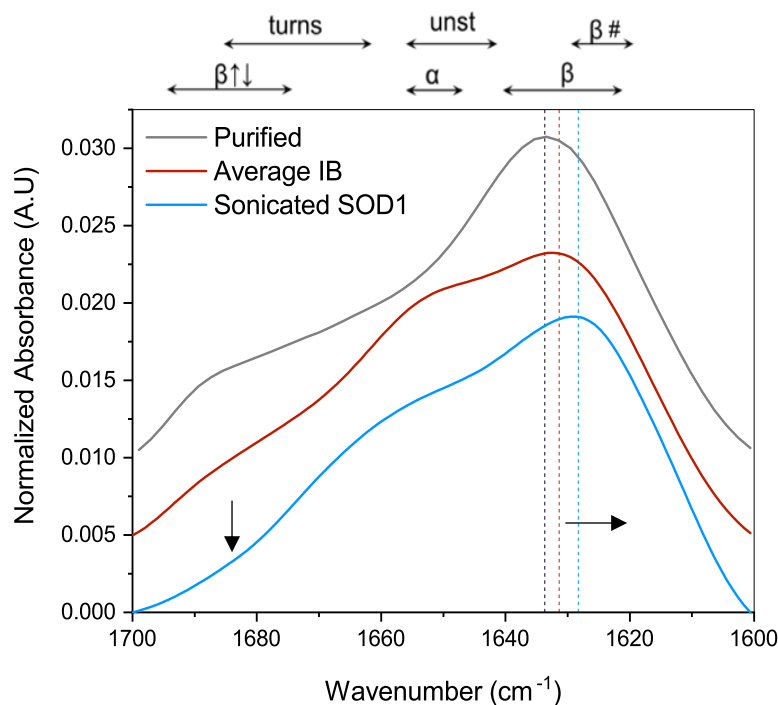


Figure 2.5: FTIR of mutant Adnectin IBs expressed overnight exhibit significant native features. ATR-FTIR spectra in the amide I region for Adnectin mutants, and purified parent Adnectin. Spectral regions for antiparallel β -sheet ($\beta\updownarrow$), β -turn, α -helix (α), unstructured (unst), β -sheet (β), and β -amyloid with strong intermolecular hydrogen bonds ($\beta\#$) are indicated. Spectra are staggered vertically for easier visualization. Dashed lines represent the center of main β -sheet peak for each spectrum. Arrows represent direction of spectral changes described in text.

While these two main markers of amyloid proteins are clearly demonstrated in the absorbance spectrum for sonicated SOD1⁹³, they are much less apparent in the average Adnectin IB spectra (Fig 2.5). Despite this, a small right-shift is observed for an average absorbance spectrum for Adnectin IBs compared a spectrum for purified protein, thus it is possible that there is a small portion of protein in the IB forming fibril-like structures. However, β 2M, a fibril-forming protein with a similar native-fold, exhibits a much larger shift upon fibril formation verified by AFM visualization⁹⁴ which further supports that it is unlikely that amyloid is a main component of the IBs studied here.

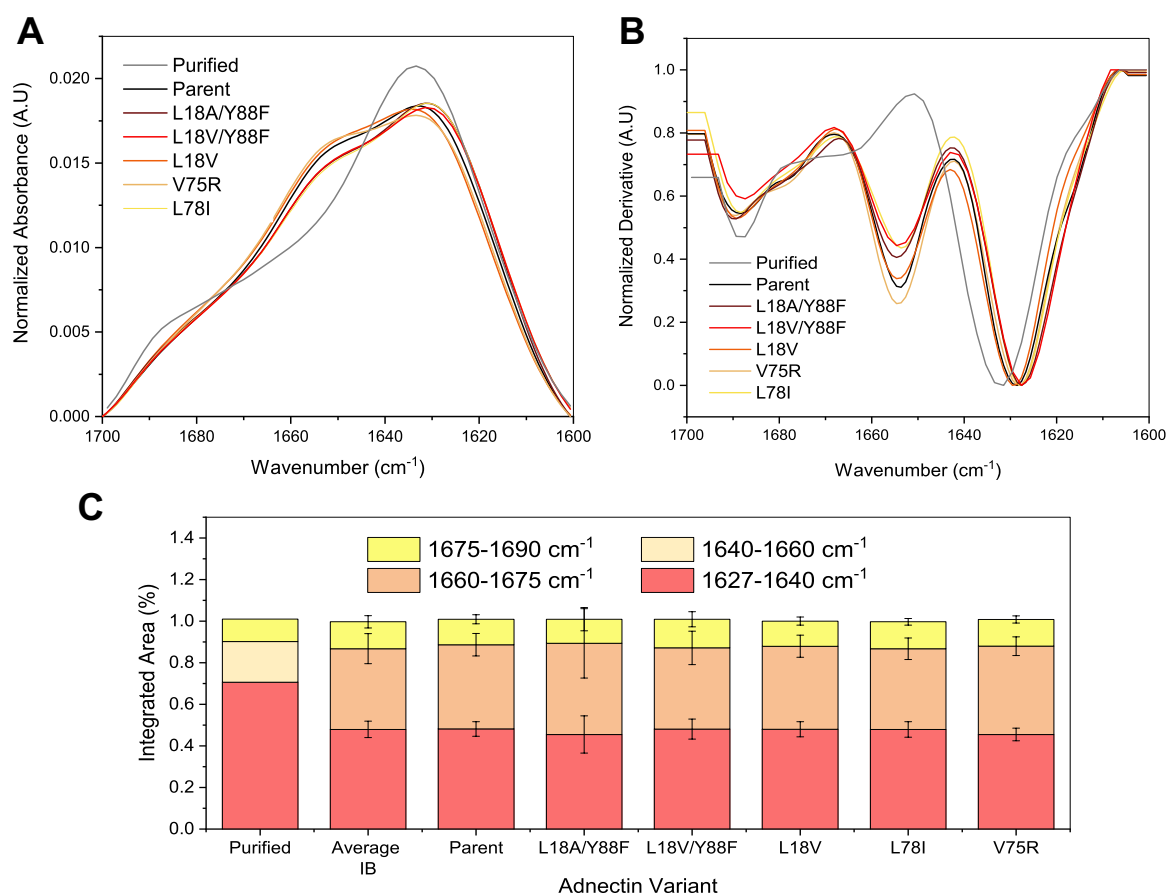


Figure 2.6: Point mutations do not have significant effects on the FTIR spectra of Adnectin IBs demonstrating similar aggregate structure. Each spectrum is the average of 3 technical replicates for 2 separate culture growths. Each absorbance spectrum is normalized to an area of 1 (A), before calculating the second derivative (B). Integrated area from three component multiple peak fitting (C) performed using OriginPro 2018.

Another notable result is that Adnectin IBs exhibit similar absorbance and derivative spectra as well as integrated areas from multiple peak fitting despite varying solubilities and predicted aggregation propensity patterns (Fig 2.6, Fig B.3, Table A.1-A.2). One key spectral region exhibiting small variation is between 1642-1657 cm⁻¹ corresponding to random coil or α -helical structures (Fig. 2.6A,B)^{73,95}. As these measurements were performed on impure samples, it is possible that this spectral variation is a result of other sample components. To address this, a densitometric analysis of protein composition was performed. This analysis was performed for SOD1 IBs due to the availability of an SDS-PAGE gel for every sample analyzed using FTIR.

Using this analysis, it was determined that the relative amount of the main contaminant, outer membrane protein F (OmpF) in the sample does not have a statistically significant correlation with the amount of signal in the 1642-1657 cm^{-1} region ($R^2 = 0.1873$, $p\text{-value}=0.0567$) (Fig. 2.7). As Adnectins are present at a similar concentration in IB samples, it can be assumed that spectral differences in this region are most likely a true result of small differences in IB structure.

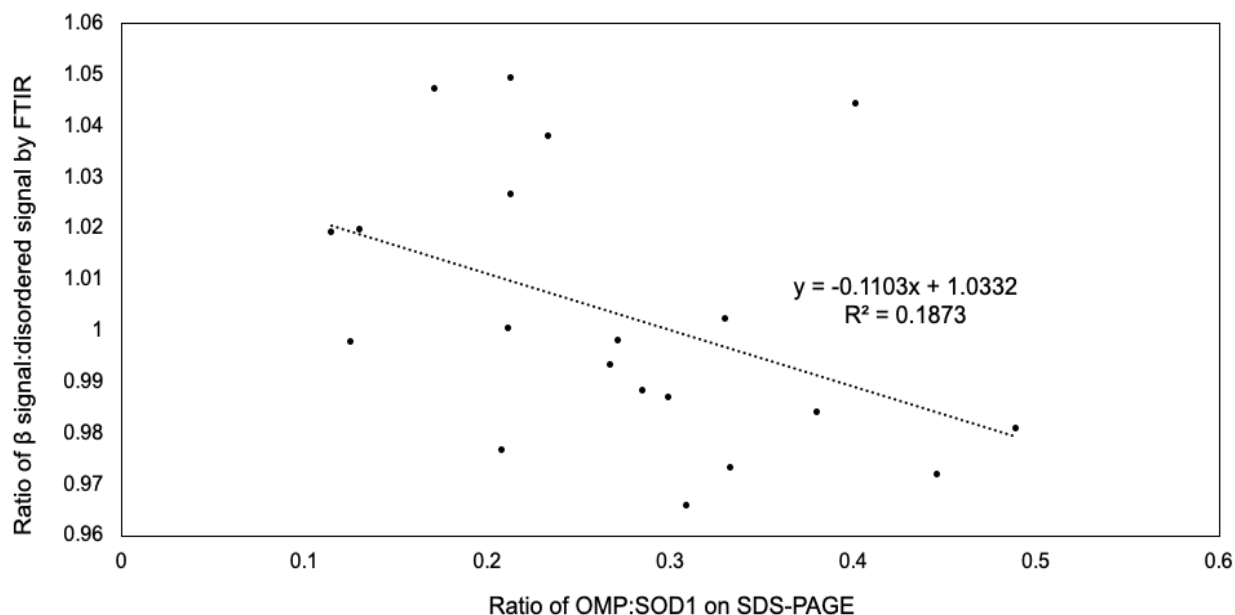


Figure 2.7: FTIR spectral features are attributed to overexpressed protein and not contaminants. Poor correlation between disordered peak (1653 cm^{-1}) intensity in FTIR analysis and amount of OmpF relative to SOD1 present in an IB sample as determined by densitometry using GelAnalyzer⁹⁶ software. Each individual point represents an average protein determination from 2 SDS-PAGE gels and an average disordered signal ratio from 3 technical replicates run by FTIR.

The last region of interest in the FTIR absorbance spectra is between 1611 cm^{-1} - 1630 cm^{-1} where signal typically increases when a protein is forming fibrillar aggregates. As mentioned previously, average Adnectin IBs do not show a large shift towards signal in this region. However, certain point mutants, L78I and L18A/Y88F do show larger right shifts in comparison to other mutants. To investigate whether this aligns with computationally predicted aggregation-prone regions, we used AmylPred2, a consensus predictor that evaluates amyloid propensity using 10 separate algorithms available online⁹⁷. Here, we evaluate amyloid propensity as the sum of individual methods which consider a residue to be aggregation-prone. Interestingly, L78I is

computationally predicted to be more amyloidogenic specifically at the C-terminus while L18A/Y88F is predicted to be less amyloidogenic specifically at the N-terminus (Fig. 2.8). It is unsurprising that predictions do not align with observed data as previous work has demonstrated the failures of sequenced-based solubility predictors for Adnectins⁹⁸ as well as point-mutants of SOD1⁹⁹. It is possible the shift observed may not be attributed to amyloid formation and simply be caused by a more minor conformational rearrangement upon packing of native-monomers into the IB. Therefore, for more concrete conclusions to be made, this should be investigated further using other amyloid detection methods such as limited proteolysis and urea solubilization as well as high-resolution structure determination with qHDX.

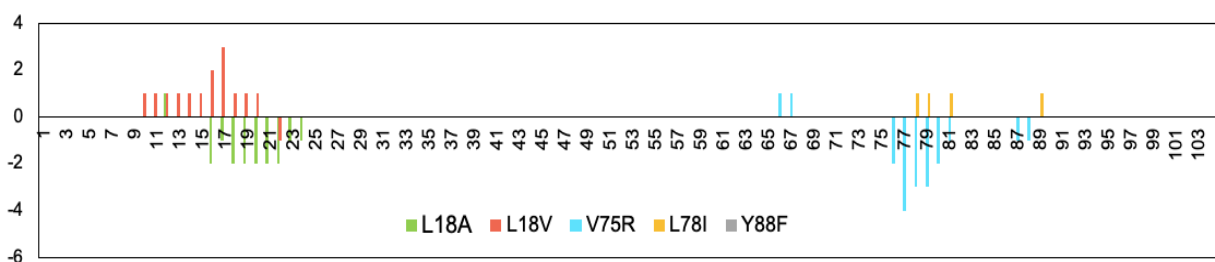


Figure 2.8: Predicted change in consensus amyloid aggregation propensity. Difference in predicted aggregation propensity as a result of individual mutations evaluated using AmylPred2. AmylPred2 score for a single residue evaluated as the sum of individual predictors that consider that residue aggregation prone. Positive scores indicate that a residue is predicted to be more prone to form amyloid fibrils whereas negative scores indicate a residue is predicted to be less prone to form amyloid.

2.3.4 Protease digestion reveals similar protected regions despite varying levels of compactness

To further address the possibility of an amyloid core in Adnectin IBs, protease digestion with proteinase K (PK) was performed. PK resistance has been used extensively to map amyloid-forming regions of various disease associated proteins due to its preference towards globular disordered or disordered conformations over dense cross-beta¹⁰⁰. However, the degree of resistance to digestion for amyloid fibers formed by different proteins is variable^{101,102}. Therefore, it is crucial to compare to a purified form of your protein of interest. From this experiment it was revealed that purified parent Adnectin has considerable resistance to PK with >85% of the full length protein remaining after 30 minutes of digestion (Table 2.2). This is similar to what was observed for purified, human light chain domains with a similar immunoglobulin fold¹⁰³. In

contrast, IBs only have approximately 50% of the full length protein remaining after 30 minutes which corroborates FTIR data that there is a significant portion of unstructured protein in the aggregate.

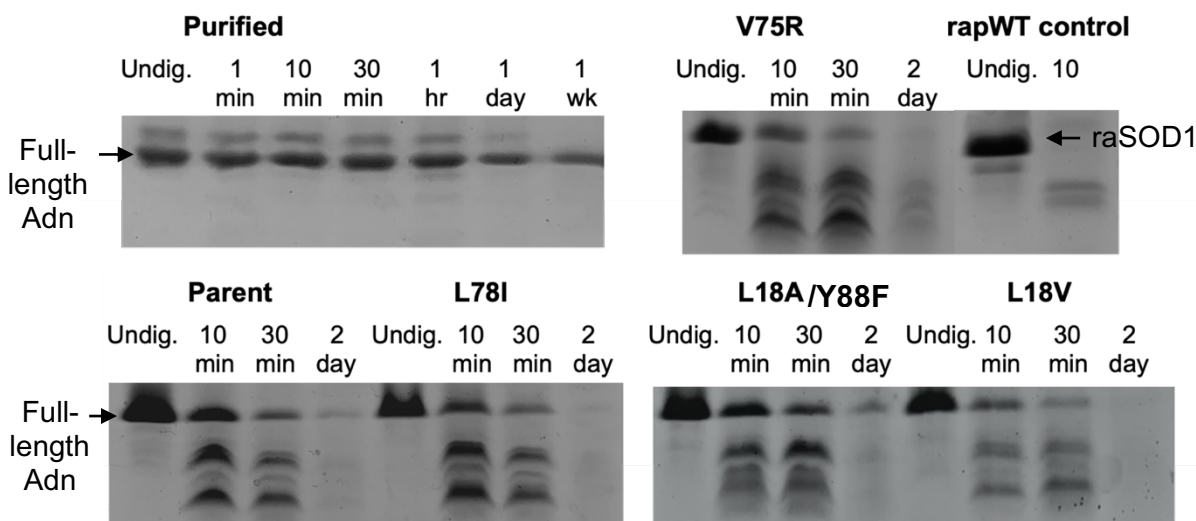


Figure 2.9: IBs show greater sensitivity to proteinase K digestion than purified Adnectin. Tris-Tricine gel analysis of purified Adnectin (2 mg/mL) and mutant IBs undigested and digested with 1 mg/mL proteinase K for specified time increments. Digestion was inhibited with 1 mM PMSF. Reduced apo SOD1 (rapWT) undigested and digested for 10 minutes is shown to verify proteinase K activity.

Additional pieces of structural information that can be visualized on these gels are that fragments released from IBs are of similar sizes for all point mutants (~4.8 and 6.7 kDa) (Fig. 2.9). Thus, the possibility should be addressed that these fragments represent amyloidogenic segments of the Adnectin IB. To investigate this further, CR binding was performed on full length IBs and digested fragments. CR is a sulfonated azo dye that is able to specifically bind amyloid aggregates with cross β -sheet structure in a manner detectable by UV-visible spectroscopy. Spectral changes indicative of CR binding include a shift in the main CR absorbance peak from approximately 490 nm to 521 nm and the appearance of a shoulder at a typical maximum of 541 nm¹⁰⁴. Upon the addition of CR to Adnectin IBs, all mutants show a shift in the main CR absorbance peak and the appearance of a small shoulder (Fig. 2.10A-E). Furthermore, cells that were not induced with IPTG to produce Adnectins, did not exhibit these same changes indicating that the binding is occurring

Table 2.2: Percentage of full-length Adnectin remaining after 30 minutes of PK digestion.

Variant	% Full Length
Purified	87 %
Parent	52 %
L18A/Y88F	50 %
L18V	35 %
V75R	46%
L78I	44 %

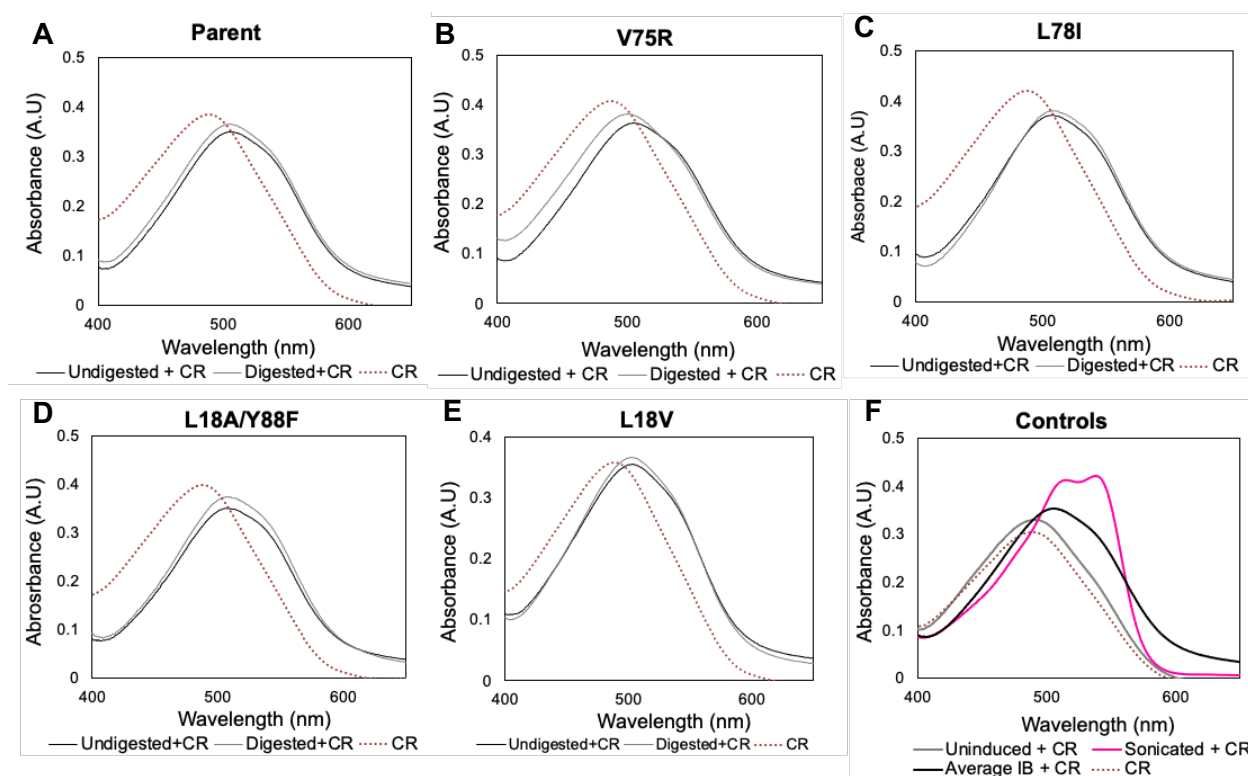


Figure 2.10 Adnectin IBs and protein fragments released upon proteinase K digestion weakly bind congo red dye. Spectra of protein bound dye for mutant IBs (A-E) was evaluated by the difference between protein+CR and protein absorbance spectra. Experiments were undigested and digested with 0.1 mg/mL proteinase K for 10 minutes. Digestion was inhibited with 1 mM PMSF. CR difference spectra shown for uninduced cells and sonicated amyloid SOD1 as controls (F).

specifically to the IB structure (Fig. 2.10F). However, when compared to sonicated SOD1, known to form amyloid ⁹³, the CR spectral changes are much less drastic. To investigate whether a

disordered protein “shell” was preventing stronger CR binding, as seen for sonicated SOD1, a CR binding assay was performed for all mutants after 10 minutes of PK digestion. It was found that digestions do not bind CR different from full length IBs indicating that although some protein may have CR positive cross β -sheet structure, it is not a main component of the IB aggregate.

2.3.5 Chemical solubilization of Adnectin IBs expressed at 37°C, overnight.

Solubilization profiles of Adnectin IBs were determined using optical density measurements at 350 nm. It was observed that IBs were increasingly soluble with increasing urea concentrations measured over a range of 0.5-8 M with complete solubilization occurring between 6 and 7 M (Fig 2.11A). Based on previously published work, it was assumed that solubilization followed a two state mechanism in which the protein is either in an aggregated state that contributes to absorbance or a soluble state which does not^{105,106}. In line with this assumption, the data were fit on OriginPro 2018 using a linear denaturant-dependence of stability. All mutants demonstrated similar solubilization profiles with fitted midpoints of solubilization (C_{mid}) falling between 2.6 and 3.6 M urea (Fig 2.11A, Table 2.3). Furthermore biological replicates demonstrated consistent solubilization profiles (Fig. C.1). Interestingly, the C_{mid} values obtained from the fitting analysis correlate reasonably well with the computationally predicted $\Delta\Delta G_{F-U}$ by FoldX ($R^2=0.808$, $p=0.015$) supporting the conclusion that variants with a lower thermodynamic stability may form a more stable aggregate potentially due to more amyloid-like interactions (Fig. 2.11B, Table C.1). Conversely, denaturation parameters do not correlate with experimentally measured aggregation propensity ($R^2 = 0.0191$, $p=0.794$) (Fig. 2.11C) in contrast to results for A β point mutants¹⁰⁷.

Table 2.3: Equilibrium denaturation parameters of Adnectin IBs. Data fit to a linear extrapolation model (See Section 2.6.6) using OriginPro 2018.

Variant	C_{mid} (M)	m (M ⁻¹)
Parent	3.02 +/- 0.159	0.546 +/- 0.080
L18A/Y88F	3.55 +/- 0.066	0.771 +/- 0.064
L18V/Y88F	3.28 +/- 0.087	0.691 +/- 0.072
L18V	3.04 +/- 0.110	0.532 +/- 0.053
V75R	3.18 +/- 0.058	0.649 +/- 0.040
L78I	2.78 +/- 0.092	0.600 +/- 0.042

The behaviour of pure Adnectin in urea was visualized using fluorescence spectroscopy. Proteins contain intrinsic fluorophores such as tryptophan and tyrosine which are sensitive to their environment; an increase in solvent exposure upon urea-induced unfolding causes a shift in their emission spectra ^{108,109}. It should be noted that IB solubilization measured by light scattering and protein denaturation measured by fluorescence are inherently distinct characteristics; the light scattering measurement does not provide information on the fraction of protein that may be folded after solubilization. However, difficulties with measurement and interpretation due to multiple sample components make obtaining fluorescence spectra for IBs troublesome. Furthermore, this comparison can still provide some insight on the sensitivity of IB versus protein structure to denaturant concentration. Pure Adnectin molecules exhibit a λ_{max} of 309 nm at 0M urea and do not exhibit any significant spectral differences until > 6M urea (Fig 2.11D). Even at 8M, spectra show a significant amount of the original fluorescence peak at 306 nm without any major changes in intensity. However, there is evidence of a new λ_{max} of ~357 nm that could represent a shift in the tryptophan chemical environment attributed to higher dynamics and some population of unfolded protein. This indicates that like its ancestor ¹⁰FN₃ and other immunoglobulin proteins ^{110,111}, the Adnectin fold is stable at even high concentrations of urea. IB solubilization measured here also differs from that seen for fibrils formed by A β ; Halverson *et al.* demonstrate extreme resistance to denaturation even in high concentrations of strong denaturant guanidinium thiocyanate ¹¹².

Equilibrium solubilization experiments also provide an *m* value, which generally reports on the cooperativity of the unfolding transition. For globular proteins this parameter correlates very well with the change in solvent accessible area upon unfolding ¹¹³. For a protein with 105 residues, the predicted value is expected to be 1.35. By analogy, one could expect that an aggregate which exposes more protein surface to solvent upon disaggregation would have a higher *m*-value, however, there is a lack of literature available supporting the transferability of this trend. Nonetheless, it is evident that Adnectin IBs exhibit a less cooperative solubilization than what is expected for the monomeric protein indicating that the dissociation of aggregates is not as simple as a transition from the native to a fully unfolded state. Upon searching for correlations, it was found that the *m*-values for IB solubilization correlate moderately with FTIR β -peak position

($R^2=0.699$, $p=0.038$) (Fig. 2.12) indicating that IBs containing more ordered, intermolecular β -structure have a higher cooperativity of unfolding.

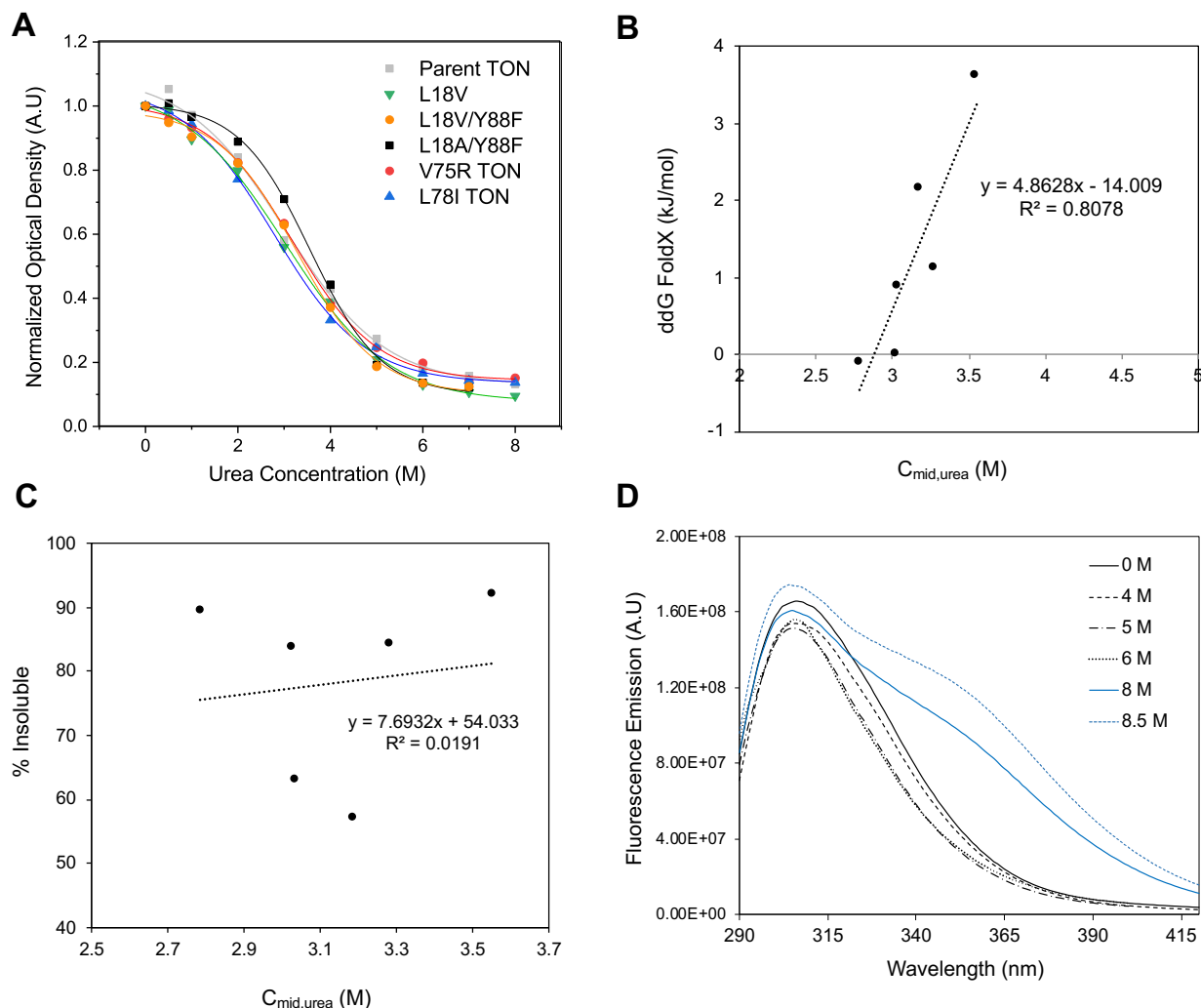


Figure 2.11 Urea solubilization profiles demonstrating full solubilization of Adnectin IBs at 7M urea despite high thermodynamic stability of pure Adnectin. A) Solubilization profiles of mutant Adnectin IBs. Optical density at 350 nm is normalized and shown as a fraction of initial signal intensity. B) Scatter plot of computationally predicted $\Delta\Delta G_{F-U}$ by FoldX with C_{mid} obtained by fitting urea solubilization profiles showing statistically significant correlation ($R^2=0.808$, $p=0.015$) C) Scatter plot of Adnectin aggregation propensity (% insoluble) with C_{mid} obtained by fitting urea solubilization profiles showing a lack of statistically significant correlation ($R^2 = 0.0191$, $p=0.794$) D) Fluorescence emission spectrum of purified Adnectin at 0M, 4M and 8.5 M urea demonstrating significant native structure even at high concentrations of urea. Excitation was at 277 nm with a slit width of 3 nm, emission was recorded from 290-420 nm with a slit width of 5 nm.

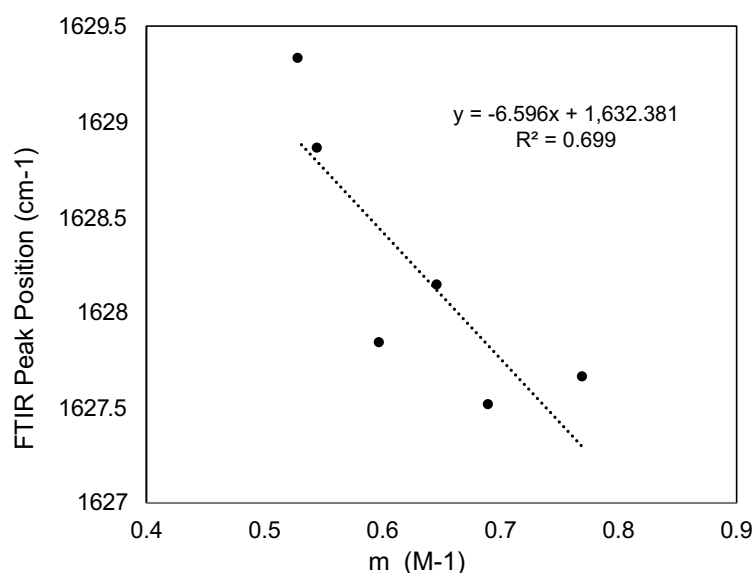


Figure 2.12: Correlation between IB denaturation cooperativity and FTIR β -peak position. IBs containing increased ordered, intermolecular β -structure have a higher cooperativity of unfolding within 95% confidence ($p=0.038$).

In summary, equilibrium denaturation supports FTIR and protease digestion conclusions that Adnectin IBs contain an ensemble of structures with some contributions from native and amyloid structures. The differences observed between purified and aggregated Adnectins can be explained by a number of possible scenarios. It is possible that monomers within the IB are present in a less stable conformation making them more susceptible to urea induced unfolding and release from the aggregate. However, turbidity measurements do not provide the same level of detailed information as fluorescence measurements on the conformational state of a protein, therefore it is also possible that fully folded monomers or small soluble oligomers are being released from the aggregate. This scenario is supported by similar data obtained for IBs formed by asparaginase expressed at 37 °C. Despite IBs exhibiting a lower midpoint of denaturation than purified protein^{31,114}, they still display significant enzymatic activity indicating a native conformation³¹. Further exploration of the state in which protein is released from an Adnectin IB could include size exclusion chromatography of soluble fragments¹¹⁵, a more extensive description of which is included in Chapter 5.2.

2.4 Conclusions

Through analysis with low resolution structural analysis techniques, it was determined that Adnectin IBs contain considerable native structure, some disordered structures and some amyloid conformations. FTIR spectra maintain some of the native anti-parallel signal ($>1680\text{ cm}^{-1}$) while also showing an increase in random coil signal. Moreover, IBs are moderately resistant to proteinase K digestion, resulting in similar fragments of ~ 4.8 and 6.7 kDa . It was determined that the protease resistant fragments do not bind CR as strongly as amyloid fibrils. Thus, despite being less protease resistant than purified protein, the protein may have a native-like conformation with some regions being locally perturbed and therefore more susceptible to digestion. Finally, urea solubilization was performed to investigate the stability of IBs. All mutant IBs were solubilized using urea despite $^{10}\text{FN}_3$ domains being stable in high concentrations of urea. Midpoints of solubilization fell between 2.6 and 3.6 M urea with full solubilization occurring between 6 and 7 M . Interestingly, these midpoints correlated well with the predicted changes in stability caused by each mutation indicating that the stability of the native state is inversely related to the stability of the IB. It is important to note that one should use caution when interpreting solubilization results as turbidity measurements cannot explicitly determine whether folded or unfolded monomers or small soluble oligomers are being released from the aggregate. In conclusion, while the low-resolution techniques used within this chapter can provide simple and rapid structure detection techniques for purified aggregates, they are more complicated when applied to cellular aggregates. Furthermore, a direct explanation of aggregation determinants often remains elusive. Therefore, further studies using residue-specific methods are warranted.

