

Proteolysis and Mass Spectrometry for Structural Characterization of Adnectin Inclusion Bodies

by

Xiaoyue Liu

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, 2024

© Xiaoyue Liu 2024

Author's Declaration

I hereby declare that I am the sole author of this 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.

Abstract

Inclusion bodies (IBs) are cellular aggregates that commonly form upon overexpression of protein in heterologous hosts. The formation of IBs is of broad interest for protein production for research, medical or biotechnological applications, and soluble or insoluble forms such as self-immobilized catalysts, drug delivery systems, and functional protein release systems. These aggregates are known to be heterogeneous, containing different conformations of the constituent protein. Adnectins are engineered binding proteins with high global thermostability and expression levels but varying propensity to form IBs, providing an attractive system for studying IB formation. A single-point mutation and two sets of multiple-point mutations with increasing charged residues were designed to alter the Adnectin IB structure. Fourier Transform Infrared spectroscopy (FTIR) showed an increased amyloid-like signal for the four-point mutations. To further analyze IB structure, we combined protease digestion with mass spectrometry (MS). Proteinase K, as a non-specific protease, can serve as an analytical tool by conducting parallel experiments digesting unlabeled and ^{15}N -labelled IBs, followed by matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) MS. This procedure facilitated the identification of digested fragments. Comparing the fragments from the Adnectin variant IBs reveals both similarities between the IBs as well as different protease-accessible sites indicating structural differences. This new procedure for protease digestion analysis provides a valuable complementary tool to help overcome challenges in characterizing heterogeneous aggregation processes and advance the understanding and, ultimately, control of IB structure.

Acknowledgments

I would like to thank my supervisor, Dr. Elizabeth Meiering, for providing me the opportunity to work in her group and pursue both undergraduate and graduate research.

I would also like to thank all the members of the Meiering lab, past and present, for their support throughout my research.

Finally, I would like to express my appreciation to my parents and family for their support and continuous encouragement throughout my studies.

Table of Contents

<i>Author's Declaration</i>	<i>ii</i>
<i>Abstract</i>	<i>iii</i>
<i>Acknowledgments</i>	<i>iv</i>
<i>List of Figures</i>	<i>vii</i>
<i>List of Abbreviations.....</i>	<i>viii</i>
Chapter 1: Introduction	1
1.1 Protein Aggregation	1
1.1.1 Protein folding, misfolding and aggregation.....	1
1.1.2 Protein aggregation pathways.....	2
1.1.3 IB formation and structure.....	5
1.2 10th Fibronectin Type III Domain and Adnectins	8
1.2.1 Unfolding, folding and biological function of ¹⁰ FN3.....	8
1.2.2 Adnectin and rationally designed variants.....	11
1.3 Protease Digestion as a Structure Characterization Tool	12
1.3.1 Trypsin.....	13
1.3.2 Proteinase K (PK).....	13
1.4 Mass Spectrometry.....	14
1.5 Research Objectives	16
Chapter 2: Methods	18
2.1 Sample Preparation.....	18
2.1.1 Cell growth and protein expression	18
2.1.2 IB preparation.....	18
2.2 Low-resolution Methods.....	19
2.2.1 Chemical denaturation	19
2.2.2 Attenuated Total Reflectance (ATR)-FTIR	20
2.3 Method development for IB analysis by proteolysis and mass spectrometry.....	21
2.3.1 Determination of conditions for proteolytic digestion of IBs	22
2.3.2 Digestion of Adnectin variants IBs and MS analysis.....	23
2.3.3 Control experiments on pure protein and denatured IBs	28
2.3.4 Replicate samples	29
2.3.5 Discussion and conclusion of the method development	29
Chapter 3: Results and Discussion.....	30
3.1 Chemical denaturation	30
3.2 ATR-FTIR revealing secondary structure composition.....	32
3.3 Proteolysis and MS.....	33
3.3.1 Post-translational modification of 4 Adnectin variant IBs.....	33
3.3.2 Proteolysis during lysing process	36
3.3.3 Trypsin cut sites on variant IBs	38
3.3.4 PK digestion reveals information on protected structure in IBs	38

<i>Chapter 4: Conclusions and Future Work</i>	49
4.1 Conclusions	49
4.2 Future Work	50
<i>References</i>	51
<i>Appendix</i>	57
<i>Appendix-Chapter 2</i>	57
<i>Appendix-Chapter 3</i>	59

List of Figures

Figure 1.1: Protein folding and aggregation landscape.....	2
Figure 1.2: Generating ensembles of dynamic misfolding proteins on the scale of the flexibility of the precursors.....	3
Figure 1.3: Domain-swapping assembly illustration.....	4
Figure 1.4: IB composition.....	7
Figure 1.5: Fibronectin and the ¹⁰ FNIII.....	9
Figure 1.6: Φ-values analysis of FN10 domain.....	10
Figure 1.7: Domain-swapped dimers and corresponding assembled FN3 oligomers.....	11
Figure 1.8: Adnectin variants and IB formation.....	12
Figure 1.9: Secondary structure component in Adnectin IBs.....	17
Figure 2.1: ATR-FTIR data analysis workflow.....	21
Figure 2.2: Tricine-SDS gel analysis for PK digestion condition determination.....	22
Figure 2.3: “Mass shift” in MS spectrum.....	24
Figure 2.4: Parallel experiment PK digestion on unlabelled and ¹⁵ N-labelled Parent IB.....	25
Figure 2.5: The sample preparation workflow for ESI-MS/(MS) analysis.....	26
Figure 2.6: ESI/MS of pure Parent Adnectin and solubilized Parent IBs.....	27
Figure 2.7: Example of fragment identification by ESI/MS.....	28
Figure 3.1: Chemical denaturation profiles of 4 Adnectin variants IBs.....	30
Figure 3.2: FTIR spectra revealing secondary structure components of Adnectin IBs.....	33
Figure 3.3: PTMs of 4 Adnectin Variants IBs.....	35
Figure 3.4: Fragments identified from four Adnectin variants IBs samples.....	37
Figure 3.5: Trypsin cut sites observed in Adnectin variant IBs from MS analysis.....	38
Figure 3.6: Structure analysis from PK digestion on Parent Adnectin IBs.....	40
Figure 3.7: MALDI spectra of PK digestion control experiments.....	42
Figure 3.8: PK digestion profile of SDS denatured IBs.....	43
Figure 3.9: PK digestion profile on 4 Adnectin variants IBs.	44
Figure 3.10: PK cut sites of 4 Adnectin IBs show on Adnectin native structure.....	46
Figure A2.1: ESI-MS/MS for fragment identification.....	57
Figure A2.2: Fragments identification of MALDI analysis, reproducibility analysis.....	58
Figure A3.1: Figure A3.1: PTM analysis of Parent, V75R, RR, RRGR IB.....	59
Figure A3.2: SDS-PAGE analysis of 4 Adnectin variant IBs.	60
Figure A3.3: ZipperDB and Tango aggregation propensity prediction.....	61
Figure A3.4: qHDX protection of RR and RRGR IB compared to Parent IB.....	62
Figure A3.5: qHDX protection of V75R IB compared to Parent IB.....	62

List of Abbreviations

A β	Amyloid-beta
AD	Alzheimer's Disease
ALS	Amyotrophic Lateral Sclerosis
ATR	Attenuated Total Reflectance
α Syn	α -Synuclein
β_2 m	β_2 -microglobulin
CryoEM	Cryo-Electron Microscopy
DCTB	2-[(2E)-3-(4-tert-butylphenyl)-2-methylprop-2-enylidene] malononitrile
DHB	2,5-dihydroxybenzoic acid
DMSO	Dimethylsulfoxide
DSC	Differential Scanning Calorimetry
EDTA	Ethylenediaminetetraacetic Acid
ESI	Electrospray Ionization
HCCA	α -cyano-4-hydroxycinnamic acid
FA	Formic Acid
FTIR	Fourier Transform Infrared Spectrometry
GdnHCl	Guanidine Hydrochloride
IBs	Inclusion Bodies
IDP	Intrinsically Disordered Proteins
IDR	Intrinsically Disordered Region
IPTG	isopropyl β -D-1-thiogalactopyranoside
LB	Luria Broth
LLPS	Liquid-Liquid Phase Separation
MALDI	Matrix-Assisted Laser Desorption/Ionization
MAP	Methionine Aminopeptidase
MQ	Milli-Q
MW	Molecular Weights
nOG	N-Octylglucoside
OD	Optical Density
PK	Proteinase K
PMSF	Phenylmethanesulphonyl Fluoride
PTM	Post-Translational Modification
qHDX	Quenched Hydrogen/Deuterium Exchange
SA	Sinapinic Acid
SDS-PAGE	Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis
SOD1	Superoxide Dismutase 1
ssNMR	Solid-State Nuclear Magnetic Resonance
^{10} F n 3	10th Fibronectin Type III Domain
TEN	Tris, EDTA, NaCl
THF	Tetrahydrofuran
TOF	Time-Of-Flight
Q	Quadrupole

Chapter 1: Introduction

1.1 Protein Aggregation

1.1.1 Protein folding, misfolding and aggregation

Proteins are macromolecules found in all organisms on Earth. These polymers are synthesized as linear chains of individual amino acids covalently linked by peptide bonds. This linear arrangement of the polypeptide is called the protein primary structure. For globular proteins, the polypeptide folds to adopt a native structure and achieve their function. The protein folding process can happen co-translationally or after the synthesized protein leave the ribosome¹. Protein folding is also a highly complex and heterogeneous process, relying on the cooperation of many weak non-covalent interactions, including the formation of hydrogen bonds, ionic bonds, hydrophobic effect, and van der Waals forces². However, as the thermodynamic hypothesis states, the native structure of a protein is encoded by its amino acid sequence³.

Protein folding has been investigated extensively, resulting in models that describe this process. Generally, the models describe the native protein structure is relevant to the primary sequence, independent from the folding pathway, and the native state of a protein generally has the lowest Gibbs energy. Even though none of the models can fully explain the entire process of the various protein species, they all provide a better understanding of the factors that may influence the process and lead to better models. In one of the recent models shown in (Figure 1.1, left), the polypeptide chain is considered to explore the energy surface of the funnel as it follows several downhill routes toward the correct native structure⁴.

Importantly, errors can occur during folding, leading to protein misfolding. The misfolding phenomenon can result in loss of biological function and can have a toxic effect on living cells⁵. Both internal and external factors can cause misfolding. Internal factors can influence protein folding in several ways. A protein may pass through several intermediate folding states to achieve native structure, and some intermediate states possess interactions that are not present in the protein native state of the protein. Falling into these kinetic traps can lead to protein misfolding⁵. Mutations in the primary sequence are another main internal factor of protein misfolding, as they influence

the folding process and affect the folded structure of a protein. Examples of external factors include the malfunction of protein-folding chaperones, high cellular stress, and temperature^{6,7}. A common feature of misfolded protein is the exposure of hydrophobic regions. One of the main driving forces of protein folding is hydrophobic collapse, which usually results in a hydrophobic core in the natively folded protein⁴. The exposure of the hydrophobic region of proteins could greatly affect their natural biological function as well as disfunction. Interactions between these misfolded proteins can be thermodynamically favorable, leading to the formation of a variety of stable protein-associated states known as protein aggregates (Figure 1.1, right). Similar to protein folding, protein aggregation is driven by various non-covalent forces and involves the formation of aggregation intermediates⁴. Among all kinds of investigated aggregates, amyloid fibrils are commonly recognized as stable aggregate structures, typically a cross- β fold that occupies the energy minima of the protein aggregation energy landscape⁴.

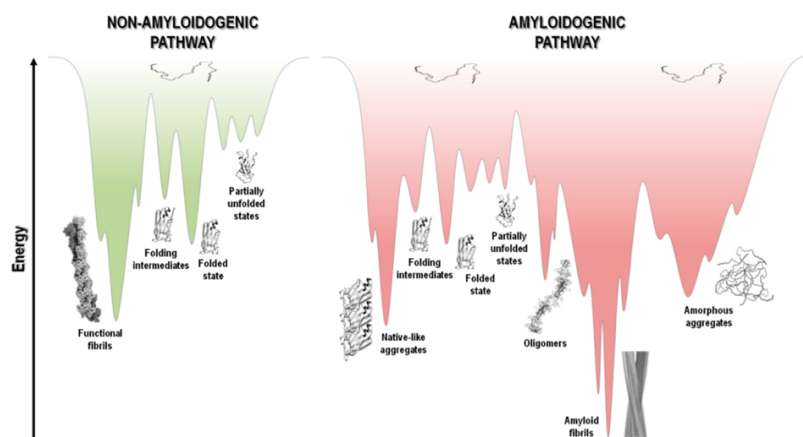


Figure 1.1: Protein folding and aggregation landscape. The non-amyloidogenic landscape is presented in green on the left, and the amyloidogenic landscape is presented in red on the right. The local minima represent possible stable folded or aggregated states⁴.

1.1.2 Protein aggregation pathways

Protein aggregation usually results in an insoluble protein assembly. Aggregated species are typically toxic and commonly found in neurodegenerative diseases, such as amyotrophic lateral sclerosis (ALS), Parkinson's disease, Alzheimer's disease (AD) and prion disease⁸. Therefore, it is of great importance to study the aggregation process. However, the aggregation process is highly

complicated and heterogeneous and could vary between precursor proteins under different conditions⁹. Environmental conditions, including temperature, pH, and stress, can greatly influence protein aggregation, resulting in heterogeneous pathways⁴. This means *in vivo* and *in vitro* aggregation can be significantly different. Compared with the complex cellular environment, *in vitro* aggregation conditions are easier to control and have been extensively studied¹⁰. Research has confirmed that the final aggregates of the same precursor can vary with experimental conditions. Despite the differences, it is commonly recognized that the aggregation process can be classified into several stages: initiation, elongation, and reaching the stable final structure¹¹.

Proteins can contain flexible and dynamic regions that enable them to perform biological functions¹². However, protein flexibility can also initiate protein aggregation⁹. A classical aggregation mechanism starts with the interaction between the dynamic region of the monomers. This nucleation then seeds the formation of soluble intermediates⁹. These oligomers undergo further aggregation processes and eventually reach a relatively stable aggregate structure. Figure 1.2 describes various aggregation pathways starting from protein precursors with different flexibility⁹. On the flexibility scale, the intrinsically disordered proteins (IDP) are the most dynamic, including α -synuclein (α Syn), tau, and amyloid-beta ($A\beta$), followed by proteins with the intrinsically disordered region (IDR), such as prion proteins, and fully folded protein are considered having lowest flexibility such as transthyretin and β_2 -microglobulin (β_2 m)⁹.

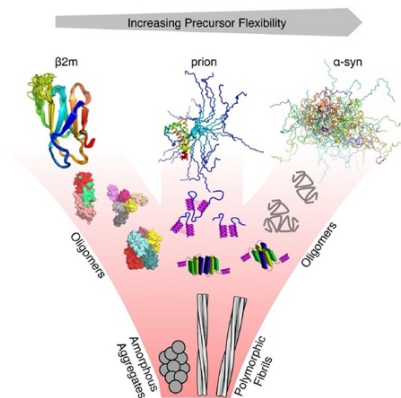


Figure 1.2: Generating ensembles of dynamic misfolding proteins on the scale of the flexibility of the precursors. A globular protein with local fluctuation may preserve native features in aggregates, and IDPs could have a greater tendency to form amyloid-like structures⁹.

For some of the well-folded proteins, one of the oligomerization processes is known as domain-swapping (Figure 1.3). This process allows the monomeric proteins to exchange a strand or domain to generate multimers. It can also result in an open-ended aggregation. This mechanism has been found in both *in vivo* and *in vitro*¹³.



Figure 1.3: Domain-swapping assembly illustration. The blue and green are swap domains, and loops form a β -sheet structure¹⁴.

Studies have been performed experimentally and theoretically to investigate protein aggregation pathways. However, describing the species involved in the different stages of aggregation in terms of individual conformations is challenging since the proteins undergo conformation changes under different timescales. For example, the initiation of aggregation of an IDP could happen in a millisecond timescale¹⁵. Other proteins with a tight fold and less dynamic regions tend to be more resilient to aggregation. However, the aggregation may still be initiated by factors such as under misfolding conditions, proteolysis, or fusing to an aggregation-prone tag¹⁶. It has been proposed that the elongation phase is not only about the increase of the oligomer size or molecular weight but also about concomitant structural transitions. Such structural transitions can occur at all stages usually reach the final stable amyloid-like aggregates. Researchers can resolve some aggregate structures using high resolution by cryo-electron microscopy (CryoEM) and solid-state nuclear magnetic resonance (ssNMR)¹⁷. These techniques are powerful in providing atomic resolution structural information of homogenous aggregates. The solved amyloid or amyloid-like structures have been summarized in the amyloid atlas¹⁷. However, cryoEM and ssNMR biophysical methods has are both restricted to analyzing homogeneous structures with limited local fluctuations¹⁸. Therefore, many regions of aggregates are disordered and so cannot be characterized by these methods.