Chapter 3

High-resolution structural analysis of Adnectin IBs using quenched amide hydrogen-deuterium exchange NMR.

Statement of contribution

Description of qHDX (Section 3.1.2) was modified from Schaefer *et al*¹. Some of the qHDX data in this chapter (data contributing to Figure 3.5) have been published previously¹¹⁶. SOD1 resonance assignments included for comparison were performed by Dalia Naser. SOD1 rate constants included for comparison were acquired and analyzed by Dalia Naser and Michael Tarasca. SDS-page analysis and FTIR spectra for washed Adnectin IBs were performed in collaboration with Xiaoyue Liu as part of an undergraduate research project.

3.1 Introduction

3.1.1 Nuclear magnetic resonance spectroscopy

Nuclear Magnetic Resonance (NMR) Spectroscopy is a useful tool for determining the structure and dynamics of protein molecules in solution. NMR spectroscopy measures the response of nuclei with non-zero nuclear spin to radiofrequency pulses experienced in a uniform magnetic field^{117,118}. When pulsed, the magnetic moment of an individual nucleus will resonate at a specific frequency which is dependent on its chemical environment^{117,118}. This frequency is converted to a universal measure independent of the magnetic field strength known as chemical shift through the comparison to a standard reference frequency. The most commonly studied nuclei for biological molecules are hydrogen, nitrogen and carbon. However, carbon and nitrogen are both not naturally abundant in their NMR observable isotopes, ¹³C and ¹⁵N, respectively. Therefore, proteins can be uniformly labelled with these isotopes during the cell growth and expression stage.

For large molecules such as proteins, multidimensional NMR, in which magnetization is transferred from one nucleus type to another, is often required to resolve signals from specific

nucleus types ^{117,118}. For example, simple two-dimensional (2D) ¹H-¹⁵N heteronuclear quantum spin correlation (HSQC) spectra can provide a great deal of information without time-intensive structure determination experiments. These spectra result in a cross-peak for every individual amide in a protein backbone and thus have one peak for every amino acid residue with the exception of proline ^{117,118}. These types of experiments can be used to measure a protein's stability and dynamics as well as indicate whether a protein is properly folded or undergoing self-association ¹¹⁹. By adding an additional carbon dimension and running standard triple resonance experiments, these processes can be pinpointed to specific residues and/or an ensemble of molecular structures can be generated. Historically, solution NMR was limited in terms of studying aggregation-prone proteins. Self-association causes molecules to tumble slower in solution resulting in quick dispersion of NMR signal ¹²⁰. Additionally, intermediate species in the aggregation process are transient and difficult to identify with common NMR experiments ¹²¹. For these reasons, modifications such as quenched hydrogen deuterium exchange (qHDX) NMR have been developed to study aggregating proteins ¹¹⁶.

3.1.2 Quenched hydrogen deuterium exchange NMR

qHDX with solution NMR readout is a powerful but also little-used method to obtain high resolution structural information on large cellular complexes or aggregates. Traditional HDX provides data on native protein structure and dynamics by measuring the extent of protection for individual protein amides (NHs) in a protein structure against exchange with solvent (D₂O) over time. However, this method cannot be performed on large assemblies directly as they are not receptive to NMR measurements; instead, protection may be measured using an adapted qHDX method. Upon suspending aggregate samples in D₂O, NHs that are inaccessible to solvent are protected against exchange, while solvent accessible amides are converted to deuterated amides (ND). Exchange is then quenched by flash freezing and the assembly is dissolved in dimethyl sulfoxide (DMSO), a solvent that slows exchange rates ~four orders of magnitude compared to aqueous solutions ¹²². DMSO also effectively denatures the protein by lowering the free energy of the unfolded state relative to the native state ¹²³. Thus, the originally too-large for traditional ¹⁵N-based solution NMR assembly becomes an amenable size; the majority of residues may be analyzable for protein subunits smaller than ~15 kDa. As deuterium is not observed in a ¹H-¹⁵N

HSQC spectrum, only NHs that were buried or H-bond protected in the assembly, and therefore did not exchange with D₂O, will be observed. Analysis of HSQC peaks reveals which specific residues were involved in protective structure and which were exposed, providing a high-resolution analysis of the protein assembly. The observed protection pattern reports on the average structure of all populated protein conformations in the aggregate. Extensive background on qHDX was described by Hoshino *et al.*¹²⁴ and methodological details for samples of pure proteins aggregated *in vitro* were described by Alexandrescu¹²⁵.

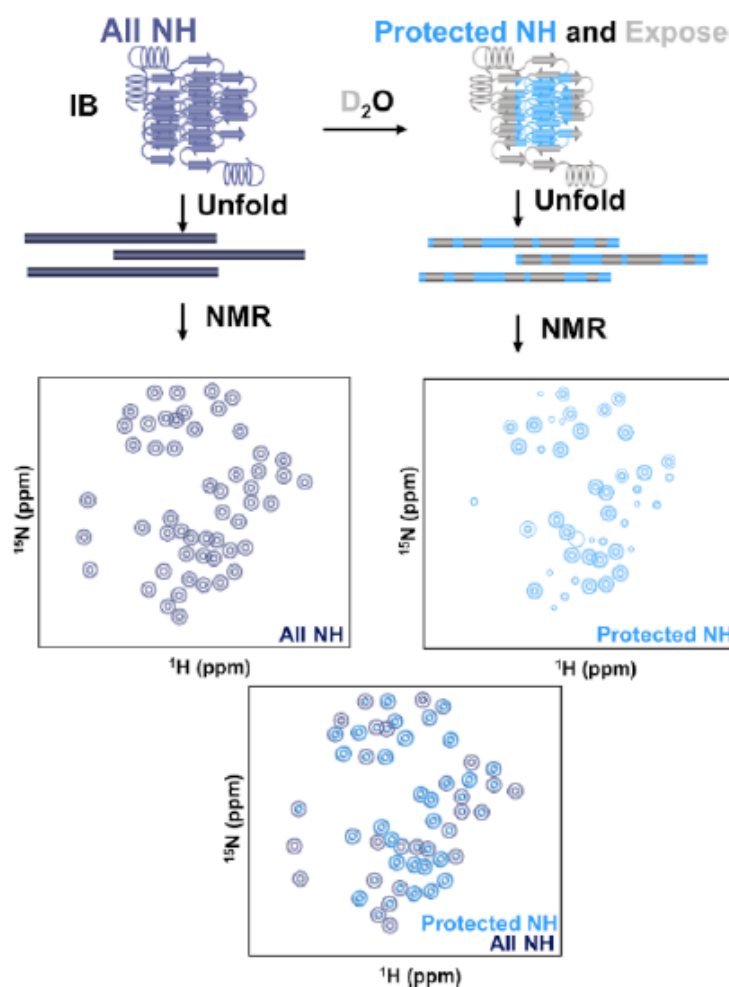


Figure 3.1 Quenched hydrogen-deuterium exchange NMR for analysis of inclusion body structure. Inclusion bodies are incubated in D₂O to allow surface exposed amide protons to exchange for deuterons. Both exchanged and protonated control proteins are unfolded then analyzed via a ¹H-¹⁵N HSQC NMR experiment. Differences in cross-peak signal intensities are correlated to the degree of solvent-protection in the aggregate structure. Figure reproduced from Tarasca *et al.*¹¹⁶.

This method was developed independently by the Alexandrescu ³⁹ and Goto ³⁸ groups in the early 2000s, and it was initially used to determine structural details of *in vitro* amyloid. Alexandrescu applied qHDX to cold shock protein A (CspA) fibrils that were made through acid-denaturing the purified protein. His work elegantly showed that the entirety of the protein was structured in the fibrils, despite their unfolded origins. Soon after, the Goto group reported mapping the core of β 2-microglobulin amyloid fibrils. Similar to CspA, all amides, even those natively found in loop regions, showed strong protection against exchange, indicating the whole protein was involved in the fibril structure. In addition, the protection was compared to that of purified native β 2-microglobulin and found to be remarkably similar in the core, but varied in the loops and termini, suggesting an extensive hydrogen bonding network underlying the rigidity and structural stability of the fibrils. Further application of qHDX to amyloid- β fibrils made from purified protein segments showed, notably, that increasing the length of the aggregating fragment by as little as two residues drastically changed the fibril structure ^{126–128}.

Some years later, the application of qHDX to cellular IBs was reported by Wang *et al.* in 2008 ⁴⁰. The cellular IBs were washed several times during their preparation to make largely pure samples. In a remarkable contrast to the preceding studies of *in vitro* aggregates, Wang and coworkers observed protection of only a small proportion of amides in IBs formed by several structurally diverse proteins. Based on the qHDX and complementary experiments, they inferred that the proteins formed small stretches of amyloid and all other residues were unprotected. This result - the first qHDX for an *in vivo* aggregate - showed that cellular aggregation may be very different from *in vitro* aggregation. This was furthered by their 2010 study, where one protein was aggregated in different ways (heat, concentration, acid, amyloid fibrils, and IB) and its structure was analyzed ¹²⁹. The IB did not resemble the others, further highlighting the importance of *in vivo* studies. Strikingly, the IB resembled the soluble protein more closely than the other aggregates. qHDX studies on washed amyloid- β IBs again showed a new view by revealing the presence of a new, α -helical, aggregation-prone species ¹³⁰.

For some time, qHDX was only used sparingly, predominantly for purified protein aggregates ^{131,132}, until a recent publication on SOD1 cellular IB structure by Naser *et al.* ⁴¹. This study differed from previous work in key aspects. First, the IBs were not washed or purified in any

way and incurred minimal processing before amide exchange, thus minimizing the chances of structural alterations. Second, the study was conducted on a diverse set of ALS-associated mutant SOD1s, containing point mutations dispersed throughout the native protein structure and having a wide range of stabilities and other biophysical parameters. Prior to this work, no high-resolution study on *in vivo* point mutant aggregates had been reported. Remarkably, almost all the aggregates had indistinguishable structures, despite their native differences. Notably, amide protection was observed throughout the whole protein, corresponding to both natively protected regions as well as loops that are not protected in the native structure. However, some of these loops were shown to be amyloidogenic in predictions and other experiments^{133,134}. These results revealed a more complicated picture of aggregation: that the SOD1 IB was an ensemble of different structures that can form via multiple aggregation pathways. The extension of this method to investigate a set of Adnectin IBs will provide insight on how a higher thermostability affects IB structure and aggregation patterns.

2.1.3 Detergent washing

Bacterial IBs, while composed of predominantly overexpressed protein, also contain lipids, nucleic acids, and components of the cell membrane. A major drawback to studying cell-derived IBs is that for many methods, including NMR, these contaminants also give measurable signal. This provides the challenge of discerning which signal comes from your protein of interest and what is a result of contaminating materials. For use in biotechnological applications such as tissue engineering, immobilized catalysts and drug delivery systems, it is common for IBs to be washed with mild solubilization agents. This serves the purpose of removing contaminants, in particular membrane proteins, and has been shown to increase the recovery of bioactive proteins. For the

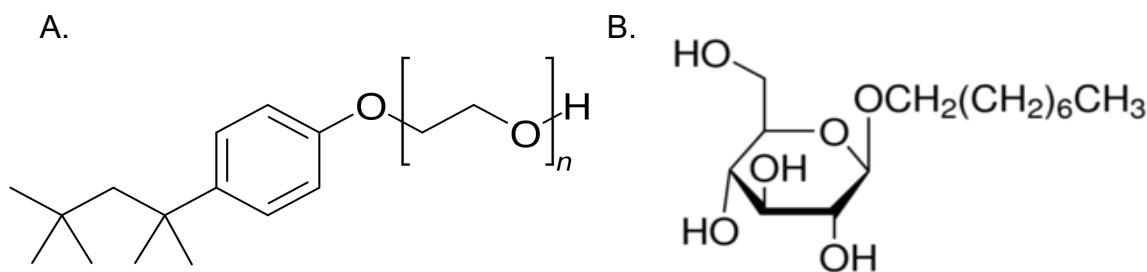


Figure 3.2: Chemical Structures of mild, non-ionic detergents used to wash bacterial IBs. A) Triton-X 100 B) N-Octyl glucoside.

purpose of this thesis, washing with mild, non-ionic detergents Triton-X 100 and N-octyl glucoside (NOG) (Fig. 2.2) was performed to 1) investigate the effects of contaminant proteins, lipids and nucleic acids, 2) assess the stability of Adnectin IBs towards mild solubilization agents, and 3) investigate the possibility of a more structured and potentially amyloid core.

3.1.3 Mutant Selection

Adnectin mutants for qHDX analysis were selected on the basis of low-resolution analyses performed in Chapter 2, the availability of previous experimental characterization of the variant, and the structural relevance to known association mechanisms. For these reasons, qHDX was performed on the parent Adnectin and point-mutants V75R and L78I which are less/more aggregation prone than the parent, respectively. These mutations are of relevance because they are located in the F-strand (adjacent to the FG loop) and the FG loop respectively. As mentioned previously, β -strand G domain-swapping has been observed in the crystal structures for several Tencon proteins (immunoglobulin-fold) resulting in hexamers and larger order multimeric structures (Fig 1.4) ⁵⁹. In particular, the large, left-handed, helical structure provides an interesting potential self-association mechanism as it is open-ended with the possibility of forming very long strand-swapped chains. The plausibility of this mechanism is further strengthened by HDX studies indicating that mutants displaying higher aggregation propensities have greater access to an aggregation-prone conformation of the F and G-strands (discussed in further detail section 2.1.2) ⁷². It should be noted that while qHDX may not be able to discern between protection induced by native hydrogen bonding and hydrogen bonding that occurs within a domain-swapped structure, it can likely discern between this and other amyloid aggregation patterns.

3.2 Methods

3.2.1 qHDX-NMR sample preparation

Adnectin inclusion body samples were grown and prepared following the procedures described in section 2.2.1 and 2.2.2. For NMR experiments, protein is grown with 0.5 g/L labelled $^{15}\text{NH}_4\text{Cl}$. Duplicate flasks of non-isotopically labelled protein (grown with 0.5 g/L $^{14}\text{NH}_4\text{Cl}$) are always grown concurrently for each mutant and expression time.

Sample aliquots were thawed and centrifuged at 16,300 g for 15 minutes. The pellets were resuspended in 5 mL of H₂O or D₂O, for protonated reference and quenched HDX experiments (qHDX), respectively. After incubation for 1-hour at room temperature, the sample was flash frozen in liquid nitrogen and lyophilized for 48 hours. Samples of parent Adnectin were also incubated for 2 minutes and 4 hours at room temperature, and 1 week and 2 weeks at 4 °C. The longer incubations were performed at lower temperature to coincide with conditions in previous experiment and allow for comparison to previous data⁴⁰ while also slowing structural changes in the IBs, potential bacterial growth and sample degradation. For plotting sample signal intensities as a function of time in D₂O, amide intrinsic exchange rates were calculated using a Server Program for Hydrogen Exchange Rate Estimation (SPHERE)^{122,135} at 20 °C and 4 °C and used to convert exchange time at 4 °C to the corresponding time at 20°C.

Lyophilized samples were dissolved in a buffer solution containing 95% DMSO-d₆, 5% D₂O, 25 mM DTT and a small volume of 5% DCA to achieve a final pD ($pD = pH_{read} + 0.4$) of 5.5. The amount of DCA to add to ¹⁵N-labelled samples was predetermined using the unlabelled sample prepared in parallel. Sample pHs were confirmed by measuring pH after all spectra were obtained.

3.2.3 qHDX NMR experiments

Quenched HDX experiments on adnectin IBs were performed using an adaptation qHDX methods described previously for purified protein and IB^{40,129}. All qHDX NMR experiments were conducted at a temperature of 19 °C using a Bruker Avance 600 MHz spectrometer locked to DMSO-d₆. For qHDX experiments, lyophilized sample dissolution and setup time were consistently kept to 10 mins. A ¹H-1D NMR spectrum was then acquired for the qHDX sample, immediately followed by nine sequential ¹H-¹⁵N HSQC experiments. The acquisition time for each spectrum was 20 minutes with 4096 points collected in the direct dimension. The relaxation delay used was 0.8 seconds and 8 scans were performed. The qHDX sample was then stored at room temperature for one week, after which three ¹H-¹⁵N HSQC spectra were acquired to measure the equilibrium H-D sample signal intensities. Smoothed peak intensities of the initial HSQC were used as a measure of the protection of each residue in the aggregate (described below). All qHDX exchange experiments except the parent time course in D₂O were performed in duplicate, and

reported protection (see NMR Data Analysis below) is the average of the two independent culture growths (i.e. biological replicates).

Fully protonated control samples contained 5% (v/v) H₂O instead of D₂O and were made to a pH of 5.5. A ¹H 1D NMR experiment followed by a ¹H-¹⁵N HSQC experiment were performed on a protonated sample for every culture growth analyzed. Spectra were processed using Topspin software (Bruker).

3.2.3 qHDX NMR analysis

The full set of HSQC spectra *versus* time was used to fit the intrinsic exchange rate constant of the unfolded protein in DMSO-d₆ (k_{int}) for individual residues. Processed spectra were analyzed using the Collaborative Computer Project for NMR (CCPNMR) software (V2.4) to measure the decreasing intensity of each crosspeak with time. These intensities were then fit in the software to a single exponential decay with an offset to obtain k_{int} . For each residue the k_{int} values were averaged across biological replicates and used to correct data and obtain the intensity for the time of dissolution from the fitted line.

As protein concentration in IBs and other components in the unwashed IBs are variable, this must be carefully accounted for. Accordingly, we report the fraction of amide protected in a D₂O sample relative to fully protonated H₂O sample using the following procedure. The smoothed intensity for each residue in the first HSQC was scaled for protein concentration by dividing by the average amide signal in the sample.

Intensity of Residue i amide in D₂O sample

$$= \text{Smoothed Intensity}_{\text{Residue } i} / (\text{AvgIntensity}_{\text{All}}) \quad \text{Eq. 3.1}$$

To minimize technical and experimental variations for H₂O signal intensities, signals in H₂O samples were averaged. Thus, after the intensity for each residue in the H₂O spectrum was scaled for concentration using the average amide signal for each sample, the value was also averaged across the set of mutant spectra.

Intensity of Residue i amide in H₂O

$$= [(\text{Intensity}_{\text{Residue } i} / (\text{AvgIntensity}_{\text{All residues}})_{\text{sample } y}] \times \text{Averaged over H}_2\text{O spectra of all mutants} \quad \text{Eq. 3.2}$$

Since the average amide D₂O signal for a sample is a combination of protein concentration and the extent of signal lost to exchange, the extent of exchange must be accounted for or else the fraction protected may be erroneously high. Accordingly, the extent of exchange for the qHDX samples was calculated as the ratio of the intensity of the ¹H 1D amide region (7.5-8.5 ppm) of the D₂O sample divided by the intensity of the corresponding H₂O sample (from the same growth), where these ¹H spectra were scaled for protein concentration using the methyl region; the average ratio for all samples was 0.5. Thus, fraction amide protected was calculated as:

Fraction of Residue *i* amide protected

$$= 0.5 \times \text{Intensity of Residue } i \text{ amide in D}_2\text{O sample} / \text{Avg Intensity of Residue } i \text{ amide in H}_2\text{O}$$

Eq. 3.3

3.2.4 Sequence specific resonance assignments

For making sequence specific resonance assignments Adnectin 6199_CO1 (FG loop sequence: DYRDYC) was grown as described above except with 2 g/L ¹³C-glucose in order to produce uniformly ¹³C-¹⁵N-labelled inclusion bodies. Labelled IBs were isolated and prepared analogous to protonated control samples. NMR experiments were performed at 25°C on a Bruker Avance 600 MHz spectrometer locked to DMSO-d₆. Standard spectra for protein specific residue assignments, HNCO, HN(CA)CO, HNCACB, CBCA(CO)NH, and HN(CA)NNH, were acquired according to Bruker pulse sequences and parameter sets. Assignments were made using Computer Aided Resonance Assignment (CARA) software with reference to previous assignments made for unfolded proteins ¹¹⁶.

3.2.5 Detergent Washing

IBs were washed as previously described using 0.1% Triton-X 100 and 1% N-octylglucoside ¹³⁶. In brief, IBs were harvested by centrifugation at 16,300 x g for 15 minutes. Washing was accomplished via resuspension in detergent and incubation for 5 minutes. Washed IBs were pelleted by centrifugation at various combinations of speed and time. Detergent was removed by centrifugation and resuspension in TEN buffer.

3.3 Results and Discussion

3.3.1 Experimental considerations for applying qHDX to study Adnectin IBs

For the analysis of Adnectin IBs using qHDX, there were a few key experiments that were performed to ensure the method could indeed be adapted to this new system. For traditional NMR samples, a high protein concentration is required. qHDX has the same requirement and poses an additional challenge of contaminating sample components as the IBs are not purified from cellular lysate. Therefore, it is important that your protein of interest is present in high enough concentration in the IB sample to give ^1H - ^{15}N HSQC cross-peaks that have good signal to noise and are distinct from NMR signals of contaminating proteins. Proteins (and their point-mutants) which will give good quality spectra for qHDX can be predicted through SDS-PAGE gel estimation of the relative amounts of your protein of interest and contaminating proteins. However, this ultimately must be confirmed through spectral acquisition as other factors such as sample viscosity, conformational exchange and other interfering sample components (lipids, nucleic acids) can also affect spectral quality. For Adnectins it was determined that IBs expressed overnight were ideal for qHDX analysis as peaks in the protonated spectrum were intense enough to lose ~50% of the protonated signal intensity and still have reliable signal to noise for spectral analysis (Fig 3.3).

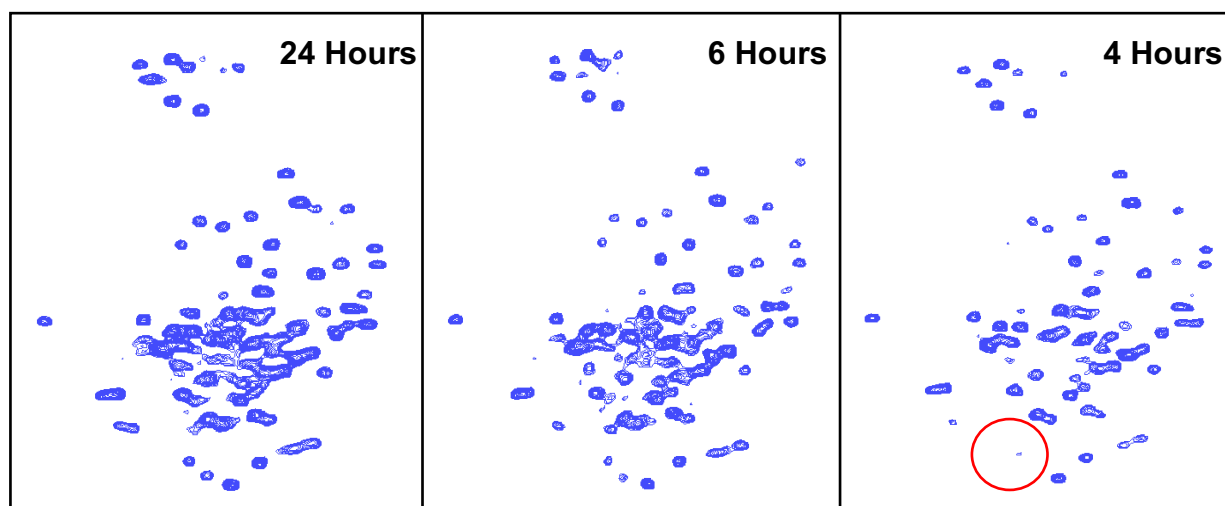
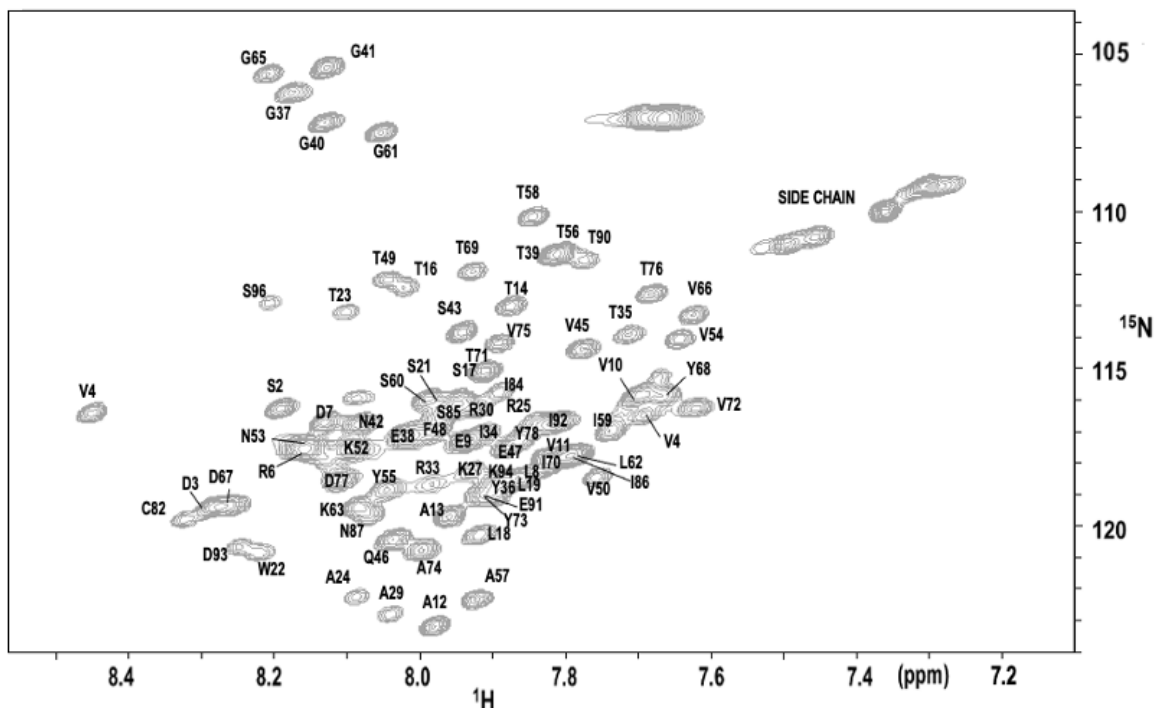


Figure 3.3: Protonated ^1H - ^{15}N HSQC spectra for V75R Adnectin IBs expressed for varying amounts of time post-induction with IPTG. Peak intensities and spectral quality are decreased at shorter expression times resulting in decreased signal to noise. Red circle highlights peaks that are of equal intensity to noise in the spectrum and therefore unusable for qHDX.

The second consideration is that for NMR acquisition, the aggregate is dissolved in DMSO and therefore unfolded. This results in a crowded spectrum with signal concentrated in the random coil region (7.5-8.5 ppm in the ^1H dimension). To obtain valuable structural information from this method, it is essential to have sufficient well-resolved peaks that are dispersed throughout the primary sequence of your protein. This ensures that one can confidently pinpoint which peaks are affected by exchange with D_2O . Although unfolded proteins typically give similar spectral patterns, assignments are highly dependent on protein primary sequence¹¹⁶. Therefore protein sequences with high % composition of isolated residue types (Ala, Gly, Ser/Thr) and little to no sequence repeats are more suited to this method. Using this logic, it was predicted that Adnectins would have a sufficiently resolved spectra as $\sim 1/3$ of the residues are of these residue types. Once initial spectral considerations were confirmed, sequence-specific resonance assignments were acquired using standard three-dimensional (3D) experiments (Fig 3.3A, Table D.1). Intrinsic rates of exchange in DMSO were also analyzed to help confirm resonance assignments and ensure experimental parameters used previously for SOD1 sufficiently slowed forward exchange during NMR acquisition (Fig. D.1). In total, 78 of 92 non-proline residues were able to be assigned and 52 were isolated enough for use in qHDX (Fig 3.4); these isolated residues are found throughout the primary sequence making Adnectins highly amenable to the qHDX method (Fig. 3.4B).

Interestingly, many of the residues that could not be assigned were concentrated in aggregation prone areas of the protein such as the FG and BC loops. A possible explanation for this is that the signals for these residues are much weaker due to line broadening caused by alternate conformations and/or self-association. This argument is strengthened, particularly for residues closer to the N-terminus, by ^{13}C chemical shifts that demonstrate potential residual structure despite being in DMSO. It was previously determined for SOD1 that standard deviations for $\text{C}\alpha/\text{C}\beta$ shifts are very similar between an amino acid type and are not influenced by primary sequence effects^{99,116}. Upon comparison, it was found that Adnectin residues exhibit slightly larger deviations in chemical shift for $\text{C}\alpha/\text{C}\beta$ specifically (Table 3.1) which could be attributed to a more structured chemical environment. When further examined, it was found that the residues exhibiting variations in $\text{C}\alpha$ chemical shift are often found in stretches, many of which are also located in charged, loop regions (Table 3.2). Unlike aqueous solutions, DMSO preferentially interacts with hydrophobic side chains over polar or charged residues. Thus, it is possible that the deviations and

A



B

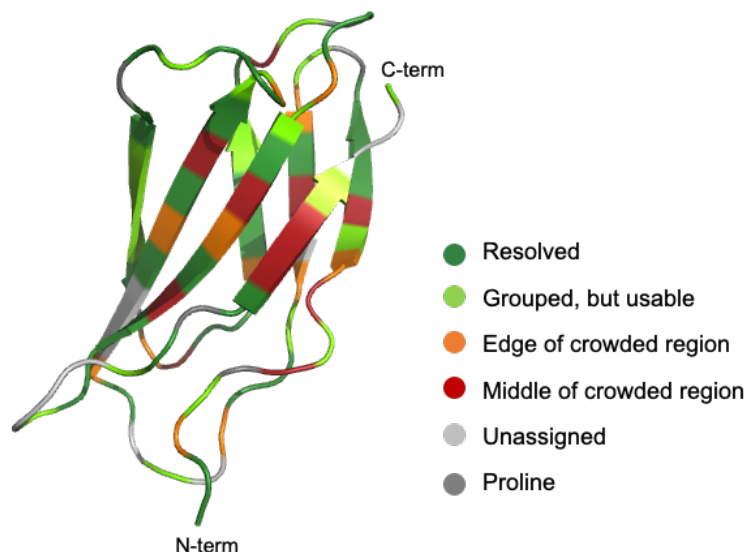


Figure 3.4: ^1H - ^{15}N -2D-HSQC spectrum of Adnectin 6199-CO1 in 95% DMSO/5% H_2O , pH 5.5, with backbone amide assignments. A) Sequence specific resonance assignments made at 25°C as described in section 3.2.4. Assignments labelled using the single letter amino acid code and residue number. Unassigned cross-peaks for sidechains labelled accordingly. B) Ribbon diagram of an Adnectin monomer colored by resolution of ^1H - ^{15}N -2D-HSQC cross-peaks. Green indicates that the residue is usable in qHDX analysis, red/orange indicates that the residue is in a crowded region and therefore unusable for qHDX analysis. Light gray indicates the residue is unassigned and dark gray indicates a proline residue with no HSQC cross-peak.

potential line-broadening are a consequence of solvent effects upon dissolution in DMSO. Ultimately these charged sequence effects should not have negative consequences on experimental results as each exchanged spectrum will be compared to an average protonated spectra for the same point mutant.

Table 3.1: Comparison of average deviations in chemical shift for each atom type for proteins unfolded in DMSO. SOD1 assignments were made by Dalia Naser.

Residue Type	Average Standard Deviation (ppm)	
	SOD1	Adnectin
C	0.434	0.428
C α	0.351	0.489
C β	0.205	0.261
N	0.114	0.089
H	1.37	1.18

Table 3.2: Residues exhibiting deviations from average chemical shift for C α /C β carbons. Average chemical shift includes assignments made for SOD1 and Adnectin. Residues shown are at least 1 SD away from average.

C α Deviations			C β Deviations		
Residue	Deviation from average (ppm)	SD for that res type (ppm)	Residue	Deviation from average (ppm)	SD for that res type (ppm)
V4	-1.87	0.705	V1	0.529	0.252
T14	-1.72	0.770	S2	0.260	0.190
T16	1.74	0.770	T16	-0.31	0.162
S17	1.40	0.732	S17	-0.388	0.190
A24	1.46	0.571	L18	-0.463	0.245
R25	0.959	0.663	L19	-0.319	0.245
R30	0.887	0.663	S21	-0.388	0.190
S43	-1.92	0.732	A24	-0.358	0.280
V50	-2.08	0.705	A29	-0.358	0.280
L62	-0.728	0.387	N42	0.322	0.173
K63	-0.879	0.794	A57	0.435	0.280
K96	-1.82	0.794	K63	-0.660	0.402
			I70	0.453	0.271
			A74	0.363	0.280
			I84	0.453	0.271

3.3.2 qHDX protection for mature Adnectin IBs, expressed at 37°C overnight

To further refine the structural details of Adnectin IBs, qHDX was performed on IBs formed by parent Adnectin as well as two point mutants, V75R (more soluble than parent) and L78I (less soluble than parent) after overnight expression. The measured qHDX data indicate extensive protection throughout the protein sequence for all three variants analyzed (Fig 3.5A). Despite the availability of resonance assignments, protection is not shown for residues G2, S3, and V4 due to the ambiguity caused by multiple species (Fig. D.2). N-terminal gluconoylation is a common post-translational modification for Adnectins⁹⁸ and while present in a comparable amount across IB samples, may influence peak intensities in unmeasurable ways.

While the observed protection pattern aligns with results observed for native-like SOD1 IBs, it contrasts with other previously reported data showing prominent amyloid structure by CR binding, x-ray diffraction and amide protection only for short stretches of sequence which were predicted to form steric zippers⁴⁰. As protection against amide exchange is generally a result of hydrogen bonding¹³⁷, it is expected that a natively-folded, β -sheet protein like Adnectin would display a significant degree of protection throughout the primary sequence in hydrogen-bonded β -strand residues (Fig 3.5B)¹³⁸. However, protection of residues located in loops such as the BC and FG loops is much higher than previously observed for purified Adnectin and similarly folded β 2M and therefore much higher than would be expected if these structures were still exhibiting typical dynamics^{38,72}. To further analyze the cause of the extensive protection, qHDX data was compared to computationally predicted factors influencing aggregation such as hydrophobicity (Fig. 3.5C), β -strand propensity, de-solvation energy and steric-zipper propensity (Fig 3.5D). The lack of a strong correlation with any of the aforementioned predictions indicates that a large structural transition allowing only short stretches of the protein sequence to participate in aggregation is unlikely.

Although differences in protection between mutants are predominantly within error (Fig. D.3) stretches of residues exhibiting significant differences increase confidence that they are true differences and provide meaningful insights to the effect of mutation. These differences are more pronounced for V75R than L78I which becomes increasingly evident when viewed on a ribbon structure (Fig 3.5E). V75R exhibits some small decreases in protection near the region of mutation

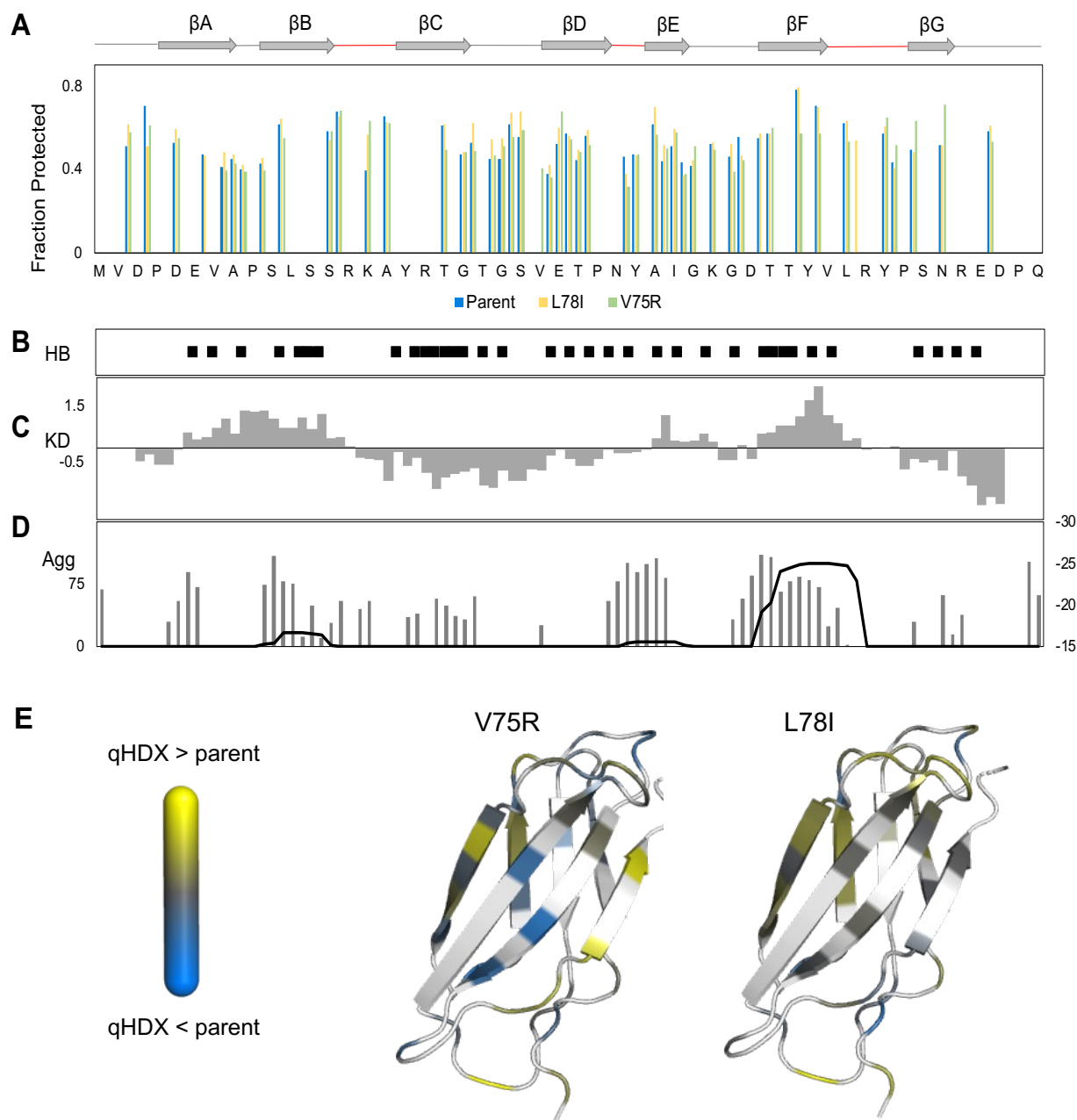


Figure 3.5: Residue-level structure of Adnectin IBs from qHDX analysis. A) Fraction amide protected for parent Adnectin and two point-mutants. Each bar is an average of two biological replicates. Gaps indicate areas of no data due to an unassigned or “crowded” residue. B) Hydrogen-bonded residues observed in the crystal structure of WT ¹⁰FN₃ ¹³⁸. C) Kyte-Doolittle hydrophathy index of parent Adnectin ¹³⁹. D) Tango ¹⁴⁰ (line) and ZipperDB ¹³³ (bars) predicted aggregation/amyloid propensity respectively, for the parent Adnectin sequence. E) % difference in qHDX protection compared to parent Adnectin for two point mutants colored in a gradient of blue_gray_yellow representing a range of -50% to 50% change. White indicates no data.

(residues 72-77). This aligns with the predicted effect of a positively charged arginine residue pointing out into solution ie. a decrease in aggregation-propensity of the F-strand by preventing hydrophobic association in this region. This is accompanied by concurrent increases in protection for exterior strands of the β -sandwich indicating a shift in the ensemble of structures present (Fig. 3.5E). In particular, the G-strand (residues 81-87) shows increased protection. It is possible that this region is simply more favored for preferential hydrophobic association. However, it is also compelling that changes are seen in this region due to evidence of an aggregation-prone state for the purified protein; purified V75R demonstrates increased HDX protection of the G-strand due to a lack of conformational changes and persistence of native hydrogen bonding ⁷². Despite these small changes, Adnectins display very high stability, and high tolerance to mutation so it may not be surprising that a single point-mutation does not have such a dramatic structural effect. It is likely that multiple mutations would be required to cause a larger shift in the ensemble such that a change in structure is more clearly evident by qHDX.

In addition to measuring amide protection patterns, qHDX experiments can be used to study exchange kinetics by measuring protection for different D₂O exchange times. While in some aggregates this results in a single exponential decay ³⁹, others are better described by a period of fast initial exchange followed by a slower phase or plateau ^{40,41,129,141}. Adnectin IBs exhibit the latter behaviour, sometimes termed biphasic or multiexponential, indicating there is conformational heterogeneity in the aggregate. Here we analyze five timepoints of exchange, 2, 60, 240, 1760 and 3770 minutes. The resultant data was fit to a double exponential decay using OriginPro 2018 to confirm this is the case for each individual peak (Fig. 3.6, Table D.2). While, the fits can be analyzed as a whole for general interpretation the individual rate constants may be unreliable because of limited data points and structural changes in the aggregate that could occur over long incubation times. Furthermore, while it appears the exchange has plateaued, the samples were not shown to be fully exchanged after even the longest incubation in D₂O. Thus, an additional even slower phase cannot be excluded, despite not being observed on the timescale of this experiment. Varying explanations have been reported to explain this phenomenon. One model proposes that the entire aggregate may be ordered with outer edges being able to exchange more quickly than the same residues located in the core of the aggregate, protected from solvent ¹⁴¹. Others maintain that biphasic exchange is a result of a fast-exchanging amorphous disordered population coexisting with a more structured or potentially amyloid-like core ^{41,129}. The second

model is more likely in this case as it is supported by the significant contribution of disordered signals in the FTIR absorbance spectra of Adnectin IBs (Fig. 2.5).

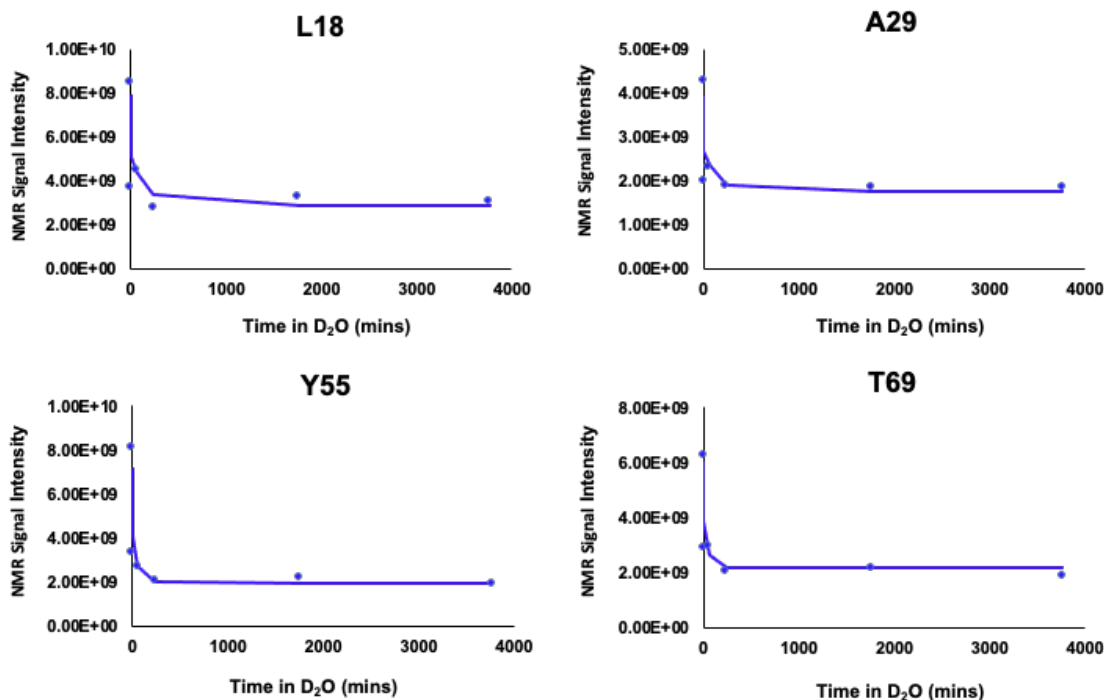


Figure 3.6 Biphasic Exchange of IB amide protons in D₂O. Aliquots of IBs from the same culture expressing parent Adnectin grown at 37 °C and harvested at 18 hours post-induction with IPTG were incubated in D₂O for 0 (protonated), 2, 60, and 240 minutes at room temperature and for 1 and 2 weeks at 4 °C (converted to 1760 and 3770 minutes respectively, at 20 °C using SPHERE^{122,135}). Raw data corrected for protein concentration shown as solid circles, data fit to a double exponential decay with an offset (Amide peak intensity = $A_1e^{(-x/t1)} + A_2e^{(-x/t2)} + \text{offset}$) shown as a solid line. Data was fit with OriginPro 2018.

3.3.3 qHDX protection for Adnectin IBs expressed for 6 hours at 37°C

The effect of expression time was investigated by measuring qHDX protection for the same three Adnectin variants grown for a shorter expression time post induction with IPTG. SDS-page gel analysis of Adnectin solubility reveals that over time, soluble protein is recruited to the IB (Fig. A.2). Therefore we proposed that the investigation of an earlier aggregate may provide a more clear structural interpretation of the maturation process. Furthermore, it is possible that qHDX protection “matures” or “averages out” as expression time increases making it difficult to discern differences between point-mutants, due to misfolded protein packing in many diverse ways. Upon

comparison of 6 hour and overnight V75R, it was found that qHDX protection is largely similar for both time points when error bars are considered. Some small differences are observed for residues 12-16 and 18-30, parts of the A-strand and the B-strand/BC loop respectively. At 6 hours these strands both exhibit intermediate levels of protection, whereas at overnight time points the B-strand and BC loop residues have increased relative protection (Fig. 3.7A). For native $^{10}\text{FN}_3$ domains, residues equivalent to 17-23 ('SLLISWS') have been identified as a potential cryptic binding site capable of initiating fibrillogenesis through association with the B-strand of another monomer ⁵⁵. Therefore, it is possible that at overnight time points this aggregation pathway is becoming more accessible. For L78I, smaller differences are observed when comparing 6 hour and overnight IBs (average difference of 11% vs 16% for V75R) indicating that these IBs may have reached their "mature" structure sooner than those formed by V75R (Fig. 3.7B). However, residues 18, 23, 24, and 26 still exhibit the same general trend with some of the increased protection for L78I expressed ON that is more evident in V75R. As the V75R mutation has a much larger solubilizing effect, it is expected that it would have a greater influence on aggregation pathways, specifically shifting aggregation away from the C-terminus and making N-terminal aggregation pathways more favorable. Finally upon comparison of V75R to L78I both expressed for 6 hours the protection patterns display the smallest differences indicating that early aggregates are likely formed through similar pathways.

Structural changes throughout the lifetime of an IB have been demonstrated using other low resolution methods such as urea denaturation and proteinase K digestion ³⁴. These methods have the advantage of being able to investigate aggregates expressed for as little as 1 hour post induction given it can be confirmed (by SDS-page or other methods) that the turbidity changes are a result of your protein of interest. Exploring shorter expression times using qHDX could be extremely advantageous for providing insight on residues that "seed" aggregation or identifying toxic precursors to aggregation. Strategies for adapting this method include removing contaminant proteins using washing with mild ionic detergents (to be explored here) or through selective isotopic labelling of your protein of interest using a media switch protocol (being explored by other members of the Meiring group) ⁹⁹.

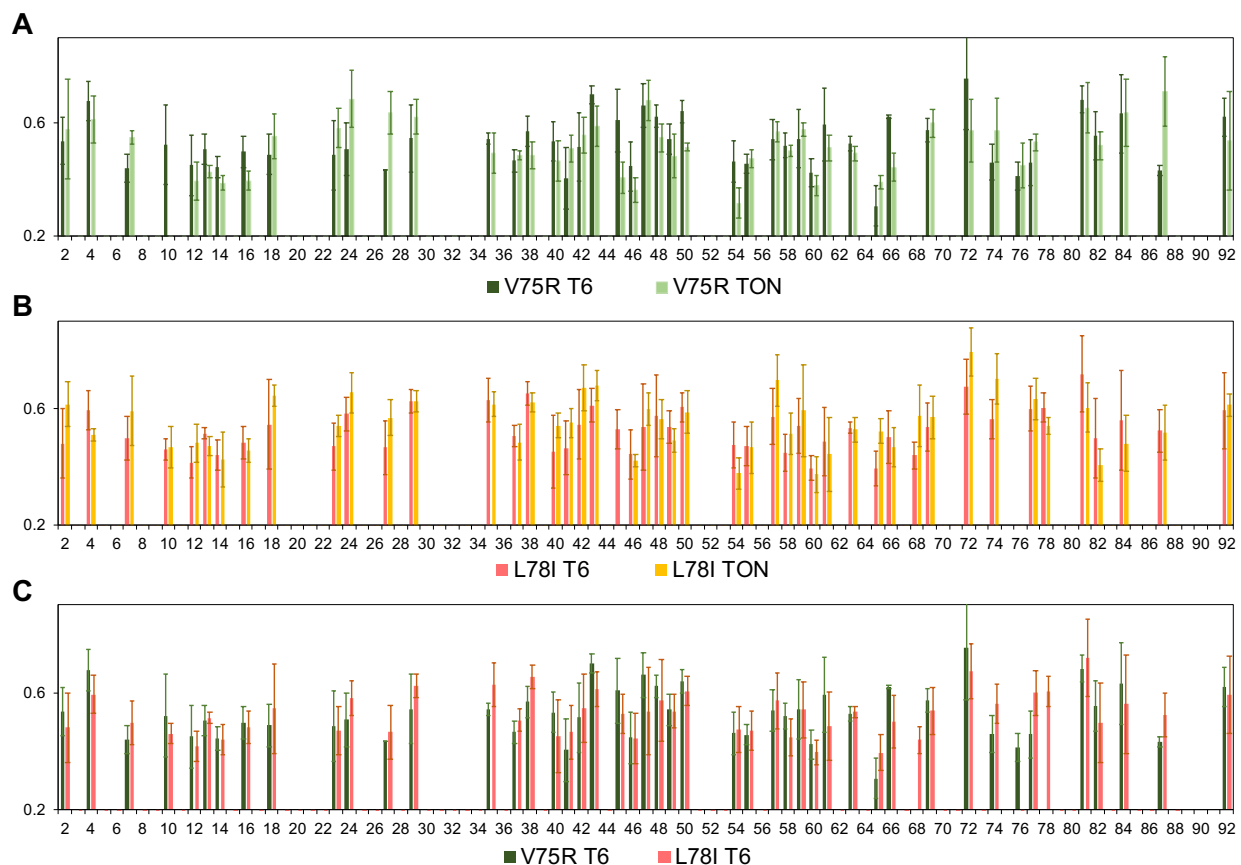


Figure 3.7: Fraction amide protection for Adnectin IBs expressed 6 and 18 hours post induction. Point-mutants are compared to the qHDX protection for the same mutation, expressed overnight (18 hrs) and for 6 hours at 37 °C (A and B) and to each other, each expressed for 6 hours at 37 °C (C). Each bar is an average of two biological replicates. Gaps indicate areas of no data due to an unassigned or “crowded” residue. Error bars represent propagated error from the average of two biological replicates.

3.3.4 Detergent washing of Adnectin IBs

Mild, non-ionic detergents such as Triton-X 100 and N-octyl glucoside (NOG) are used in many purification and protein preparation protocols. Washing with mild detergents is also performed during purification of IBs for structural studies to decrease signal from contaminating proteins. However, the structural impact of these detergents on aggregating proteins has been little explored. Here, a detailed investigation on whether this method can reliably be used for qHDX sample preparation. First it was investigated how detergents impacted IB sample composition through SDS-PAGE analysis (Fig. 3.8). It was found that washing with Triton-X 100 at maximum speed was the most successful at removing OmpF which is the main contaminant of these samples.

NOG washing was also successful in removing OMPF, however, this resulted in a more considerable loss in Adnectin and thus did not significantly alter the relative sample contributions.

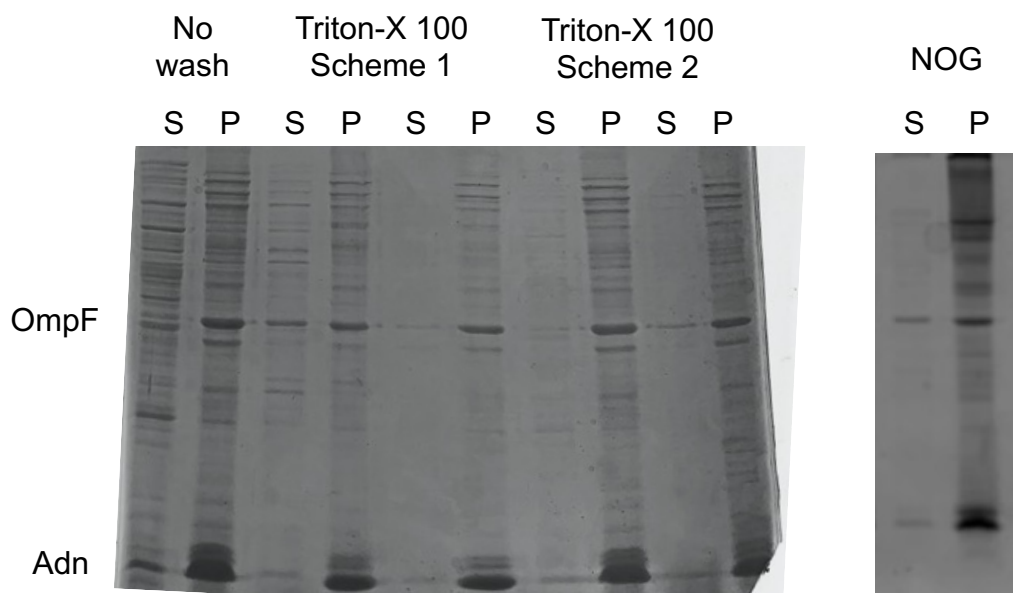


Figure 3.8: SDS-PAGE gel analysis of washed Adnectin IBs. IBs were resuspended in detergent specified for 5 minutes. Triton scheme 1 indicates a 15 minute spin at 16300g followed by removal of detergent with an identical spin. Triton scheme 2 indicates a 15 minute spin at 8000g followed by removal of detergent with a 15 minute spin at 16300g. NOG washing indicates a 5 minute spin at 12000g followed by removal of detergent with an identical spin. Supernatant (S) and pellet (P) from each spin are shown. Triton-X washes illustrated were performed on L78I expressed overnight, NOG was performed on V75R expressed overnight.

FTIR spectroscopy was also used to analyze the effects of detergent washing on the three Adnectin variants previously studied using qHDX. It was found that the impact of detergents was largely consistent between variants, detergent types and centrifugation speeds. Washed IBs exhibit decreased disordered signal at $\sim 1653 \text{ cm}^{-1}$ in addition to a right shift of the β -sheet absorbance peak (Fig. 3.9, Fig. B.4-B.6). This type of peak shift is often accompanied by a decrease in signal above 1680 cm^{-1} which is indicative of increased parallel cross β -sheet or amyloid character. However, in this case, significant high wavenumber signal remains meaning any fibrils formed still have distinct anti-parallel structure. A possible hypothesis is that detergent molecules are disrupting the structure of unfolded or misfolded species in the amorphous “shell” of the IB. Due to Adnectins fast *in vitro* refolding rate¹⁴², monomers may quickly refold and associate with existing native-like fibrils, potentially in a domain-swapped state. This theory aligns with previous

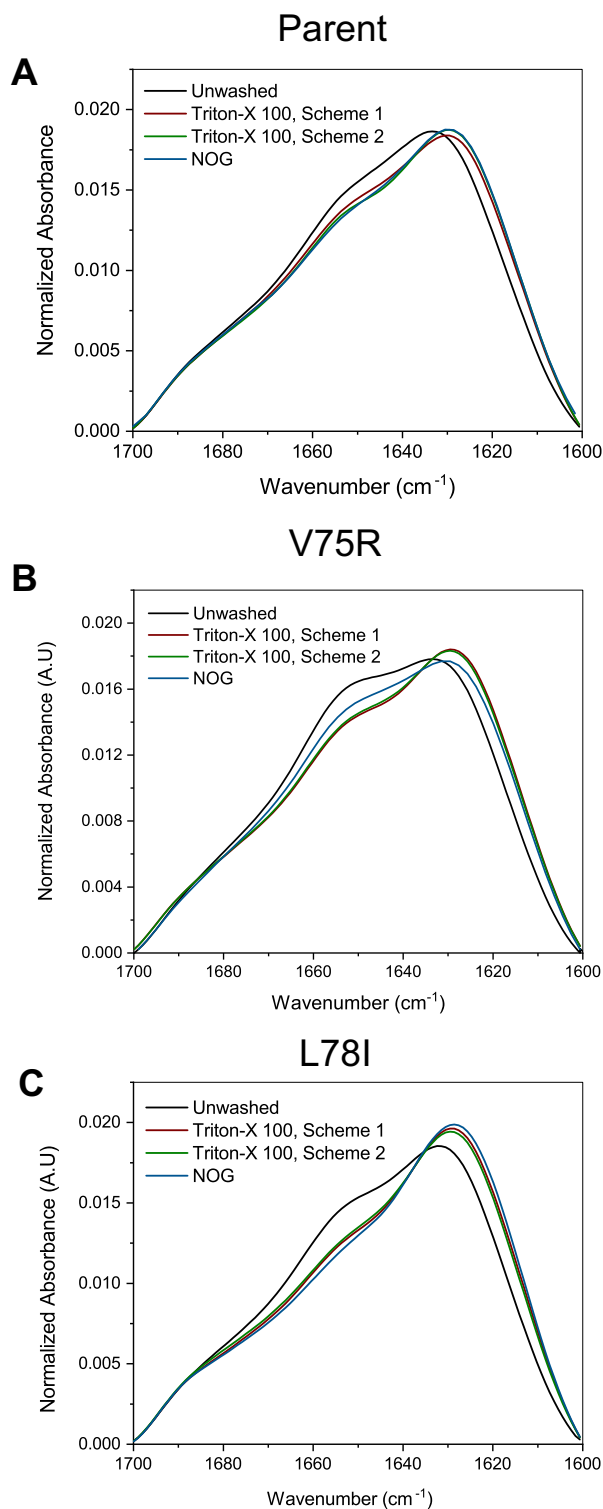


Figure 3.9: FTIR absorbance spectra of detergent washed IBs. Triton scheme 1 indicates a 15 minute spin at 16300g followed by removal of detergent with an identical spin. Triton scheme 2 indicates a 15 minute spin at 8000g followed by removal of detergent with a 15 minute spin at 16300g. NOG washing indicates a 5 minute spin at 12000g followed by removal of detergent with an identical spin. Washes were performed for parent (A), L78I (B), and V75R (C) IBs.

observation that β -sheet peak position correlates with oligomer size, with larger oligomers shifting to lower wavenumber. This theory of fibril extension could be investigated further using electron microscopy.

Triton-X 100 was chosen for pre-treating samples for NMR analysis due to the somewhat conflicting reports of structural changes induced by NOG^{143–145}. It was found that despite only small changes in sample composition, detergent washing was successful in decreasing the intensity of many contaminant peaks, such as the peak at 8.25/109 ppm which is known to be OmpF and others of unknown origin at 7.75/112.5 ppm and 7.65/110 ppm (chemical shift positions given in the format ^1H shift/ ^{15}N shift) (Fig. 3.10A). Furthermore, the resolution of the central crowded region of the spectra improved greatly. This is crucial for qHDX analysis as it is important to compare intensity changes at the center of a given peak. Taken together, this indicates that washing

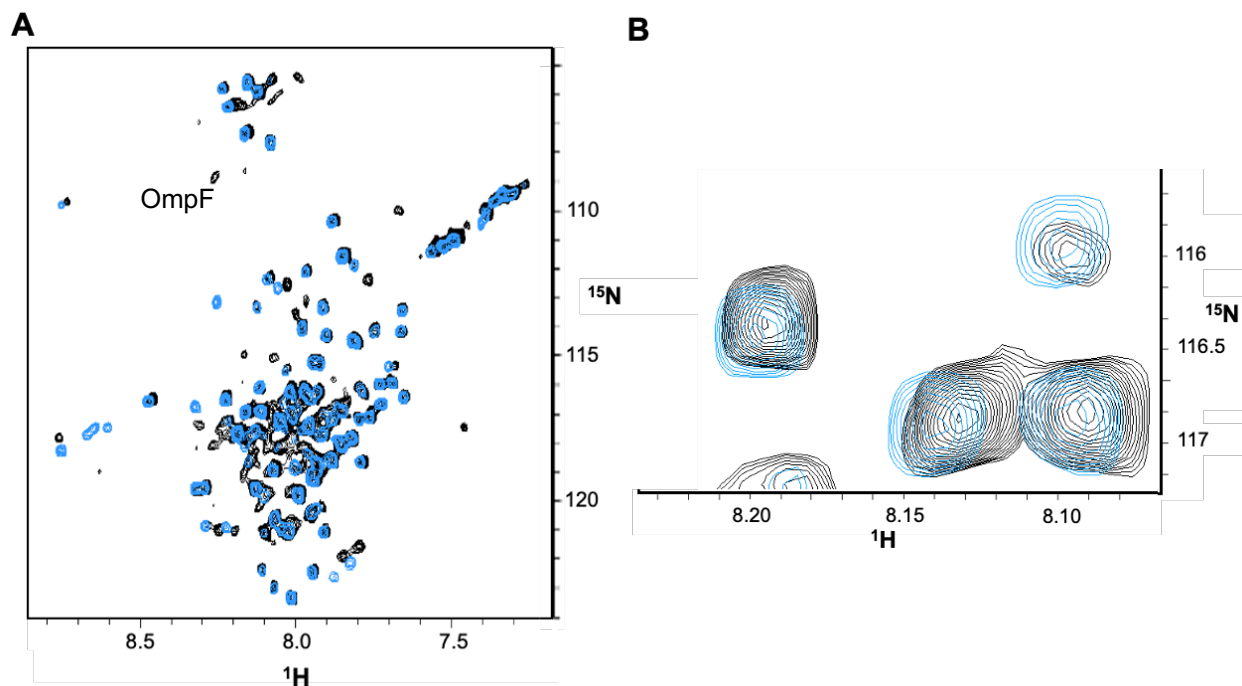


Figure 3.10: Improved ^1H - ^{15}N HSQC spectral resolution upon washing with Triton-X 100. A) Overlay of protonated spectra for unwashed (black) and washed (blue) L78I expressed for 6 hours post-induction. IBs were resuspended in detergent specified for 5 minutes. IB was collected by centrifuging 15 minutes at 16300 g followed by removal of detergent with an identical spin. B) Zoom in of HSQC peaks demonstrating improved resolution and more uniform peak shapes.

of IBs is a viable strategy for adapting qHDX to study proteins that are present at a lower relative amount in the insoluble fraction of a cellular sample provided that washing does not induce any major structural transitions in the IB.

A further investigation using qHDX was performed to complement FTIR structural analysis of washed IBs. Both variants retained their extensive protection providing additional support that at least the ordered fraction of the ensemble is not undergoing large structural transitions (Fig. 3.11). Notably, through comparison of the protection patterns for unwashed and washed variants it was found that the areas where protection ratios differed from 1.0 were similar for V75R and L78I. However, the direction of these changes were not always the same, meaning that some regions becoming more protected upon washing for V75R became less protected upon washing for L78I and vice versa (Fig 3.11). A potential explanation for this is that fibrils present in each IB

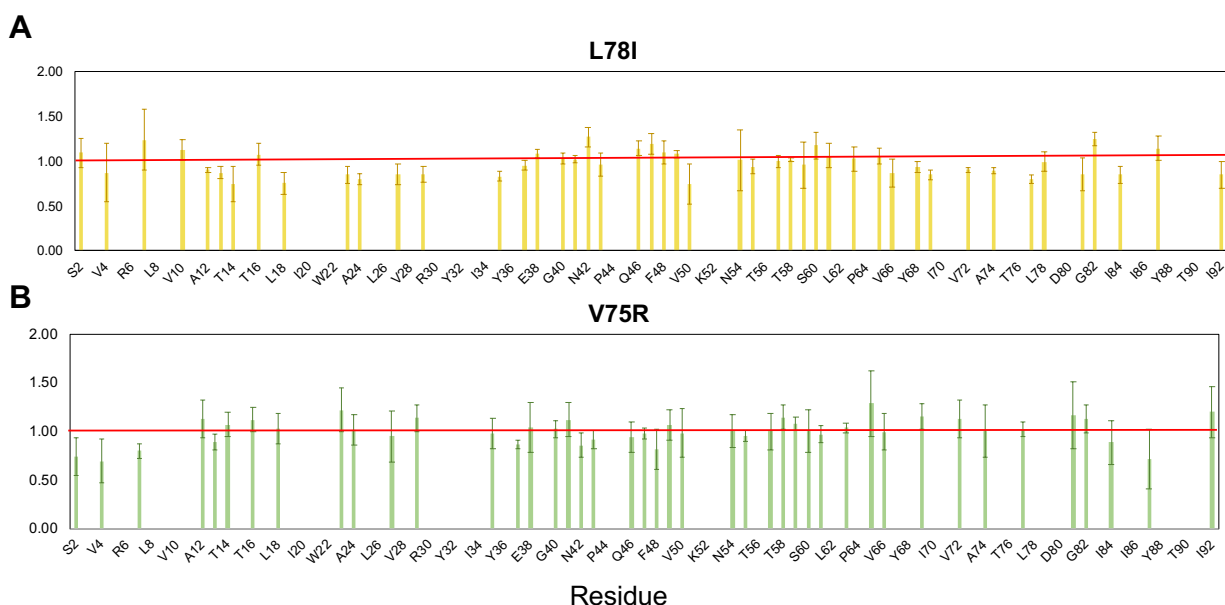


Figure 3.11: Protection ratios for comparing IB structure unwashed vs. washed with Triton-X 100. Ratio of the fraction amide protected for each mutants IB washed with Triton-X 100 to the average fraction amide protected for that mutant, unwashed from the same growth. The red line represents a value of 1.0, any data points above this line have a higher observed protection upon washing, and any points below this line have a lower observed protection. Error bars represent propagated error on the average protection for two biological replicates.

do not have the same core residues and thus if fibril extension is occurring upon washing the original ‘seeds’ differ between mutants. Unfortunately this indicates structural changes induced by detergent may be highly protein dependent. While washing may be appropriate for use in qHDX

studies, any detergent-induced structural changes should be assessed by one or more additional methods such as FTIR and one should take caution when interpreting small changes in IB structure.

3.4 Conclusions

In summary, qHDX can be easily applied to study the properties of mature Adnectin IBs at a residue-specific level. The results presented here including 1) the biphasic exchange profile and 2) extensive protection throughout the primary sequence beyond residues expected to undergo hydrophobic association or form steric zippers, provide compelling support for a heterogeneous aggregate with significant native like structures and some disordered components. This aligns extremely well with data from the low resolution studies presented previously in Chapter 2. Furthermore, small changes in the protection pattern upon mutation and maturation time can be explained by a native-like aggregation model. First, when a charged mutation such as V75R is introduced to a protein's scaffold, aggregation of that region is repelled and protection is shifted towards other more hydrophobic patches of the protein's surface. Second, comparing structures with varying expression time from 6 hours to overnight post induction with IPTG, reveals some evidence that a N-terminal aggregation pathway focused in the B strand and BC loop becomes more accessible/favorable with increased expression time. However, the changes observed are small and should not be over-interpreted. Thus to obtain a more confident view of the residue level-structure of Adnectin IBs more dramatic mutation or difference in expression time may be required. For investigating the effect of shorter expression times or mutations that result in a much lower fraction of Adnectin in the insoluble fraction, washing with a non-ionic detergent may be employed to remove contaminant proteins and improve spectral quality. A preliminary assessment of structural changes caused by detergents indicates that non-ionic detergents do not have a significant effect on the structure of Adnectin IBs. Although this response may be highly protein dependent, so when applying this strategy to a new system, it is recommended to use a lower-resolution method such as FTIR to ensure there are no large effects induced by washing.

Chapter 4

Structural engineering of Adnectin inclusion bodies.

Statement of contribution

Template scripts for performing MD simulations were graciously provided by Upneet Bala. All other design aspects including python script writing for analysis of multiple sequence alignments and MD simulations were performed by Anna Schaefer.

4.1 Introduction

4.1.1 Context

In Chapter 3, the effects of single, previously studied point mutations were investigated in terms of their residue-level IB structure using qHDX. It was determined that these point mutants all displayed evidence of extensive protection pointing to a structural ensemble containing significant native-like structure. To strengthen this ensemble hypothesis, we employed a reverse approach. A mutant with a disrupted hydrophobic core was designed with the goal of promoting amyloid aggregation, and investigating how a non-native aggregate would respond to the techniques used in previous chapters. To confirm that it was indeed the destabilizing core residues and not a charge dependent effect at the N-terminus, a second mutant was designed with the goal of decreasing the N-terminal aggregation propensity of a stable, properly folded mutant. This chapter will outline the rationale design, prediction accuracy and resulting aggregate structure of these mutants.