Another aggregation process that has been studied in great detail recently is liquid-liquid phase separation (LLPS), which has been reported to be an unconventional aggregation pathway¹⁹. LLPS is a phase transition process initiated by protein condensate forming from a homogenous solution, resulting in a two-phase solution: condensed and diluted. The condensed phase later passes through a phase transition to a solid phase, which contains the insoluble aggregates. Like the conventional aggregation pathway described above, the LLPS has also been observed *in vivo* and *in vitro*. Phase-separated membrane-less organelles in cells can be important in regulating cellular functions²⁰. However, LLPS is also reported to be relevant to various human diseases, such as neurodegenerative diseases and cancer. Compared to cellular LLPS, LLPS *in vitro* is easily observable by controlling protein concentration and buffer conditions. However, as a new concept of protein aggregation, the detailed mechanisms of LLPS are not yet fully understood.

1.1.3 IB formation and structure

Inclusions can broadly refer to all kinds of cellular aggregates, which range from protein aggregates associated with neurodegenerative diseases to those associated with the overexpression of recombinant proteins in *E. coli*. The latter densely packed insoluble intracellular protein aggregates formed in the cytoplasm of cells by overexpressed recombinant proteins are known as inclusion bodies (IBs). Due to the large size and complexity of IBs²¹, the mechanisms of formation and structures of IBs remain poorly defined.

IBs are a nice system for systematically investigating mechanisms of cellular aggregation. Because of the ease of growth of *E. coli*, defined genetics, and facile control of cell growth conditions. Different cell strains have also been derived for various purposes. Generally, the derived strains can produce target proteins with high efficiency or increase the tolerance for producing toxic proteins. It is also easy to design point mutations on the target protein to understand molecular aggregation mechanisms under the exact control of cellular environments. The growth conditions, such as time and temperature, are easily controlled. Early studies on IBs targeted preventing the forming of IBs in cells since it was considered a problem during protein production²². However, in recent years, there has been an increase in interest in IBs because of their potential for biomedical and biotechnological applications. IBs have been investigated to serve as self-immobilized catalysts, drug delivery systems, and functional protein release systems²³.

The complexity of IB formation lies not only in the in-cell environment but also in the varying cellular conditions during the IB formation process, with increasing cellular stress. The continual changing of the cellular environment could cause IB formation to vary significantly from other aggregation pathways. Studies have been conducted to capture the initiation of IB formation, but it is difficult to perform experimentally. The early-stage aggregates are too small to visualize by EM and also hard to monitor by other biophysical methods due to the strong interference by other cellular components caused by the low abundance of IB²⁴. A recent study confirmed the existence of dense, rod-shaped IBs in the early stage of aggregation. Using an auto-fluorescent protein GFP⁺ as a model protein for analyzing IB formation, this study successfully captured and analyzed the early stages of IB formation²⁴. The structure of the rod-shaped aggregates is found to be porous in morphology, and multiple aggregates are found in each cell²⁴. However, it has been found that there are differences between the IBs formed by various precursor proteins. Over decades, studies with both theoretical and experimental methods have been conducted to find the correlations between precursor protein properties and their ability to form IBs²⁵. The studies mainly focus on protein folding kinetics and native monomer solubility and stability. Research has shown that protein *in vitro* solubility correlates with their propensity for IB formation²². Proteins with efficient folding pathways may have less probability of forming IBs²². Therefore, initiating IB formation is generally accepted to be in competition with protein folding, and the final IB structure may be highly dependent on the precursor folding pathway and solubility²⁶. Over the decades of exploration, knowledge of IB composition has greatly increased, and evidence has shown that native-like structures, amyloid-like structures, as well as other native folding and aggregation intermediates are present in the IB structure (Figure 3). Such heterogeneity is generally observed in IB structures, and the final IB structure results from the competition of native folding and aggregation.

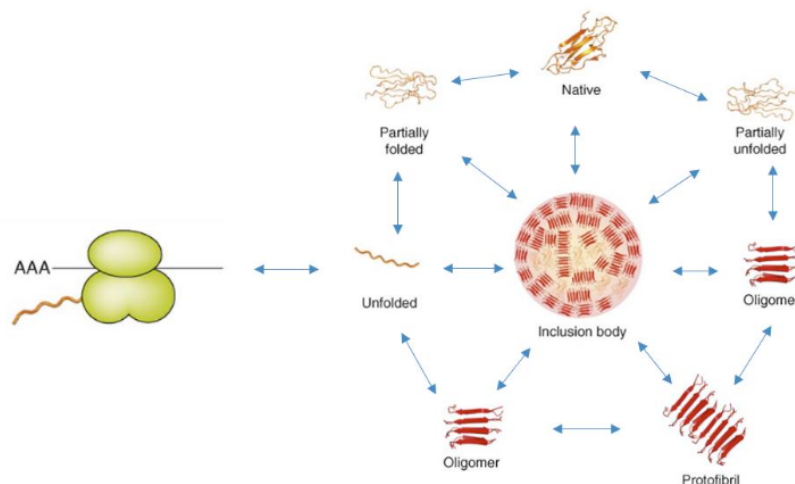


Figure 1.4: IB composition. Under high cellular stress, recombinant proteins may aggregate into higher-order structures. These large, dense aggregates can consist of different forms of the expressed protein, such as native, unfolded, and partially folded structures, along with their various assemblies²¹.

Methods for characterizing aggregates, in general, can also be used for IBs structure analysis. However, high-resolution biophysical methods for analyzing insoluble protein complexes, such as cryoEM and ssNMR, cannot comprehensively determine the structure of IBs. Nonetheless, they still provide valuable insights into the ordered region. It has allowed the determination of amyloid-like components in various IBs and further revealed the high similarity between IBs and their *in vitro* aggregates²⁷. Fourier Transform Infrared Spectrometry (FTIR) is a common method for revealing the secondary structure and heterogeneity of IBs. Interleukin-1 β , human growth hormone, β_2 m, and human Superoxide Dismutase 1 (SOD1) IB are all reported to have native-like components^{28–30}. Some of the studies further dive into high quality IBs analysis such as native protein-releasing systems, which possess similar mechanisms to secretory granules³¹. Unlike the amyloid-like structure, the native component is hard to observe directly by high-resolution methods. Combining moderate-resolution methods such as FTIR with quenched hydrogen/deuterium exchange (qHDX)-NMR, the association of native-like SOD1 monomers in IBs has been revealed. The understanding of IB structure was greatly enhanced by the development of various structural characterization techniques. However, characterizing IB structure at high resolution is still challenging owing to the heterogeneity and lack of high-resolution methods. To conquer the challenge, studies on different protein models should be performed, and additional methods need to be developed, as will be explored further in this thesis.

1.2 10th Fibronectin Type III Domain and Adnectins

A model system is required to understand IB formation. Adnectins, known as monobodies, are a family of engineered therapeutic proteins based on the 10th fibronectin type III domain (¹⁰Fn3)³². Adnectins have been designed to bind with high affinity and specificity to a wide variety of therapeutical targets for medical and biotechnological applications³². Applications of Adnectins have been developed in the field of biotherapeutics and biosensors³³. With the development of Adnectins for such applications, the study of their production, storage, and aggregation in bacteria has become of great interest³³. Adnectins also serve as a powerful system for unravelling specific molecular determinants of IBs formation owing to the existing knowledge of their aggregation behavior and propensity³⁴. To better understand the folding and aggregation mechanism of Adnectin in IBs, it may be valuable to compare to the well-studied biological structure and function of ¹⁰Fn3.

1.2.1 Unfolding, folding and biological function of ¹⁰FN3

The ¹⁰FN3 is an evolutionarily conserved, stable β -sandwich fold single-domain protein without disulphide bonds³⁵. The folding of ¹⁰FN3 is an extremely fast process that is hard to capture in cells. Many investigations were performed by simulations, and experiments to analyze the folding kinetics of ¹⁰FN3 are mainly performed *in vitro* with the assistance of denaturants. Studies have focused on the unfolding, refolding, and dynamics of ¹⁰FN3³⁵. Thus, it is a well-characterized model.

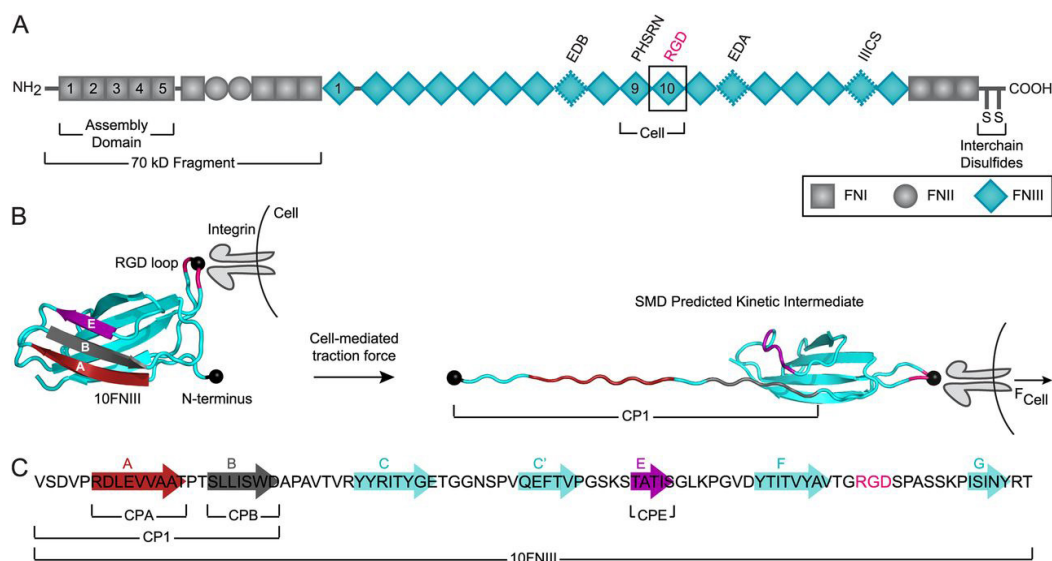


Figure 1.5: Fibronectin and the ¹⁰FNIII. A) Multidomain fibronectin and ⁹FN3, ¹⁰FN3 are responsible for cell-binding. B) ¹⁰FN3 monomer under cell-mediated traction force, a kinetic intermediate with unfolded N-terminus, is predicted by simulations. C) Sequence of ¹⁰FN3 with highlighted β -strands in native structure³⁵.

Biologically, fibronectin is a multidomain protein that exists as elastic fibrils in the extracellular matrix and undergoes reversible unfolding to perform its function. Among the domains in FN3, ¹⁰FN3 is believed to be the first to unfold due to its low mechanical stability. During the unfolding of ¹⁰FN3, ‘cryptic binding sites’ are revealed to facilitate the aggregation of the individual domains to fibrils³⁵. This mechanical unfolding has been extensively studied to gain insight into the initiation of fibrillogenesis. It has been confirmed that two distinct unfolding pathways exist under mechanical stress, tested experimentally and in computation. One is the releasing of the A and B strands, and the G strand is detached first for the other. On the other hand, the cellular-mediated unfolding of FN3 for performing biological function has also been simulated. As shown in Figure 1.5, this prediction indicates that the unfolding intermediate will cause the exposure of β -strands A but not strand E³⁵.

¹⁰FN3 is also known to be able to refold after being denatured by strong denaturants, which allows the investigation of its refolding in a controlled environment³⁶. A three-state transition has been proposed to describe the refolding of ¹⁰FN3, and one intermediate state during the refolding has been found under low-concentration of guanidine isothiocyanate³⁶. Later, Φ -value analysis

revealed more details of the intermediate structure, that the C, D, E, and F β -strands form an extended nucleus, and among these strands, β D and β E have lower Φ -values compared to C and F strands³⁷ (Figure 1.6). The higher Φ -value represents the residues are involved in a more stable structure.

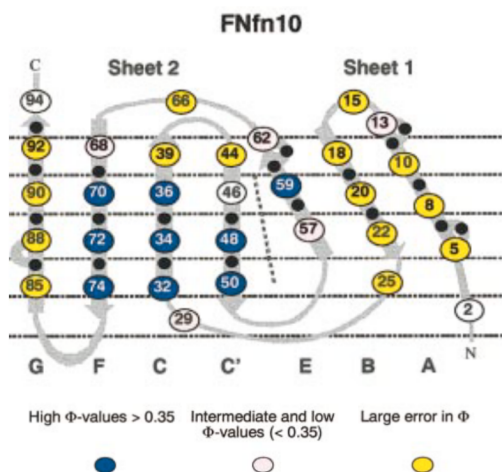


Figure 1.6: Φ -values analysis of FN10 domain. Different color-coded residue numbers representing different Φ -value ranges are presented in the schematic model³⁷.

Interestingly, the multimeric forms of ¹⁰FN3 have been observed by X-ray crystallography. These structures are mainly formed by domain swapping; as mentioned in 1.1.2, the β -strand A or β -strand G is capable of initiating the swapping³⁸. Figure 1.7 below shows the observed crystal structure of hexameric oligomers associated with G-strand and A-strand swapped dimers and corresponding higher-order multimers³⁹. The G-strand swapped dimers have overall high similarities with differences caused by various FG loop lengths. Size exclusion chromatography showed that the oligomerization was present before the crystallization process. The domain swapping can be extended to an open-ended aggregation of the monobodies, as illustrated in Section 1.1.2³⁹.

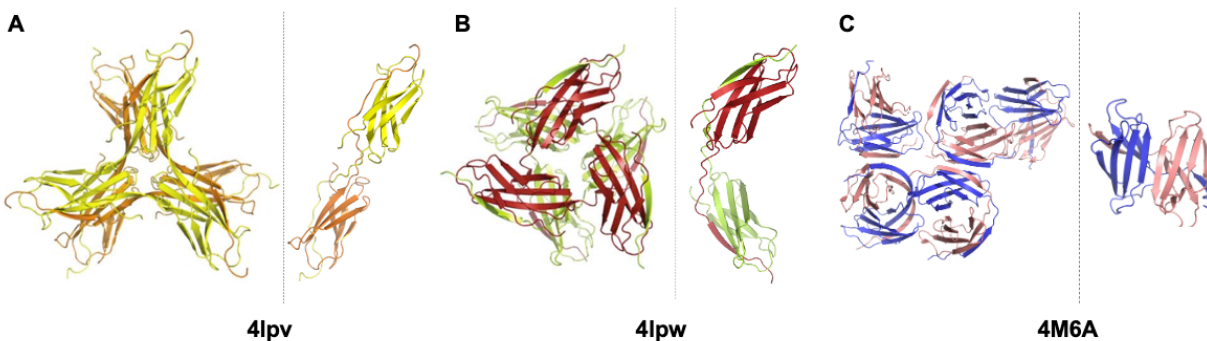


Figure 1.7: Domain-swapped dimers and corresponding assembled FN3 oligomers. For each structure, the oligomers are presented on the left, and dimers are presented on the right. A) 4LPV and B) 4LPW are G-strand swapped hexamers with 8 and 9 residues in the FG loop. C) 4M6A represents a tightly packed A-strand swapped dimer^{38,39}.

1.2.2 Adnectin and rationally designed variants

Adnectins are highly similar to ¹⁰FN3 in primary sequence and protein structure. The differences are focused on the binding loops for their designed biotechnical applications³². Adnectins have favorable characteristics of high thermostability and solubility and lack of disulfide bonds. Therefore, the solubility, structural dynamics, and production yield of Adnectin and its variants have been widely studied. Particularly, the effects of mutations on the FG loop and their relative solubility have been characterized³⁴. Experiments on IB formation by FG loop variants have also been performed. It has been reported that the monomer solubility could be relevant to their IB formation propensity²⁶. After that, more detailed structural characterization on Adnectin variant IBs has been performed. Research on IB chemical stability, secondary structure characterization, and high resolution to residue level protection pattern against backbone HDX has been conducted. However, in this fast-folding, high-stability model, most single-point mutations have little impact on their IB formation and structure. Multi-point mutations have been rationally designed to investigate the possibility of manipulating IB structure by primary sequence mutations. V75R is a well-characterized mutant that is confirmed to have a higher solubility compared to the Parent Adnectin. In the crystallized Parent Adnectin, a G-strand swapped dimer was found. Therefore, an increased surface charge modulates monomer solubility, and IB structure was hypothesized. In Figure 5, the design rationale of two other mutants is illustrated. One mutant RR has two surface

point mutations that lie on β A, V11R, and β B, L19R. The mutant aims to keep the monomer native fold but prevent N terminal aggregation. Another variant, RRGGR, has two additional mutations in the hydrophobic core, W22G and I34R; these additional mutations are predicted to disrupt the native fold. Initial experiments have been conducted to determine the effects of these mutations on IB formation. Chemical denaturation, FTIR, and qHDX are all widely used to characterize IB structure. These methods have provided valuable information. However, additional methods should be applied to better understand the IB structure, as discussed further below.