4.1.2 Protein engineering and design

Many globular proteins exist at a fine balance between their folded and unfolded states, meaning they are only marginally stable (Fig 4.1). Thus, in many cases, there is substantial room for stability improvement without impairing the protein's function¹⁴⁶. However, this is not trivial as most mutations are destabilizing and the rare stabilizing mutations can interfere with a proteins function or cause aggregation^{147,148}. Protein folding is governed by both thermodynamic and

kinetic control, although significantly more is known about the molecular features governing thermodynamic stability ¹⁴⁹. While there is no universally accepted strategy, stabilization of a folded protein is often focused on introducing interactions such as hydrogen bonds, and Van der Waal's interactions, promoting a compact, hydrophobic core and comparison to libraries of previously determined stability data. Conversely, destabilization can be predicted by evaluating the removal of these interactions. The approach applied here will use computational predictors of stability and aggregation followed by production and experimental characterization. This strategy has the advantage of being able to screen multiple mutations in a relatively short amount of time. Furthermore, while they often fall short in terms of predicting stabilizing mutations, it is in our favor that computational prediction tools have much better reported success for identifying destabilizing mutations.

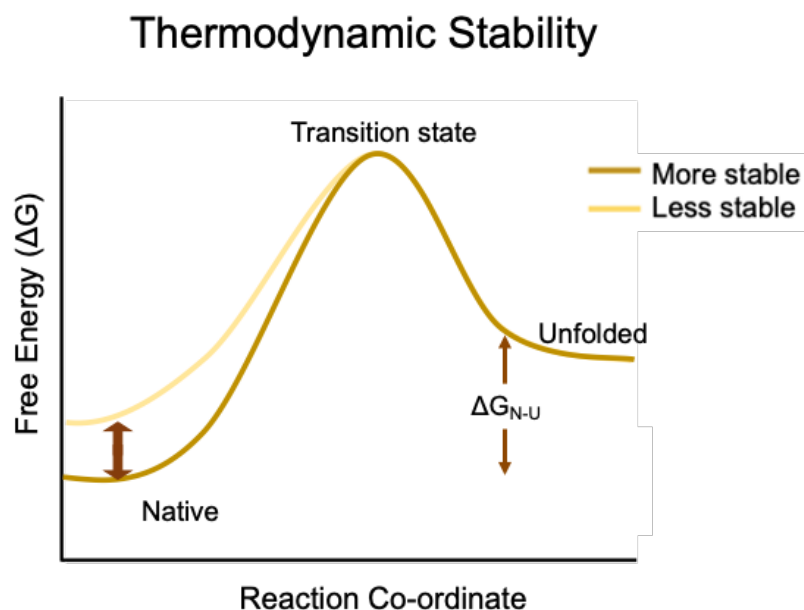


Figure 4.1: Free energy profile representing altered thermodynamic stability of a protein.

4.1.3 Computational stability predictors

Computational stability predictors evaluate the effect of a point-mutation by using unique scoring algorithms to evaluate the change in Gibbs free energy of unfolding, $\Delta\Delta G$. There are a great number of these tools available that can be grouped into categories based on their scoring functions. These categories include scoring functions based on molecular mechanics, statistical

functions, empirical data, physics-based force fields/potentials, and those using machine learning based on physiochemical features or statistical scoring functions. For simplicity, only those techniques applied here will be described further, a comprehensive description of computational stability predictors and their performance can be found elsewhere¹⁵⁰. Predictors to be used in this study were evaluated based on their statistical performance metrics; representative predictors of categories displaying high Spearman rank and Matthews correlation coefficients were chosen. Multiple predictors were used in hopes of obtaining a better sampling of destabilizing mutations. First, FoldX is a force field based on empirical energy calculations¹⁵¹. The $\Delta\Delta G$ of a mutation is calculated using a function that considers van der Waals contributions of all atoms and the difference in solvation energies for polar and non-polar groups, respectively, when going from the folded to the unfolded state. It has been demonstrated that this algorithm performs best when comparing the hypothetical differences in unfolding energy between two well-resolved structures, ie. between a wild-type variant and a mutated protein, rather than evaluating the ΔG of a single protein¹⁵². Secondly DFIRE2.0 is an all-atom statistical energy function that compares to a reference state of distance-scaled, finite ideal gasses¹⁵³. Finally, two machine-learning based algorithms were chosen as they both demonstrated relatively high success compared to other methods. MAESTRO is an ensemble method which integrates statistical scoring functions with biochemical properties such as secondary structure, solvent accessible area, change in hydrophobicity, and the change in isoelectric point to train two separate machine learning strategies¹⁵⁴. Similarly, PoPMuSiC employs a machine learning strategy, trained by statistical parameters, however, in this case a signal neural network is trained by a larger set of parameters (24) with weighting coefficients that depend on solvent accessibility of the mutated residue¹⁵⁵.

4.1.4 Aggregation propensity predictors

In addition to tailoring stability, this chapter will focus on deliberately altering aggregation properties. Recent strategies of controlling and understanding aggregation have often relied on solubility and/or aggregation prediction algorithms¹⁵⁶. Mutations are introduced into predicted aggregation-prone regions (APR) of the protein typically in attempt to decrease self-association. This is relevant for the design of proteins produced for the pharmaceutical industry and IB aggregates produced for biotechnological purposes. While this strategy has been effective in many cases, its performance is limited by the accuracy of APR identification. The earliest and arguably

the most simple method of predicting aggregation propensity of a protein sequence is based on properties of the amino acids that make up that sequence and not the order in which they appear. The most well-known example of this is the Chiti-Dobson solubility prediction algorithm which is based on the weighted sum of the change in hydrophobicity, secondary structure propensity and net charge. A similar style of predictor, known as a sliding-window predictor, evaluates the same parameters albeit in short sequences of residues so that the order but not the relative orientation of residues is considered. These methods include the well-established PROSOL II, TANGO, and WALTZ, tend to overpredict APR as they do not take into account that the sequence may be buried in the protein and participating in interactions within the hydrophobic core ¹⁵⁷. To combat this issue, the newer approach is to take into account 3D structure and conformational dynamics ¹⁵⁸. Based on this information, a diverse combination of predictors will be used herein when evaluating the potential effects of a point-mutation.

4.2 Methods

4.2.1 Adnectin alignment construction and solubility analysis.

Solubilizing mutations were chosen using a rational similar to that previously described for V75R ⁷². The FN₃ multiple sequence alignment 'cd00063' was retrieved from the Conserved Domain Database (<https://www.ncbi.nlm.nih.gov/cdd>). The β -strand A sequences were extracted independently and used in the construction of hypothetical FN₃ sequences with the wild-type ¹⁰FN₃ sequence at all non- β -strand A positions. A similar library was constructed for β -strand B sequences. Analysis of co-variation was completed using python scripts.

The amino acid sequences constructed as described above, were submitted to the CamSol Intrinsic server (<http://www-mvsoftware.ch.cam.ac.uk/index.php/camsolintrinsic>) ¹⁵⁹ with a pH of 7.0 and to the Aggrescan web server (<http://bioinf.uab.es/aggrescan/>) ¹⁶⁰.

4.2.3 Aggregation propensity and stability prediction for engineered Adnectin mutants

Stability changes resulting from point mutations were predicted using DFIRE2.0 ¹⁵³ (Stand-alone executable), Schrodinger Maestro (<https://pbwww.services.came.sbg.ac.at/maestro/web/maestro>), PoPMuSiC (<https://soft.dezyme.com>) ¹⁵⁵ and FoldX stability function (Stand-alone executable) ¹⁵¹.

Input parent Adnectin structure was created with homology modelling software SwissModel ⁶⁰ by Julia Steckner. Mutant structures were generated using the FoldX BuildModel function. Scores for each predictor were normalized on a scale of 0 to 1 using Eqn. 4.1 and summed for a maximum score of 4.0.

$$\text{Normalized score} = (\text{Prediction}_{\text{Residue } i} - \text{Prediction}_{\text{Lowest}}) / (\text{Prediction}_{\text{Highest}} - \text{Prediction}_{\text{Lowest}})$$

Eqn 4.1

Aggregation prone regions were evaluated using the techniques mentioned in section 4.2.1 along with sequence-based method Tango ¹⁴⁰, with a pH of 7.0, temperature of 298.15 K, ionic strength of 0.02 M and protein concentration of 1M and the ZipperDB for amyloid propensity prediction ¹³³.

4.2.4 Protein expression and experimental characterization

Adnectin inclusion body samples were grown and prepared following the procedures described in section 2.2.1 and 2.2.2. For NMR experiments, protein is grown with 0.5 g/L labelled ¹⁵NH₄Cl. Duplicate flasks of non-isotopically labelled protein (grown with 0.5 g/L ¹⁴NH₄Cl) are always grown concurrently for each mutant and expression time.

FTIR, chemical denaturation and protease digestion experiments were carried out as described in sections 2.1.3-2.1.6. qHDX data acquisition and analysis were performed as in section 3.2.

4.2.5 Implicit solvent molecular dynamics simulations and analysis.

The starting structure of monomeric Adnectin was a homology model constructed using SwissModel ⁶⁰. Mutations were introduced using UCSF Chimera ¹⁶¹. Optimization of initial hydrogen position and subsequent energy minimization was performed using Chimera's built-in minimization function; this consisted of 100 iterations of steepest descent followed by 10 conjugate gradient minimization steps using the Amber ff14SB force field ¹⁶². All subsequent simulation was performed using Amber with a ff19SB force field and an "optimal" point-charge (OPC) water model ¹⁶³. The salt concentration was set to 0.15M. After energy minimization of the solvent and solute, the system is heated up to 310.15 K in 100 ps. This was followed by 11.2 ns of equilibration with restraints of varying strengths applied to backbone and hydrogen atoms

separately. Finally, 100 ns of production was performed using a time step of 2 fs. The SHAKE algorithm ¹⁶⁴ is employed to fix all bonds involving hydrogen atoms. Langevin dynamics ¹⁶⁵ with a collision frequency of 1.0 ps⁻¹ was applied to regulate temperature. The trajectory was saved every 2 ps for a total of 50,000 frames analyzed.

MD analysis was performed using custom python scripts and python modules such as MDTraj ¹⁶⁶. Spatial aggregation propensity was calculated using Eqn 4.2, as described previously by Voynov *et al.* ¹⁶⁷. In brief, SAA is the ‘solvent accessible area’ of side chain atoms contained within radius R from the central atom calculated at each saved frame of the simulation. SAA of side chain of fully exposed residue is the SAA of side chains of the residue (X) in the fully extended conformation of tripeptide ‘Ala-X-Ala’ taken from Samanta, Bahadur and Chakrabart ¹⁶⁸. Residue hydrophobicity obtained from Black and Mould, scale is normalized such that glycine hydrophobicity is 0 ¹⁶⁹. For this analysis a radius (R) of 5 Å was used.

$$\left(\text{Spatial-aggregation-propensity (SAP)} \right)_{\text{atom } i} = \sum_{\text{Simulation Average}} \left\{ \sum_{\text{Residues with at least one side chain atom within R from atom } i} \left(\frac{\text{SAA of side chain atoms within radius R}}{\text{SAA of side chain atoms of fully exposed residue}} \times \text{Residue Hydrophobicity} \right) \right\}$$

Eqn. 4.2

4.3 Construct Design

4.3.1 Rationale

Amyloid formation is a well-known aggregation mechanism due to its involvement in many neurodegenerative diseases. The formation of amyloid often involves structural rearrangements so that short segments of the protein are assemble into a repeating cross β-sheet orientation. While this is often studied *in vitro* for disease proteins such as Aβ, amylin, and α-synuclein ^{170–173}, this structure has also been identified in IBs derived from bacterial cultures ^{40,46}. In both cases, amyloid formation can be identified by NMR methods. When studied by qHDX, amyloid is revealed through short segments of the sequences being highly protected against exchange, while others are fully exchanged with solvent. This result is strengthened by protection correlating well with computational prediction⁴⁰ and fibrils observed through cryo-EM and ssNMR ^{171,173}. Fibril segments studied through ssNMR have also shown high correlation with amyloidogenic segments

predicted computationally ¹⁷⁴. These results all contrast with what is observed for the IBs studied in the Meiering lab. Therefore, we attribute the extensive protection to a native-like structure present in the aggregate. To strengthen this conclusion, we aim to demonstrate what an amyloid qHDX protection pattern would look like for an Adnectin. All amyloid-forming IBs previously studied by qHDX were destabilized through some mechanism, indicating that amyloid formation is increased when those segments are exposed. Therefore, we aim to promote amyloid aggregation by creating a largely destabilized Adnectin mutant.

4.3.2 Stability and solubility prediction

Destabilizing mutations for Adnectin molecules were predicted using a combination of results from FoldX ¹⁵¹, MAESTRO ¹⁷⁵, PoPMuSiC ¹⁵⁵, and DFIRE2.0 ¹⁵³. As the only structure available is a homology model and not a high resolution crystal structure, multiple computational prediction strategies must be applied to obtain increased confidence. A mutation sensitivity profile reveals that hydrophobic residues involved in folding are the least tolerant to any mutation (Fig.4.2A). This is expected as many participate in the protein's folding nucleus ¹⁷⁶ and are highly conserved in the conserved domain database (CDD) FN₃ alignment ('cd00063')¹⁷⁷. These residues were therefore chosen as potential sites for further investigation. At each position, mutation to a charged residue (Lys, Arg, Glu, Asp) and to β -sheet breakers (Pro, Gly) was evaluated. Normalized scores reveal that the single mutation predicted to be the most destabilizing to an Adnectin scaffold is W22G (Fig. 4.2B, Table E.1). Based on previous reports that the fibronectin folding nucleus is quite tolerant to mutation ¹⁷⁸, a second candidate was also chosen to increase effectiveness. In

Table 4.1 Sequence conservation of hydrophobic core residues for FN₃ domain.

Residue	Conservation in the CDD FN ₃ Alignment
L18	18.4%
I20	18.3%
W22	95.4%
Y32	83.6%
I34	28.9%
I59	21.6%
I70	15.3%
V72	64.6%

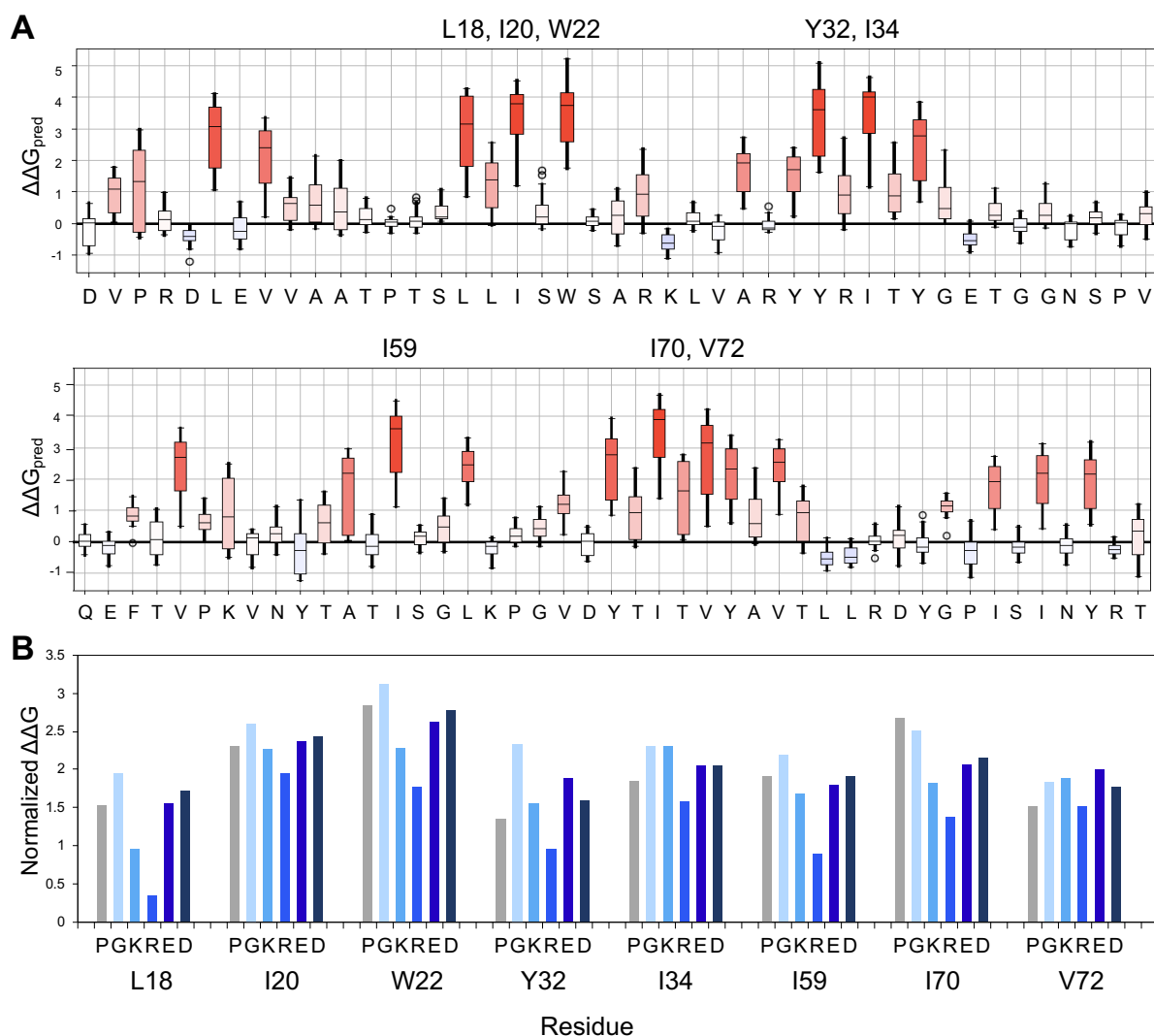


Figure 4.2: Computational predictions of (de)stabilizing mutations. A) MAESTRO¹⁷⁵ mutation sensitivity profile demonstrating residue positions which have high/low tolerance to mutation. Each box represents the distribution of $\Delta\Delta G$ values upon mutating to each possible amino acid residue. Boxes are colored in a gradient from blue to red based on the median value where blue represents a residue prone to stabilizing mutations and red represents a residue prone to destabilizing mutation. B) Normalized $\Delta\Delta G$ prediction scores (Table E.1) for the core residues identified as being sensitive to destabilizing mutations. Scores were evaluated using predictions from FoldX¹⁵¹, MAESTRO¹⁷⁵, DFIRE2.0¹⁵³, and PoPMuSiC¹⁵⁵ with a maximum possible score of 4.0. See Eqn 4.1 (Section 4.2.3) for calculating normalized scores.

designing the second mutation, amenability to qHDX analysis was also evaluated. To avoid ambiguity in assignment transfer, the second destabilizing mutation should be of a different residue type and not proline. Aggregation propensity was also considered; to maintain the native

aggregation/amyloid-forming propensity of the C-terminus, mutation candidates were focussed at the N-terminus. It was found that a mutation at position 34 is more destabilizing when in combination with W22G and thus I34R was chosen as the second candidate (Fig. E.1, Table E.2).

After thermostability was determined, additional mutations were evaluated to tailor the aggregation propensity of the proposed mutant. The goal was to create a model system displaying similar protection to the IBs investigated by Wang *et al.*⁴⁰, ie. qHDX protection of just one limited region of amyloid-forming residues. To do so, additional charged mutations V11R and L19R were added at the N-terminus in an attempt to eliminate aggregation of that region. These mutations were chosen using a similar strategy as was previously described for V75R⁷². From a multiple sequence alignment of the 100 most/least soluble sequences in the CDD FN₃ alignment ('cd00063')¹⁷⁷, it is apparent that residues pointing in towards the hydrophobic core are highly conserved between soluble and insoluble sequences. In contrast, residues pointing out towards solvent (odd residue numbers) display much higher tolerance to mutation including a preference for charged residues in soluble sequences (Fig. 4.3B,D) than insoluble sequences (Fig. 4.3A,C). Since the destabilizing mutations described above are expected to have a moderately solubilizing effect on the B-strand, the A-strand was targeted first. In doing so, V11R was chosen as the most common solubilizing, charged mutation in the A-strand. The complementary mutation L19R was chosen for three reasons: 1) when position 11 contains a charged arginine residue, position 19 is the most common position for another charged residue in the CDD alignment (Fig. E.1B), 2) given the minimal changes in protection seen for V75R, it is likely that more than one point-mutation is required to have a large effect on solubility and 3) previous available knowledge of arginine gatekeepers⁶⁷. These mutations were also studied independently of W22G/I34R to assess whether any changes in IB structure were caused by decreased thermostability or increased surface charge.

4.3.3 Conformational stability investigation using molecular dynamics

Molecular dynamics simulations using implicit solvent were also performed in an attempt to validate the previous computational stability predictions and perform a preliminary evaluation of potential conformations experienced by the proposed multi-point mutants. Simulations were performed for parent Adnectins as a point of comparison. Using a strategy developed by Bernhardt

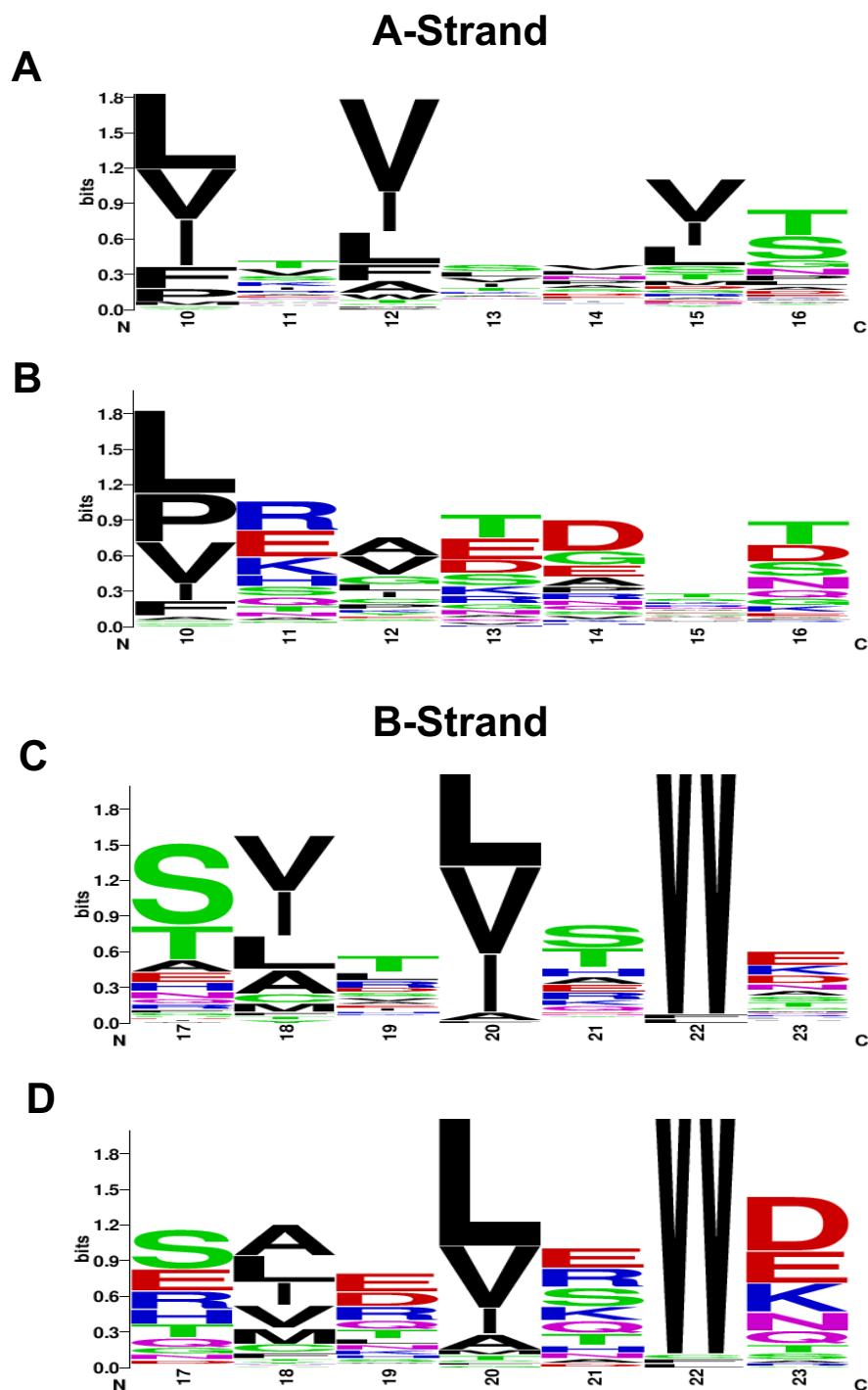


Figure 4.3: Amino acid sequence ‘logos’ (created using WebLogo¹⁷⁹) for insoluble and soluble FN₃ β -strand sequences. Logos were constructed using the 100 least/most soluble sequences from the CDD multiple sequence alignment ‘cd00063’¹⁷⁷ for strand A (A, B, respectively) and strand B (C, D, respectively) as predicted by a summation of CamSol¹⁵⁹ and Aggrescan¹⁶⁰ predictions. The height of each letter is proportional to the frequency of the corresponding amino acid at that position.

Trout's research group¹⁶⁷, the exposure of hydrophobic aggregation-prone regions was evaluated. It was found that the A/B strand arginine mutations were successful in decreasing the spatial aggregation propensity of these strands relative to parent Adnectin (Fig. 4.4A). Furthermore, secondary structure propensity for each frame of the simulation was evaluated using MDTraj's DSSP function (Fig 4.4B)^{166,180}. These preliminary estimations reveal that RRRG exhibits weaker β -strand propensity than parent Adnectin, while RR maintains a largely native-like secondary structure. Both of these observations gave confidence that the designs would be successful and thus we proceeded to experimental characterization.

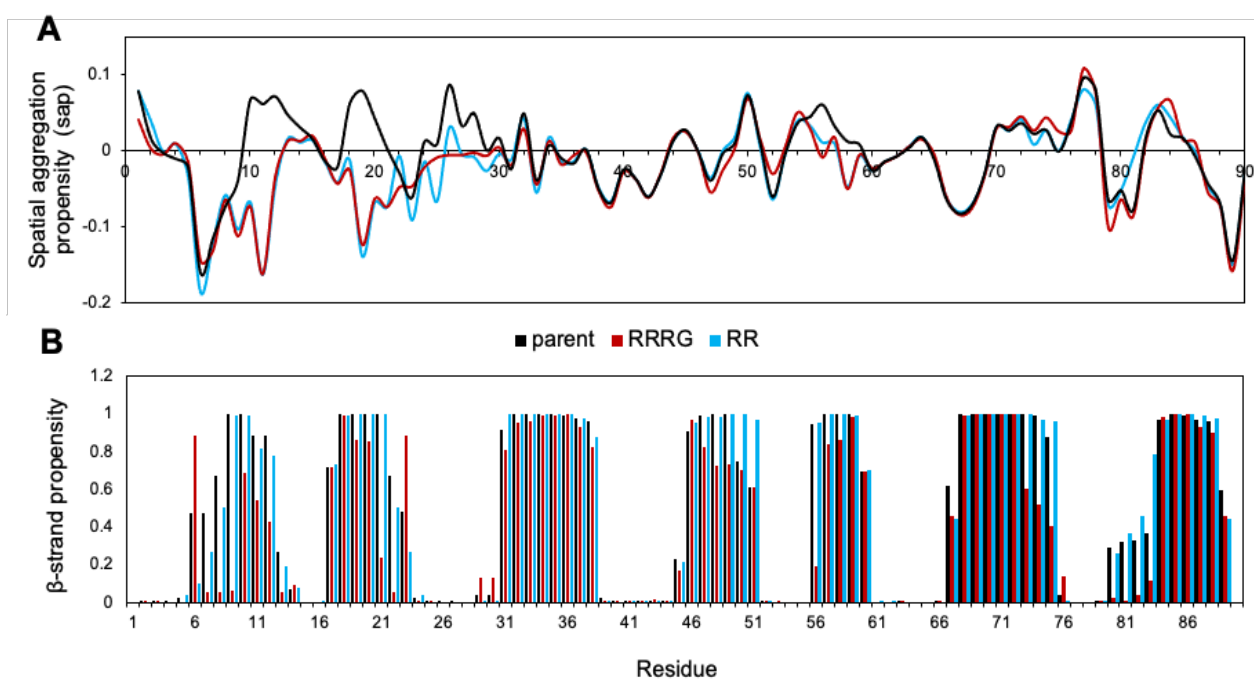


Figure 4.4: Conformational analysis of MD simulations for Adnectin variants. A) Spatial aggregation propensity measuring the dynamic exposure of adjacent hydrophobic amino acids forming an aggregation hot-spot¹⁶⁷. Values greater than 0 indicate a hydrophobic path, values less than 0 correspond to polar, hydrophilic residues. B) Secondary structure propensity calculated using DSSP 2.2.0^{166,180}.

4.4 Experimental results and discussion

4.4.1 Expression levels and aggregation Propensity

As mentioned in Chapter 2, the first step of studying aggregates and their properties is to evaluate the expression and aggregation propensity of a new construct. This ensures that there is

enough protein in the insoluble fraction to be studied by structural analysis methods. It was found that in both media types there was no soluble expression of RRRG (Fig.4.5A). In addition, the overall expression of RRRG in LB media was quite low. This can be explained by increased proteasomal expression and activity in LB causing quicker removal of unstructured and aggregated protein ⁸⁶. When expressed in M9 minimal media, however, the aggregation propensity remains very high and a significant amount of protein remains in the insoluble fraction. Surprisingly RR is

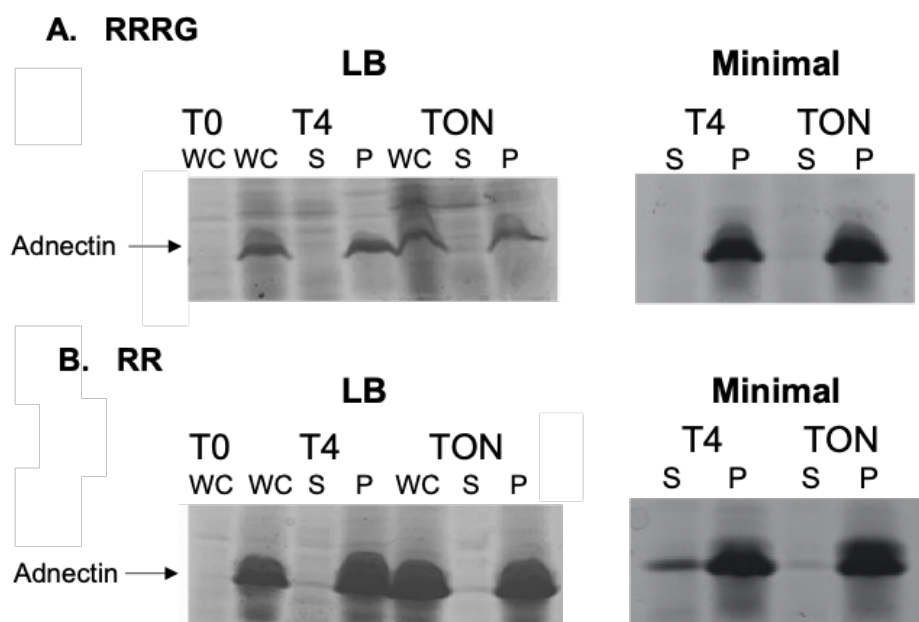


Figure 4.5: Soluble and insoluble expression levels of engineered Adnectin mutants in different media types by SDS-page gel analysis. A) Expression of RRRG in LB and M9 minimal media at 0, 4 and 24 hours post-induction with IPTG (T0, T4, and TON, respectively). B) Expression of RR in LB and M9 minimal media at 0, 4 and 24 hours post-induction with IPTG (T0, T4, and TON, respectively). Whole cell (WC) fraction was centrifuged to obtain soluble (S) and insoluble (P) material. T0 was omitted for M9 growths as LB growths showed no leaky expression.

extremely aggregation prone in both media types despite the expected solubilizing effect of surface exposed arginine residues (Fig. 4.5B). If the main pathway of aggregation is through the F-strand and FG-loop, it is possible that charged residues in the N-terminal strands would not have a large effect on solubility. However, it was unexpected that this construct would aggregate more than the parent Adnectin. Thus, we were hopeful that additional structural investigation would provide insight on this result.

4.4.2 FTIR and urea solubilization reveal subtle changes in secondary structure and stability of mutant IBs

FTIR analysis was completed for the new Adnectin aggregates in the same manner as applied previously. It was found that both RRRG and RR displayed similar absorbance profiles to parent Adnectin and other point-mutants (Fig. 4.6, 2.6). However, RRRG has considerably increased signal at 1653 cm^{-1} corresponding to increased random coil protein (Fig. 4.6A). This aligns with the expected changes caused by an extremely disrupted core that prevents proper secondary and tertiary structure from forming. Further small differences are revealed upon the examination of the second derivative spectra. The main β -sheet absorbance peak for RRRG shifts from 1628 cm^{-1} to 1625.6 cm^{-1} possibly indicating an increased contribution from amyloid species (Fig. 4.6B). While a small shift in this β peak is also seen for RR, when compared to parent, there is an additional peak shape change (loss of shoulder at 1620 cm^{-1}) making it difficult to conclude whether there is an increase in cross β species.

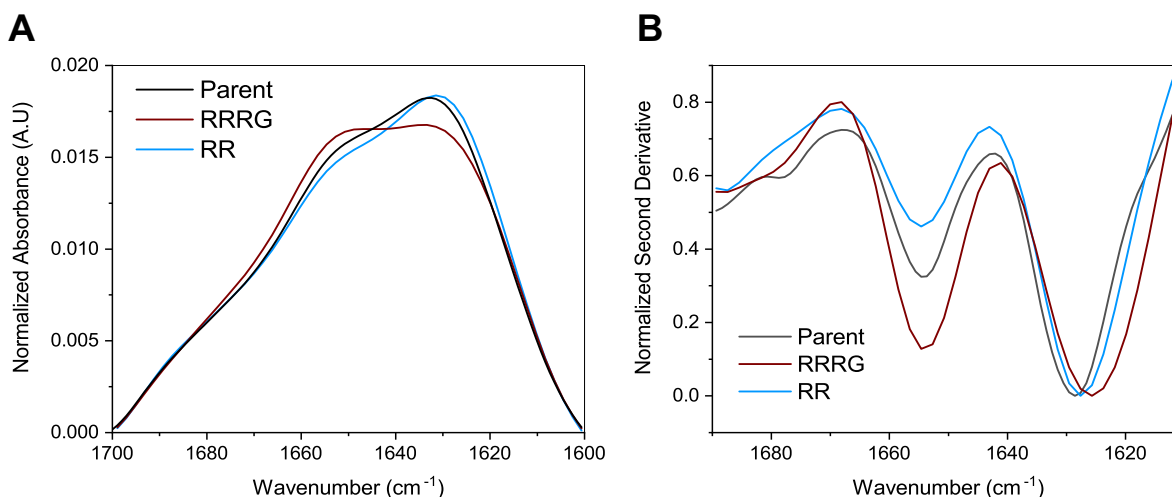


Figure 4.6: FTIR spectral comparison of parent and engineered Adnectin IBs. Area normalized ATR-FTIR spectra in the amide I region for IBs formed by engineered Adnectin mutants, and parent Adnectin. Each absorbance spectrum (A) is a normalized to an area of 1, before calculating the second derivative (B).