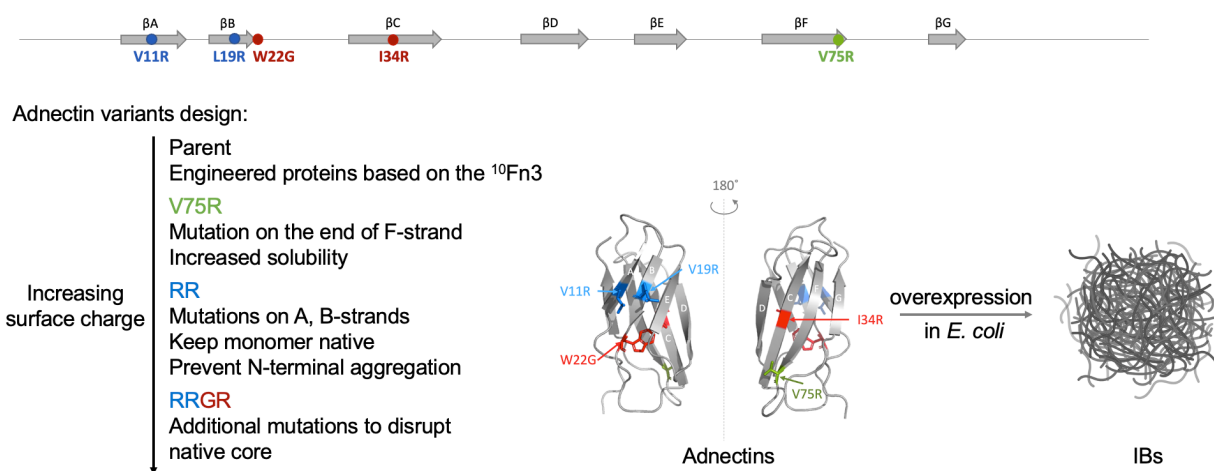


Figure 1.8: Adnectin variants and IB formation. Designing rationale is listed on the left, with mutations on the ribbon structure present on the top. The native fold of the Adnectin ribbon structure is presented with mutation points labelled. The ribbon structure is presented in two side with a 180° rotation. One side β A, β B and β E, the other side includes β C, β D, β F, and β G.

1.3 Protease Digestion as a Structure Characterization Tool

Proteolysis is enzyme-catalyzed hydrolysis of peptide bonds between amino acid residues. In organisms, proteolysis is essential for converting nutrients and is important in regulating cellular activity by participating in post-translational modifications⁴⁰. In the laboratory, it is widely used to digest large protein complexes into smaller fragments for MS analysis. It is a standard procedure in bottom-up proteomics and various methods that involve MS analysis⁴¹. Based on different protease properties, their different usage has been developed.

1.3.1 Trypsin

Trypsin is a serine protease with high specificity for cleavage at the carboxyl side of Lys and Arg. The binding mechanism by which trypsin recognizes the basic side chains has been solidly established. As a serine protease, trypsin contains a catalytic triad involving aspartate, histidine, and serine⁴². Research also confirmed that trypsin cannot digest lysine and arginine derivatives⁴³. Therefore, studies on localizing specific modifications of these two residues have been conducted and adopted for other covalent labelling and cross-linking methods for protein structure synthesis.

Trypsin is commonly used in proteomics; the specificity of trypsin has a great advantage for MS analysis. Tryptic fragments are readily analyzable by commonly available MS methods. Therefore, trypsin has been commonly used to cleave the intact proteins into an MS-amenable fragment mixture. Trypsin is also used to extract Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis (SDS-PAGE) separated proteins, combined with MS analysis, which allows for the accurate detection of proteins and their isoforms⁴⁴. Other methods based on trypsin specificity have been developed, such as residue-specific covalent labelling. Chemicals like N-hydroxysuccinimide ester and formaldehyde can serve as probes specifically covalently linked to Lys residues⁴³. The modified Lys residues will increase the fragment masses, and covalently linked Lys cannot be recognized by trypsin, indicating the Lys residues are exposed to solvents. This strategy has been used to analyze intracellular conformational changes. Even without the help of chemical labelling, the trypsin digestion mechanism can reveal some extent of structural information of proteins. Since the digestion process requires Lys or Arg to be exposed, the buried residues are inaccessible to trypsin digestion. It has been reported that some amyloid fibrils are resistant to trypsin digestion⁴⁵. Moreover, trypsin can reveal a protease-resistant fragment *in vitro*⁴⁶. analysis of the fragment produced can be inferred to the overall structure of the large protein complex.

1.3.2 Proteinase K (PK)

Proteinase K (PK) is a serine endopeptidase similar to trypsin but differs in its broad specificity of digestion sites. This protease is commonly used for DNA and RNA extraction because it is powerful at degrading proteins into small fragments under strong denaturing conditions. It has

been reported that PK prefers cleaving after hydrophobic residues⁴⁷. However, PK nevertheless has a much less specific cleavage pattern than trypsin. This lack of specificity makes PK particularly useful for analyzing aggregate resistance to proteolytic digestion and detecting protease-resistant fibril cores, as the PK can degrade most exposed protein regions while leaving the highly stable fibril cores intact⁴⁸. Consequently, PK is widely used in biochemical research to analyze protein aggregate structures and study protein aggregation.

PK digestion of protein aggregates provides valuable information both kinetically and structurally. The availability of the cleavage sites in substrate protein aggregates is essential for kinetics⁴⁹. It is confirmed that PK has high activity with the presence of detergents and denaturants. Later, research showed that higher relative activity in the presence of these additives was caused by the significant increase in the exposure of the substrate binding sites when they were unfolded. It is also a result of the high stability of the PK structure and fragmented PK remaining active for performing proteolysis. With this basic understanding, the relative protease resistance of aggregate can be inferred by measuring the turbidity changes under the same digestion conditions. The outstanding difference in the kinetic time course of aggregation digestion proved the accessibility of the substrate could outweigh the PK preference of hydrophobic residues. Therefore, PK digestion can provide a potentially high-resolution technique in structural characterization. The cut sites of PK digestion on aggregates can be directly interpreted in terms of residue exposure. Longer fragments can be inferred to be a protected or buried region in the aggregate.

1.4 Mass Spectrometry

After proteolytic digestion, a method for identifying fragments is required. A turbidity scan or SDS-PAGE is a low-resolution method for visualizing the differences in aggregate digestion. MS can be a powerful tool for providing high-resolution results. In general, MS is an analytical method for detecting and separating ions based on their mass-to-charge ratio and providing numeric results. Based on this fundamental principle, numerous mass spectrometers have been developed for different purposes. There was a barrier to protein analysis with MS until revolutionary soft ionization methods were developed. Since then, MS has been a popular technique for large-scale

proteomics and protein structure analysis, complementing other biophysical methods. MS analysis includes several steps: sample introduction, MS analysis, ion detection, and data interpretation.

The ion source determines the samples that the spectrometer can analyze. Early mass spectrometers were limited to analyzing gas-phase ions. The development of ionization methods, such as matrix-assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI)^{50,51}, greatly broadened the application of MS. These two methods can both convert ions to gas phase with limited fragmentation of macromolecules. Under certain conditions, MALDI can even retain the analyte structure for other imaging methods⁵². MALDI uses a certain matrix to transform macromolecules into ions. The mechanism is complicated and still not fully understood. However, it is certain that different matrices can be chosen for different analytes depending on their chemical properties and mass range for analysis. Commonly used matrices for protein and peptide analysis are sinapinic acid (SA), α -cyano-4-hydroxycinnamic acid (HCCA) and 2,5-dihydroxybenzoic acid (DHB)⁵³. Typically, proteins are converted to a singly or doubly charged ion. The detection mass range for protein analysis can cover anything from lower than 1 kDa to higher than 100 kDa. For ESI, the ionization of a protein usually happens under denaturing conditions. The charge state of a denatured protein can easily reach 10 or more charges, generating an envelope of charge distribution. ESI is also commonly combined with a liquid chromatography system to further increase the resolution by the analysis of sample hydrophobicity. Both MALDI and ESI have a wide range of applications. They are capable of large-scale proteomics, metabolomics, structural biology, and pharmaceutical analysis.

Another crucial component in a mass spectrometer is the ion analyzer. They have the same fundamental principle of separating ions by their mass-to-charge ratio. However, various versions exist with different bases for separation: time-of-flight (TOF) by ion flight time, quadrupole by the ions stability of the oscillation trajectories in electric fields, ion trap and magnetic sector by the ions oscillation frequency⁵⁴. Among these ion separation methods, orbitrap, as a novel ion trap analyzer, has a very high resolution. A tandem or hybrid analyzer with more than one detection methods can further combine their strength and improve the performance of the analysis⁵⁵.

The two mass spectrometers used in this project are MALDI-TOF and ESI-Quadrupole (Q)-orbitrap, with different combinations of ion source and mass analyzer. Both instruments have proven to be powerful protein analytical tools. They have been widely used as structural characterization tools in various biochemical methods: proteolysis, cross-linking, or HDX⁵⁶.

1.5 Research Objectives

The research presented here aimed to structurally characterize IBs using conventional low-resolution methods and a novel approach involving proteolysis and MS analysis. Adnectin is the model system we use to analyze IB structure through chemical denaturation and FTIR. These low-resolution methods will provide insights into the stability and secondary structure components of the IBs.

Comparing the pure protein and IBs FTIR spectra (Figure 1.9) reveals similarities. IBs display significant native and β -sheet content and possess a significant unstructured signal. The β band shifts to a lower wavenumber, which is a characteristic of fibrils. Among the 4 Adnectin variants, the FTIR spectra show differences; notably, the RRGR IB has a considerably higher unstructured signal and a significant β band shift to a lower wavenumber. To further investigate structural differences, chemical denaturation using urea and guanidine hydrochloride (GdnHCl) will be performed. FTIR analysis of variant IBs in Milli-Q (MQ) will examine the potential influence of buffer environment changes on IB structure. Additional methods are required to obtain more detailed information on IB structure.

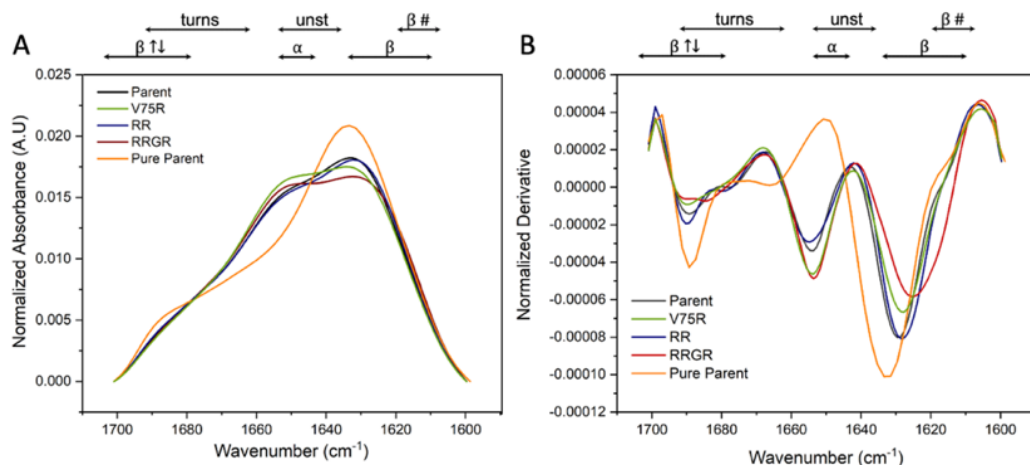


Figure 1.9: Secondary structure component in Adnectin IBs. Comparison of IBs to pure Parent Adnectin indicating native features in IBs. A) Absorbance spectra in the amide I region normalized to an area of 1. B) Second derivative calculated by OriginPro2018.

This thesis introduces a new method that combines proteolysis with MS analysis. This technique involves conducting parallel proteolysis experiments on both unlabeled and ^{15}N -labeled IBs, followed by MS analysis to identify protected fragments within the variant IBs. This approach will enable the identification of proteolytic fragments despite the presence of contaminants in the IBs. Several rationally designed Adnectin mutants will be used to investigate the IB structure. By combining conventional low-resolution methods with this newly developed technique, the study aims to characterize the heterogeneous aggregate structure of IBs. This research seeks to enhance the understanding of IB structure. This study aims to reveal the IBs structure and verify the effects of point mutations on mature IB structure and, ultimately, seek the control of IBs structure for medical and biotechnological applications.

Chapter 2: Methods

2.1 Sample Preparation

2.1.1 Cell growth and protein expression

Adnectin variants were expressed in *E. coli* BL21(DE3) pLysS cells containing pET-9d vector. All solutions were made using MQ ultrapure water, and all chemicals were purchased from *BioShop* unless specified. The glycerol stocks were added to 50ml Luria Broth (LB) media using 50 $\mu\text{g mL}^{-1}$ Kanamycin and 34 $\mu\text{g mL}^{-1}$ chloramphenicol with shaking at 37°C 225 RPM overnight. The cells were then collected by centrifugation at 4,500 xg for 10 minutes at 4°C then resuspended in 5 mL of the supernatant and transferred to a 4L flask containing 1L 1xM9 salts (12.8g $\text{Na}_2\text{HPO}_4 \cdot \text{H}_2\text{O}$, 3.0g KH_2PO_4 , 500mg NaCl) which was shaken overnight. 20mL 20% Glucose, 1mL 100mM CaCl_2 , 2mL 1M MgSO_4 , 5mL 100mg/mL $^{14}\text{NH}_4\text{Cl}$ (or $^{15}\text{NH}_4\text{Cl}$ for isotopically labelled samples), 1mL trace metal solution (3 μM $(\text{NH}_4)_6\text{Mo}_7\text{O}_{24}$, 400 μM H_3BO_3 , 30 μM CoCl_2 , 10 μM CuSO_4 , 80 μM MnCl_2 and 10 μM ZnCl_2) were added, alongside 50 $\mu\text{g/mL}$ Kanamycin and 34 $\mu\text{g mL}^{-1}$ chloramphenicol. As the optical density reaches 0.7 ± 0.1 at 600nm (OD_{600}), the protein expression is initiated by the addition of 1 mM isopropyl β -D-1-thiogalactopyranoside (IPTG) incubation at the same condition (37°C, 225 rpm) for an overnight. The cells are harvested at 4,500 xg for 15 minutes at 4°C. For parallel labelled experiments, the duplicate flask of ^{15}N labelled growth was performed simultaneously with the unlabelled growth.

2.1.2 IB preparation

For each 500 mL culture pelleted cells, 15 mL of TEN buffer (containing 20 mM tris (hydroxymethyl) aminomethane (Tris), 1 mM ethylenediaminetetraacetic acid (EDTA), and 100 mM sodium chloride (NaCl)) pH 8.1 or TE buffer (containing 10 mM Tris pH 8.0; 1 mM EDTA) pH 8.0 was added. For MS analysis of fragmentation during lysing, 1 mM phenylmethylsulphonyl fluoride (PMSF) was added to the samples, which is aimed at limiting the proteolytic digestion during the lysing process. The cell pellets were immediately resuspended. Then, in a freeze-thaw cycle, the mixture was flash-frozen in liquid nitrogen and thawed in a 37 °C water bath. 1.5 mg DNase I (*Sigma-Aldrich*) was added to every 15 mL mixture and inverted 25 times for full mixing.

The samples were then incubated at room temperature for 30 minutes. Three additional freeze-thaw cycles were conducted. The insoluble fraction containing IBs was preserved in the pellets after centrifuging at 16,300 xg at 4 °C for 15 minutes. Each pellet from the 15 mL mixture was resuspended in 4 mL TEN buffer for the samples lysed in TEN buffer or 4 mL MQ water for the samples lysed in TE buffer. The approximate concentration of Adnectins in IB samples is 5 mg/mL. The samples were aliquoted and flash-frozen at -80°C. Further washing treatments n-Octylglucoside (nOG) (*Thermo Scientific*) were also conducted and served as a purification technique for IBs. IB samples were pelleted and resuspended in the same volume of 1% nOG. After 10 minutes of incubation at room temperature, the mixture was centrifuged at 20,000 xg, and the pellets were resuspended in their original buffer.

2.2 Low-resolution Methods

2.2.1 Chemical denaturation

Chemical denaturation is a well-established method for characterizing the unfolding mechanism and chemical stability against denaturants of pure proteins or protein aggregates. For purified proteins, it is common to use a fluorometer to monitor the environmental changes of the fluorescent side chains. Tryptophan and tyrosine residues have distinct fluorescence wavelengths with specific excitation wavelengths. A two-state, one-step unfolding mechanism was fitted to a model in Eq 1⁵⁷. In this equation, C_{mid} represents the midpoint denaturant concentration for unfolding, and the m value corresponds to the cooperativity of the protein unfolding against the denaturant.

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

Eq.1

However, the unfolding mechanism of protein aggregates is believed to be a much more complicated process. The uncertainty of the position of fluorescent residues in aggregate structure and the contaminants in IBs greatly complicates the interpretation. Thus, turbidity measurement is commonly used for aggregate denaturation. The decrease in turbidity leads to a lower optical

density at 350 nm, which can be measured by a microplate reader. Urea and GdnHCl were used to perform the chemical denaturation. Adnectin IBs were diluted 10-fold in 20 mM Tris buffer (pH 8.1) with either urea or GdnHCl concentrations 0-7 M. The solubilization kinetics was measured for 3 hours. However, it is observed that the denaturing process could take longer than 3 hours to reach the equilibrium for some denaturant concentrations. Therefore, an overnight measurement was taken on the next day for 1 hour, and the last 20 minutes absorbance values were averaged to plot the denaturation profile.

2.2.2 Attenuated Total Reflectance (ATR)-FTIR

FTIR has been proven valid for measuring the secondary structure of protein aggregates. Previous studies have used this method for various protein aggregates, including heterogeneous samples like IBs⁵⁸. The analysis can be done with both solid powder and in solution. Due to the uncertainty of structural changes of IBs during the lyophilization process, using ATR crystal and optimized workflow, the measurement can be adapted to the liquid phase. The experiment was performed with a Tensor 37 FTIR spectrometer (*Bruker Optics, Germany*) installed with Bio-ATR II cell with a ZnSe crystal under constant N₂ gas flow at 25°C. The detection ranges from 4000-1200 cm⁻¹. For each sample, a set of 256 air scans was performed, followed by 4 sets for each pre-sample buffer, sample, and post-sample buffer scan. Before loading the post-sample buffer, the ATR crystal was extensively flushed to remove the residue sample in the cell. However, some residue samples could still attach to the crystal, influencing both sample and post-sample buffer measurements. Buffer extraction was done using the fourth scan of the sample subtracted by the fourth scan of the post-sample buffer to minimize the influence of the residue samples on the crystal and ensure the sample temperature. The workflow of data analysis is presented in Figure 2.1. The initial data processing for steps B and C was done on OPUS 6.5 software. Further analysis, including baseline correction between 1700 cm⁻¹-1600 cm⁻¹, normalization by area, and the taking of second derivatives, was performed on OriginPro2018. The wavenumber ranges were assigned to different secondary components in the literature. Intermolecular β -sheet (#): 1620-1625 cm⁻¹, β -sheet: 1628-1630 cm⁻¹, unstructured: 1640-1660 cm⁻¹, turns 1660-1680 cm⁻¹, and antiparallel β -sheet: 1680-1690 cm⁻¹⁵⁹. However, the secondary structure peak assignment could shift for different variants, and it is believed that the intensity of the peak is not the only factor that provides structural information. It is still difficult to have a reliable model for accurate quantification, and the interpretation must be

done cautiously. For the four Adnectin variants IBs, the measurement was performed both in TEN buffer (Section 1.5) and in MQ water (Section 3.2). The effect of ionic strength on IBs structure will be discussed.

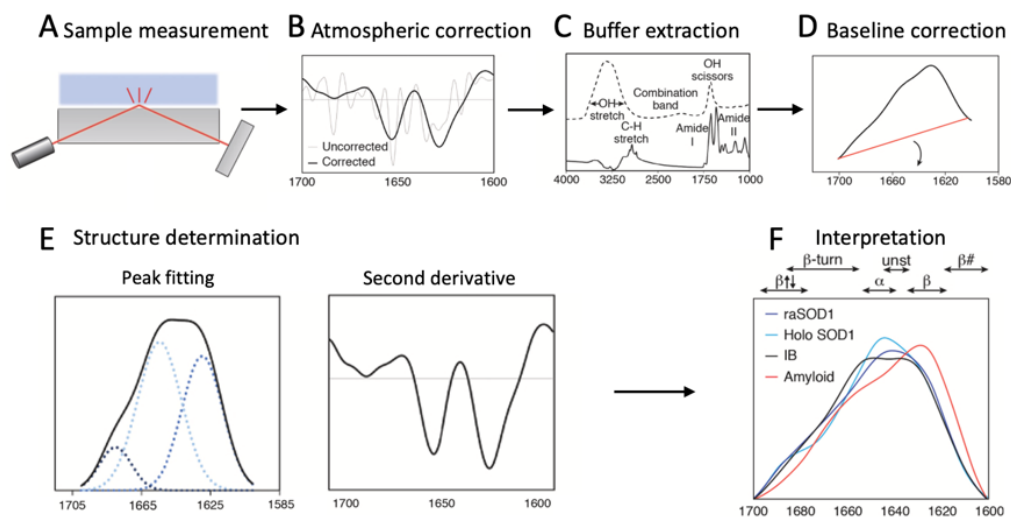


Figure 2.1: ATR-FTIR data analysis workflow. After the sample measurement, the absorbance spectrum can be obtained. The data processing includes atmospheric correction, buffer extraction, baseline correction, structure determination and interpretation⁵⁸.

2.3 Method development for IB analysis by proteolysis and mass spectrometry

As discussed in Section 1.3, proteolysis has long been used as a tool for structural characterization of protein aggregates. However, this method has not been combined with high-resolution MS to analyze IBs structure. The challenges include the impurity, insolubility, and heterogeneity of IBs. The development of a method for obtaining high-resolution detection of the protease digestion sites of IBs is presented in this Section. The digestion mainly focuses on IBs formed by variant Adnectins, but for accurate interpretation, digestion experiments were also performed on samples of pure Adnectin and denatured Adnectin IBs.

2.3.1 Determination of conditions for proteolytic digestion of IBs

Since the susceptibility of different aggregates to proteolytic digestion can vary widely, it is critical to perform test experiments to determine suitable digestion conditions for Adnectin variant IBs. The concentrations of protease and IBs and the digestion time should be optimized for MS analysis. The conditions of 0.1 or 0.5 mg/mL PK were tested with IBs diluted 1:10 in MQ water for 5 minutes. Digestion was then inhibited by adding 1 mM or 5 mM phenylmethanesulphonyl fluoride (PMSF) followed by incubating for 15 minutes. The samples were then centrifuged at 21,100g for 10 minutes, the supernatant was removed, and the pellet was resuspended in the same amount of MQ water as the removed supernatant. It was found that 0.1 mg/mL of PK did not produce significant fragments detected by Tris-Tricine SDS-PAGE gels, and 0.5 mg/mL of PK produced a clear digestion pattern (Figure 2.2A). The fragments are from the pellet fraction. Another important factor for this experiment is the digestion time. In (Figure 2.2B), different digestion times were tried, and the fragment intensity was the highest for 5 minutes of digestion. It could be that the intensity decreases at longer times due to the fragments being further digested by proteases. Thus, the digestion condition for subsequent experiments was determined, which is 5 minutes of digestion with 0.5 mg/mL PK or trypsin.

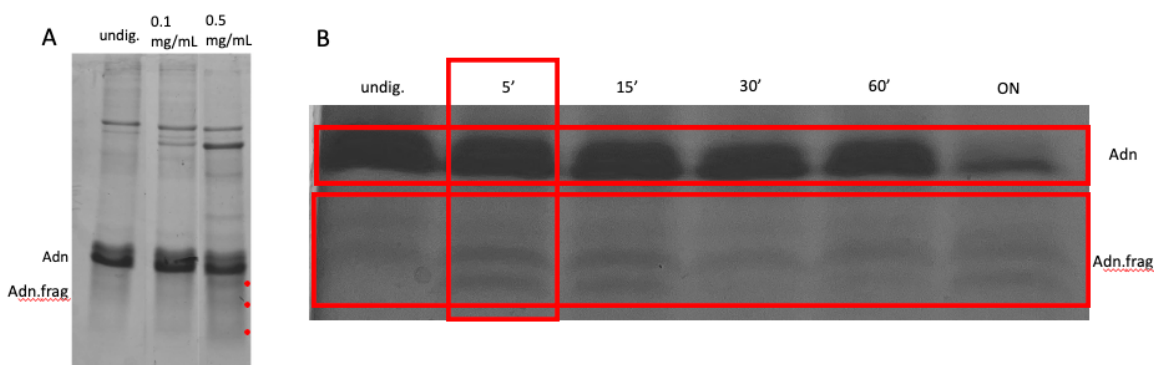


Figure 2.2: Tricine-SDS gel analysis for PK digestion condition determination. A) Digestion concentration determination. From left to right: undigested Parent Adnectin IB, 0.1 mg/mL PK digested pellets, and 0.5 mg/mL PK digested pellets. Red dots showing fragments. B) Digestion time determination. From left to right: increasing digestion time. Red boxes highlighting full-length Adnectin and Adnectin fragments.

2.3.2 Digestion of Adnectin variants IBs and MS analysis

2.3.2.1 Parallel proteolysis of unlabelled and ^{15}N -labelled IBs

Identifying the digested fragment sizes requires a higher resolution method than SDS-PAGE. MS is known to provide an accurate mass. However, it is challenging to identify the proteolytic fragments originating from the overexpressed protein in the IB since contaminant proteins may also undergo proteolysis and generate fragments, which could be a factor in restricting the MS analysis of IBs. MS, as a high-resolution method, also possesses relatively high sensitivity. The spectrum produced from the IB fraction is complicated, and it is challenging to determine whether the fragments produced from IBs are contaminant proteins. Furthermore, a high concentration of proteases is added to the sample, and autolysis of the proteases could also affect spectra analysis.

Here, I describe a new method developed for identifying fragments of recombinant proteins in IBs. The method is based on a parallel digestion experiment for unlabeled and ^{15}N -labelled IBs. The digestion was performed simultaneously for 10-fold diluted parallel-grown unlabeled and ^{15}N -labelled IBs with 0.5 mg/mL PK and inhibited with 5 mM PMSF for 15 minutes of incubation. In principle, the parallel experiment can provide an identical proteolysis process for all the proteins in the IBs sample, including protease autolysis. The MS spectra of the unlabeled and ^{15}N -labelled IBs are expected to have a “mass shift” (Figure 2.3). This mass shift can not only help with identifying fragments arising from the proteins in the IB but can also help determine the amino acid sequence of the fragment by measuring the mass differences of unlabeled and ^{15}N -labelled peptides. Different fragment species may have a different residue composition with varying amounts of nitrogen in the amino acid side chains. Thus, the fragment identification relies on both mass and the number of nitrogen atoms. As the primary sequence is known for the overexpressed recombinant protein, masses can be calculated for all possible peptide fragments with the natural isotope or with ^{15}N . For example, a peptide fragment YRTEIDKPSQH HHHHHH will have 2059.1887 with natural isotope or 2091.9460 with ^{15}N . The unlabelled and ^{15}N mass could not match the possible fragment mass from Adnectin, which belongs to cellular contaminants.

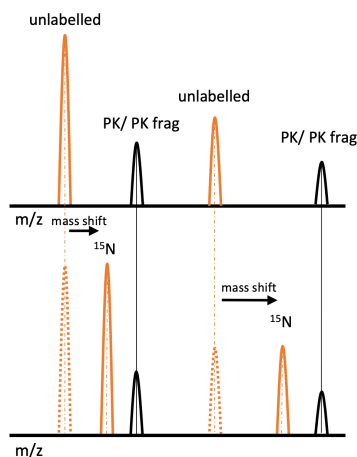


Figure 2.3: “Mass shift” in MS spectrum. The upper figure represents the MS spectrum of PK digestion of unlabelled samples. The lower figure represents a mass shift for the ^{15}N -labelled peptide fragments. The black peaks with no mass difference between the two spectra are expected to be the PK or PK self-digested fragments.

2.3.2.2 MS analysis for identifying the fragment species

Two MS instruments were used in this project for PK digested fragment identification, MALDI-TOF and ESI-Quadrupole(Q)-MS/(MS). By combining results from these two instruments, we were able to identify many fragments based on their masses. The detailed procedure and instrument setup are described in this Section.

For preparing samples for MALDI-TOF analysis, the undigested IBs samples were diluted 25-fold in MQ water. The 10-fold diluted digested IBs were further diluted 2.5-fold with MQ water. The samples are mixed with the saturated HCCA matrix in 1:1 acetonitrile/ H_2O +0.1% trifluoroacetic acid at a 1:1 ratio. After fully mixing, 1 μL of mixture was spotted on the stainless-steel plate for MALDI analysis. Linear positive ion mode with a detection range of 2-20 kDa was used for the fragment analysis. The calibration was done using $[(\text{CsI}_3)_n\text{Cs}]^+$ cluster ion series⁶⁰. These ions were generated by freshly prepared CsI_3 and 2-[(2E)-3-(4-tert-butylphenyl)-2-methylprop-2-enylidene] malononitrile (DCTB) in tetrahydrofuran (THF) with MALDI. A range of 2-13 kDa can be calibrated with m/z accuracy < 10ppm. For each IBs sample, a sum of 5000 laser shots was obtained. The MS spectrum was automatically generated by *flexAnalysis 3.4* (Bruker). A set of

examples of undigested unlabelled and ^{15}N -labelled samples is provided (Figure 2.4). The experimental fragment masses were also obtained from *flexAnalysis 3.4* (Bruker). Full-length Adnectin variant sequence and experimental fragment masses were imported in *protein calculator v3.4*. Using *scan MS results*, a list of potential fragment sequences was exported. After manual comparison with the calculated theoretical ^{15}N -labelled masses and comparing them with ^{15}N -labelled experimental data, a majority of the m/z peaks could be identified. However, for some peaks, it was hard to assign to fragments because the theoretical mass differences of candidate fragment ions were too close to be identified. To confirm the MALDI-based identification of fragments, ESI-MS was used to obtain a higher-resolution mass spectrum with isotopic resolution to identify the fragments that have similar masses.

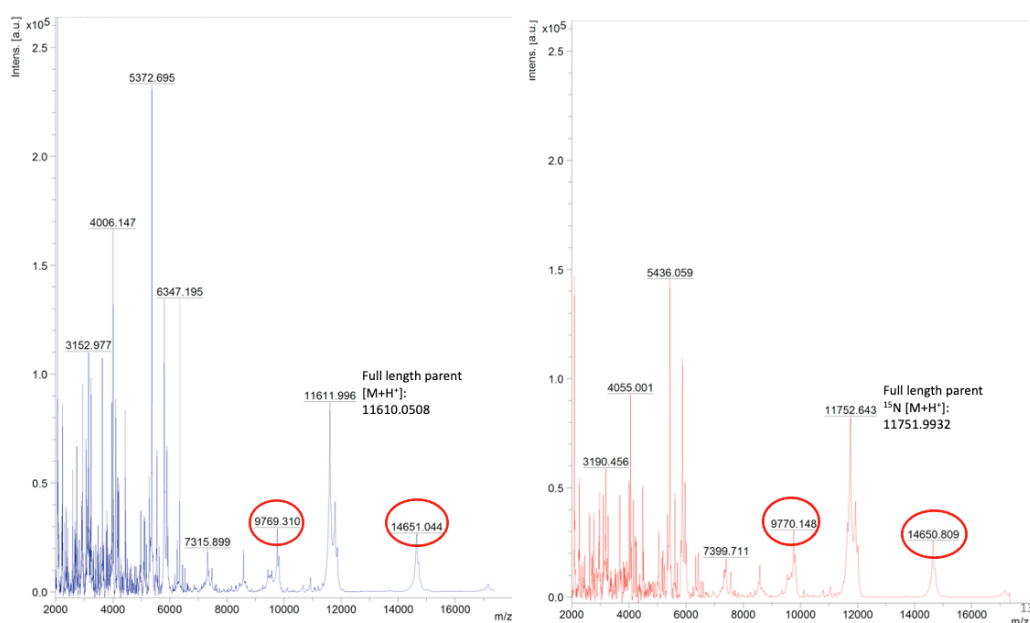


Figure 2.4: Parallel experiment PK digestion on unlabelled and ^{15}N -labelled Parent IB. The red circle highlights the peaks that did not shift, indicating PK or PK fragment ions.