Urea solubilization of mutant IBs was also performed as in Chapter 2. It was found that both RRRG and RR have similar C_{mid} to parent IBs in that they are both lower than would be expected for the unfolding of native, monomeric Adnectin (Fig. 4.7). Interestingly, RRRG also has a lower C_{mid}

than would be expected given the trend that less stable variants have increasing C_{mid} due to stabilizing and potentially amyloidogenic interactions (Table 4.2). Thus, it is likely that the high stability of Adnectins means that while possible, amyloid formation is restricted from being a major pathway of aggregation. Furthermore, both engineered variants have higher m -values than parent IBs or any other point-mutant studied indicating a greater cooperativity of unfolding (Table 4.2). As described in section 2.3.5, an aggregate which exposes more protein surface to solvent upon disaggregation would have a higher m -value; this could be possible through a conformational change in either or both the aggregated or non-aggregated state (ie. state not contributing to turbidity). While it is difficult to quantify whether an amyloid conformation would have a lower solvent accessible area than the native protein due to variable lengths of protein participating in the fibril core, it is very plausible that at least RRRG would have a less compact non-aggregated state than the previously studied mutants due to destabilizing interactions in its core.

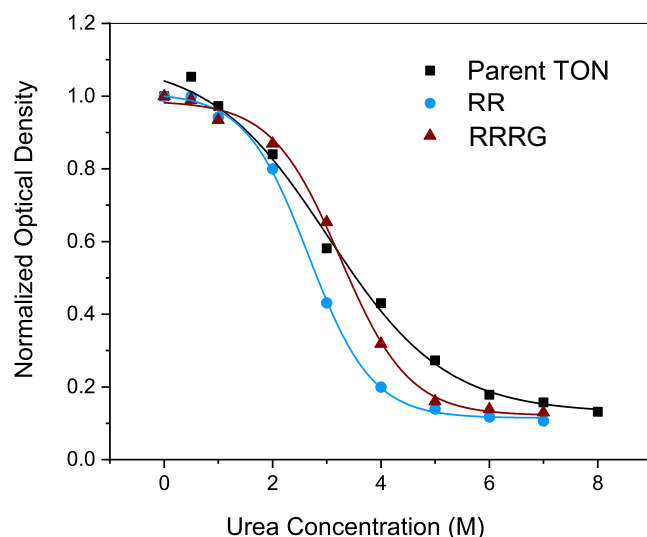


Figure 4.7: Urea solubilization profiles of parent Adnectin and engineered Adnectin variants. Fitting of solubilization profiles to a two-state denaturation model indicate midpoints of denaturation for RRRG and RR are similar to parent IBs. However, both engineered variants have increased m -values indicating increased cooperativity of disaggregation.

Table 4.2: Equilibrium denaturation parameters of parent and engineered Adnectin IBs. Data fit to a linear extrapolation model (See Section 2.6.6) using OriginPro 2018.

Variant	C_{mid} (M)	m (kcal/mol)
Parent	3.02 +/- 0.159	0.546 +/- 0.080
RRRG	3.25 +/- 0.060	0.937 +/- 0.082
RR	2.66 +/- 0.034	1.011 +/- 0.053

4.4.3 Residue-level structure by qHDX

To further investigate the residue-level structure and evaluate specific contributions from amyloid structures, qHDX was performed on the engineered variants. Due to multiple mutations, peak shifts/new assignments were evaluated before analysis. As the mutations were concentrated to a region within the N-terminal region (residues 11-22), chemical shifts were observed for more residues than previously anticipated (Fig. E.2) ¹¹⁶. However, many isolated peaks were easily tracked. Unfortunately residues A12, A13 and L18 are very close to one another in the ^1H - ^{15}N HSQC spectrum making peak tracking non-intuitive (Fig. E.2). Luckily, intrinsic rate constants demonstrate trends based on residue type; leucine residues tend to have slower intrinsic rates of exchange than alanine residues. Therefore, A12 and A13 were discerned from L18 based on concerted peak movements (upfield in the nitrogen dimension) and consistently faster intrinsic rate constants (Fig. E.3).

qHDX results reveal more defined structural changes (ie. outside error bars) with respect to parent Adnectin and the other point-mutants studied in previous chapters (Fig. E.4). Both variants show a lower overall protection by ^1H -NMR spectra than the previously studied Adnectin IBs indicating that there is a larger amount of the ‘fast-exchanging’ population typically associated with less-structured protein (Fig. E.5). However, with 1D NMR spectra, it is impossible to identify whether this is true within a monomer or whether there is an increased fraction of fully unfolded protein. More detailed study of RRRG demonstrates lower normalized protection values than parent Adnectin IBs in the most but not all regions of the protein, indicating a true change in the structural ensemble. However, near regions predicted to be the most amyloidogenic (Fig. 4.8A),

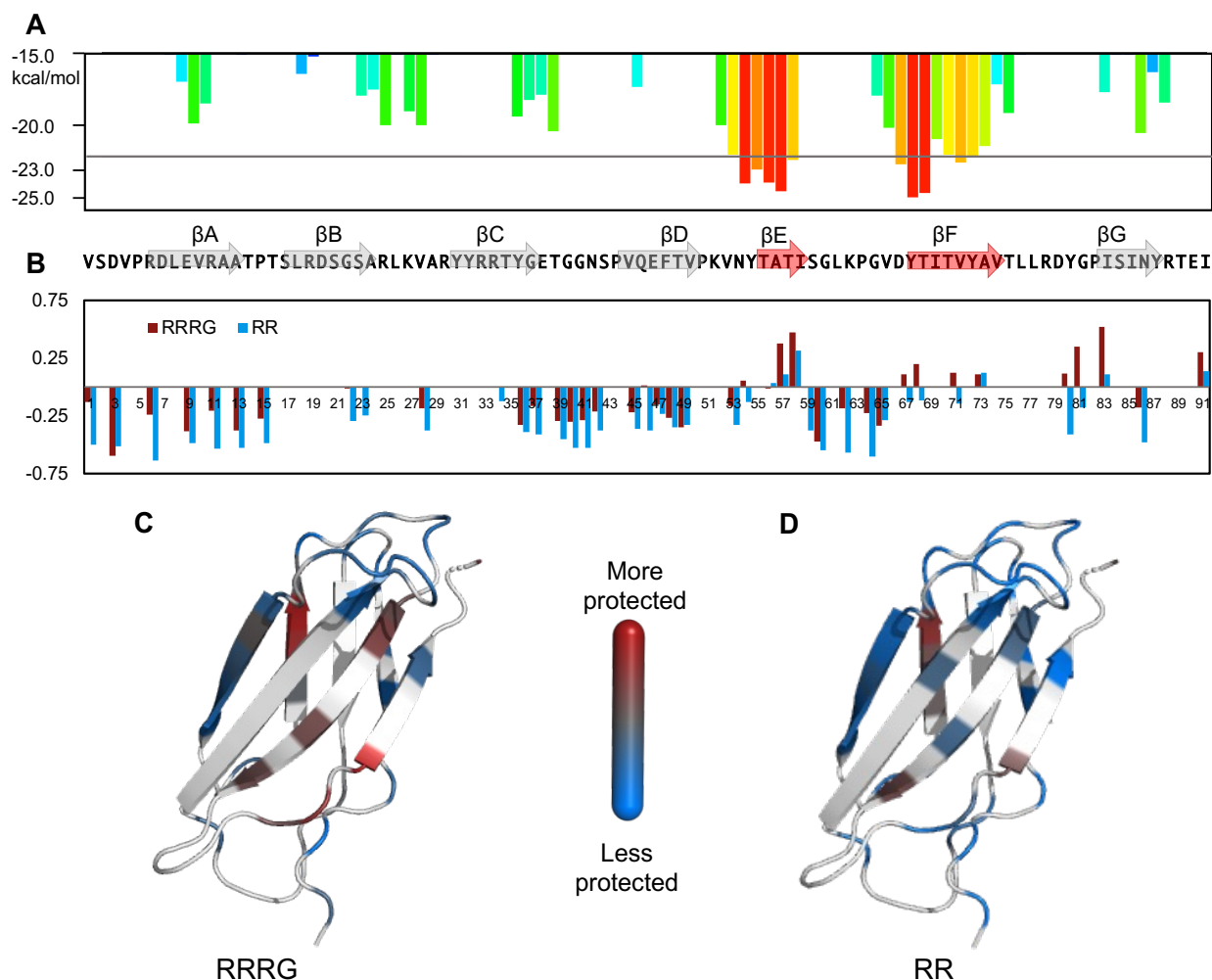


Figure 4.8: Average qHDX protection of engineered Adnectin IBs compared to parent when expressed at 37 °C overnight. A) ZipperDB ¹³³ aggregation propensity for RRRG Adnectin. Regions with a Rosetta energy score lower than -23 kcal mol⁻¹ are highlighted in orange/red and are predicted to assemble into steric zippers. B) Ratio of engineered Adnectin qHDX protection relative to the parent qHDX protection. A ratio greater/less than 1 indicates increased/decreased protection relative to the parent respectively. C/D) % difference in qHDX protection for RRRG (C) and RR (D) protection relative to the average parent qHDX protection. % difference in amide protection shown on a ribbon structure using color gradient of blue_gray_red representing a range of -50% to 50% change. White indicates regions of no data.

protection is increased as we hypothesized would be the case for a more destabilized protein (Fig. 4.8B,C). Interestingly, RR also exhibits some of these same changes, with even more pronounced decreased in the A-strand, CD, and EF loop regions accompanied by less prominent increases in the aggregation-prone regions (Fig 4.8B,C). This implies that the RR monomers making up the IB

are less tightly packed with one another and/or there has been a decrease in native hydrogen-bonding leading to a looser, but not necessarily more amyloid, aggregate. This is supported by urea solubilization data where RR shows the lowest C_{mid} value of any point-mutant studied herein. Therefore, it is likely that the neighboring arginine residues had a more destabilizing effect than was predicted by computational tools. This could be investigated by removing one of the residues or relocating one to a different aggregation-prone strand (F-strand). Alternatively, the folding behaviour and thermostability of purified RR could be investigated to make a more concrete mechanistic conclusion.

4.5 Conclusions

In summary, the data presented in this chapter supports previous hypotheses that the ensemble of structures that make up an IB structure can be rationally altered. Using a combination of computational stability and aggregation propensity predictors, an Adnectin variant was designed with structural properties more similar to those seen in previous IB studies⁴⁰. However, predicting and subsequently altering aggregation patterns remains non-trivial given the currently available techniques. The second engineered variant studied here behaved quite different than expected with a much lower overall extent of protection in qHDX studies and a lower stability towards urea solubilization while maintaining a high aggregation propensity and a large degree of native secondary structure as measured by FTIR. This information is pertinent for those looking to engineer biotherapeutics, especially those based on an immunoglobulin scaffold. First, it supports that a combined approach truly does have potential for successful protein design. However, predicting the effects of multiple mutations are not always trivial, and can affect the thermostability of proteins in cells in ways not predicted by computational techniques.

Chapter 5

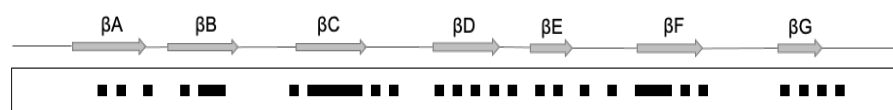
Conclusions and future directions.

5.1 Conclusions

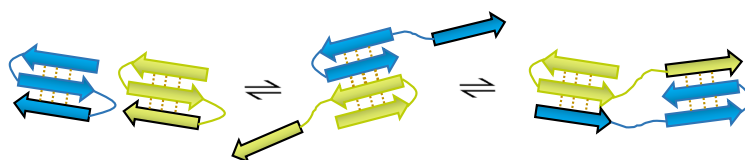
In summary, this thesis provides a comprehensive analysis of Adnectin IB structure through several different methods such as FTIR spectroscopy, protease digestion, amyloid-specific dye (CR) binding, urea solubilization and qHDX with NMR readout. FTIR reveals that Adnectin IBs contain considerable native structure, some disordered structures and some amyloid conformations, as they maintain significant anti-parallel signal without the significant right shift of the lower β -sheet component, typically attributed to amyloid fibril formation ⁴⁵. Additionally, it was determined that although portions of the aggregate were moderately resistant to protease digestion, these resistant regions did not bind strongly to CR dye. This supports that there is both unstructured protein and a considerable contribution from protein in a native-like state. Finally, urea solubilization confirmed that the point-mutant IBs were not only extremely similar in their secondary and tertiary structures but also in their stabilities unlike what was reported for other proteins ¹⁰⁵.

The structure of Adnectin IB aggregates was also investigated using qHDX. This method has been sparingly applied to cellular aggregates until a recent advancement of the method by other students in the Meiering group to study IBs formed by reduced apo SOD1 ¹¹⁶. The studies outlined within this thesis serve as a proof of concept that this method can be straightforwardly applied to other systems as well as an exploration of how IB structure may differ for a protein with such high thermostability. In agreement with the results of Chapter 2, parent, L78I (less soluble than parent) and V75R (more soluble than parent) display remarkably similar protection patterns. This aligns well with results from SOD1 point-mutation studies and can be explained by the ensemble model; differences in aggregate structure by qHDX are subtle as no single mutation can block all self-association and aggregation pathways ⁴¹. Notably, the subtle differences in the aggregate structures are concentrated to areas of interest as identified by studies of pure Adnectin ⁷², wild-type

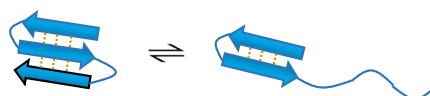
fibronectin⁵⁵ and Tencon⁵⁹ as well as areas predicted to be aggregation-prone/amyloidogenic by computational predictors¹³³. This contributes to an ensemble-view of an IB containing natively folded Adnectin, domain-swapped species, misfolded/locally unfolded structures and some amyloid-like fibrils (Fig. 5.1).



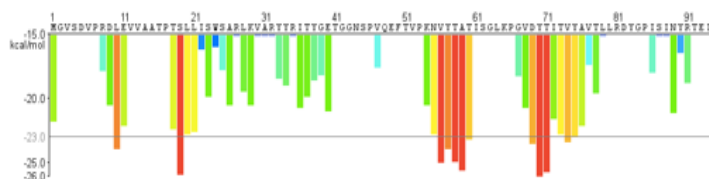
Pathway 1:
Higher order species made
of natively **folded** Adn



Pathway 2: Domain swapped
folded Adn monomers



Pathway 3: Local Unfolding



Pathway 4: Amyloid formation via
predicted regions

Figure 5.1: Adnectin IB aggregation involves multiple pathways including native and non-native interactions. Native secondary structure and qHDX protection results from an ensemble of natively folded Adnectin (Path 1), potentially domain swapped species in a near-native fold as reported for Tencon variants⁵⁹ (Path 2), as well as misfolded/locally unfolded species exposing hydrophobic binding sites⁵⁵ (Path 3) or amyloidogenic segments¹³³ (Path 4).

Finally, to help advance our knowledge of the molecular determinants of aggregation and deliberately alter the ensemble of structures, engineered variants with multiple point-mutations were rationally designed. A mutant with a disrupted hydrophobic core was designed with the goal of specifically preventing native-like association mechanisms and thereby promoting amyloid aggregation. Furthermore, two mutations in the middle of a proposed self-association state and charged mutations in the N-terminal strands, were made in hopes that alternate aggregation pathways were blocked. It was found that as in previous chapters average aggregate structure was still somewhat resilient to mutations. Differences in qHDX revealed that these designed mutations resulted in less extensive protection throughout the sequence with protection maintained mainly in the regions predicted to be amyloidogenic. However, other regions/pathways were not knocked out entirely. This confirmed hypotheses that amyloid was likely a contributor to Adnectin IB formation, but could only be a main pathway once others were prevented. This knowledge is critical for those who wish to design therapeutics based off the fibronectin scaffold, and ultimately contributes to our overall understanding of aggregation mechanisms.

5.2 Future Experiments

Size exclusion chromatography of urea solubilization products

Urea solubilization is able to provide information on the relative stabilities and heterogeneity of IB aggregates. Here we determine that midpoints of solubilization remain similar despite point-mutations and correlate reasonably well with the predicted stabilities of each native monomer. Another useful piece of information would be the identification of species released from the aggregate. Soluble but folded monomers, soluber oligomeric species (dimers and trimers) as well as unfolded species may be in the soluble fraction of a urea denaturation experiment, and the presence and relative amounts of these species may differ between Adnectin variants. To study this, size exclusion chromatography following a modified version of the protocol applied by Wang and Colón¹¹⁵ or Chow *et al.*¹⁸¹ may be used. Because elution volume is not only dependent on size, but also on hydrodynamic volume, through comparison of the elution volumes for solubilized aggregate peaks to pure monomeric Adnectin, one could in theory elucidate the conformations as well as their relative amounts that are released from Adnectin IBs.

Limited proteolysis and mass spectrometry

The Tris-Glycine SDS-page analysis performed for this thesis reveal that mutant IBs contain protease-resistant segments of similar size. To further pinpoint what regions of the protein sequence are involved in these fragments, mass spectrometry (MS) should be performed. The composition of the main resistant fragments could be identified using MALDI-TOF MS. This has been applied previously to aggregates produced *in vitro* from pure protein^{182–184} and aggregation intermediates¹⁸⁵ but has not yet been applied to study cellular samples. Considerations will need to be taken to ensure that fragments can be linked to the protein of interest and not a cellular contaminant.

Investigate soluble precursors of aggregation

As discussed in section 3.3.3, studying early aggregates is of great interest for elucidating what species are involved in initiating aggregation. Another avenue of pursuing this it through the investigation of the soluble cellular fraction to try and identify small soluble oligomers that are precursors to aggregate formation. Potential methods of study include FTIR and NMR spectroscopy given that there is sufficient soluble Adnectin in the soluble portion of the cell lysate to outcompete signal from other contaminants.

Thorough characterization of purified engineered variants

Deciphering mechanisms of aggregation relies heavily on information gained from the study of purified protein. At the time of this thesis, engineered variants RRRG and RR have not yet been thoroughly characterized in their purified form. Through NMR analysis of charged mutation V75R, decreased self-association is revealed by additional cross-peaks and more uniform peak shape and intensity⁷². Initial NMR studies of pure RR (unpublished) reveal that the double charged mutant does not achieve the same solubilizing affect as V75R despite similar design strategies. To better predict how these mutants may be behaving in a cellular environment, it would be advantageous to study these variants further through thermal denaturation data obtained using differential scanning calorimetry, secondary structure data obtained using FTIR spectroscopy and structural/dynamics data obtained using NMR spectroscopy, in particular HDX.

Additional molecular dynamics simulations

Finally, computational studies have shown great advances in recent years in identifying aggregation-prone conformations of many globular proteins^{186–188}. Unfortunately the simulations performed here are restricted by the quality of the starting PDB structure and the implicit solvent model used. Accurate interpretations of structural data could be improved by more detailed molecular dynamics simulations with a more accurate starting structure (ie. crystal or NMR structure) and an explicit solvent model.

References

1. Schaefer, A., Naser, D., Siebeneichler, B., Tarasca, M. V & Meiering, E. Methodological advances and strategies for high resolution structure determination of cellular protein aggregates. *J. Biol. Chem.* (2022).
2. Lehinger, A. L., Nelson, D. L. & Cox, M. M. *Principles of Biochemistry*. W.H Freeman (2013).
3. Anfinsen, C. B., Haber, E., Selajf, M. & White, F. H. The Kinetics of Formation of Native Ribonuclease During Oxidation of the Reduced Polypeptide. *Proc. Natl. Acad. Sci.* **47**, 1309–1314 (1961).
4. Anfinsen, C. B. Principles that Govern the Folding of Protein Chains. *Sci. New Ser.* **181**, 223–230 (1973).
5. Levinthal, C. How to fold graciously. in *Proceedings of a meeting held at Alberton house, Monticello, Illinois* (eds. DeBrunner, P., Tsibris, J. & Munck, E.) 22–24 (University of Illinois Press, 1969). doi:10.4236/ajmb.2012.23019
6. Levinthal, C. Are there pathways to protein folding? *J. Chim. Phys.* **65**, 44–45 (1968).
7. Baker, D. & Agard, D. A. Kinetics versus Thermodynamics in Protein Folding. *Biochemistry* **33**, 7505–7509 (1994).
8. Dill, K. A. & Chan, H. S. From Levinthal to Pathways to Funnels: The “New View” of Protein Folding Kinetics. *Nat. Struct. Biol.* **4**, 10 (1997).
9. Goldbeck, R. A., Thomas, Y. G., Chen, E., Esquerra, R. M. & Kliger, D. S. Multiple pathways on a protein-folding energy landscape: Kinetic evidence. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 2782–2787 (1999).
10. Best, R. B. Atomistic molecular simulations of protein folding. *Current Opinion in Structural Biology* **22**, 52–61 (2012).
11. Moreno-Gonzalez, I. & Soto, C. Misfolded protein aggregates: Mechanisms, structures and potential for disease transmission. *Seminars in Cell and Developmental Biology* **22**, 482–487 (2011).
12. Ramirez-Alvarado, M. Principles of Protein Misfolding. *Prog. Mol. Biol. Transl. Sci.* **84**, 115–160 (2008).
13. Fersht, A. *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*. (W.H. Freeman and Company, 1999).
14. Marinelli, P. *et al.* Global Protein Stabilization Does Not Suffice to Prevent Amyloid Fibril Formation. *ACS Chem. Biol.* **13**, 2094–2105 (2018).
15. Iadanza, M. G., Jackson, M. P., Hewitt, E. W., Ranson, N. A. & Radford, S. E. A New Era for Understanding Amyloid Structures and Disease. *Nat. Rev. Mol. Cell Biol.* **19**, 755–773 (2018).
16. Chiti, F. & Dobson, C. M. Protein Misfolding, Functional Amyloid, and Human Disease. *Annu. Rev. Biochem.* **75**, 333–366 (2006).
17. Chiti, F. & Dobson, C. M. Amyloid formation by globular proteins under native conditions. *Nat. Chem. Biol.* **5**, 15–22 (2009).
18. Tyedmers, J., Mogk, A. & Bukau, B. Cellular strategies for controlling protein aggregation. *Nat. Rev. Mol. Biol.* **11**, 777–788 (2010).
19. Sunde, M. *et al.* Common core structure of amyloid fibrils by synchrotron X-ray diffraction. *J. Mol. Biol.* **273**, 729–739 (1997).
20. Chatani, E. & Yamamoto, N. Recent progress on understanding the mechanisms of amyloid nucleation. *Biophys. Rev.* **10**, 527–534 (2018).
21. Bowden, G. A., Paredes, A. M. & Georgiou, G. Structure and Morphology of Protein Inclusion Bodies in Escherichia Coli. *Nat. Biotechnol.* **9**, 725–730 (1991).
22. Fink, A. L. Protein aggregation: folding aggregates, inclusion bodies and amyloid. *Fold. Des.* **3**, R9–R23 (1998).
23. Rinas, U. *et al.* Bacterial Inclusion Bodies: Discovering Their Better Half. *Trends Biochem. Sci.*

- 42**, 726–737 (2017).
24. Huang, D. *et al.* Antiparallel β -Sheet Structure within the C-Terminal Region of 42-Residue Alzheimer's Amyloid- β Peptides When They Form 150-kDa Oligomers. *J. Mol. Biol.* **427**, 2319–2328 (2015).
 25. Stroud, J. C., Liu, C., Teng, P. K. & Eisenberg, D. Toxic fibrillar oligomers of amyloid- β have cross- β structure. *Proc. Natl. Acad. Sci.* **109**, 7717–7722 (2012).
 26. Reinke, A. A. & Gestwicki, J. E. Insight into amyloid structure using chemical probes. *Chem. Biol. Drug Des.* **77**, 399–411 (2011).
 27. Sulatsky, M. I. *et al.* Effect of the fluorescent probes ThT and ANS on the mature amyloid fibrils. *Prion* **14**, 67–75 (2020).
 28. Yakupova, E. I., Bobyleva, L. G., Vikhlyantsev, I. M. & Bobylev, A. G. Congo Red and amyloids: history and relationship. *Biosci. Rep.* **39**, 1–22 (2019).
 29. Biancalana, M. & Koide, S. Molecular mechanism of Thioflavin-T binding to amyloid fibrils. *Biochim. Biophys. Acta - Proteins Proteomics* **1804**, 1405–1412 (2010).
 30. Masino, L. *et al.* Domain architecture of the polyglutamine protein ataxin-3: a globular domain followed by a flexible tail. *FEBS Lett.* **549**, 21–25 (2003).
 31. Singh, A., Upadhyay, V., Singh, A. & Panda, A. K. Structure-Function Relationship of Inclusion Bodies of a Multimeric Protein. *Front. Microbiol.* **11**, 1–10 (2020).
 32. Yoshinaga, S. *et al.* Preserved proteinase K-resistant core after amplification of alpha-synuclein aggregates: Implication to disease-related structural study. *Biochem. Biophys. Res. Commun.* **522**, 655–661 (2020).
 33. Polverino De Laureto, P. *et al.* Protein Aggregation and Amyloid Fibril Formation by an SH3 Domain Probed by Limited Proteolysis. *J. Mol. Biol.* **334**, 129–141 (2003).
 34. Upadhyay, A. K., Murmu, A., Singh, A. & Panda, A. K. Kinetics of Inclusion Body Formation and Its Correlation with the Characteristics of Protein Aggregates in *Escherichia coli*. *PLoS One* **7**, 33951 (2012).
 35. Pelton, J. T. & Mclean, L. R. Spectroscopic Methods for Analysis of Protein Secondary Structure. *Anal. Biochem.* **277**, 167–176 (2000).
 36. Sarroukh, R., Goormaghtigh, E., Ruysschaert, J. M. & Raussens, V. ATR-FTIR: A 'rejuvenated' tool to investigate amyloid proteins. *Biochim. Biophys. Acta - Biomembr.* **1828**, 2328–2338 (2013).
 37. Milošević, J., Prodanović, R. & Polović, N. On the Protein Fibrillation Pathway: Oligomer Intermediates Detection Using ATR-FTIR Spectroscopy. *Molecules* **26**, 970 (15 pages) (2021).
 38. Hoshino, M. *et al.* Mapping the core of the 2-microglobulin amyloid fibril by H/D exchange. *Nat Struct Biol* **9**, 332–336 (2002).
 39. Alexandrescu, A. An NMR-based quenched hydrogen exchange investigation of model amyloid fibrils formed by cold shock protein A. *Pacific Symp. Biocomput.* **78**, 67–78 (2001).
 40. Wang, L., Maji, S. K., Sawaya, M. R., Eisenberg, D. & Riek, R. Bacterial Inclusion Bodies Contain Amyloid-Like Structure. *PLoS Biol* **6**, 1791–1801 (2008).
 41. Naser, D. *et al.* High Resolution NMR H/D Exchange of Human Superoxide Dismutase Inclusion Bodies Reveals Significant Native Features Despite Structural Heterogeneity. *Angew. Chemie - Int. Ed.* **61**, e202112645 (2022).
 42. Tycko, R. Molecular Structure of Aggregated Amyloid- β : Insights from Solid-State Nuclear Magnetic Resonance. *Cold Spring Harb. Perspect. Med.* **6**, (2016).
 43. Seuring, C. *et al.* The three-dimensional structure of human β -endorphin amyloid fibrils. *Nat. Struct. Mol. Biol.* **27**, 1178–1184 (2020).
 44. Oberg, K., Chrnyk, B. A., Wetzel, R. & Fink, A. L. Nativelike Secondary Structure in Interleukin-13 Inclusion Bodies by Attenuated Total Reflectance FTIR. *Biochemistry* **33**, 2628–2634 (1994).
 45. Zandomenighi, G., Krebs, M. R. H., McCammon, M. G. & Fändrich, M. FTIR reveals structural differences between native β -sheet proteins and amyloid fibrils. *Protein Sci.* **13**, 3314–3321

- (2004).
46. Carrió, M., González-Montalbán, N., Vera, A., Villaverde, A. & Ventura, S. Amyloid-like properties of Bacterial Inclusion Bodies. *J. Mol. Biol.* **347**, 1025–1037 (2005).
 47. Upadhyay, A. *et al.* Complex Inclusion Bodies and Defective Proteosome Hubs in Neurodegenerative Disease: New Clues, New Challenges. *Neurosci.* **28**, 271–282 (2022).
 48. Navarro, S., Carija, A., Muñoz-Torrero, D. & Ventura, S. A fast and specific method to screen for intracellular amyloid inhibitors using bacterial model systems. *Eur. J. Med. Chem.* **121**, 785–792 (2016).
 49. S. Pouplana *et al.* Thioflavin-S Staining of Bacterial Inclusion Bodies for the Fast, Simple, and Inexpensive Screening of Amyloid Aggregation. *Curr. Med. Chem.* **21**, 1152–1159 (2014).
 50. Villar-Piqué, A., Espargaró, A., Sabaté, R., de Groot, N. S. & Ventura, S. Using bacterial inclusion bodies to screen for amyloid aggregation inhibitors. *Microb. Cell Fact.* **11**, 55 (2012).
 51. Espargaró, A., Villar-Piqué, A., Sabaté, R. & Ventura, S. Yeast prions form infectious amyloid inclusion bodies in bacteria. *Microb. Cell Fact.* **11**, (2012).
 52. Cano-Garrido, O. *et al.* Supramolecular organization of protein-releasing functional amyloids solved in bacterial inclusion bodies. *Acta Biomater.* **9**, 6134–6142 (2012).
 53. Virtudes Céspedes, M. *et al.* Engineering Secretory Amyloids for Remote and Highly Selective Destruction of Metastatic Foci. *Adv. Mater.* **32**, 1907348–1907348 (2020).
 54. Zollinger, A. J. & Smith, M. L. Fibronectin, the extracellular glue. *Matrix Biol.* **60–61**, 27–37 (2017).
 55. Gee, E. P. S., Yü Ksel, D., Stultz, C. M. & Ingber, D. E. SLLISWD sequence in the 10FNIII domain initiates fibronectin fibrillogenesis. *J. Biol. Chem.* **288**, 21329–21340 (2013).
 56. Sechler, J. L. *et al.* A novel fibronectin binding site required for fibronectin fibril growth during matrix assembly. *J. Cell Biol.* **154**, (2001).
 57. Neale, D. B. *et al.* Aligned Networks of Engineered Fibrillar Fibronectin Guide Cellular Orientation and Motility. *Small Struct.* **2**, 2000137 (2021).
 58. Chandler, P. G. & Buckle, A. M. Development and Differentiation in Monobodies Based on the Fibronectin Type 3 Domain. *Cells* **9**, 610 (2020).
 59. Teplyakov, A. *et al.* C-terminal β -strand swapping in a consensus-derived fibronectin Type III scaffold. *Proteins Struct. Funct. Bioinforma.* **82**, 1359–1369 (2014).
 60. Arnold, K., Bordoli, L., Rgen Kopp, J. & Schwede, T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22**, 195–201 (2006).
 61. Delano, W. L. The PyMOL Molecular Graphics System. (2002).
 62. Donnelly, D. J. *et al.* Synthesis and Biologic Evaluation of a Novel 18 F-Labeled Adnectin as a PET Radioligand for Imaging PD-L1 Expression. *J Nucl Med* **59**, 529–535 (2018).
 63. Lipovsek, D. Adnectins: engineered target-binding protein therapeutics. *Protein Eng. Des. Sel.* **24**, 3–9 (2011).
 64. Mitchell, T. *et al.* Pharmacologic profile of the adnectin BMS-962476, a small protein biologic alternative to PCSK9 antibodies for low-density lipoprotein lowering. *J. Pharmacol. Exp. Ther.* **350**, 412–424 (2014).
 65. Ackermann, M., Morse, B. A., Delventhal, V., Carvajal, I. M. & Konerding, M. A. Anti-VEGFR2 and anti-IGF-1R-Adnectins inhibit Ewing’s sarcoma A673-xenograft growth and normalize tumor vascular architecture. *Angiogenesis* **15**, 685–695 (2012).
 66. Cota, E., Hamill, S. J., Fowler, S. B. & Clarke, J. Two proteins with the same structure respond very differently to mutation: The role of plasticity in protein stability. *J. Mol. Biol.* **302**, 713–725 (2000).
 67. Trainor, K. *et al.* Ensemble Modeling and Intracellular Aggregation of an Engineered Immunoglobulin-Like Domain. *J. Mol. Biol.* **428**, 1365–1374 (2016).
 68. Trainor, K. Adnectin Solubility and Dynamics. (University of Waterloo, 2019).
 69. Rennella, E. *et al.* Equilibrium unfolding thermodynamics of β 2-microglobulin analyzed through native-state H/D exchange. *Biophys. J.* **96**, 169–179 (2009).