The Thermos Q-Exactive mass spectrometer was used for fragment verification with high resolution. Compared to MALDI-TOF, ESI-Q-Exactive has a much lower tolerance for insoluble samples, salts, and detergents and requires higher sample purity. Therefore, a solubilization and purification process was conducted before ESI analysis. The sample preparation workflow is presented in Figure 2.5. The isolated insoluble fraction of the digested sample was first dissolved by 70% formic acid (FA) and then diluted 2-fold with MQ water to meet the requirement for the

centriprep® (EMD Millipore) centrifugal membrane filter compatibility. The initial MW cutoff was 10 kDa, and the flowthrough (FT) from each centrifugation (3000 xg, 30 minutes) was collected and then transferred to a 3 kDa MW cutoff centriprep® device. The FT from the second device of every centrifugation is then transferred to an Amicon® Ultra Centrifugal Filter, 3 kDa MW cutoff. After total purification and concentration of 3 hours, with rebalancing by transferring FT and adding MQ water for each 30 minutes, the concentrates from the centriprep® 3 kDa MW cutoff device and Amicon® 3 kDa MW cutoff device are collected for ESI-MS/(MS) analysis. The concentrates in MQ water were mixed with MeOH and FA to a final ratio of H₂O: MeOH: FA of 1:1:0.1 with the final peptide concentration lower than 10 µM.

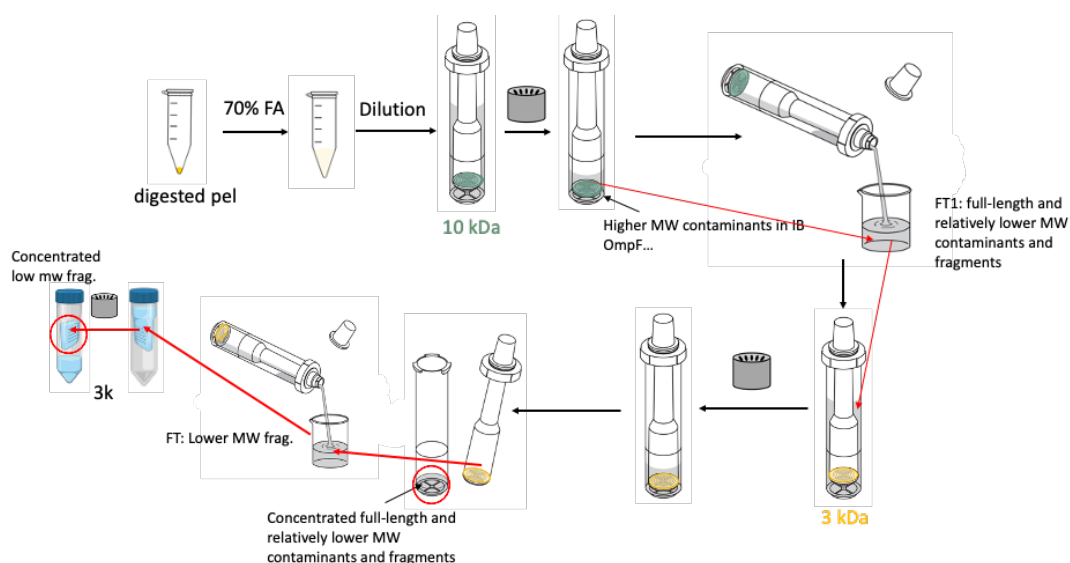


Figure 2.5: The sample preparation workflow for ESI-MS/(MS) analysis. The digested pellet of the IBs was solubilized by 70% formic acid and further purified and concentrated. Red circles highlight samples for ESI-MS analysis.

An initial trial on solubilized undigested parent IBs was performed, and the concentrates from the first 3 kDa cutoff devices were used for ESI/MS analysis. Surprisingly, a significant amount of Parent Adnectin protein went through the 10 kDa cutoff device and observed by ESI/MS detection (Figure 2.6). The comparison of ESI-MS spectra of pure Parent Adnectin and solubilized Adnectin IBs shows the solubilized IBs spectrum has high similarities to pure proteins, after strong acid solubilization and purification, Parent Adnectin from IBs also presents a charge distribution from 9 to 15 charge states. However, both the raw m/z and deconvoluted mass spectrum showed that

solubilized IBs had more peaks, possibly due to contaminants in the IBs sample. The comparison also proved it is a valid workflow to obtain solubilized protein from IBs. With this workflow, it is possible to perform higher-resolution detection with ESI/MS on the samples from IBs and potentially identify fragment species.

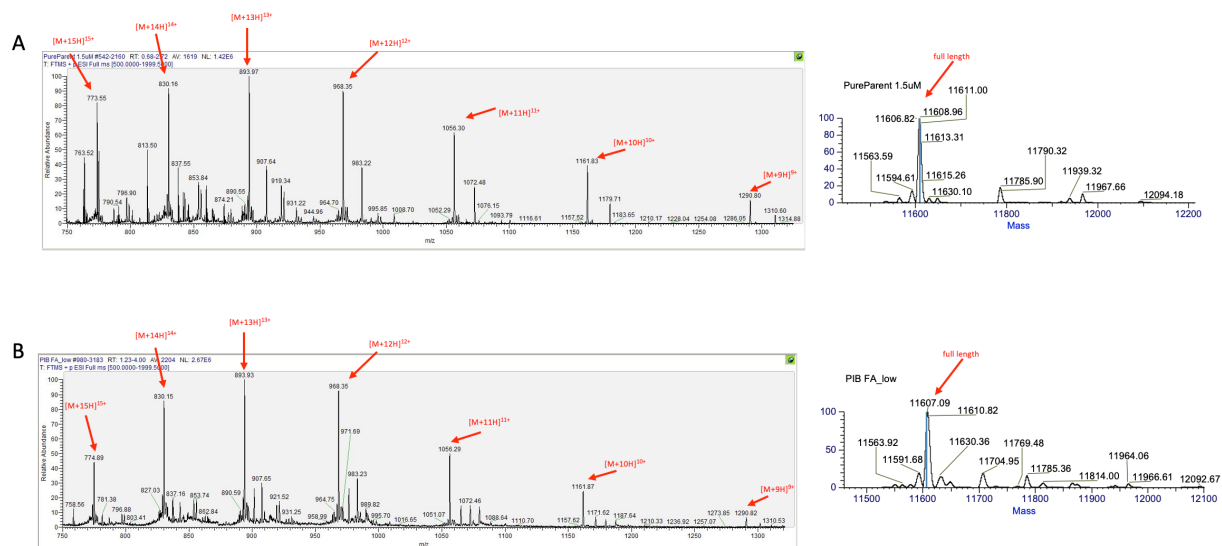


Figure 2.6: ESI/MS of pure Parent Adnectin and solubilized Parent IBs. A) Pure Parent Adnectin m/z and deconvoluted mass spectra. B) Solubilized Parent IB sample m/z and deconvoluted mass spectra, sample preparation followed by the ESI/MS sample preparation workflow. Both m/z spectra show a charge distribution from 9 to 15 charge states. 17,500 resolution @m/z200.

With the highest resolution of 140,000 @ m/z 200 and scanning from 135 to 2000 m/z, a significant number of peptide ions were detected from the solubilized digested IBs, the theoretical monoisotopic m/z values for peptides were calculated, and the theoretical isotopic distribution was generated by *Xcalibur* (Thermo). By comparing the theoretical monoisotopic m/z and the theoretical isotopic distribution to the experimentally obtained spectra, the exact identification of fragment species was performed (Figure 2.7). ESI-MS/MS was conducted for fragment identification when the theoretical m/z ratio differences were lower than 0.03. (Figure A2.1)

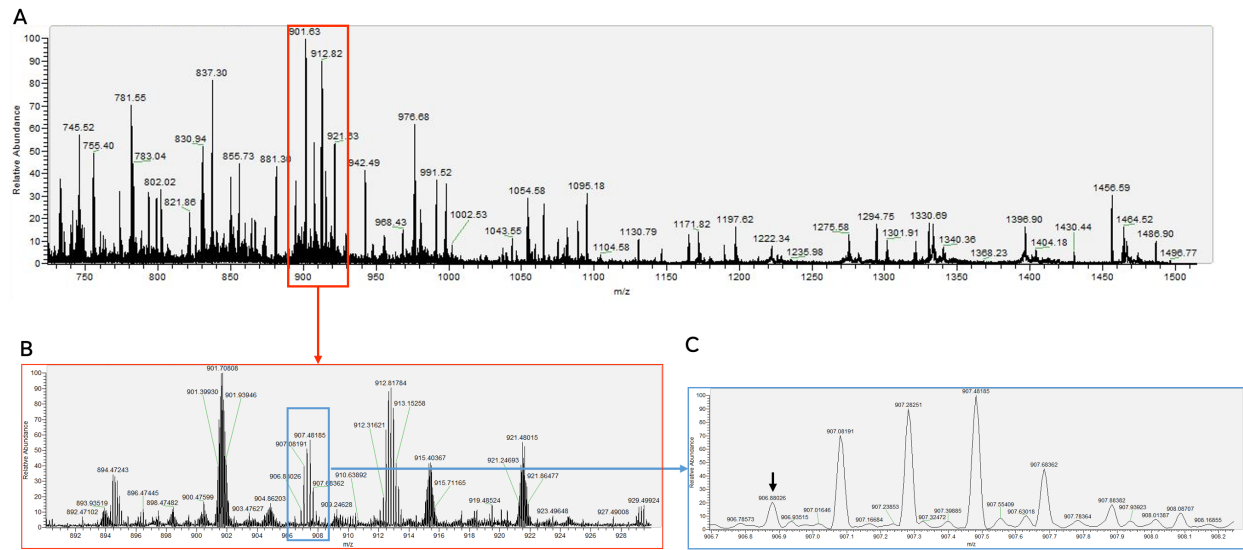


Figure 2.7: Example of fragment identification by ESI/MS. A) High-resolution mode (14,000 @m/z200), PK digested m/z spectrum. B) Zooming in, the isotopic distribution of full-length Adnectin and proteolytic fragments can be observed. C) Isotopic distribution of an example fragment ion. An experimental monoisotopic mass can be obtained from the lowest peak, as the black arrow highlights. The fragment identified here is EVRAATPTSLRISWSARLKVARYRITYGETGGNSPVQEF with 5 charges, with a theoretical monoisotopic m/z of 906.87782.

2.3.3 Control experiments on pure protein and denatured IBs

The digestion of pure and denatured protein was performed to interpret the digested fragments more thoroughly and obtain structural information. Pure Parent Adnectin was diluted in MQ water to a similar concentration as Adnectin in IB samples, and the same digestion experiments were conducted. The digested sample was diluted 2.5-fold without centrifugation and then mixed with a freshly prepared HCCA matrix at a 1:1 ratio.

Another control experiment was the PK digestion of denatured IBs. The unlabelled and ^{15}N -labelled Parent IB were dissolved in 2% w/v SDS and heated at 90 °C for 15 minutes. The PK digestion started immediately after the solution was lowered to room temperature. Since SDS strongly interferes with the MS signal, acetone precipitation isolated the digested fragments. After 5 minutes of digestion, PMSF was added, and samples were incubated for 15 minutes. 800 μL of cold (-20°C) acetone was added to 200 μL of digested samples. The samples were incubated at -

20°C for two days, the precipitated fragments were isolated by centrifugation at 1000 xg for 5 minutes, and the pellets were dried for 20 minutes and then resuspended in 200 µL of MQ water. Then, two more precipitation processes were performed. The final precipitated fractions were insoluble in water. Thus, only MALDI-TOF was used to identify the digested fragments.

2.3.4 Replicate samples

The protein aggregation process could vary between replicates. Therefore, different final IB structures could present for repeated biological growth. All the PMSF uninhibited, lysed, undigested IB samples as well as PK digested samples for each variant IB were performed for four biological replicates. The MALDI analysis has overall high reproducibility between the biological replicates (Figure A2.2). The fragments that were detected in at least 3 of the 4 biological replicates were recorded and used for structural interpretation (Section 3.3.4). One additional ESI/MS analysis was conducted to verify the fragments. A set of IBs samples prepared with PMSF and subsequently digested by PK for verifying the cellular protease influence for PK digestion analysis. Two biological replicates were performed for nOG-washed IBs.

2.3.5 Discussion and conclusion of the method development

The method combining proteolysis and two types of MS analysis allowed the identification of many proteolytic fragments. However, fragment identification was mainly based on MALDI analysis since the preparation procedure for ESI/MS could result in the loss of fragments, and the strong acid solubilization could induce further fragmentation. Here, the newly developed method was introduced for the first time, allowing the identification of proteolysis fragments from heterogeneous aggregates in the presence of cellular contaminants. Detailed results and interpretation are presented in Chapter 3. The results show how this method ultimately provided information on protease accessibility for various IBs, and combined with other structural characterization methods, can be used as a tool for analyzing protein conformations in IBs.

Chapter 3: Results and Discussion

Urea denaturation data from section 3.1 was acquired in collaboration with Grace DelMistro as part of an undergraduate research project. The denaturation experiment was optimized with Bruna Siebeneichler and Grace DelMistro from previous protocol written by Anna Schaefer and Dalia Naser. Unpublished qHDX results from Anna Schaefer and differential scanning calorimetry (DSC) data from Julia Steckner are included in section 3.3.4.2 for discussion.

3.1 Chemical denaturation

The denaturation profiles of the 4 Adnectin variants were determined by measuring OD₃₅₀. Both urea and GdnHCl were used to analyze the relative stability of Adnectin variant IBs. As described in Chapter 2, the denaturation kinetics were also measured. The differences between 3 hours of urea denaturation and ON denaturation profiles are presented in Figure 3.1. The data for each variant IBs are from two growths, each with four technical replicates. The urea denaturation profiles were fitted on OriginPro2018 with a denaturant-dependent equation, as mentioned in Chapter 2. The midpoint of urea solubilization was decreased for the ON measurement, indicating that, at least for some denaturant concentrations, 3 hours was insufficient to reach the equilibrium. It is also clear that there is a significant difference between the denaturants. While the urea denaturation profiles are able to fit by a two-state model, the GdnHCl denaturation profile cannot fit into the same two-state unfolding model.

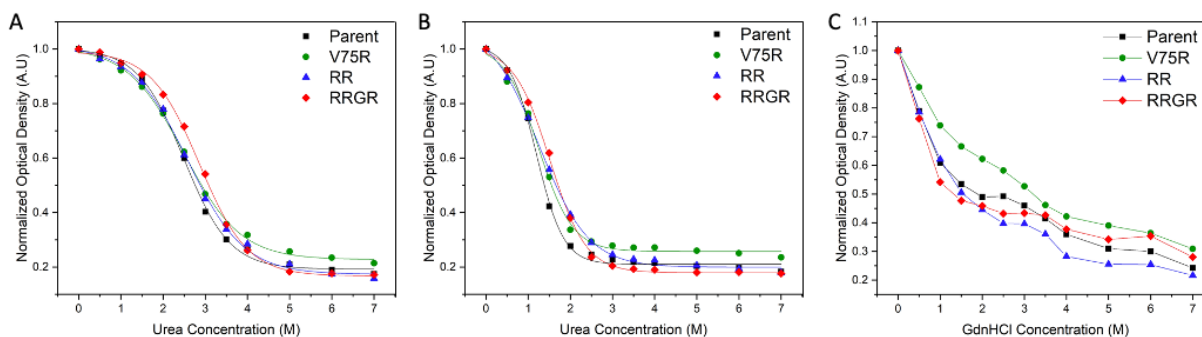


Figure 3.1: Chemical denaturation profiles of 4 Adnectin variants IBs. OD₃₅₀ nm is normalized to the initial signal intensity at 0M denaturants. A) 3 hours urea denaturation profiles. B) ON urea denaturation profiles. C) ON GdnHCl denaturation profiles.

GdnHCl and urea are both widely used denaturants, but they operate through different mechanisms to disrupt protein structures. Urea primarily denatures proteins by interfering with hydrogen bonds within the protein aggregates and the water molecule arrangement around the protein. This disruption weakens the hydrogen bonds that stabilize the protein structures, leading to unfolding. Urea also affects hydrophobic interactions. It can increase the solubility of hydrophobic residues in water and thus promote protein unfolding. The process can also result in a progressive loss of protein structure. At the aggregate level, this could cause global solubilization or a release of protein monomers and lower turbidity by reducing the aggregate amount or aggregate sizes. On the other hand, GdnHCl disrupts hydrogen bonds and interferes with ionic interactions within the protein. It can directly interact with the side chains, stabilizing the unfolded state through charge interactions.

The guanidinium ion is particularly effective at disrupting protein structures due to its ability to form multiple hydrogen bonds, and it possesses a large planar structure that can interact with the protein backbone and side chains. This usually leads to a more complete and rapid unfolding compared to urea. GdnHCl also increases the solubility of hydrophobic residues, similar to urea, but to a greater extent due to its strong ionic nature. This can explain that for GdnHCl, the intensity drops significantly at even low concentrations. For GdnHCl denaturation, the 4 Adnectin IBs went through a fast denaturation at low concentrations, and then the slope of decline decreased significantly. Even between 6 and 7 M, the intensity was still decreasing. This could be a stabilization effect of the ionic strength provided by GdnHCl.

It is important to remember the potential mechanism differences between the two denaturants. The differences in denaturation mechanism can also have variable effects on different mutants, especially charge mutations. From the urea denaturation, the Parent, V75R, and RR IBs had relatively similar C_{mid} , and the RRGR was more stable. This meets the expectation that by disrupting the core formation of the Adnectin monomer, RRGR could adopt an amyloid-like structure that increases its stability. However, in GdnHCl denaturation, the turbidity of RRGR IBs decreased the most with concentrations 0.5-1.5 M. This could be due to the three additional Arg residues. The ionic strength on the mutated sidechains could be the major denaturation factor for RRGR IBs at low concentrations. For 2M and higher concentrations, RRGR reached a steady state. This could be a result of the stable amyloid-like component resisting the denaturation. V75R IB,

which was much more unstable in urea, exhibited a high stability in GdnHCl. The additional charge would involve a stable and buried structure that prevented the charge effect introduced by Gdn⁺. The Parent and RR IB had little difference in urea denaturation. The differences lie in the higher concentrations of GdnHCl; the increasing ionic strength could influence the two additional charges in RR more.

In summary, the relative stability of IBs is highly dependent on denaturant properties. In general, RRGR is relatively more stable in urea and high concentrations of GdnHCl. This once more suggests there could be a significant structure difference in RRGR IBs and complements previous FTIR data. The other 3 variants are not significantly different, except V75R had a strong resistance against GdnHCl.

3.2 ATR-FTIR revealing secondary structure composition

As shown in Section 1.5, FTIR spectra can reveal secondary structure components in IBs. The Adnectin IBs have high similarities to the purified Parent Adnectin, with most differences in the unstructured region and a peak shift representing intermolecular β (#)-sheet features. The experiment was performed with IBs in TEN buffer. However, for MS analysis, a low salt condition is required to avoid signal suppression. The IBs digestion was performed in MQ water, as mentioned in Section 2.3.

The ionic strength difference could affect the IB structure. FTIR analysis with buffer subtraction can provide information on the potential influence of buffers on IB structure. Figure 3.2 shows the FTIR spectra of the 4 Adnectin variant IBs in MQ water. From the spectra, we can see that the general shape of each variant spectrum shows no significant difference compared to previous spectra of IBs in the TEN buffer (Figure 1.9). The spectra still show antiparallel β -sheet, unstructured β -sheet, and intermolecular β (#)-sheet features. The RRGR IB still exhibits distinct amyloid-like components, with a peak at 1622 cm⁻¹, which is significantly lower than the other 3 variants. Parent, V75R and RR have the peak at a higher wave number. For V75R, in particular, there is a significant decrease in unstructured components. The difference in the strength of ionic interactions could have effects on the V75R IB structure, which complements the denaturation

profiles. In summary, the overall structure of the 4 Adnectin variants in MQ water still exhibits native-like, unstructured, and amyloid-like structures. The differences between the variants of IBs are also preserved in the MQ water.

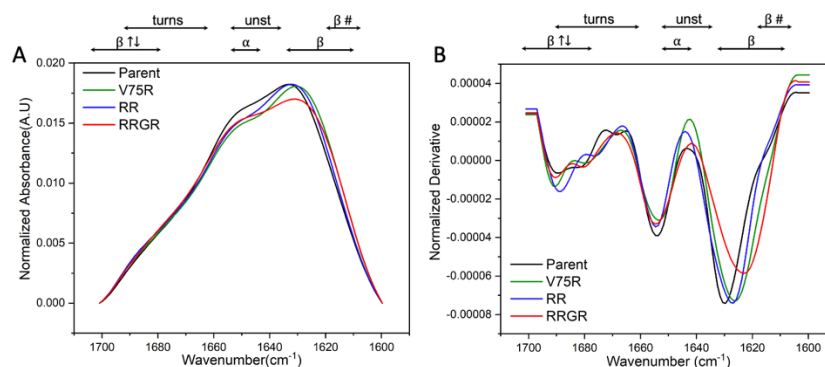


Figure 3.2: FTIR spectra revealing secondary structure components of Adnectin IBs. Each spectrum is the average of 2 technical replicates from 2 growths. Secondary assignments are presented above the spectra. A) Normalized absorbance spectra in the amide I region. The absorbance spectra, focused on the amide I region (1600-1700 cm⁻¹), are normalized to an area of 1, allowing for comparison of secondary structure content among the different variants. B) Second derivative spectra. The second derivatives of the absorbance spectra, calculated using OriginPro2018, providing more detailed insights into the secondary structural elements present in the IBs.

3.3 Proteolysis and MS

3.3.1 Post-translational modification of 4 Adnectin variant IBs

Post-translational modifications (PTMs) refer to a chemical group covalently linked to a protein after it is synthesized on the ribosome. For Adnectin, N-terminal gluconoylation is a common PTM that coexists with phosphogluconoylation after methionine cleavage^{34,61}. It has been reported that gluconoylation can readily occur in BL21(DE3)⁶². The process is reported as the methionine cleavage on the N-terminus by methionine aminopeptidase (MAP), followed by adding the gluconoyl group or phosphogluconoyl group⁶³. The gluconoyl group can also be phosphorylated spontaneously, and the phosphate group can also be reversibly cleaved off by the host cell phosphatase⁶³. These PTMs could affect protein structure and function, and it has been previously discussed that PTM can influence IB formation or solubility⁶⁴. However, it has been found that the

relative intensity for PTM protein and unmodified protein bands in SDS-PAGE is the same for the soluble and IB fractions of Adnectins, which suggests the PTM has little effect on IB formation propensity^{34,61}. It is still uncertain if these PTMs impact the IB structure. The composition of the two distinguishable bands on SDS gels was also discussed. There are several possible molecular weights (MW) for the Adnectin variant after the PTM, which are the full-length Adnectin, the full-length Adnectin with Met uncleaved (+131 Da), the full-length Adnectin with gluconoyl group (+178 Da), and the full-length Adnectin with phosphogluconoyl group (+258 Da)⁶⁵. The slight difference in the MW is unlikely to result in distinguishable bands on gels. Therefore, the PTM likely interferes with SDS binding and causes the separation of two bands. Due to the poor resolution of SDS page analysis, it is difficult to investigate further.

MS is a powerful method for analyzing PTMs. The four possible final MWs after PTM have all been confirmed in Adnectin variants IBs in this project. However, the sample preparation for the MS analysis potentially influences the PTM analysis. Phosphogluconoylation, gluconoylation, and phosphorylation can occur without cellular enzymes⁶³. It is clear that among the four Adnectin variants IBs, Parent, V75R, and RR IBs have three distinct peaks on MS spectra: full-length, gluconoylated, and phosphogluconoylated. The full-length Adnectin always appears with the highest intensity, followed by the gluconoylation (+178), and the phosphogluconoylation (+258) appears to be the lowest intensity (Figure A3.1). Due to the lack of knowledge of the ionization process, it is not possible to determine their relative abundance by their intensity. However, comparing the changes in the relative intensity of the PTMs between the different variants IBs can still be valid since the same PTM on different variants overall possesses similar chemical properties for ionization, and the same MS analysis process was performed. The data show that V75R IBs have a relatively higher gluconoylation fraction than Parent IBs and RR IBs. The pattern is markedly different for RRGR IBs, which is the only case where the uncleaved methionine (+131 Da) species is detected, constituting approximately half of the intensity compared to the full-length RRGR (Figure 3.3). A low-intensity phosphogluconoylation (+258) peak is also present in RRGR IBs.

Due to the abovementioned uncertainties and restrictions, it is difficult to deduce the causes of the PTMs and their potential to interfere IBs structures. However, the results observed can correspond

well to the SDS-PAGE analysis, which support that the two bands are caused by SDS banding and not the slight MW differences. RRGR without a distinct gluconoylation and phosphogluconoylation in MALDI also presents just one clear band on the gel (Figure A3.2). This observation can be relevant to the aggregation kinetics. As mentioned earlier, the cleavage of a Met residue significantly differs from the other PTMs. It is the only PTM that is required to take place in a cellular environment with the presence of MAP and potentially before the aggregation occurs, which results in the N-terminus being buried in the IBs⁶⁶. It is also reported that the cleavage of the Met group with MAP is an efficient process⁶³. It suggests that the other mutants underwent a cleavage of N-terminal Met before gluconoylation or phosphogluconoylation, then aggregated. This suggests the aggregation process took place much faster for RRGR than for the other variants. This finding is consistent with the previous result that during the growth of RRGR IBs, there is very little soluble RRGR for different growth times. This interpretation is reasonable because the RRGR is the only variant designed to strongly disrupt the native monomer core formation. These results provide additional evidence for the hypothesis that the monomer stability and its ability to achieve native folding can strongly alter the aggregation kinetics. The proteins that are not able to achieve a native state could prefer aggregation with significantly higher kinetics.

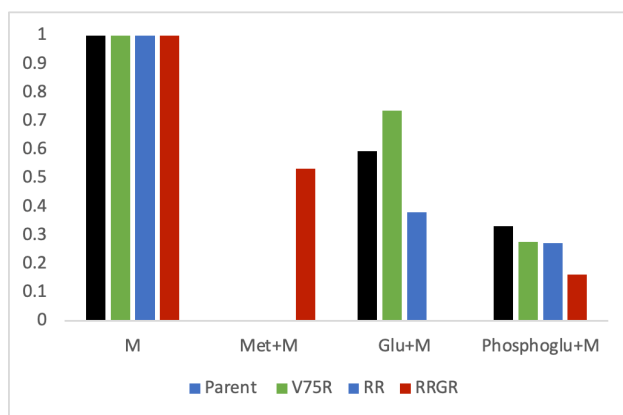


Figure 3.3: PTMs of 4 Adnectin Variants IBs. The bar chart shows the relative abundance of different PTMs normalized to the full-length intensity of each Adnectin variant. The following PTMs are presented. M: Represents the full-length Adnectin with the N-terminal Met cleaved off. Met+M: Indicates full-length Adnectin with the N-terminal Met intact. Glu+M: Refers to gluconoylated full-length Adnectin after the N-terminal Met has been cleaved off. Phosphoglu+M: Represents phosphogluconoylated full-length Adnectin after the N-terminal Met has been cleaved off.

3.3.2 Proteolysis during lysing process

It is well established that IBs can contain proteolytic fragments, which can influence the use of IBs for protein production⁶⁷. Therefore, efforts have been made to reduce the composition of the proportion of proteolytic fragments in IBs and increase protein yield. Knockout of various proteases has been a standard and effective method for lowering proteolysis of target proteins. The cell strain used in this project, *E. coli* BL21(DE3) pLysS has been designed to be deficient in lon protease and outer membrane protease OmpT expression. It has strongly enhanced the ability to reach a higher expression level, and adding pLysS improved the tolerance for toxic protein production. However, the proteolysis in the host cell protease can never be fully inhibited. For Adnectin variants IBs, the proteolytic digestion by the host cell protease has not been studied extensively. Despite faint bands below the full-length Adnectin sometimes being observed (Figure 2.2), the full-length Adnectin always was the predominant band compared to insoluble contaminants and other smaller MW proteins or peptides contaminants. It was also uncertain if the lower MW bands were the proteolytic fragments from the IBs. Understanding the composition of IBs, including proteolytic fragments from the host cell protease, is central to this project. Fortunately, fragments could be identified using the parallel labelling methods developed here, as described in Chapter 2.

Analysis of proteolytic cleavage sites was first performed for IBs prepared without the addition of PMSF during cell lysis (Section 2.1.2). Figure 3.4 shows the identified fragments with regular freeze-thaw lysing. Next, lysing with PMSF was conducted to further determine the source of the fragments, as described in Section 2.1.2. The results showed a significant decrease in MS signal for most fragments, suggesting that proteolysis by the host cell protease occurred during the lysing process. Moreover, digestion inhibited by the PMSF mainly follows aromatic residues Tyr and Phe. This indicates that serine proteases that possess chymotrypsin-like properties could dominate proteolytic digestion during IB sample preparation. However, some fragments are still present, even lysing with PMSF. This suggests that a trace of the proteolysis of the full-length protein occurred before harvest, or the protease could not be fully inhibited by PMSF. Previous research found that PMSF and other inhibitors did not fully eliminate the presence of fragments in IBs formed in BL21(DE3) and K12⁶⁷. The digestion sites for the IBs in the research were found mainly

after L and R, and the paper concluded the fragmentation was not caused by common proteases such as Lon, ClpP, OmpT, and FtsH. The experiments also ruled out the possibility of the fragmentation of the mRNA of the target protein. However, the different cut sites and the fact that PMSF largely prevents fragmentation of the Adnectin IBs suggest that different proteolysis may occur during IB sample preparation and vary between the recombinant proteins. Therefore, the proteases excluded in the prior research could still be valid for our Adnectin IBs. More experiments are required to investigate the detailed proteolysis processes that occur during the preparation of IBs.

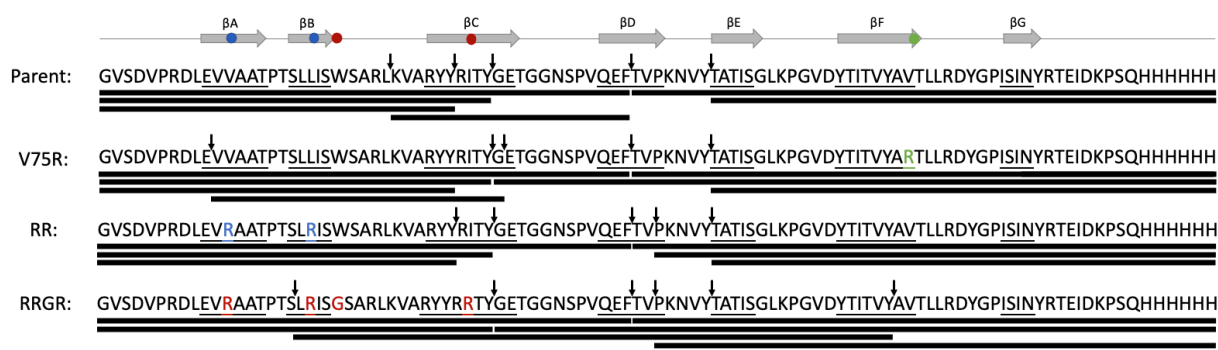


Figure 3.4: Fragments identified from four Adnectin variants IBs samples. The β -strands of the native structure are shown as ribbon arrows above the sequence and are underlined within the primary sequence. Sequences are listed for each variant, with color-coded residues showing point mutations, and the black bars represent fragments identified and mapped under the sequences. The black arrows above each sequence show the cut sites.

Despite the lack of understanding of the proteolysis by host cell protease, information could still be interpreted as IBs structural information. Since most of the fragments were suppressed by adding 1 mM PMSF during lysing, it suggests that the proteases were digesting the mature IBs after 18 hr growth during lysing instead of during the growth. Therefore, the fragments identified could be interpreted as structural information of IBs. Due to the lack of information on the host cell protease, the interpretation should be cautious. From Figure 3.4, the fragments identified are shown below each variant sequence. In general, the cut sites are highly similar between the 4 variants IBs. The major cut site is located in the βD after F, with both N-terminal and C-terminal fragments present. Another 2 common cut sites for all 4 variants are on βC 36Y/37G and before βE 55Y/56T. This could suggest unstructured regions near these sites. However, there are also differences in the cut sites. Due to the lack of understanding of the host cell protease, it is hard to

interpret other fragments and their correlation with IBs structure. Further discussion will be provided in later Sections.

3.3.3 Trypsin cut sites on variant IBs

The 4 Adnectin variants IBs underwent trypsin digestion and MALDI analysis. The cut sites can be identified without parallel experiments. Trypsin is highly specific for the cleave after the Arg and Lys groups, except before proline. For Parent and V75R IBs, the experimentally cut sites were found at every theoretical possible cut site. However, for RR and RRGR IBs, the cut site after the Arg mutation on the β B and on the BC loop was not observed (Figure 3.5). It could suggest that the BC loop of RR and RRGR were more resistant to trypsin digestion. The additional R on the β A for RR and RRGR has become a cut site, as well as the mutation of V75R. These findings increased the possibility that the β B could be extensively protected in the IBs structure. Further interpretation of the relative preference on the cut sites and the relative protease resistance requires quantification of the fragment amounts, but MS analysis has limitations on doing exact quantification. Further interpretation will be made by combining the PK digestion results in the following Sections.