70. Gebauer, M. & Skerra, A. Engineered Protein Scaffolds as Next-Generation Therapeutics. *Annu. Rev. Pharmacol. Toxicol.* **60**, 391–415 (2020).
71. Elgundi, Z., Reslan, M., Cruz, E., Sifniotis, V. & Kayser, V. The state-of-play and future of antibody therapeutics. *Adv. Drug Deliv. Rev.* **122**, 2–19 (2017).
72. Trainor, K. *et al.* Design for Solubility May Reveal Induction of Amide Hydrogen/Deuterium Exchange by Protein Self-Association. *J. Mol. Biol.* **434**, (2022).
73. Barth, A. Infrared spectroscopy of proteins. *Biochim. Biophys. Acta - Bioenerg.* **1767**, 1073–1101 (2007).
74. Wilkosz, N. *et al.* Molecular spectroscopic markers of abnormal protein aggregation. *Molecules* **25**, 1–26 (2020).
75. Goormaghtigh, E., Cabiaux, V. & Ruyschaert, J. M. Determination of Soluble and Membrane Protein Structure by Fourier Transform Infrared Spectroscopy. in *Physicochemical Methods in the Study of Biomembranes. Subcellular Biochemistry* (eds. Hilderson, H. . & Ralston, G. B.) **23**, (Springer, 1994).
76. Cerf, E. *et al.* Antiparallel β -sheet: a signature structure of the oligomeric amyloid β -peptide. *Biochem J* **421**, 415–423 (2009).
77. Zhou, L. & Kurouski, D. Structural Characterization of Individual α -Synuclein Oligomers Formed at Different Stages of Protein Aggregation by Atomic Force Microscope Infrared Spectroscopy. . *Anal. Chem.* **92**, 6806–6810 (2020).
78. Rinas, U., Hoffmann, F., Betiku, E., Estapé, D. & Marten, S. Inclusion body anatomy and functioning of chaperone-mediated in vivo inclusion body disassembly during high-level recombinant protein production in *Escherichia coli*. *J. Biotechnol.* **127**, 244–257 (2007).
79. Bennion, B. J. & Daggett, V. *The molecular basis for the chemical denaturation of proteins by urea. PNAS* **100**, (2003).
80. Oberg, K. A. & Fink, A. L. Methods for collecting and analyzing attenuated total reflectance FTIR spectra of proteins in solution. *Tech. Protein Chem.* **6**, 475–484 (1995).
81. Yang, H., Yang, S., Kong, J., Dong, A. & Yu, S. Obtaining information about protein secondary structures in aqueous solution using Fourier transform IR spectroscopy. *Nat. Protoc.* **10**, 382–396 (2015).
82. Klunk, W. E., Pettegrew, J. W. & Abraham, D. J. Quantitative Evaluation of Congo Red Binding to Amyloid-like Proteins with a Beta-pleated Sheet Conformation'. *J. Histochem. Cytochemistry* **37**, 1273–1281 (1989).
83. Li, Z., Nimtz, M. & Rinas, U. *The metabolic potential of Escherichia coli BL21 in defined and rich medium.* (2014).
84. Jiang, X.-R., Wang, H., Shen, R. & Chen, G.-Q. Engineering the bacterial shapes for enhanced inclusion bodies accumulation. (2015). doi:10.1016/j.ymben.2015.03.017
85. Massiah, M. A., Wright, K. M. & Du, H. Obtaining soluble folded proteins from inclusion bodies using sarkosyl, triton X-100, and CHAPS: Application to LB and M9 minimal media. *Curr. Protoc. Protein Sci.* **2016**, 6.13.1-6.13.24 (2016).
86. Schmidt, A. *et al.* The quantitative and condition-dependent *Escherichia coli* proteome. **34**, (2016).
87. Miller, L. M., Bourassa, M. W. & Smith, R. J. FTIR spectroscopic imaging of protein aggregation in living cells. *Biochim. Biophys. Acta* **1828**, 2339–2346 (2013).
88. Levitt, M. & Greer, J. Automatic identification of secondary structure in globular proteins. *J. Mol. Biol.* **114**, 181–239 (1977).
89. Dong, A., Huang, P. & Caughey, W. S. Protein Secondary Structures in Water from Second-Derivative Amide I Infrared Spectra. *Biochemistry* **29**, 3303–3308 (1990).
90. Shivu, B. *et al.* Distinct β -Sheet Structure in Protein Aggregates Determined by ATR–FTIR Spectroscopy. *Biochemistry* **52**, 5176–5183 (2013).
91. Celej, M. S. *et al.* Toxic prefibrillar α -synuclein amyloid oligomers adopt a distinctive antiparallel β -sheet structure. *Biochem J* **443**, 719–726 (2012).
92. Sarroukh, R. *et al.* Transformation of amyloid β (1-40) oligomers into fibrils is characterized by a

- major change in secondary structure. *Cell. Mol. Life Sci.* **68**, 1429–1438 (2011).
93. Stathopoulos, P. B. *et al.* Sonication of proteins causes formation of aggregates that resemble amyloid. *Protein Sci.* **13**, 3017–3027 (2004).
 94. Natalello, A. *et al.* Wild Type Beta-2 Microglobulin and DE Loop Mutants Display a Common Fibrillar Architecture. *PLoS One* **10**, (2015).
 95. Natalello, A. & Doglia, S. M. Insoluble Protein Assemblies Characterized by Fourier Transform Infrared Spectroscopy. in *Insoluble Protein Assemblies Characterized by Fourier Transform Infrared Spectroscopy* (ed. Garcia-Fruitos, E.) **1258**, (Springer Science , 2015).
 96. Lazar, I. J. & Lazar, I. S. GelAnalyzer. (2010).
 97. Tsolis, A. C., Papandreou, N. C., Iconomidou, V. A. & Hamodrakas, S. J. A Consensus Method for the Prediction of ‘Aggregation-Prone’ Peptides in Globular Proteins. *PLoS One* **8**, e54175 (2013).
 98. Trainor, K. Inclusion Body Formation by Mutants of the Tenth Human Fibronectin Type III Domain. (University of Waterloo, 2015).
 99. Naser, D. Cellular Inclusion Bodies: Structures and Mechanisms. (University of Waterloo, 2022).
 100. Balguerie, A. *et al.* Domain organization and structure±function relationship of the HET-s prion protein of *Podospora anserina*. *EMBO J.* **22**, 2071–2081 (2003).
 101. Schönfelder, J. *et al.* Amyloid Protease resistance of ex vivo amyloid fibrils implies the proteolytic selection of disease-associated fibril morphologies Protease resistance of ex vivo amyloid fibrils implies the proteolytic selection of disease-associated fibril morphologies. *Amyloid* **28**, 243–251 (2021).
 102. Tanaka, G. *et al.* Sequence- and seed-structure-dependent polymorphic fibrils of alpha-synuclein. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1865**, 1410–1420 (2019).
 103. Morgan, G. J. & Kelly, J. W. The Kinetic Stability of a Full-Length Antibody Light Chain Dimer Determines whether Endoproteolysis Can Release Amyloidogenic Variable Domains. *J. Mol. Biol.* **428**, 4280–4297 (2016).
 104. Espargaró, A. *et al.* On the Binding of Congo Red to Amyloid Fibrils. *Angew. Chemie* **132**, 8181–8184 (2020).
 105. Espargaró, A., Sabaté, R. & Ventura, S. Kinetic and thermodynamic stability of bacterial intracellular aggregates. *FEBS Lett.* **582**, 3669–3673 (2008).
 106. Baldwin, A. J. *et al.* Metastability of Native Proteins and the Phenomenon of Amyloid Formation. *J. Am. Chem. Soc.* **133**, 42 (2011).
 107. Villar-Pique, A. *et al.* The effect of amyloidogenic peptides on bacterial aging correlates with their intrinsic aggregation propensity. *J. Mol. Biol.* **421**, 270–281 (2012).
 108. Eftink, M. R. The use of fluorescence methods to monitor unfolding transitions in proteins. *Biophys. J.* **66**, 482–501 (1994).
 109. Vivian, J. T. & Callis, P. R. Mechanisms of tryptophan fluorescence shifts in proteins. *Biophys. J.* **80**, 2093–2109 (2001).
 110. Koide, A., Jordan, M. R., Horner, S. R., Batori, V. & Koide, S. Stabilization of a Fibronectin Type III Domain by the Removal of Unfavorable Electrostatic Interactions on the Protein Surface †. (2001). doi:10.1021/bi010916y
 111. Dumoulin, M. *et al.* Single-domain antibody fragments with high conformational stability. *Protein Sci.* **11**, 500–515 (2002).
 112. Halverson, K., Fraser, P. E., Kirschner, D. A. & Lansbury, P. T. Molecular Determinants of Amyloid Deposition in Alzheimer’s Disease: Conformational Studies of Synthetic β -Protein Fragments¹. *Biochemistry* **29**, 2639–2644 (1990).
 113. Myers, J. K., Pace, C. N., Scholtz, J. M. & Scholtz, M. *m values and heat capacity changes: Relation to changes in accessible surface areas of protein unfolding.* *Protein Science* **4**, (Cambridge University Press, 1995).
 114. Upadhyay, A. K. *et al.* Refolding and purification of recombinant L-asparaginase from inclusion bodies of *E. coli* into active tetrameric protein. *Front. Microbiol.* **5**, 486 (2014).

115. Wang, L. & Colon, W. Urea-induced denaturation of apolipoprotein serum amyloid A reveals marginal stability of hexamer. *Protein Sci.* **14**, 1811–1817 (2005).
116. Tarasca, M. V, Naser, D., Schaefer, A., Soule, T. G. & Meiering, E. M. Quenched hydrogen-deuterium amide exchange optimization for high-resolution structural analysis of cellular protein aggregates. *Anal. Biochem.* **652**, (2022).
117. Rule, G. & Hitchens, T. K. *The Fundamentals of Protein NMR Spectroscopy*. (Springer Netherlands, 2006).
118. Doucleff, M., Hatcher-Skeers, M. & Crane, N. J. *Pocket Guide to Biomolecular NMR*. (Springer Berlin Heidelberg, 2011). doi:<https://doi.org/10.1007/978-3-642-16251-0>
119. Dyson, H. J. & Wright, P. E. Unfolded proteins and protein folding studied by NMR. *Chem. Rev.* **104**, 3607–3622 (2004).
120. Karamanos, T. K., Kalverda, A. P., Thompson, G. S. & Radford, S. E. Mechanisms of amyloid formation revealed by solution NMR. *Prog Nucl Magn Reson Spectrosc.* **88–89**, 86–104 (2015).
121. Baldwin, A. J. & Kay, L. E. NMR spectroscopy brings invisible protein states into focus. (2009). doi:10.1038/nchembio.238
122. Zhang, Y.-Z., Paterson, Y. & Roder, H. Rapid amide proton exchange rates in peptides and proteins measured by solvent quenching and two-dimensional NMR. *Protein Sci.* **4**, 804–814 (1995).
123. Arakawa, T., Kita, Y. & Timasheff, S. N. Protein precipitation and denaturation by dimethyl sulfoxide. *Biophys. Chem.* **131**, 62–70 (2007).
124. Hoshino, M., Katou, H., Yamaguchi, K.-I. & Goto, Y. Dimethylsulfoxide-quenched hydrogen/deuterium exchange method to study amyloid fibril structure. **1768**, 1886–1899 (2007).
125. Alexandrescu, A. T. Quenched hydrogen exchange NMR of amyloid fibrils. *Methods Mol. Biol.* **1345**, 211–222 (2016).
126. Ippel, J. H., Olofsson, A., Rgen Schleucher, J., Lundgren, E. & Wijmenga, S. S. Probing solvent accessibility of amyloid fibrils by solution NMR spectroscopy. *PNAS* **99**, 2021 (2002).
127. Whittemore, N. A. *et al.* Hydrogen-Deuterium (H/D) Exchange Mapping of A 1-40 Amyloid Fibril Secondary Structure Using Nuclear Magnetic Resonance Spectroscopy. *Biochemistry* **44**, 4434–4441 (2005).
128. Olofsson, A., Sauer-Eriksson, A. E. & Öhman, A. The Solvent Protection of Alzheimer Amyloid- β -(1–42) Fibrils as Determined by Solution NMR Spectroscopy. *J. Biol. Chem.* **281**, 477–483 (2006).
129. Wang, L., Schubert, D., Sawaya, M. R., Eisenberg, D. & Riek, R. Multidimensional structure-activity relationship of a protein in its aggregated states. *Angew. Chemie - Int. Ed.* **49**, 3904–3908 (2010).
130. Dasari, M. *et al.* Bacterial Inclusion Bodies of Alzheimer’s Disease β -Amyloid Peptides Can Be Employed To Study Native-Like Aggregation Intermediate States. *ChemBioChem* **12**, 407–423 (2011).
131. Rennella, E., Morgan, G. J., Yan, N., Kelly, J. W. & Kay, L. E. The Role of Protein Thermodynamics and Primary Structure in Fibrillogenesis of Variable Domains from Immunoglobulin Light Chains. *J. Am. Chem. Soc.* **141**, 13562 (2019).
132. Damo, S. M. *et al.* Probing the Conformation of a Prion Protein Fibril with Hydrogen Exchange. *J. Biol. Chem.* **285**, 32303–32311 (2010).
133. Thompson, M. J. *et al.* The 3D profile method for identifying fibril-forming segments of proteins. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 4074–4078 (2006).
134. Ivanova, M. I. *et al.* Aggregation-triggering segments of SOD1 fibril formation support a common pathway for familial and sporadic ALS. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 197–201 (2014).
135. Bai, Y., Milne, J., Mayne, L. & Englander, S. Primary structure effects on peptide group hydrogen exchange. *Proteins* **17**, 75–86 (1993).
136. Mishra, G. G. Structural characterization of f-ALS associated SOD1 mutant aggregates using conformation-specific antibodies. (University of Waterloo, 2020).

137. Englander, S. W., Sosnick, T. R., Englander, J. J. & Mayne, L. Mechanisms and uses of hydrogen exchange. *Curr. Opin. Struct. Biol.* **6**, 18–23 (1996).
138. Dickinson, C. D. *et al.* Crystal Structure of the Tenth Type III Cell Adhesion Module of Human Fibronectin. *J. Mol. Biol.* **236**, 1079–1092 (1994).
139. Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132 (1982).
140. Fernandez-Escamilla, A. M., Rousseau, F., Schymkowitz, J. & Serrano, L. Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat. Biotechnol.* **22**, 1302–1306 (2004).
141. Yamaguchi, K. I. *et al.* Core and Heterogeneity of β 2-Microglobulin Amyloid Fibrils as Revealed by H/D Exchange. *J. Mol. Biol.* **338**, 559–571 (2004).
142. Cota, E. & Clarke, J. Folding of beta-sandwich proteins: Three-state transition of a fibronectin type III module. *Protein Sci.* **9**, 112–120 (2000).
143. Kuroda, Y. *et al.* Effects of detergents on the secondary structures of prion protein peptides as studied by CD spectroscopy. *J. Pept. Sci.* **9**, 212–220 (2003).
144. Wasylewski, Z. & Kozik, A. Protein—Non-ionic Detergent Interaction. *Eur. J. Biochem.* **95**, 121–126 (1979).
145. Ghobadi, S., Safarian, S., Moosavi-Movahedi, A. A. & Ranjbar, B. Octyl Glucoside Induced Formation of the Molten Globule-Like State of Glutamate Dehydrogenase. *Japanese Biochem. Soc.* **130**, 671–677 (2001).
146. Goldenzweig, A. & Fleishman, S. J. Principles of Protein Stability and Their Application in Computational Design. *Annu. Rev. Biochem.* **87**, 105–129 (2018).
147. Sachsenhauser, V. & Bardwell, J. C. Directed evolution to improve protein folding in vivo. *Curr. Opin. Struct. Biol.* **48**, 117–123 (2018).
148. Broom, A., Jacobi, Z., Trainor, K. & Meiering, E. M. Computational tools help improve protein stability but with a solubility tradeoff. *J. Biol. Chem.* **292**, 14349–14361 (2017).
149. Broom, A. *et al.* Designed protein reveals structural determinants of extreme kinetic stability. *Proc. Natl. Acad. Sci.* **112**, 14605–114610 (2015).
150. Broom, A., Trainor, K., Jacobi, Z. & Meiering, E. M. Computational Modeling of Protein Stability: Quantitative Analysis Reveals Solutions to Pervasive Problems. *Structure* **28**, 717–726.e3 (2020).
151. Schymkowitz, J. *et al.* The FoldX web server: an online force field. *Nucleic Acids Res.* **33**, W382–W388 (2005).
152. Buß, O., Rudat, J. & Ochsenreither, K. FoldX as Protein Engineering Tool: Better Than Random Based Approaches? *Comput. Struct. Biotechnol. J.* **16**, 25–33 (2018).
153. Yang, Y. & Zhou, Y. Ab initio folding of terminal segments with secondary structures reveals the fine difference between two closely related all-atom statistical energy functions. *Protein Sci.* **17**, 1212–1219 (2008).
154. Laimer, J., Hofer, H., Fritz, M., Wegenkittl, S. & Lackner, P. MAESTRO - multi agent stability prediction upon point mutations. *BMC Bioinformatics* **16**, 116 (2015).
155. Dehouck, Y. *et al.* Fast and accurate predictions of protein stability changes upon mutations using statistical potentials and neural networks: PoPMuSiC-2.0. **25**, 2537–2543 (2009).
156. Ebo, J. S., Guthertz, N., Radford, S. E. & Brockwell, D. J. Using protein engineering to understand and modulate aggregation. *Current Opinion in Structural Biology* **60**, 157–166 (2020).
157. Santos, J., Pujols, J., Pallarès, I., Iglesias, V. & Ventura, S. Computational prediction of protein aggregation: Advances in proteomics, conformation-specific algorithms and biotechnological applications. *Computational and Structural Biotechnology Journal* **18**, 1403–1413 (2020).
158. Trainor, K., Broom, A. & Meiering, E. M. Exploring the relationships between protein sequence, structure and solubility. *Curr. Opin. Struct. Biol.* **42**, 136–146 (2017).
159. Sormanni, P., Aprile, F. A. & Vendruscolo, M. The CamSol method of rational design of protein mutants with enhanced solubility. *J. Mol. Biol.* **427**, 478–490 (2015).

160. Conchillo-Solé, O. *et al.* AGGREGSCAN: a server for the prediction and evaluation of ‘hot spots’ of aggregation in polypeptides. *BMC Bioinformatics* **8**, 1–17 (2007).
161. Pettersen, E. F. *et al.* UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
162. Maier, J. A. *et al.* ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput* **11**, 23 (2015).
163. Izadi, S., Anandakrishnan, R. & Onufriev, A. V. Building Water Models: A Different Approach. *J. Phys. Chem. Lett* **5**, 14 (2014).
164. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. C. Numerical integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.* **23**, 321–341 (1977).
165. Pastor, R. W., Brooks, B. R. & Szabo, A. An analysis of the accuracy of Langevin and molecular dynamics algorithms. *Mol. Phys.* **65**, 1409–1419 (1988).
166. McGibbon, R. T. *et al.* Computational Tools MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophysj* **109**, 1528–1532 (2015).
167. Voynov, V., Chennamsetty, N., Kayser, V., Helk, B. & Trout, B. L. Predictive tools for stabilization of therapeutic proteins. *MAbs* **1**, 580–582 (2009).
168. Samanta, U., Bahadur, R. P. & Chakrabarti, P. Quantifying the accessible surface area of protein residues in their local environment. *Protein Eng.* **15**, 659–667 (2002).
169. Black, S. D. & Mould, D. R. Development of hydrophobicity parameters to analyze proteins which bear post- or cotranslational modifications. *Anal. Biochem.* **193**, 72–82 (1991).
170. Mysling, S., Betzer, C., Jensen, P. H. & Jorgensen, T. J. D. Characterizing the Dynamics of α -Synuclein Oligomers Using Hydrogen/Deuterium Exchange Monitored by Mass Spectrometry. *Biochemistry* **52**, 9097–9103 (2013).
171. Vilar, M. *et al.* The fold of-synuclein fibrils. *PNAS* **105**, 8637–8642 (2008).
172. Gopalswamy, M. *et al.* Structural characterization of amyloid fibrils from the human parathyroid hormone. *Biochim. Biophys. Acta - Proteins Proteomics* **1854**, 249–257 (2015).
173. Alexandrescu, A. T. Amide Proton Solvent Protection in Amylin Fibrils Probed by Quenched Hydrogen Exchange NMR. *PLoS One* **8**, 56467 (2013).
174. Trovato, A., Chiti, F., Maritan, A. & Seno, F. Insight into the Structure of Amyloid Fibrils from the Analysis of Globular Proteins. *PLoS Comput. Biol.* **2**, e170 (2006).
175. Laimer, J., Hiebl-Flach, J., Lengauer, D. & Lackner, P. MAESTROweb: A web server for structure-based protein stability prediction. *Bioinformatics* **32**, 1414–1416 (2016).
176. Cota, E., Steward, A., Fowler, S. B. & Clarke, J. The Folding Nucleus of a Fibronectin Type III Domain is Composed of Core Residues of the Immunoglobulin-like Fold. *J. Mol. Biol.* **305**, 1185–1194 (2001).
177. Marchler-Bauer, A. *et al.* CDD: NCBI’s conserved domain database. *Nucleic Acids Res.* **D1**, D222–D226 (2014).
178. Lappalainen, I., Hurley, M. G. & Clarke, J. Plasticity Within the Obligatory Folding Nucleus of an Immunoglobulin-like Domain. *J. Mol. Biol.* **375**, 547–559 (2008).
179. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: A Sequence Logo Generator. doi:10.1101/gr.849004
180. Kabsch, W. & Sander, C. Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features.
181. Chow, C., Kurt, N., Murphy, R. M. & Cavagnero, S. Structural characterization of apomyoglobin self-associated species in aqueous buffer and urea solution. *Biophys. J.* **90**, 298–309 (2006).
182. Chan, P. K. *et al.* Structural similarity of wild-type and ALS-mutant superoxide dismutase-1 fibrils using limited proteolysis and atomic force microscopy. *Proc. Natl. Acad. Sci.* **110**, 10934–10939 (2013).
183. Kheterpal, I., Williams, A., Murphy, C., Bledsoe, B. & Wetzel, R. Structural Features of the AB Amyloid Fibril Elucidated by Limited Proteolysis. *Biochemistry* **40**, 11757–11767 (2001).

184. Myers, S. L., Thomson, N. H., Radford, S. E. & Ashcroft, A. E. Investigating the structural properties of amyloid-like fibrils formed in vitro from β 2-microglobulin using limited proteolysis and electrospray ionisation mass spectrometry. *Rapid Commun. Mass Spectrom.* **20**, 1628–1636 (2006).
185. Monit, M., Amoresano, A., Giorgetti, S., Bellotti, V. & Pucci, P. Limited proteolysis in the investigation of B2-microglobulin amyloidogenic and fibrillar states. *Biochim. Biophys. Acts* **1753**, 44–50 (2005).
186. Codina, N. *et al.* An Expanded Conformation of an Antibody Fab Region by X-Ray Scattering, Molecular Dynamics, and smFRET Identifies an Aggregation Mechanism. *J. Mol. Biol.* **431**, 1409–1425 (2019).
187. Rodrigues, J. R., Simões, C. J. V., Silva, C. G. & Brito, R. M. M. Potentially amyloidogenic conformational intermediates populate the unfolding landscape of transthyretin: Insights from molecular dynamics simulations. *Protein Sci.* **19**, 202–219 (2010).
188. Lu, P. *et al.* On the Conformational Dynamics of β -Amyloid Forming Peptides: A Computational Perspective. (2020). doi:10.3389/fbioe.2020.00532

Appendices

Appendix A

Aggregation propensity predictions and experimental data

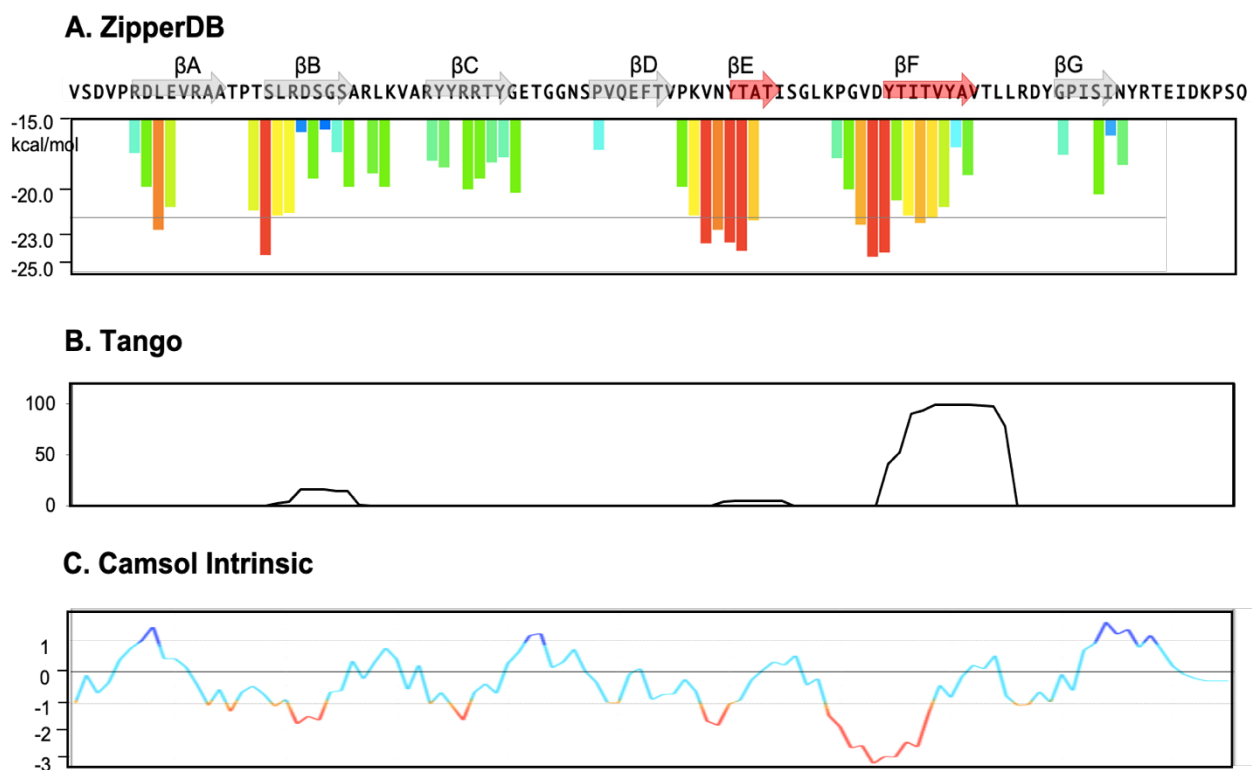


Figure A.1: Computational aggregation propensity predictions for parent Adnectin. Sequence-based predictors and ZipperDB ¹³³ (A), Tango ¹⁴⁰ (B) and Camsol Intrinsic ¹⁵⁹ (C) were used to calculate the predicted aggregation/amyloid propensity for the parent Adnectin sequence. The parent Adnectin sequence is shown with arrows indicating locations of the 7 β -strands (A-G).

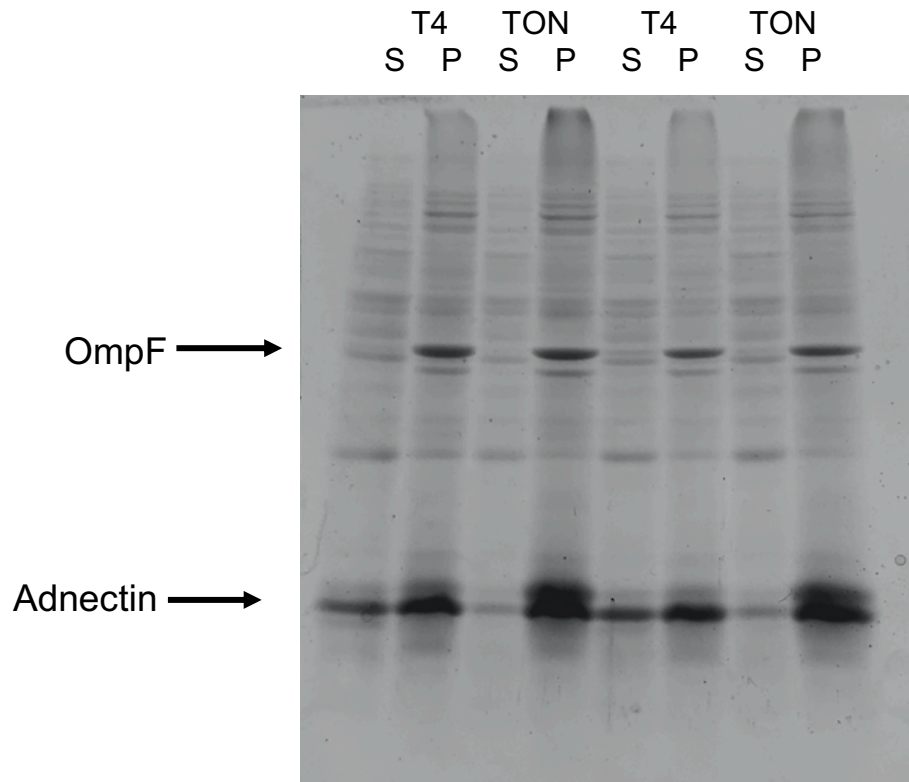


Figure A.2: SDS-page gel analysis of soluble and insoluble expression levels of Adnectin. Expression of L18A/Y88F in M9 minimal media at 4 and 24 hours post-induction with IPTG. Whole cell fraction was centrifuged to obtain soluble (S) and insoluble (P) material. Decrease in intensity of soluble band after increased expression time indicates soluble protein is being recruited into the aggregate fraction.

Table A.1 Aggregation Propensity of Adnectins grown at 37 °C in different growth media.
Aggregation propensities in LB media were acquired by Kyle Trainor.

Adnectin Variant	Aggregation Propensity			
	Luria Broth		M9 Minimal Media	
	4 h (% in IB)	24 hr (% in IB)	4 h (% in IB)	24 hr (% in IB)
Parent	59 +/- 0	88 +/- 11	60 +/- 4	84 +/- 3
L18A/Y88F	95 +/- 3	n.d.	82 +/- 3	92 +/- 1
L18V	83 +/- 2	n.d.	51 +/- 1	63 +/- 10
L18V/Y88F	61 +/- 4	n.d.	59 +/- 12	84 +/- 2
V75R	37 +/- 8	85 +/- 5	51 +/- 2	57 +/- 2
L78I	74 +/- 5	92 +/- 4	81 +/- 6	89 +/- 3
V11R/L19R	n.d.	n.d.	76 +/- 6	94 +/- 1
V11R/L19R/W22G/I34R	n.d.	n.d.	100*	100*

*No protein band visible in soluble fraction.

n.d. indicates no data.

Table A.2: Adnectin point-mutant aggregation propensity prediction scores.

Adnectin Variant	Aggrescan3D	CamSol Intrinsic
L18A/Y88F	2.0	0.2390
L18V	3.0	0.1120
L18V/Y88F	3.6	0.0719
V75R	0.1	0.5237
L78I	3.3	0.1721
V11R/L19R	-2.4	0.3474
V11R/L19R/W22G/I34R	-6.8	0.6264

Appendix B

FTIR supplemental data

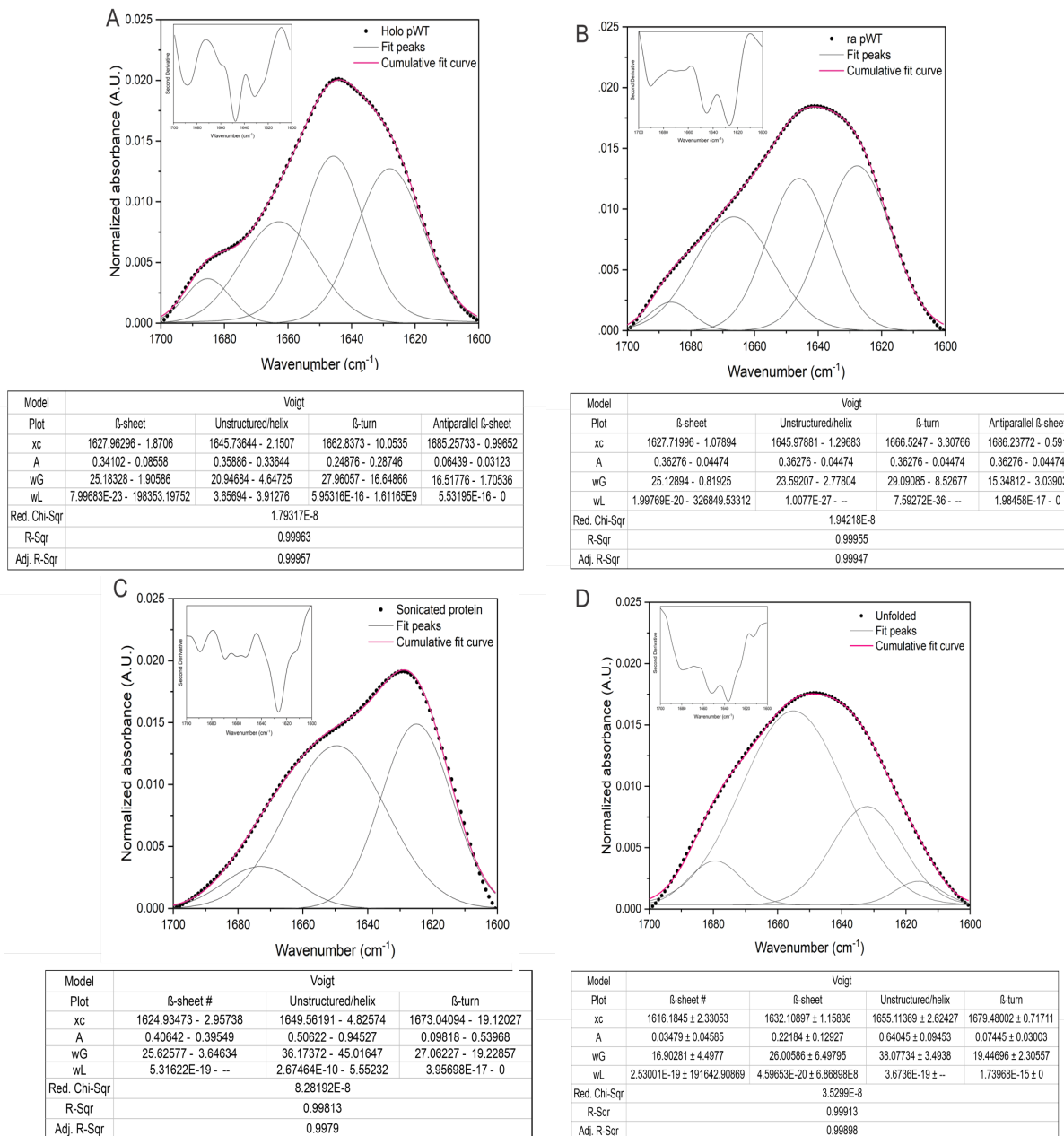


Figure B1: FTIR curve fitting analysis for pure SOD1 protein. (A) holo pWT, (B) reduced apo pWT, (C) sonicated reduced apo A4V, and (D) heat unfolded reduced apo A4V at 55 °C. Spectra correspond to Fig. 3A. A combination of native-like through unstructured conformations observed for purified proteins is compatible with the observed FTIR spectra for the IBs.