Parent	GVSDVPRDLEVVAAATPT <u>SL</u> LISWSAR <u>LK</u> VARYRITYGETGGNSPVQEFTVPKKNVYTATISGLKPGVDY <u>TI</u> TVYAVTLLRDYGPISIN <u>Y</u> RT
V75R	GVSDVPRDLEVVAAATPT <u>SL</u> LISWSAR <u>LK</u> VARYRITYGETGGNSPVQEFTVPKKNVYTATISGLKPGVDY <u>TI</u> TVYAR <u>T</u> LLRDYGPISIN <u>Y</u> RT
RR	GVSDVPRDLEVRAATPT <u>SL</u> RISWSAR <u>LK</u> VARYRITYGETGGNSPVQEFTVPKKNVYTATISGLKPGVDY <u>TI</u> TVYAVTLLRDYGPISIN <u>Y</u> RT
RRGR	GVSDVPRDLEVRAATPT <u>SL</u> RISGSAR <u>LK</u> VARYRRTYGETGGNSPVQEFTVPKKNVYTATISGLKPGVDY <u>TI</u> TVYAVTLLRDYGPISIN <u>Y</u> RT

Figure 3.5: Trypsin cut sites observed in Adnectin variant IBs from MS analysis. Red-colored R and K show cut sites were observed after digestion, and blue-colored R and K show theoretical cut sites for trypsin but were not observed after tryptic digestion. For Parent and V75R, all possible cut sites were detected experimentally. RR and RRGR have experimentally uncleaved sites that could suggest a protected region.

3.3.4 PK digestion reveals information on protected structure in IBs

In contrast to trypsin, PK is a non-specific protease, which allows obtaining more structural information without quantification of the fragments. Although it is described as having a preference for cutting after aromatic and hydrophobic residues, there is no clear evidence that it

possesses no ability to cut after other residues except before or after proline⁶⁸. In addition, PK activity is highly dependent on substrate availability⁴⁹. Based on these properties, PK has been used as a tool to probe the structure of protein aggregates⁶⁸. However, interpretation should nevertheless be done cautiously. PK has often been found to digest a few specific sites for a homogenous amyloid structure. The corresponding fragments are typically considered to be involved in forming a PK-resistant core. However, IB is a kind of aggregate with high heterogeneity. The digestion results are much more complicated. Nonetheless, the fundamental logic for interpretation still applies: the fragment represents a region with lower protease accessibility and higher protection, which could be involved in the formation of a highly compact core, and the cut sites indicate unprotected regions that are potentially less structured or disordered.

3.3.4.1 Logic of interpreting Parent Adnectin IBs digestion pattern

To further explore the IB structures, Figure 3.6 A shows the fragments identified from the Parent Adnectin IBs lysed without PMSF after 5 min PK digestion. There are significantly more fragments than without PK digestion. Cut sites are present in β A, β C, β D, β E, and the end of the β G. β B and β F are uncleaved in all fragments identified, indicating these two strands are extremely PK-resistant and highly protected in the IB structure. Some digestion sites could indicate a native-like structure, such as those before β A L8/E9, after β E at S60/G61, and after β G at N87/Y88. Surprisingly, there are no distinct cut sites in the loops. This suggests that the loops are involved in forming a stable and ordered structure in the IBs. Unlike BC, EF, and FG loops, which are uncleaved in many fragments, the CD and DE loops are only intact in some fragments. This could be interpreted in different ways as well. It could be that the regions, such as Y36-F48, which includes loop CD, are in a protected fragment that is too small (lower than 2 kDa) to be detected by the linear positive mode MS. It could also suggest that for some conformations in Parent Adnectin IBs, these regions are totally unprotected and fully degraded by PK. This could apply to the N-terminus G0-L8 and fragment T49-A57 as well. These interpretations could be that the different fragments with overlapping protected regions potentially represent heterogeneous conformations. Alternatively, it could be interpreted as a homogenous structure with varying protease accessibility in different regions. It is also possible that the smaller fragments can be formed by further digestion of the longer fragments. For example, G0-F48 could be further digested into G0-T35 or E9-F48. For this interpretation, the prominent F48/T49 cleavage site

would be the preferred cut site or more exposed to PK among these three cut sites since the two longest fragments resulting from the cutting site, G0-F48, and T49-H104, are still present in the digested sample, which means it is always the first site that protease accesses. Overall, all the interpretations support that the fragments cover protected regions and overlap more in the same region, which represents higher protection in this region.

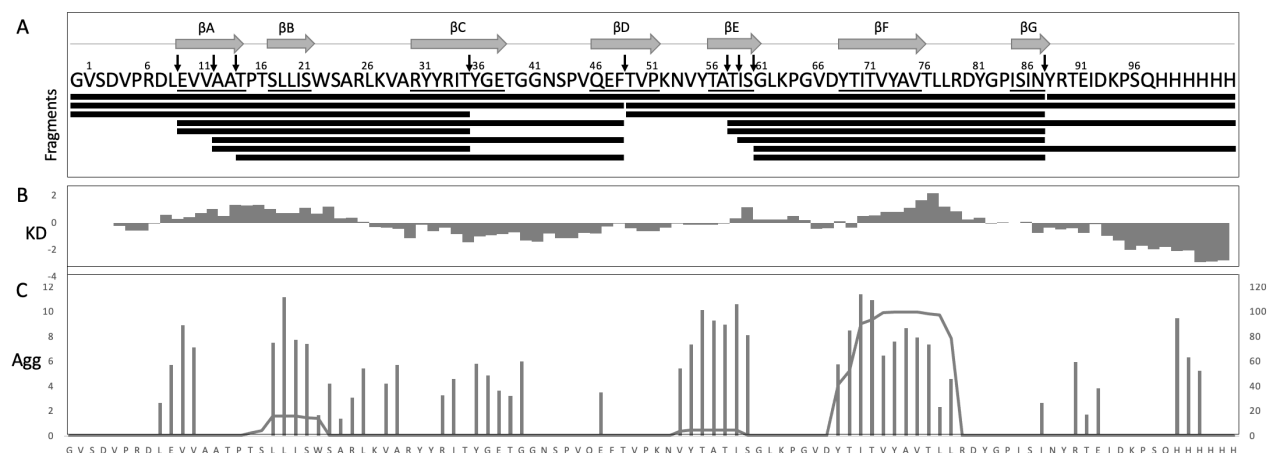


Figure 3.6: Structure analysis from PK digestion on Parent Adnectin IBs. A) PK cut sites labelled by black arrows were mapped on top of the Parent Adnectin sequence. Black bars under the primary sequence represent MS-identified fragments. The β -strands of the native structure are shown as ribbon arrows above the sequence and are underlined within the primary sequence. B) The Kyte-Doolittle hydropathy index for the Parent Adnectin sequence is displayed, illustrating the hydrophobic regions along the protein sequence⁶⁹. C) ZipperDB (gray bars, left scale)⁷⁰ and Tango (line, right scale)⁷¹ aggregation propensity prediction of the primary sequence of Parent Adnectin. Provide insights into regions of the sequence that are prone to aggregation.

We can identify that for Parent Adnectin IBs, the fragment covering the β B and BC loop is a highly protected region; another region with high protection covers the EF loop, β F, FG loop, and β G. This protection pattern partially corresponds to predictions of aggregation-prone regions. For the hydrophobicity prediction (Figure 3.6B), the β B and β F are more hydrophobic than other regions. Other aggregation predictors, such as zipper DB⁷⁰ and Tango⁷¹, also suggest that these regions have a high aggregation propensity. However, there are differences between predictions and the experimental digestion pattern in the DE loop and β E. According to Zipper DB, this region has a high propensity for aggregation; Tango also suggests that aggregation propensity exists in this region. However, as the digestion pattern shows, this region is protected in some fragments but can also be cut at three sites on the β E. The cut sites on the β E confirmed the cellular protease that

contributes to cleavage during lysis is a relatively specific protease because the β E is much more preferred for unspecific proteolysis (Fig. 3.6A), but the cleavage site was only after Y for the host cell protease (Figure 3.4). In summary, the cut sites partially correspond with the predictors. However, there are contradictions against the predictors. β E is reported to be a highly aggregation-prone region but accessible to PK.

A common challenge in almost all methods for characterizing IB structure is the similarities between native-like and amyloid-like aggregation. The driving forces of the two assemblies both involve the hydrophobic effect and hydrogen bonding. Even though FTIR shows that some native-like structures exist in IBs, the resolution is not high enough for further local structure analysis. As we see in the monomer structure, the digestion sites are concentrated mainly in the edge β -strands (A, D, E). This could be evidence of native fold. However, a direct comparison between native folds and conformation in IBs can be conducted. The digestion of the pure Parent Adnectin exhibits a very different pattern than the Parent Adnectins IBs (Figure 3.7A, Figure 2.4). Overall, pure Parent Adnectin is highly resistant to PK digestion. The fragments are barely observable on SDS-PAGE, and the fragment species are much less than for IBs, with significantly lower intensity on the MALDI spectrum. The fragment species are hard to identify because of the lack of a parallel experiment of ^{15}N labelling and a low abundance of fragments for detection by ESI (Sections 2.3.2.1, 2.3.2.2). Nevertheless, the m/z of the fragment ions detected by MALDI for pure protein significantly differ from those produced by IBs (Figure 3.7A). Thus, it is not likely that the strands that are extensively digested are forming fully native structures. Thus, the results for digestion of pure Parent support the possibility that the β A, β D, and β E are less structured in IBs than in the native protein. This interpretation also partially fits the experimentally observed three-state folding of $^{10}\text{FN3}$ discussed in Section 1.2.1. The folding intermediate shows β C and β F are significantly structured, followed by β D and β E, β B was also discussed to possess structure, and β A, β G are completely unstructured.

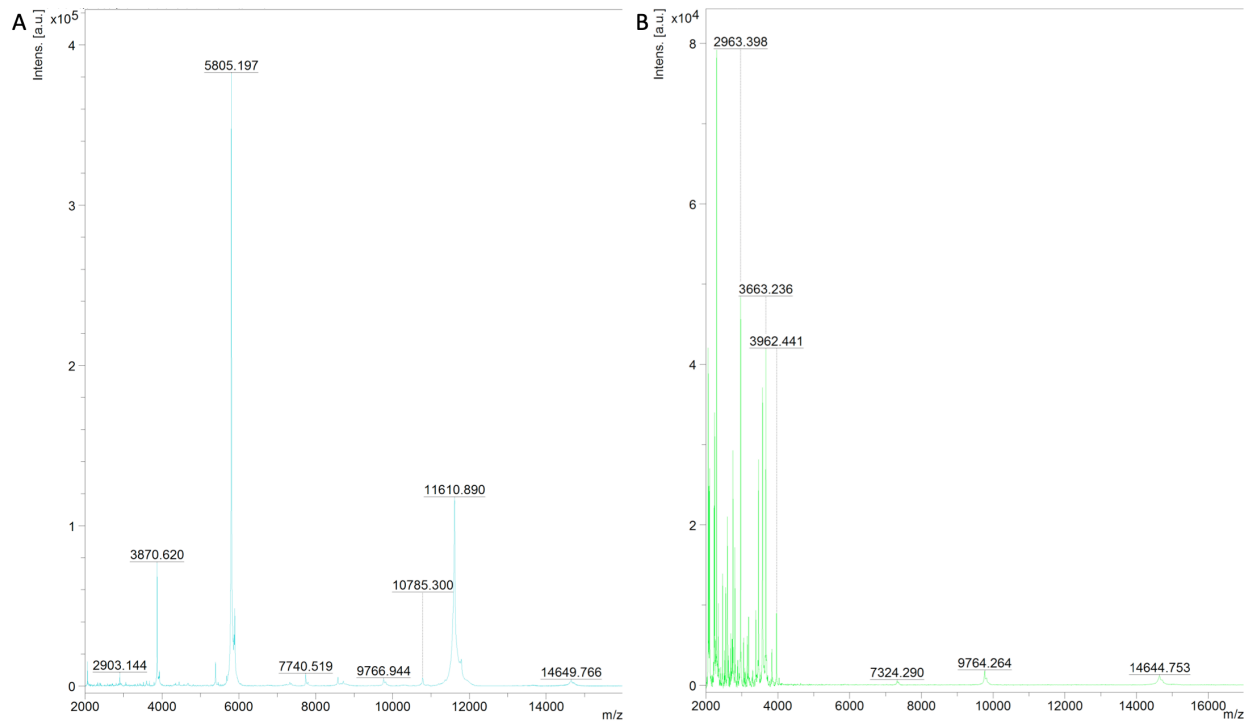


Figure 3.7: MALDI spectra of PK digestion control experiments. A) PK digestion of pure Parent protein. Singly charged, doubly charged, triply charged, and quadruply charged full-length Adnectin ions are detected, but no significant amount of peaks show digested fragments. B) PK digestion of SDS-denatured Parent IBs. No full-length Adnectin was detected, and a significant amount of fragments were present in the 2-4 kDa region.

Another extreme condition to contrast with the native-fold protein is the fully unstructured state. It is difficult to achieve a fully unfolded state, especially for a β -sheet protein with high thermostability and high resistance to denaturants like Adnectins. Previously, it was found that Adnectins go through a slow unfolding but fast refolding process, which makes it very hard to perform a digestion experiment for the unstructured protein. Fortunately, PK retains activity in the presence of detergents and denaturants, which makes it possible to perform digestion experiments under highly denaturing conditions. The digestion of SDS-unfolded Adnectin IBs showed no full-length Adnectin remaining and no detection of a fragment with MW higher than 4 kDa (Figure 3.7B). Due to the difficulty of resolubilizing acetone-precipitated fragments, the experiment was restricted to MALDI analysis. Since there are a great many fragments between 2 to 4 kDa, even with the parallel labelling method, it is difficult to assign all the peaks. However, some peaks that could be assigned to fragments indicate that the cut sites are distributed throughout the primary sequence (Figure 3.8). β B and β F were found to have cut sites, and the corresponding fragments

have relatively high intensity. However, there are still a significant amount of fragments that have intact β B and BC loops; this could be due to the SDS being deficient in fully denaturing the aggregate or protein structure. Nonetheless, this finding further confirms that the β B and β F are well protected in the IBs structure, and the cutting sites result from the exposure of a region rather than the preferences of PK.

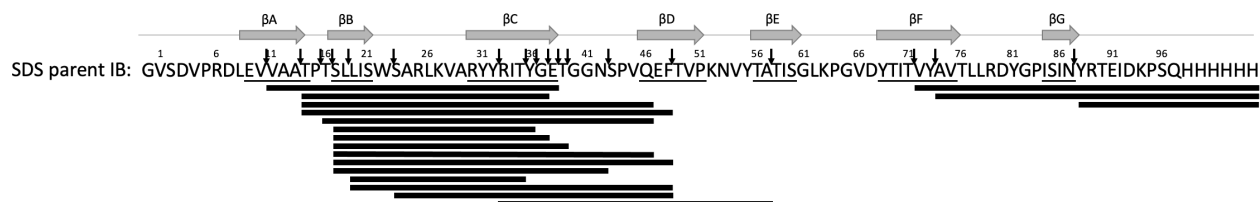


Figure 3.8: PK digestion profile of SDS denatured IBs. The β -strands of the native structure are shown as ribbon arrows above the sequence and are underlined within the primary sequence. Fragments and cut sites identified are presented by black bars under the sequence and black arrows above the sequence.

3.3.4.2 Adnectin variants IBs digestion patterns

The PK digestion pattern of 3 other Adnectin variants IBs was also analyzed using the same method and compared with Parent Adnectin. Besides the digested fragments and digestion sites revealing the protease-resistance and unprotected regions, the similarities in digestion patterns could suggest similar structure components in IBs. The differences would suggest the effect of the point mutations on IB formation. Surprisingly, the digestion patterns between the variants are similar but with some local differences. Figure 3.9 shows the digestion patterns for all four Adnectin variant IBs. Compared to Figure 3.4, we can see that some of the fragments identified from the PK digested sample were present with host cell protease digestion. Regarding the source of these fragments, one simple explanation for the same fragments that still exist after PK digestion could be that the fragments are involved in a highly compact structure and, therefore, resistant to both cellular protease and PK digestion. Another possible explanation is that they can also be produced by PK digestion. Under these circumstances, the fragments produced by the host cell protease were digested into smaller fragments, but PK also preferred these unprotected sites similar to the host cell protease.