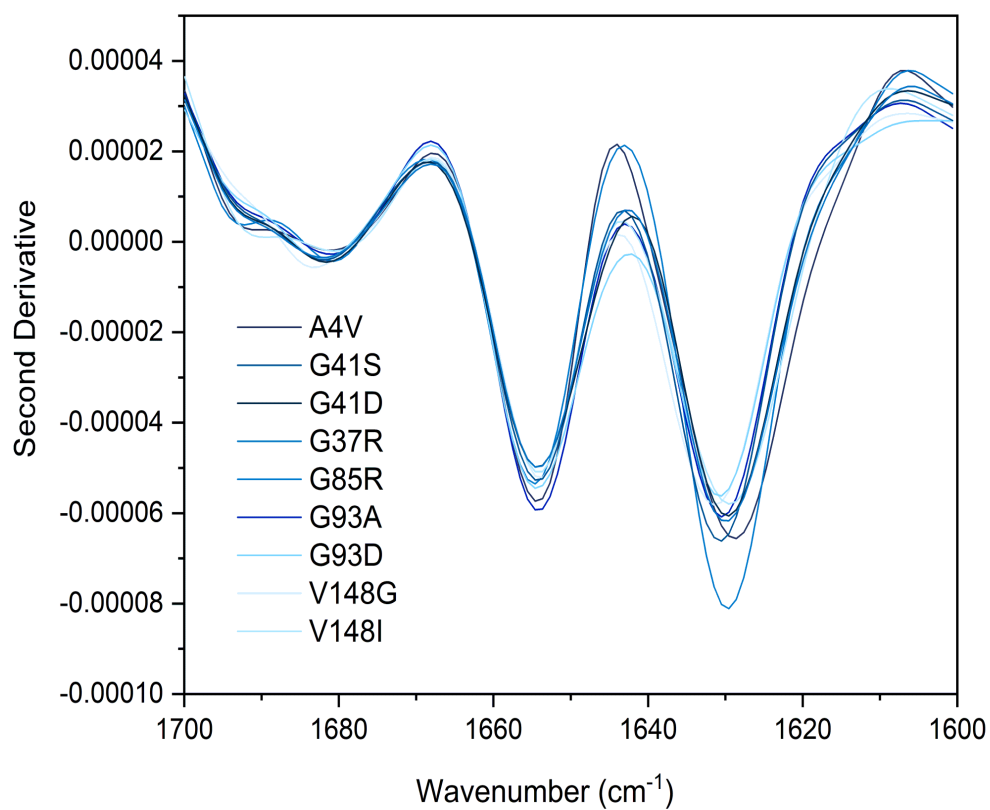


Figure B.2: FTIR second derivative spectra of mutant SOD1 IBs. Each spectrum is the average of 3 technical replicates for 2 separate culture growths. Each absorbance spectrum is normalized to an area of 1, before calculating the second derivative.

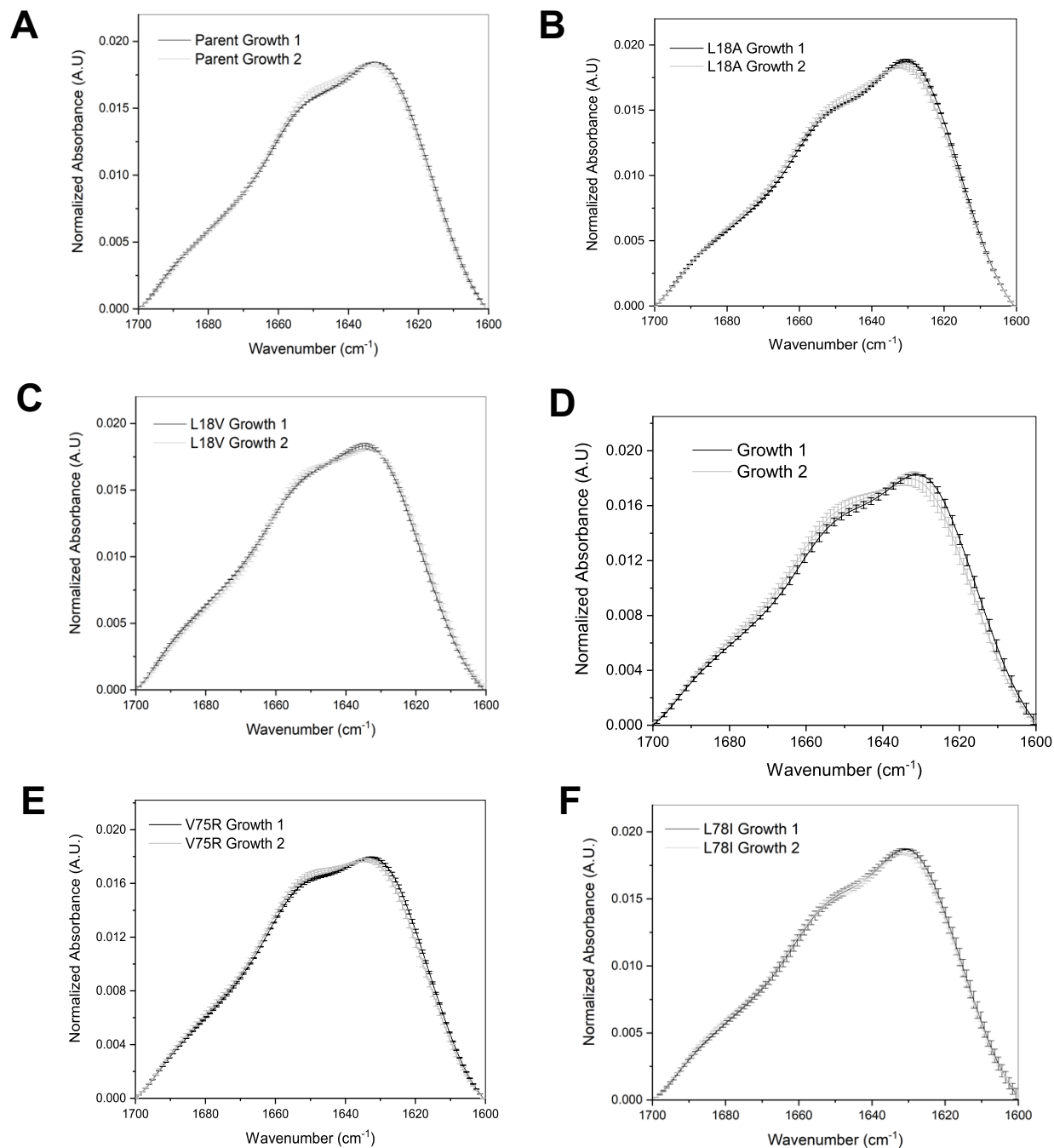


Figure B.3: FTIR absorbance spectra of sample replicates for mutant Adnectin IBs from cell cultures grown at 37 °C and harvested at 18 hours post-induction. Area normalized spectra shown are the average for 3 samples from the same growth (technical replicates), for 2 separate growths (biological replicates). Error bars are the standard deviation of technical replicates. A) Parent, B) L18A/Y88F, C) L18V, D) L18V/Y88F, E) V75R, F) L78I.

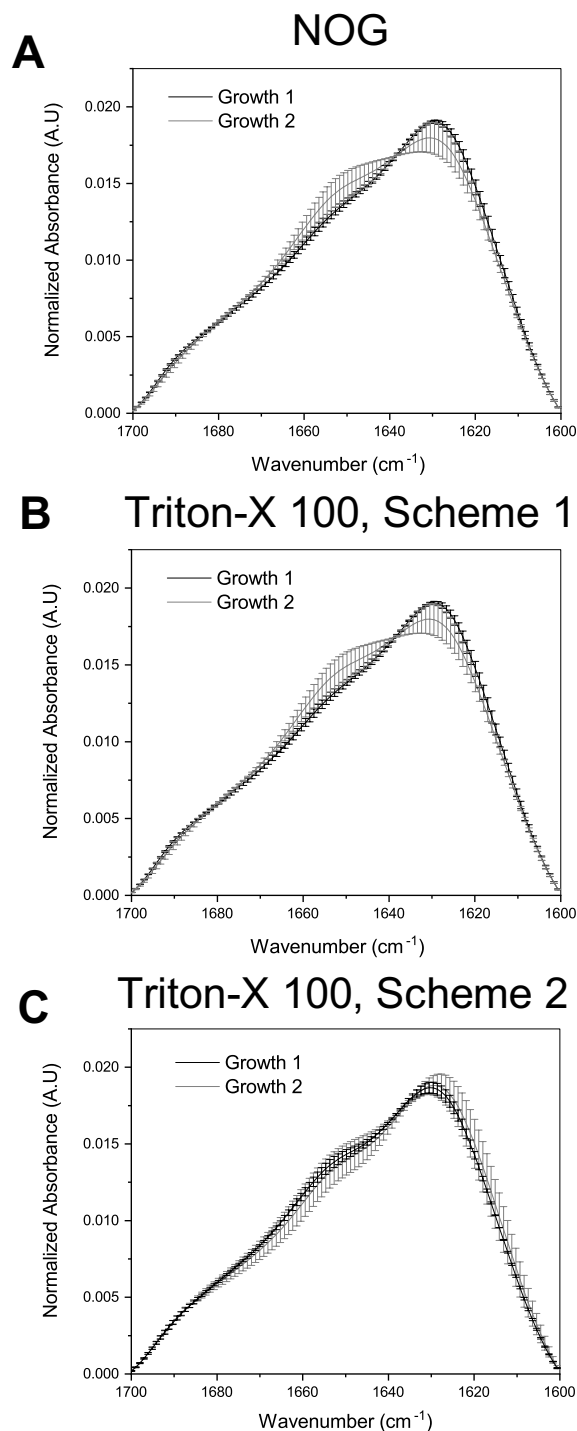


Figure B.4: FTIR absorbance spectra of sample replicates for Parent IBs washed with non-ionic detergents. Area normalized spectra shown are the average for 3 samples from the same growth (technical replicates), for 2 separate growths (biological replicates). Error bars are the standard deviation of technical replicates. Triton scheme 1 indicates a 15 minute spin at 16300g followed by removal of detergent with an identical spin. Triton scheme 2 indicates a 15 minute spin at 8000g followed by removal of detergent with a 15 minute spin at 16300g. NOG washing indicates a 5 minute spin at 12000g followed by removal of detergent with an identical spin.

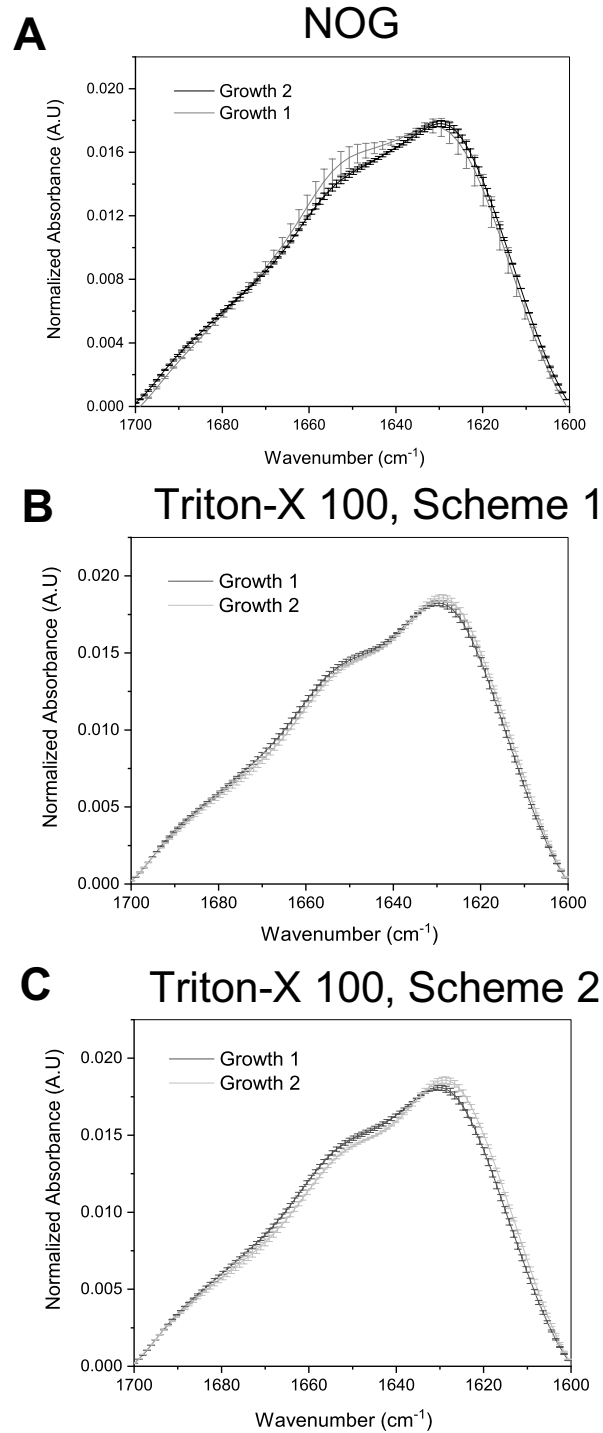


Figure B.5: FTIR absorbance spectra of sample replicates for V75R IBs washed with non-ionic detergents. Area normalized spectra shown are the average for 3 samples from the same growth (technical replicates), for 2 separate growths (biological replicates). Error bars are the standard deviation of technical replicates. Triton scheme 1 indicates a 15 minute spin at 16300g followed by removal of detergent with an identical spin. Triton scheme 2 indicates a 15 minute spin at 8000g followed by removal of detergent with a 15 minute spin at 16300g. NOG washing indicates a 5 minute spin at 12000g followed by removal of detergent with an identical spin.

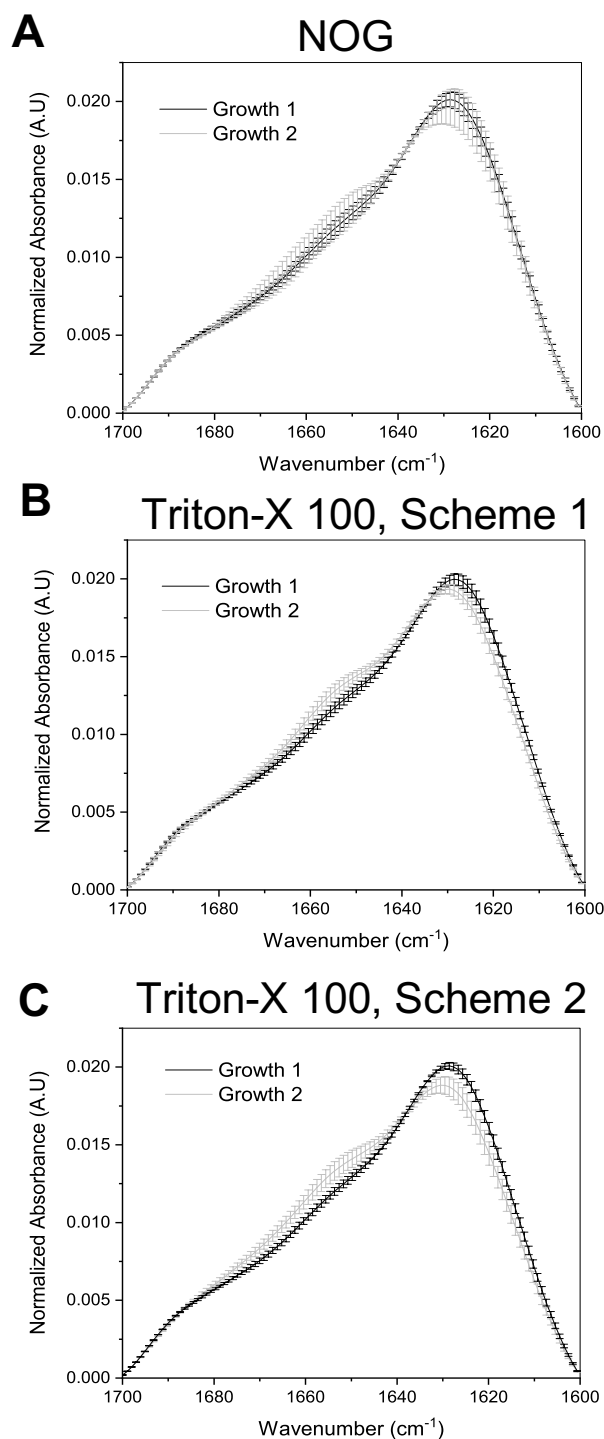


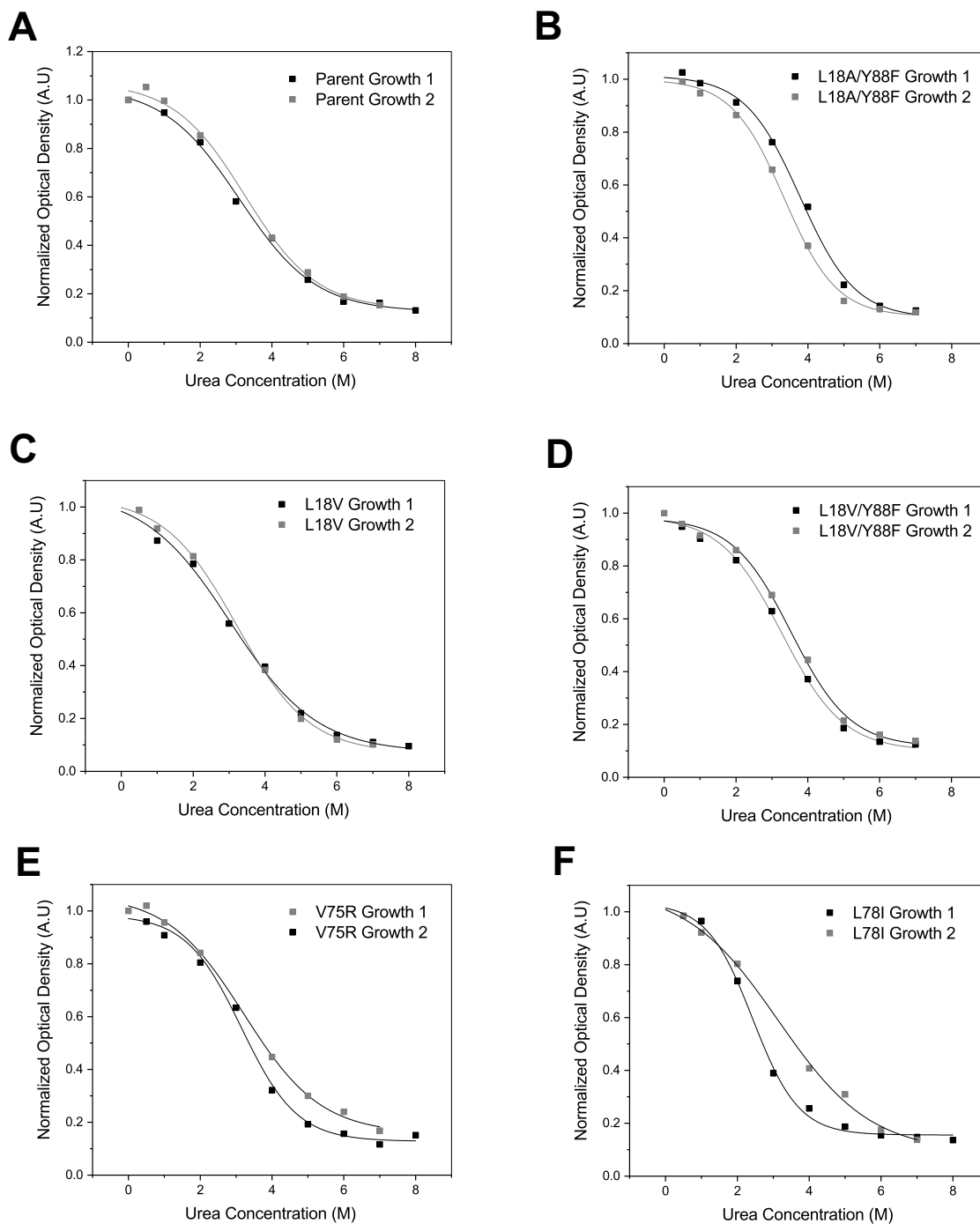
Figure B.6: FTIR absorbance spectra of sample replicates for L78I IBs washed with non-ionic detergents. Area normalized spectra shown are the average for 3 samples from the same growth (technical replicates), for 2 separate growths (biological replicates). Error bars are the standard deviation of technical replicates. Triton scheme 1 indicates a 15 minute spin at 16300g followed by removal of detergent with an identical spin. Triton scheme 2 indicates a 15 minute spin at 8000g followed by removal of detergent with a 15 minute spin at 16300g. NOG washing indicates a 5 minute spin at 12000g followed by removal of detergent with an identical spin.

Table B.1: Non-linear curve fitting results for SOD1 IB spectrum using variable numbers of components. Fitting was performed using the Voigt function on OriginPro 2018 using second derivative minimums as starting peak positions.

Parameter	Three Components				
	β -sheet	Disorder	β -turn and antiparallel β -sheet		
Xc	1629.5974 +/- 0.75657	1654.47002 +/- 0.54464	1680.00446 +/- 0.84372		
A	0.40907 +/- 0.03396	0.49328 +/- 0.04773	0.10382 +/- 0.01549		
wG	27.81021 +/- 0.63737	30.34044 +/- 2.02525	21.86983 +/- 0.985		
wL	1.2077E-15 +/- 3111305	3.90595E-13 +/- 0.01366	2.05272E-15 +/- 0		
Red. Chi-Sqr		2.80893E-8			
Adj. R-Sqr		0.99925			
Four Components					
	# β -sheet	β -sheet	Disorder	β -turn and antiparallel β -sheet	
Xc	1614.45776 +/- 1.17432	1629.92075 +/- 1.1368	1653.26974 +/- 1.64209	1678.6184 +/- 1.67475	
A	0.02974 +/- 0.03018	0.35105 +/- 0.1297	0.49732 +/- 0.14508	0.12601 +/- 0.03976	
wG	14.59957 +/- 3.3369	24.36022 +/- 3.99363	29.29647 +/- 4.38697	23.36302 +/- 1.80938	
wL	8.79127E-13 +/- 0.25034	1.1863E-9 +/- 2.07273	1.39986E-10 +/- 3.1205	1.12797E-28 +/- 0	
Red. Chi-Sqr			1.44968E-8		
Adj. R-Sqr			0.99962		
Five Components					
	# β -sheet	β -sheet	Disorder	β -turn	Antiparallel β -sheet
Xc	1616.7625 +/- 4.5632	1631.527 +/- 3.035	1653.596 +/- 3.74	1675.839 +/- 3.88	1688.093 +/- 2.10
A	0.0642 +/- 0.1209	0.3409 +/- 0.4431	0.4616 +/- 0.5095	0.1156 +/- 0.2239	0.02163 +/- 0.0498
wG	17.5998 +/- 6.6702	23.230 +/- 12.382	26.969 +/- 17.188	20.584 +/- 14.765	13.2327 +/- 5.463
wL	0.0687 +/- 5.868	2.78E-47 +/- 2E7	0.4333 +/- 4.796	2.258E-25 +/- 0	0 +/- 0
Red. Chi-Sqr			9.57271E-9		
Adj. R-Sqr			0.99975		

Appendix C

Urea Solubilization Supplemental Data



[continued on next page]

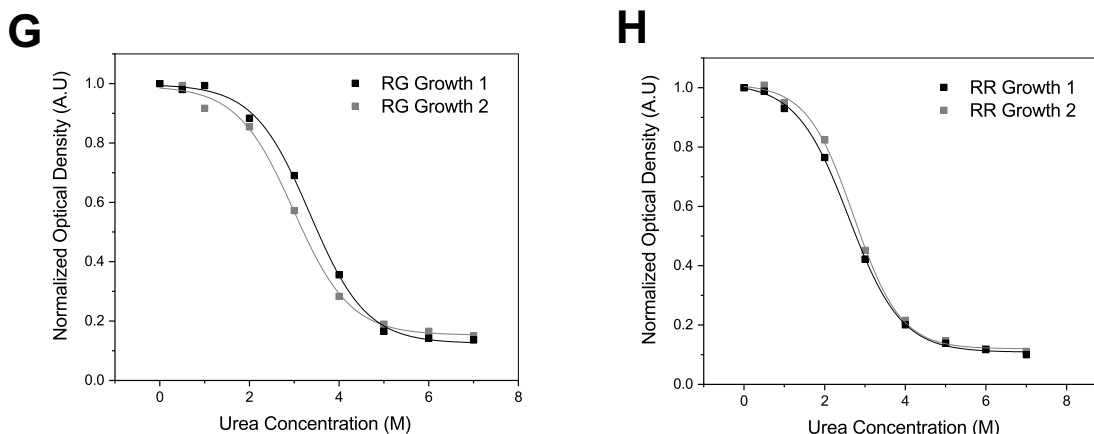


Figure C.1: Urea solubilization profiles of sample replicates for mutant Adnectin IBs from cell cultures grown at 37 °C and harvested at 18 hours post-induction. Symbols represent raw data and lines represent data fit using Origin Pro 2018. See methods section 2.2.6. Data include two samples from the same growth (technical replicates), for 2 separate growths (biological replicates). A) Parent, B) L18A/Y88F, C) L18V, D) L18V/Y88F, E) V75R, F) L78I, G) V11R/L19R H) V11R/L19R/W22G/I34R

Table C.1: Change in monomer stability relative to parent Adnectin upon mutation as predicted by FoldX¹⁵¹. Stability changes were evaluated using an Adnectin homology model created by Julia Steckner using SwissModel. Point mutations were added using the FoldX BuildModel function. Positive stability change indicates a destabilized monomer, negative indicates an increase in monomer stability.

Adnectin Variant	Computed Stability Change (kcal/mol)
L18A/Y88F	3.620
L18V	0.879
L18V/Y88F	1.113
V75R	2.157
L78I	-0.108
V11R/L19R	1.398
V11R/L19R/W22G/I34R	13.313

Appendix D

qHDX NMR supplemental data

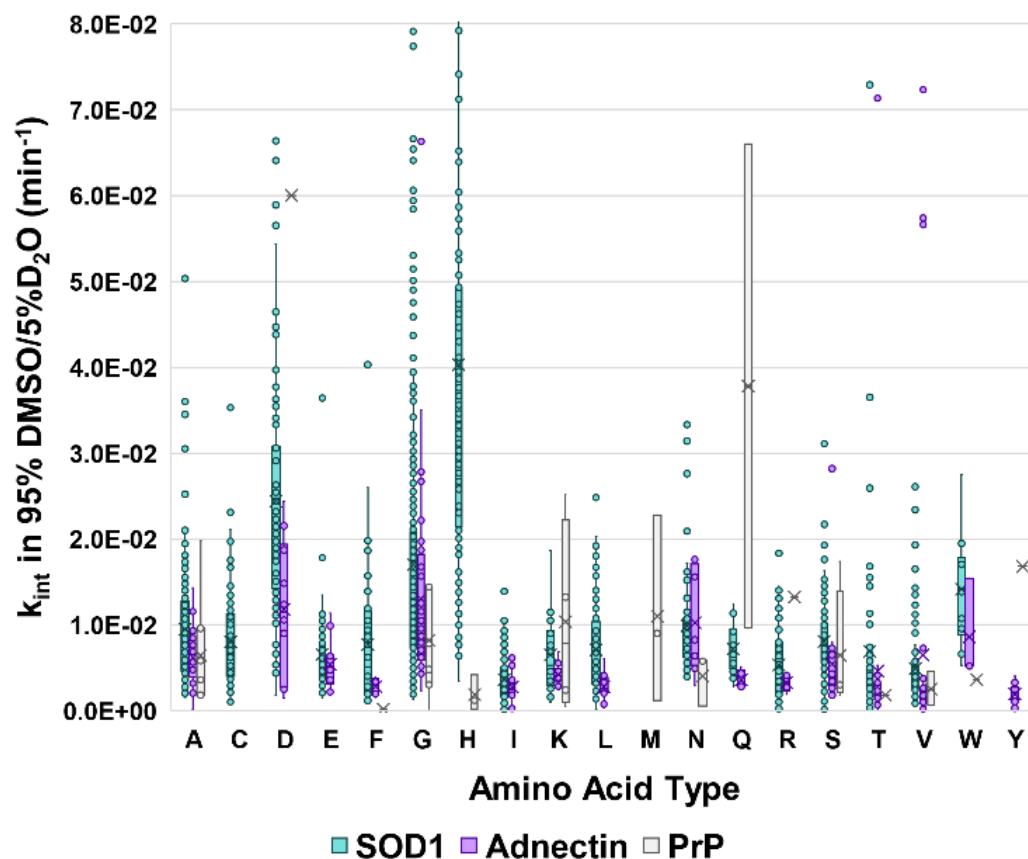


Figure D.1: Measured intrinsic rate constants of amide exchange for different amino acids. Values are plotted for SOD1 (teal) and Adnectin (purple) IB samples together with literature values for pure PrP aggregates (light grey) ¹³². Values include results for 18 SOD1 samples (two biological replicates for each of nine different mutants), six Adnectin samples, and a single mutant (P101L) for PrP ¹³². For residues with two data points, edges of the box correspond to the two data points. Note that the N-terminal Q residue is included in the plot for PrP, which accounts for the large range seen.

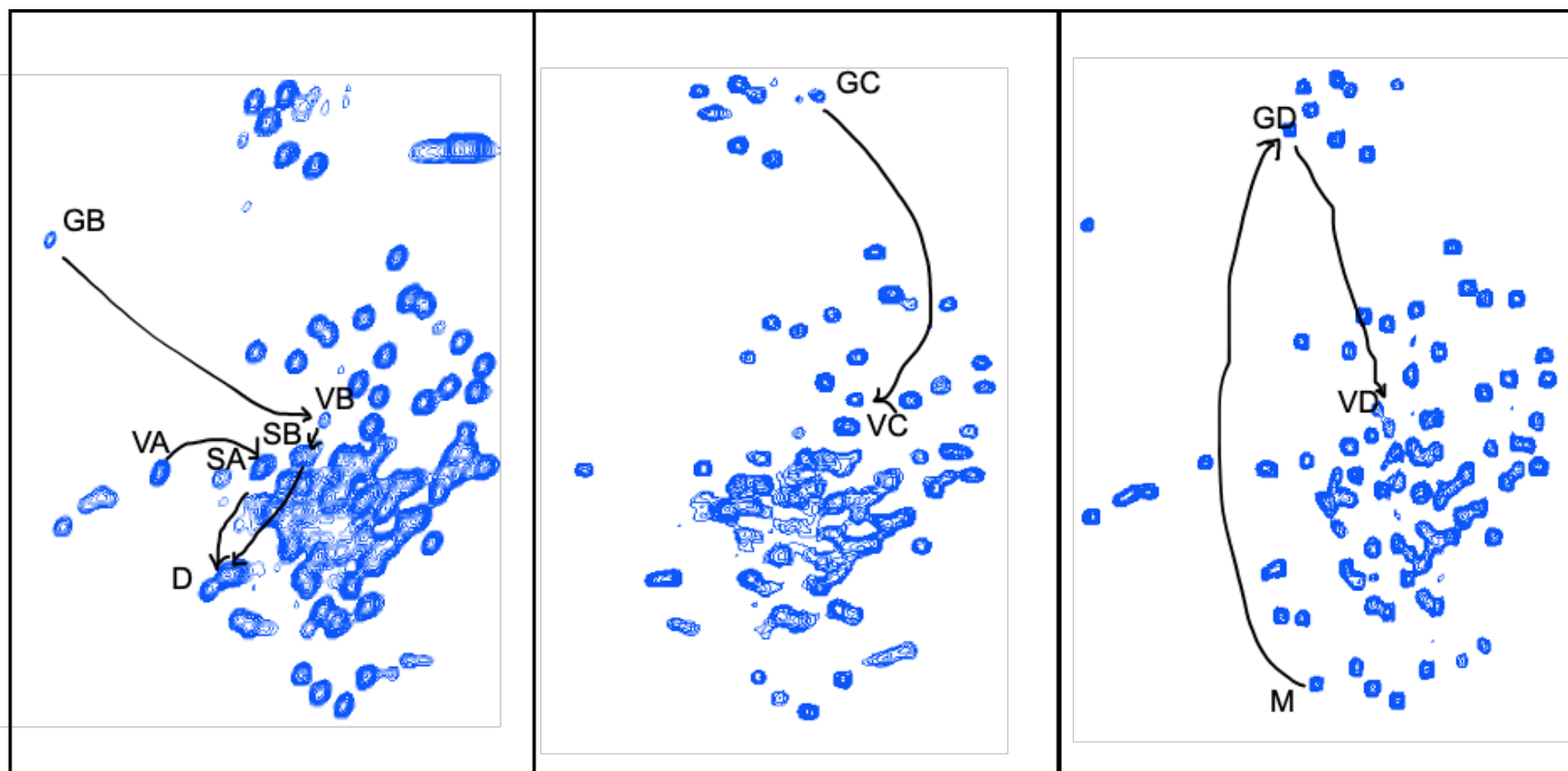


Figure D.2: ^1H - ^{15}N HSQCs showing multiple assignments for N-terminal amino acid sequence 'MGVSD'. Connectivity of peaks was determined using five standard s Bruker pulse sequences and parameter sets: HNCO, HN(CA)CO, HNCACB, CBCA(CO)NH, and HN(CA)NNH. Computer aided resonance assignment (CARA) was used for making assignments. The origin of these sets of peaks (termed A-D) was determined to be multiple species of Adnectin, assumed to be non-gluconoylated, gluconoylated x 2, and a species with N-terminal methionine uncleaved.

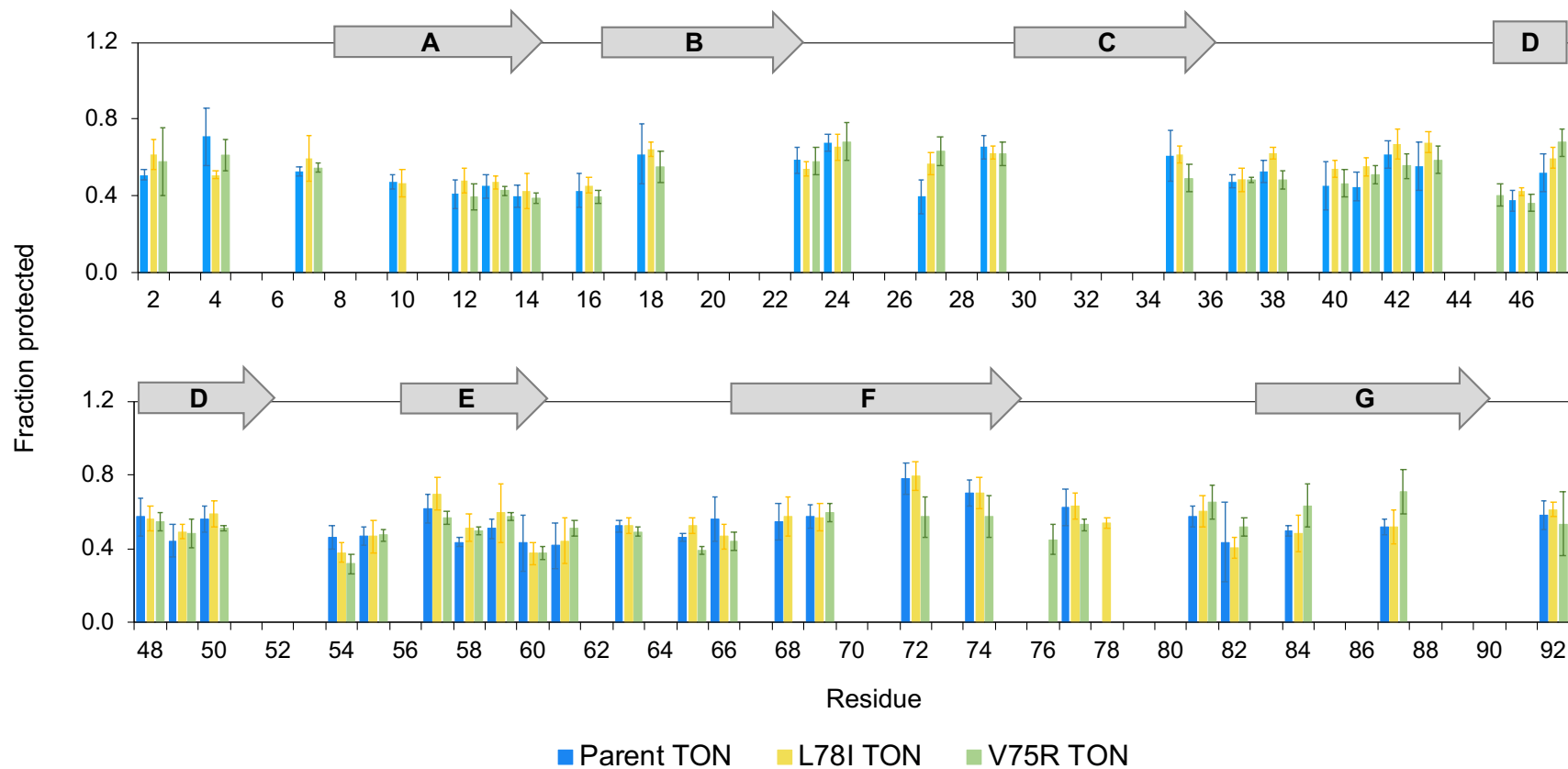


Figure D.3:Residue-level structure of Adnectin IBs from qHDX analysis. Fraction amide protected for parent Adnectin and two point-mutants, V75R (more soluble than parent) and L78I (less soluble than parent). Each bar is an average of two biological replicates, Gaps indicate areas of no data due to an unassigned or “crowded” residue. Error bars represent propagated error calculated using the standard formula:

$$\frac{\Delta z}{z} = \sqrt{\frac{\Delta x^2}{x} + \frac{\Delta y^2}{y}}$$

where z is the fraction amide protection, x is the average intensity of residue i in the D₂O sample, y is the average intensity of residue i in the H₂O sample and the Δ's represent error between technical replicates (standard deviation) of each parameter respectively

Table D.1: Resonance assignments of 6199_CO1 Adnectin (FG Loop: DYRDYC) in 95% DMSO/ 5% H₂O at pH 5.5. Resonance assignments for Adnectin IB sample grown at 37°C and harvested 18 hours post-induction. NMR experiments were performed at 25 °C for IB sample at pH 5.5 in 95% DMSO/5% H₂O/25mM DTT. CO chemical shifts were obtained from HNCO/HNCαCO experiments, and Cα and Cβ shifts were obtained from HNCαCβ/CβCα(CO)NH experiments. Unassigned residues/atoms not shown.

Residue	Assignment (ppm)				
	¹ H	¹⁵ N	Co	Cα	Cβ
1 VAL	8.445	116.376	171.054	58.044	31.163
2 SER	8.187	116.216	170.513	55.305	62.007
3 ASP	8.267	119.352	170.665	49.972	36.28
4 VAL	7.687	116.406	170.2	56.242	30.37
6 ARG	8.152	117.523	172.016	52.567	29.145
7 ASP	8.126	116.656	170.838	50.044	36.135
8 LEU	7.821	118.009	172.384	51.774	40.532
9 GLU	7.93	117.203	171.411	52.35	27.487
10 VAL	7.691	115.728	171.335	58.044	30.73
11 VAL	7.821	117.704	171.011	58.044	30.658
12 ALA	7.975	123.147	172.168	48.531	18.047
13 ALA	7.953	119.672	172.589	48.459	18.191
14 THR	7.871	112.989	169.9	57.035	67.052
16 THR	8.014	112.379	171.378	60.494	66.62
17 SER	7.909	115.093	171.4	57.035	61.359
18 LEU	7.911	120.244	173.292	52.423	40.027
19 LEU	7.86	118.431		52.423	40.171
21 SER	7.934	115.855	171.595	56.098	61.719
22 TRP	8.214	120.792	173.011	56.026	27.343
23 SER	8.096	113.153	171.84	56.963	61.359
24 ALA	8.082	122.21	173.432	50.477	17.614
25 ARG	7.926	116.151		54.08	28.641
27 LYS	7.924	118.301		52.423	30.73
29 ALA	8.036	122.759	174.092	49.612	17.614
30 ARG	7.939	116.041		54.008	28.568
34 ILE	7.911	116.965	171.584	57.755	36.712
35 THR	7.709	113.858	170.405	58.26	66.908
36 TYR	7.895	118.575		54.945	36.64
37 GLY	8.171	106.2	169.357	42.477	
38 GLU	8.009	117.089	171.897	52.423	27.632
39 THR	7.805	111.323		58.764	67.052

40 GLY	8.127	107.155		42.621	
41 GLY	8.123	105.399	169.076	42.477	
42 ASN	8.076	116.746	171.400	50.044	37.577
43 SER	7.937	113.817	169.778	53.72	61.791
45 VAL	7.773	114.328	171.681	58.188	30.658
46 GLN	8.031	120.358	171.508	52.495	27.848
47 GLU	7.871	117.466	171.346	52.495	27.92
48 PHE	7.973	116.912	171.573	54.152	37.577
49 THR	8.036	112.049	170.243	58.404	67.124
50 VAL	7.753	118.417		56.026	30.514
52 LYS	8.088	117.412	172.005	52.495	31.595
53 ASN	8.148	117.34	171.616	50.188	37.00
54 VAL	7.638	114.001	171.281	58.188	30.586
55 TYR	8.041	118.826	172.038	54.873	36.496
56 THR	7.808	111.231		58.62	67.052
57 ALA	7.916	122.256	172.935	48.747	18.407
58 THR	7.839	110.147	170.600	58.476	67.052
59 ILE	7.734	116.896	171.562	57.755	36.784
60 SER	7.985	115.953	170.978	55.954	61.863
61 GLY	8.049	107.476	170.211	42.621	
62 LEU	7.775	117.666	172.395	51.125	40.892
63 LYS	8.081	119.375	170.492	51.702	30.442
65 GLY	8.204	105.593	169.422	42.477	
66 VAL	7.621	113.255	171.335	57.9	30.947
67 ASP	8.256	119.314	170.719	50.044	36.135
68 TYR	7.654	115.831	171.216	54.44	36.712
69 THR	7.924	111.885	169.140	58.404	67.052
70 ILE	7.79	117.673	171.540	57.683	37.144
71 THR	7.906	115.034	170.373	58.692	67.052
72 VAL	7.615	116.224	171.173	58.044	30.586
73 TYR	7.896	118.929	171.303	54.44	36.784
74 ALA	7.993	120.701	172.773	48.747	18.335
75 VAL	7.887	114.114	171.422	58.188	30.586
76 THR	7.677	112.56	170.243	58.044	67.124
77 ASP	8.104	118.514	171.086	50.116	36.352
78 TYR	7.817	116.799		55.017	36.64
82 CYS	8.325	119.695	168.816	53.792	25.758
84 ILE	7.885	115.723	171.573	57.611	37.144
85 SER	7.958	116.344	170.622	55.377	61.719
86 ILE	7.79	117.722	171.216	57.611	37.072

87 ASN	8.068	119.607		50.549	37.289
89 ARG	8.12	117.838	172.005	52.783	29.001
90 THR	7.764	111.445	170.449	58.692	66.908
91 GLU	7.908	118.908	171.595	52.35	27.56
92 ILE	7.81	116.665	171.465	57.467	36.712
93 ASP	8.239	120.61	170.74	49.9	36.352
94 LYS	7.884	118.453		50.765	30.658
96 SER					
	8.203	112.867	171.032	56.098	61.575

Table D.2: Rate constants (k_{ex}) of amide exchange in D_2O for Adnectin IB aggregates fit to a double exponential decay in OriginPro 2018. Accuracy is limited due to too few data points, data omitted for residues showing significant scatter.

Residue	Fitted Parameters						
	y0	A1	t1	A2	t2	k1	k2
2 SER	4.940E+09	4.117E+09	0.0115	1.891E+09	44.778	86.851	0.0223
7 ASP	4.335E+09	5.897E+09	0.0437	2.288E+09	95.879	22.878	0.0104
10 VAL	3.411E+09	4.340E+09	0.0313	4.032E+09	91.672	31.957	0.0109
12 ALA	3.755E+09	4.540E+09	0.0547	3.237E+09	32.043	18.272	0.0312
13 ALA	4.425E+09	5.442E+09	0.0124	3.310E+09	54.572	80.486	0.0183
14 THR	2.389E+09	5.120E+09	0.0519	1.902E+09	34.240	19.275	0.0292
16 THR	1.500E+09	2.140E+09	0.0461	1.743E+09	199.334	21.691	0.0050
18 LEU	2.874E+09	2.703E+09	0.0225	2.262E+09	166.579	44.424	0.0060
23 SER	1.800E+09	1.788E+09	0.0492	8.445E+08	40.401	20.319	0.0248
24 ALA	2.535E+09	1.032E+09	0.0504	9.546E+08	112.978	19.853	0.0089
27 LYS	1.977E+09	2.859E+09	0.0501	1.719E+09	41.908	19.943	0.0239
29 ALA	1.744E+09	1.204E+09	0.0650	9.605E+08	139.148	15.391	0.0072
35 THR	2.607E+09	1.051E+09	0.0118	1.630E+09	69.508	84.487	0.0144
37 GLY	2.350E+09	3.438E+09	0.0339	1.767E+09	91.274	29.507	0.0110
40 GLU	4.055E+09	5.508E+09	0.0552	2.230E+09	50.430	18.116	0.0198
46 GLN	2.529E+09	4.971E+09	0.0484	1.959E+09	1072.056	20.680	0.0009
47 GLU	2.999E+09	2.021E+09	0.0423	1.806E+09	42.420	23.629	0.0236
48 PHE	3.168E+09	4.216E+09	0.0454	1.429E+09	73.911	22.021	0.0135
54 LYS	1.917E+09	2.803E+09	0.0172	1.463E+09	489.990	58.081	0.0020
55 TYR	1.968E+09	2.983E+09	0.0440	2.193E+09	54.909	22.710	0.0182
57 ALA	4.453E+09	2.621E+09	0.0250	2.189E+09	85.885	40.008	0.0116
58 THR	3.500E+09	4.960E+09	0.0488	2.575E+09	37.767	20.486	0.0265
59 ILE	3.739E+09	2.394E+09	0.0110	1.986E+09	45.332	90.562	0.0221

60 SER	3.163E+09	2.625E+09	0.0161	2.639E+09	91.975	61.940	0.0109
61 GLY	2.349E+09	4.663E+09	0.0449	1.641E+09	18.978	22.254	0.0527
66 VAL	2.052E+09	2.744E+09	0.0042	1.949E+09	206.562	240.794	0.0048
69 THR	2.260E+09	2.001E+09	0.0539	9.790E+08	26.000	18.557	0.0385
72 VAL	4.245E+09	8.537E+08	0.0596	2.165E+09	113.152	16.771	0.0088
74 ALA	4.329E+09	2.926E+09	0.0125	1.168E+09	77.085	80.072	0.0130
75 VAL	4.973E+09	1.724E+09	0.0710	4.149E+09	47.123	14.086	0.0212
77 LEU	2.245E+09	1.388E+09	0.0286	9.745E+08	46.298	34.999	0.0216
81 TYR	3.361E+09	2.271E+09	0.0481	1.679E+09	28.059	20.794	0.0356
82 GLY	1.253E+09	1.841E+09	0.0631	1.835E+09	91.902	15.855	0.0109
84 ILE	1.519E+09	1.317E+09	0.0266	1.154E+09	134.045	37.583	0.0075
87 ASN	1.400E+09	1.298E+09	0.0714	8.879E+08	35.877	14.009	0.0279
89 ARG	1.491E+09	2.093E+09	0.0398	1.449E+09	75.050	25.136	0.0133
92 ILE	3.050E+09	2.452E+09	0.0022	1.478E+09	0.006	451.926	162.8490

Appendix E

Computational Design and engineered variants supplemental Data

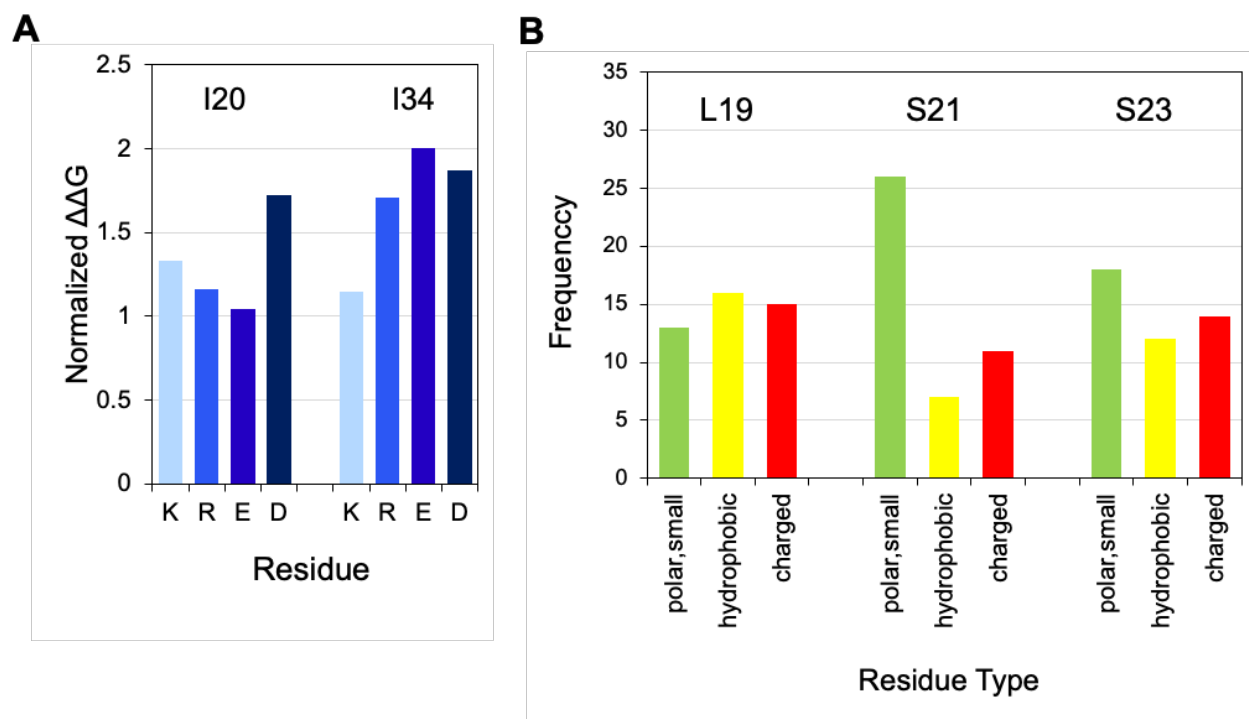


Figure E.1: Considerations for complementary point-mutations. A) Normalized $\Delta\Delta G$ prediction scores for double core mutants, the first being W22G in every case. Scores were evaluated using predictions from FoldX¹⁵¹, MAESTRO¹⁷⁵, and PoPMuSiC¹⁵⁵ with a maximum possible score of 3.0. See Eqn 4.1 (Section 4.2.3) for calculating normalized scores. B) Co-variation of amino acid residue types in the 'cd00063' FN3 alignment when arginine is present at position 11. Position 19 has the highest tolerance for charged residues.

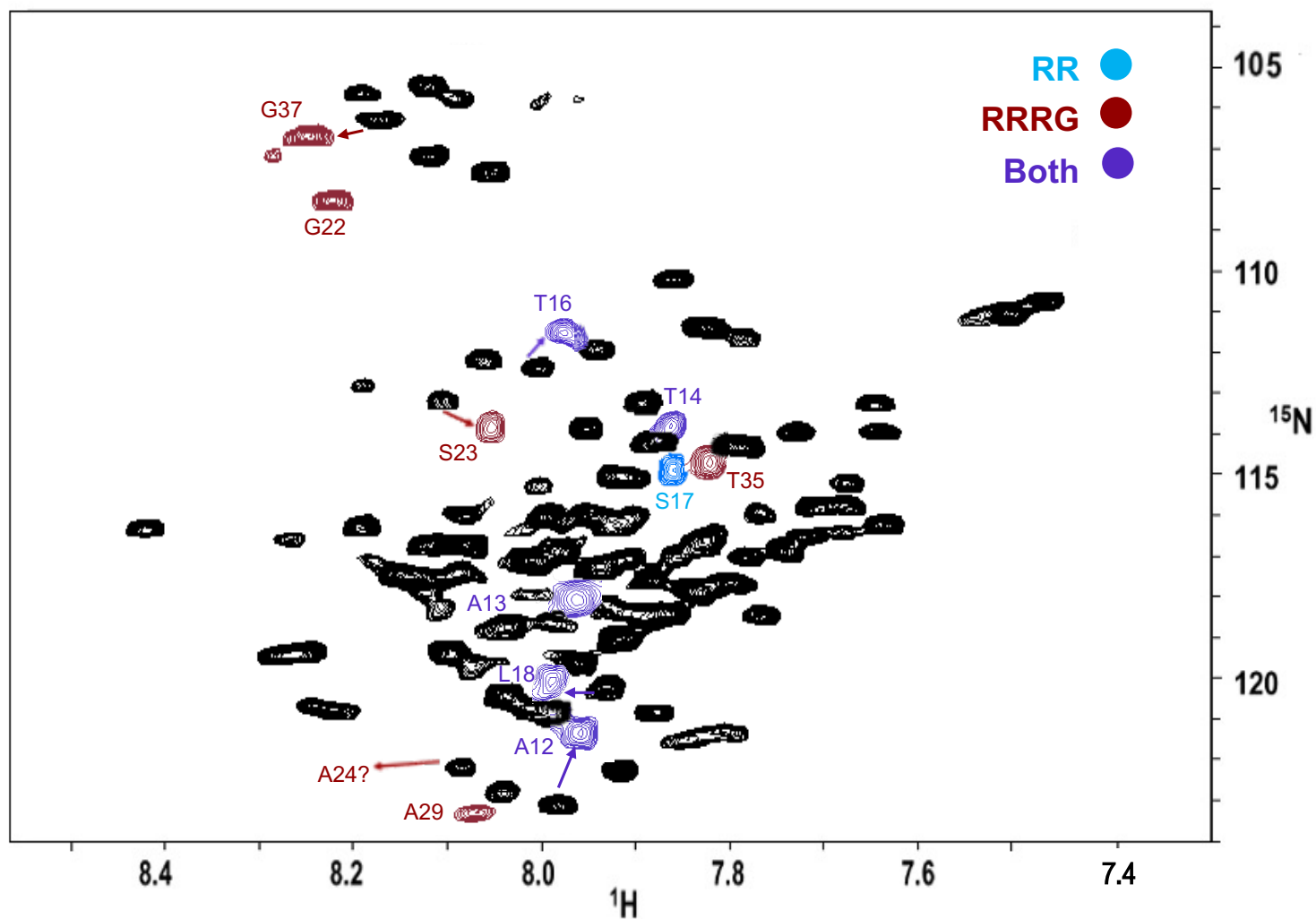


Figure E.2: ^1H - ^{15}N HSQC assignment transfers using intrinsic exchange rate constants and correlated peak movements. ^1H - ^{15}N HSQC spectrum showing new/shifted cross-peaks based on their occurrence per mutant. Peaks shifts only occurring in RR are shown in blue, shifts only occurring in RRRG are shown in red and those occurring in both mutants are shown in purple. Annotations only shown for identified residues, not all peaks were able to be tracked.

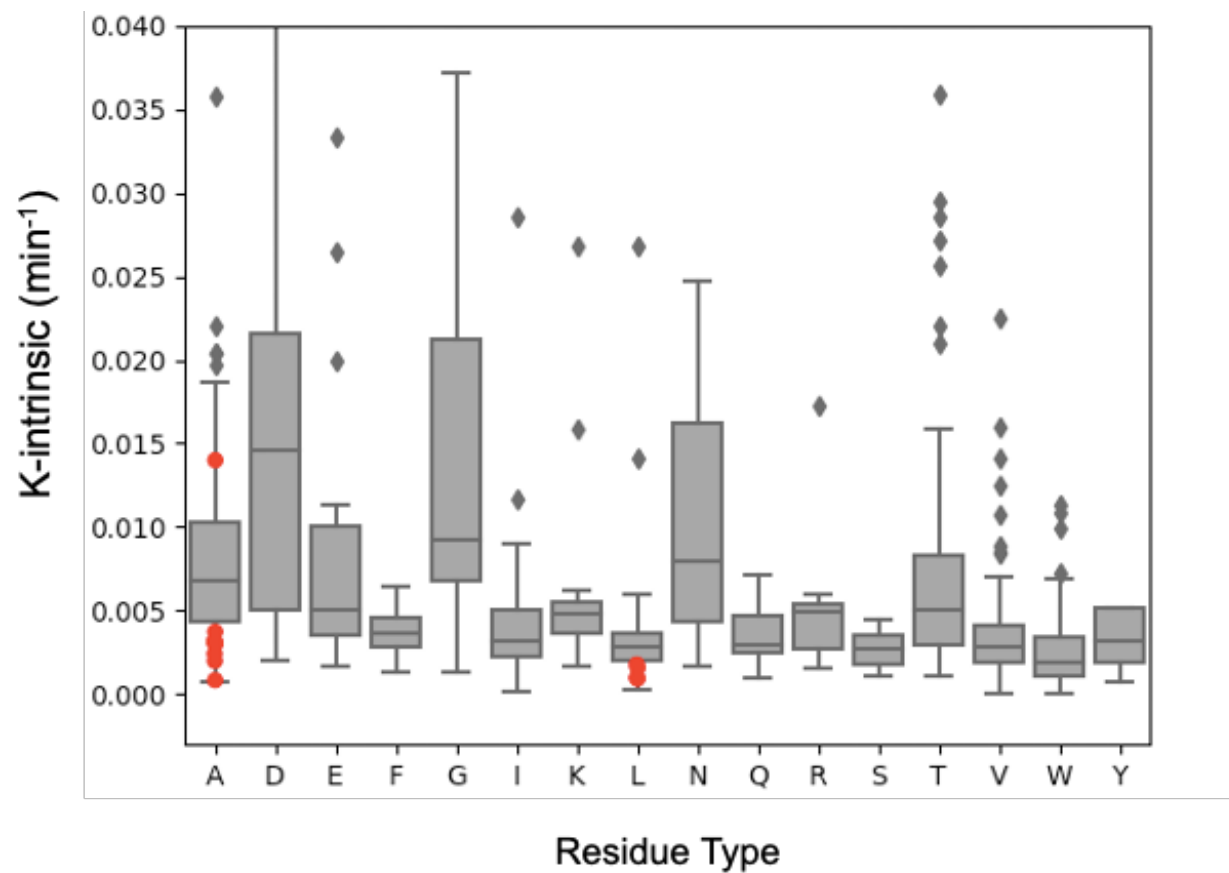


Figure E.3: Measured intrinsic rate constants (K_{int}) of amide exchange for different amino acids in Adnectin monomers unfolded in 95% DMSO/5% DCA. Values included from 12, unwashed Adnectin samples. Red dots indicate candidate peaks for residues A12, A13 and L18.

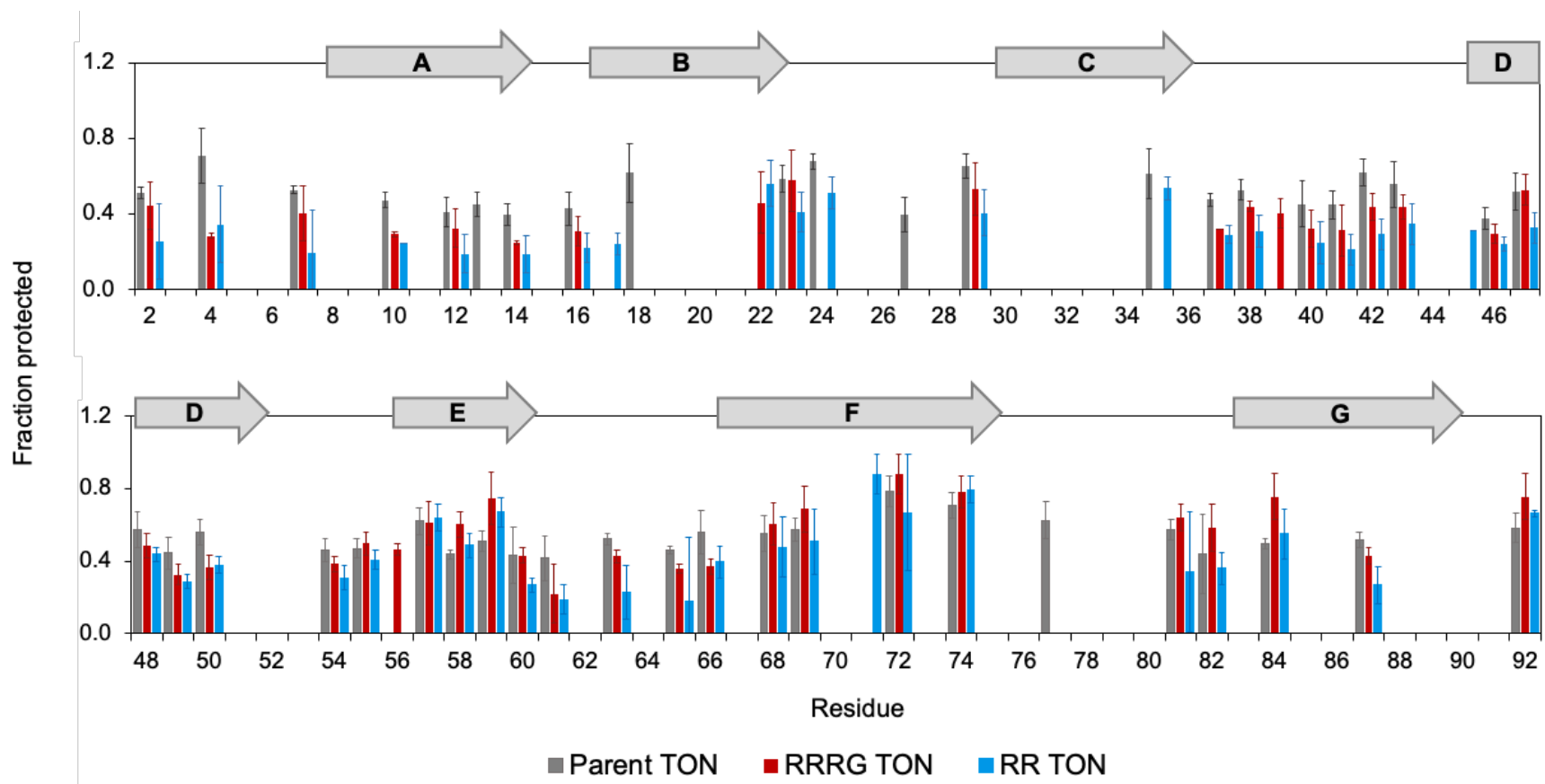


Figure E.4:Residue-level structure of Adnectin IBs from qHDX analysis. Fraction amide protected for parent Adnectin and engineered point-mutants, RRRG and RR. Each bar is an average of two biological replicates, Gaps indicate areas of no data due to an unassigned or “crowded” residue. Error bars represent propagated error calculated using the standard formula:

$$\frac{\Delta z}{z} = \sqrt{\frac{\Delta x^2}{x} + \frac{\Delta y^2}{y}}$$

where z is the fraction amide protection, x is the average intensity of residue i in the D₂O sample, y is the average intensity of residue i in the H₂O sample and the Δ's represent error between technical replicates (standard deviation) of each parameter respectively

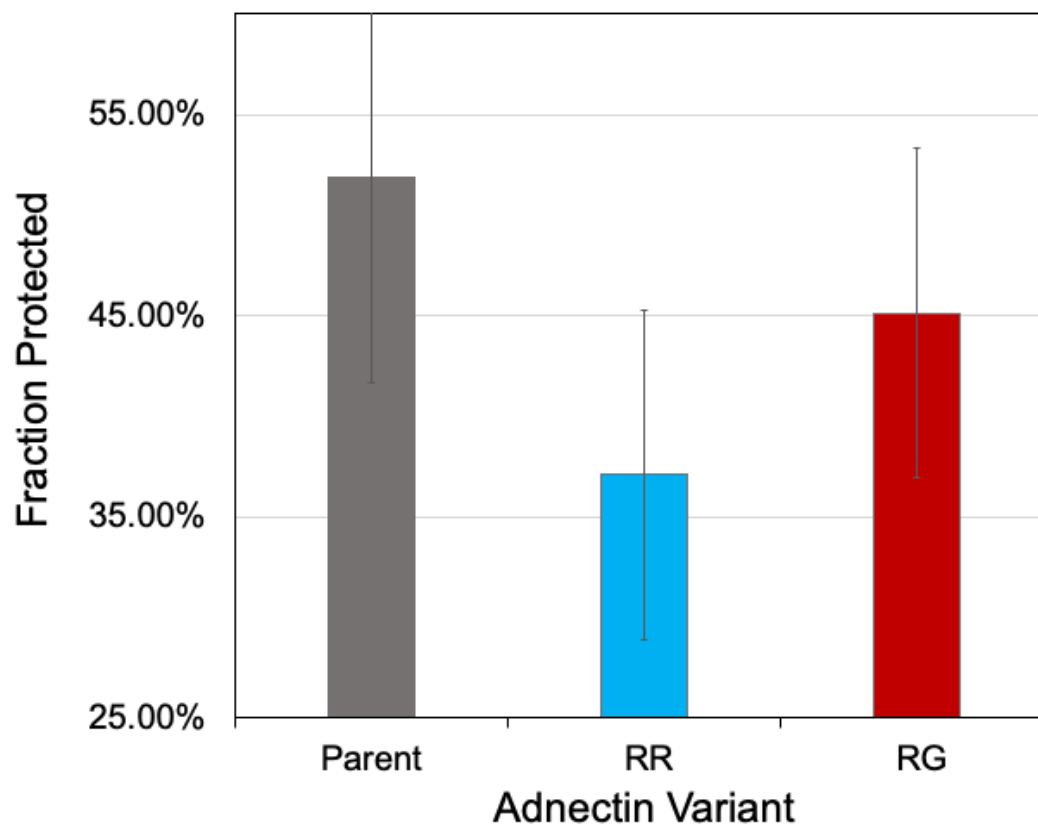


Figure E.5: Extent of overall amide exchange for Adnectin IBs in D₂O for 1 hour as determined by 1D ¹H-NMR spectra. Integrated area of amide pocket (7.98-8.5 ppm) in exchanged and protonated spectra were corrected for concentration using integrated area of methyl peaks (0.67-0.95 ppm). Error bars represent the standard deviation of two biological replicates for parent Adnectin and three biological replicates for engineered variants RR and RG.

Table E.1: Change in Adnectin monomer thermostability ($\Delta\Delta G$) upon point-mutations calculated using computational prediction tools. Individual scores were obtained using FoldX¹⁵¹, MAESTRO¹⁷⁵, DFIRE2.0¹⁵³, and PoPMuSiC¹⁵⁵ and normalized as described in section 4.2.3 for a maximum possible score of 4.0. Adnectin homology model generated by SwissModel was used as input. For FoldX and DFIRE, mutant structures were generated using FoldX BuildModel function.

Sequence position	Mutant residue Type	Computational Stability Predictor				
		FoldX	Maestro	PoPMuSiC	DFIRE	Normalized
18	P	4.71	4.082	4.39	3.326	1.528
	G	6.02	4.334	4.18	4.628	1.947
	K	3.7	3.692	2.91	3.605	0.958
	R	4.86	2.739	2.21	3.071	0.344
	E	4.41	4.082	4.04	3.956	1.549
	D	5.28	4.09	4.26	4.141	1.712
20	P	4.84	4.483	5.58	4.603	2.298
	G	5.01	4.694	5.56	5.571	2.600
	K	7.8	3.957	5.28	4.886	2.268
	R	10.24	3.653	4.29	4.462	1.956
	E	6.48	3.884	5.92	5.150	2.376
	D	6.37	3.884	5.66	5.801	2.438
22	P	8.07	4.63	5.19	6.285	2.841
	G	7.54	5.169	4.84	7.180	3.125
	K	6.67	4.07	3.95	6.795	2.289
	R	7.34	3.35	3.62	5.935	1.765
	E	6.85	4.302	4.78	6.785	2.619
	D	7.6	4.488	4.79	6.884	2.772
32	P	3.75	4.318	2.86	4.240	1.343
	G	5.2	5.022	3.98	5.600	2.329
	K	3.48	4.018	3.75	4.734	1.547
	R	5.31	3.356	2.66	4.063	0.964
	E	3.42	4.604	4.27	4.562	1.887
	D	1.58	4.374	4.2	4.349	1.599
34	P	5.1	4.275	4.59	4.049	1.845
	G	6.87	4.686	4.21	5.269	2.298

	K	10.59	4.078	4.17	5.313	2.306
	R	7.58	3.398	2.91	5.765	1.574
	E	6.16	4.062	4.75	4.904	2.058
	D	6	3.951	4.82	5.063	2.055
59	P	6.28	4.191	4.96	3.703	1.918
	G	6	4.534	4.44	5.056	2.191
	K	5.54	3.911	3.92	4.646	1.674
	R	6.36	2.963	3.21	3.462	0.894
	E	4.8	4.003	4.21	4.903	1.793
	D	4.96	4.064	4.36	5.136	1.920
70	P	7.63	4.426	5.41	5.735	2.668
	G	6.19	4.687	4.9	5.573	2.502
	K	4.66	4.09	4.3	4.782	1.818
	R	8.6	3.506	3.33	3.825	1.385
	E	5.3	3.985	4.86	5.216	2.064
	D	5.53	3.878	5.19	5.370	2.158
72	P	5.88	4.143	4.54	2.539	1.506
	G	5.58	4.574	3.84	4.214	1.835
	K	10.03	3.968	4.02	3.912	1.879
	R	15.91	3.205	2.89	3.194	1.516
	E	9.04	3.938	4.47	4.255	1.993
	D	4.8	4.14	4.34	4.328	1.761

Table E.2: Change in Adnectin W22G monomer thermostability ($\Delta\Delta G$) upon a second point-mutation calculated using computational prediction tools. Individual scores were obtained using FoldX ¹⁵¹, MAESTRO ¹⁷⁵, and PoPMuSiC ¹⁵⁵ and normalized as described in section 4.2.3 for a maximum possible score of 4.0. Mutant structures were generated using FoldX BuildModel function.

Sequence position	Mutant residue	Computational Stability Predictor			
		FoldX	Maestro	PoPMuSiC	Normalized
20	K	12.1271	4.279	11.537	1.3358
	R	12.8685	4.24	10.531	1.1634
	E	12.7364	3.973	11.704	1.0470
	D	13.269	4.081	11.965	1.7206
34	K	11.886	4.201	11.788	1.1445
	R	13.6887	4.18	10.994	1.7065
	E	12.2971	4.4048	12.113	2.0039
	D	13.0269	4.076	12.57	1.8714