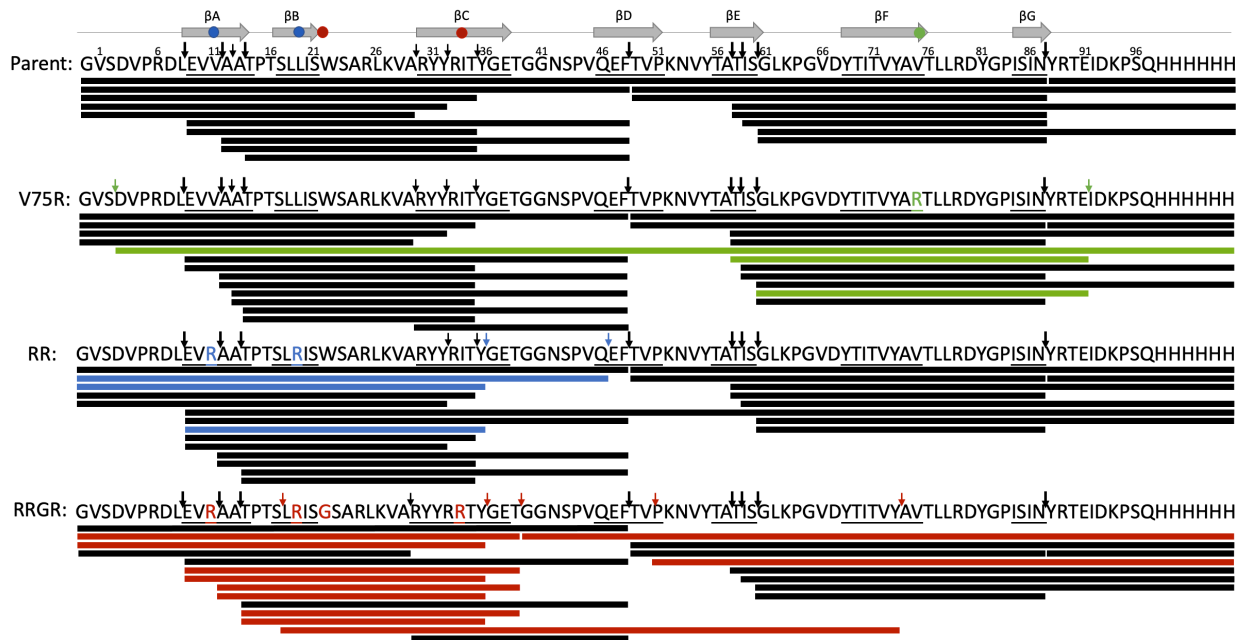


Figure 3.9: PK digestion profile on 4 Adnectin variants IBs. The PK cut sites and digested fragments are presented by vertical arrows above the primary sequence of each variant, with corresponding fragments represented by black bars beneath the sequences. Mutations in the primary sequences and ribbon structures are color-coded: V75R (green), RR (blue), and RRGR (red). The β -strands of the native structure are shown as ribbon arrows above the sequence and are underlined within the primary sequences of each variant. Point mutations specific to each variant are labelled in their respective colors. The PK cut sites unique to the mutant variants compared to the Parent Adnectin are marked by smaller colored arrows, with the corresponding fragments highlighted in the same colors. Additionally, smaller black arrows show cut sites common to more than one, but not all, of the variant IBs.

It was discussed above that the host cell protease appears to prefer cleaving following accessible Try and Phe residues. PK digestion results further support this hypothesis since cut sites are observed after a wide variety of residues for PK, including but not restricted to aromatic residues. PK also identified additional unprotected regions, which resulted from its low specificity for cleavage sites. Among the common sites cleaved by both the host cell proteases and PK, the most obvious one is F48/T49 on the β D. After PK digestion, this cutting site is still observed for all 4 Adnectin variants IBs. The cutting sites after Y32 on the β C for Parent, V75R, and RR are also present after PK digestion. However, the cleavage site after Y36, a common host cell protease cutting site for all 4 variants (Figure 3.4), was only observed with PK for RR and RRGR. For Parent and V75R, the cutting site for PK switched to after T35. This change could indicate that the

region near Y36 is overall exposed. However, PK is non-specific, so the digestion occurred after T35. These observations may be alternatively interpreted as T35 in RR and RRGR could still be involved in a protected structure, which prevented PK from accessing after T35. Another common cleavage site for host cell protease on all 4 variants IBs is Y55 at the end of the DE loop (Figure 3.4). This cut site is no longer observed after PK digestion; instead, there is cleavage before the β E, with cuts at 3 sites on the β E (Figure 3.9). The cut sites after A57, T58, and S60 appeared for all 4 variants of Adnectin IBs. Similar to Parent Adnectin IB discussed in the previous Section, this finding partially contradicts the predictors (Figure A3.3); for the 4 variants, the aggregation propensity is relatively high in this region.

Another region with a few cleavage sites is around the β A, which corresponds well to the predictors. ZipperDB and Tango show no aggregation propensity for the β A (Figure 3.6, A3.3). The predictors suggest that the aggregation propensity for Parent and V75R is high on the β B and significantly decreased after mutating V11R. However, regarding the digestion pattern, the mutations did not cause only a simple local effect (Figure A3.4). The β B for RR and RRGR is still protected from trypsin and PK digestion. This partially corresponds to the qHDX NMR results, in which the β B is not assigned fully, possibly due to the residual aggregation in Dimethylsulfoxide (DMSO)⁷². After the L19R mutation on the β B, this region was still not fully assigned. The last common cleavage site among the 4 variants is at the end of the β G.

Despite the extensive similarities, there are some notable local differences in the digestion patterns of variants. Surprisingly, the differences in the cleavage sites are not close to the mutations on both linear sequences or the native structure. V75R, with a charge mutation on the end of β F, did not exhibit an additional cleavage site near the mutation. The additional cleavage sites compared to Parent IBs are after S2 and E91. These two cleavage sites are near the chain termini and far from the mutation site in the linear sequence and 3D native folded structure (Figure 3.9, 3.10). This indicates the mutation did not cause a dramatic effect globally. However, V75R has been observed by native state HDX to be stabilized at the end of β G⁷³, which may cause additional cleaving after E91. This finding correlates well with the qHDX results. The protection pattern of V75R IBs significantly increased on β G (Figure A3.5)⁷². The melting temperature for the V75R monomer is 70.1°C, and the Parent is 76.8 °C at pH 4³⁴. These findings correspond to the decrease in protection at sites close to V75R in the native structure but increased protection in β G. However,

the decrease in protection did not reflect by additional protease cleavage near the mutation. On the other hand, the charge mutation potentially increased the stability of β G; it could be related to the negatively charged residue E91, which could also influence GdnHCl denaturation (Section 3.2), but more experiments need to be done to confirm the hypothesis.

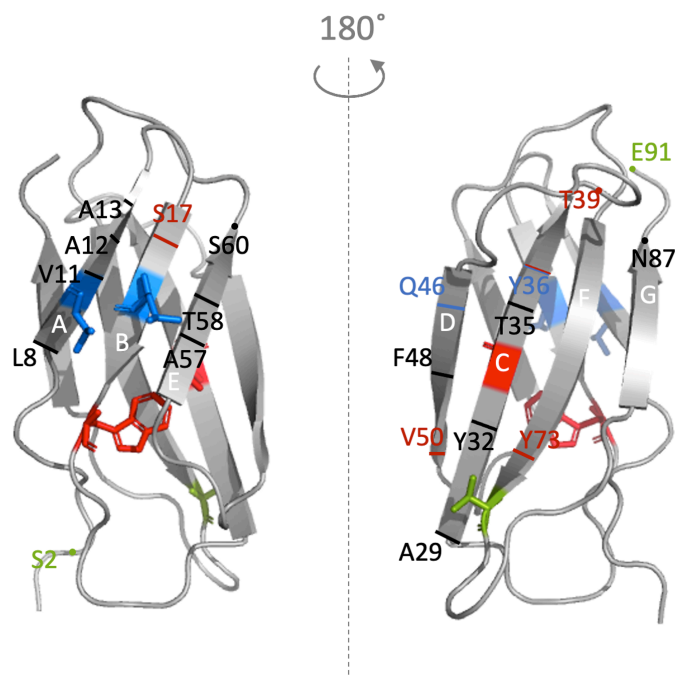


Figure 3.10: PK cut sites of 4 Adnectin IBs show on Adnectin native structure. The native structure of Adnectin is shown from two perspectives, rotated 180° to display both sides. Mutations in the variants are color-coded and shown as stick representations on the ribbon structure. PK digestion sites are represented by thin lines along the ribbon. The left view highlights the β A, β B, and β E strands, while the right view focuses on the β C, β D, β F, and β G strands. Common PK cleavage sites appeared in all four variants are shown with black lines. Variant-specific cleavage sites, differing from the Parent Adnectin, are indicated in their respective colors: V75R (green), RR (blue), and RRGR (red).

For RR and RRGR IBs, the mutation on β A and β B did not cause additional cleavage sites very close to the mutations, but the cleavage sites on β C changed. Unlike Parent, RR and RRGR have a cleavage site after Y36 (Figure 3.9), which was also observed with host cell protease cleavage (Figure 3.4). In addition, for RRGR IBs a distinct cleavage site is observed after T39 on the CD loop, and multiple additional N-terminal as well as C-terminal fragments resulted from this cut site were detected. Overall, the cleavage pattern for RRGR differs the most from the other variants.

Unfortunately, it is hard to tell the exact cause of the additional cleavage site after T39 since it is once again far from the mutation sites in the linear sequence and native folded structure (Figure 3.10). It could be a combination of decreased aggregation propensity in this region due to the 4 mutations, together with less contribution of native protein to the IBs due to the lower stability of the native protein⁷². However, the fragment species covering the CD loop still exists for RRGR digestion. This increases the possibility that this cleavage site is caused by a heterogeneous structure of IB, taking a good percentage of all heterogeneous conformations. For this conformation, T39 is more exposed to PK than Y36 and F48.

In summary, the 3 mutant IBs have more cleavage sites compared to the Parent Adnectin IB, which could suggest destabilization or heterogeneous conformations. V75R increased the protease resistance in the region of the end of β G and resulted in an additional cut site E91. The result also corresponds with the chemical denaturation that V75R is the most stable variant against GdnHCl. qHDX also suggests increased protection of β G, indicating a more stable native-like structure. RR and RRGR IBs differ from Parent IB on β C and CD loop. Both variants produced more fragment species by PK digestion. RRGR IBs, especially, have a distinct cleavage on the CD loop, which differs from the other 3 variants. This additional cut site could indicate a destabilization or a heterogeneous structure. The 4 additional mutations could cause a decrease of protection in the N-terminal and result in additional cut sites. The extended N-terminal fragments could also suggest that the extended region is involved in a heterogeneous structure distinct from the other variants. The resulting C-terminal fragments also suggest a structure protecting the common cleavage site F48. The differences between RRGR and the other three variants in chemical denaturation, FTIR, PTM, and PK digestion are expected since the mutations are made to disrupt the native fold. The aggregation pathway and IB structure of RRGR could significantly differ from Parent, V75R, and RR.

3.3.4.3 Additional experiments for further investigation

Additionally, PK digestion on PMSF-inhibited lysed IBs was performed to test further confirm PK cut sites without cellular proteolysis influencing the interpretation. For PMSF-inhibited lysing, the organic solvent was unavoidably included because the PMSF is dissolved in isopropanol. Although a low amount of 1% v/v was added (Section 2.1.2), the influence of organic solvent on IB structure

is still uncertain. Reassuringly, the fragments detected by MALDI are almost identical to the digestion on IBs lysed without PMSF, but more replicates are required to ensure there are no additional cut sites caused by involving organic solvent. However, it is not the major point to study the influence of organic solvent influence on IBs structure.

Digestion on washed variant IBs was also performed. However, the structural effect of nOG on IBs was uncertain. A previous qHDX experiment found that washed IBs had different changes for different mutant IBs⁷². Based on these observations, it was discussed that detergents could have a structural effect on IBs that may be highly protein-dependent. For digestion on washed IBs, the number of fragments detected by MALDI-MS were significantly more than for unwashed. The reason could be that the detergent washed away some contaminants, such as other proteins or lipids, and therefore, more cleavage sites are exposed for protease; it could also be the structural changes of IBs or the residue nOG affect ionization process.

Chapter 4: Conclusions and Future Work

4.1 Conclusions

IBs are aggregates formed during protein overexpression in heterologous hosts and are crucial for various applications in research, medicine, and biotechnology. Here, low-resolution methods (chemical denaturation and FTIR) provided information on IB stability and structure composition, and high-resolution methods were also used for characterizing IBs structure. Through characterizing IBs formed by different proteins, engineering IBs to produce specific structure properties can be the foundation for their practical applications.

Proteolysis and MS analysis are newly developed methods for IBs. Adnectin is used as the model protein, which directly provides structure information on the segment level from the fragments and residue level from the accessible cut sites. With chemical denaturation and FTIR analysis, the variants of Adnectin IBs have shown heterogeneity, including native-like, unstructured, and amyloid-like structures. The different charge mutations show various effects on the denaturant resistance. RRGR IB, in particular, possesses the most amyloid-like features and is the most resistant to urea but very sensitive to GdnHCl. The proteolytic digestion showed high overall similarities in the 4 Adnectin variant IBs with some local differences. For Adnectin IBs, there is no strong evidence for a well-defined protease-resistant core. However, all the fragments detected covered either the region of β B or β F, which are the central strands in the native structure. The more accessible sites are located on β A, β D, and β E, which are on the edge of a native structure. The results could still indicate a native-like structure with lower protection of the edge strands. At the same time, the well-protected region could also be involved in amyloid-like structure formation.

Proteolytic digestion is a conventional low-resolution technique for characterizing aggregate protease resistance. Using parallel experiments with ^{15}N -labelled samples, the method has been broadened to a higher resolution at the segment level and potentially at the residue level of protease-resistant regions and accessible sites. This method can be applied to other protein IB systems and allows us to analyze aggregate structures without removing contaminants. This new method for protease digestion analysis and corresponding results have great significance for

overcoming challenges in characterizing heterogeneous aggregation processes and advancing the understanding and, ultimately, control of IB structure.

4.2 Future Work

To reach the highest potential of proteolysis interpretation, more information on the IBs is still required. An MS proteomic analysis on the IB samples can be conducted⁵⁶. This analysis would require high sensitivity of the MS, and a large database is also required. With the background ions identified, quantifying fragments from IBs is also possible by using internal standards. However, the protein as internal standards should not overlap with the fragment region and stay within the CsI₃ calibrated range, ideally around 12 kDa. With the quantification, the relative accessibility of the identified cut sites can be obtained. The differences in the fragment amounts from the Adnectin variant IBs can further confirm their relative protection.

The method can also be expanded to other Adnectin variants or other IB models. Some mutations on the loop could help investigate the intermolecular association and potential structural transitions associated with the maturation of IBs. Other initial work of apomyoglobin IBs was performed in our group. With more proteolysis experiments performed, comparing the digestion profile for different proteins in IB can further help us understand the aggregation mechanism and engineer IBs for practical applications.

MS without proteolysis is also a powerful tool for characterizing protein and aggregate structures⁵⁶. Combined with other methods, such as qHDX, covalent labeling, and cross-linking, MS can provide more information on the structure of IBs.

References

- (1) Alberts, B.; Johnson, A.; Lewis, J.; Raff, M.; Roberts, K.; Walter, P. From RNA to Protein. In *Molecular Biology of the Cell. 4th edition*; Garland Science, 2002.
- (2) Newberry, R. W.; Raines, R. T. Secondary Forces in Protein Folding. *ACS Chem. Biol.* **2019**, *14* (8), 1677–1686. <https://doi.org/10.1021/acscchembio.9b00339>.
- (3) Anfinsen, C. B. Principles That Govern the Folding of Protein Chains. *Science* **1973**, *181* (4096), 223–230.
- (4) Almeida, Z. L.; Brito, R. M. Structure and Aggregation Mechanisms in Amyloids.
- (5) Ajmal, M. R. Protein Misfolding and Aggregation in Proteinopathies: Causes, Mechanism and Cellular Response. *Diseases* **2023**, *11* (1), 30. <https://doi.org/10.3390/diseases11010030>.
- (6) Vabulas, R. M.; Raychaudhuri, S.; Hayer-Hartl, M.; Hartl, F. U. Protein Folding in the Cytoplasm and the Heat Shock Response. *Cold Spring Harb. Perspect. Biol.* **2010**, *2* (12), a004390–a004390. <https://doi.org/10.1101/cshperspect.a004390>.
- (7) Hartl, F. U.; Bracher, A.; Hayer-Hartl, M. Molecular Chaperones in Protein Folding and Proteostasis. *Nature* **2011**, *475* (7356), 324–332. <https://doi.org/10.1038/nature10317>.
- (8) 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.* **2018**, *19* (12), 755–773. <https://doi.org/10.1038/s41580-018-0060-8>.
- (9) Karamanos, T. K.; Kalverda, A. P.; Radford, S. E. Generating Ensembles of Dynamic Misfolding Proteins. *Front. Neurosci.* **2022**, *16*. <https://doi.org/10.3389/fnins.2022.881534>.
- (10) Thorlaksen, C.; Schultz, H. S.; Gammelgaard, S. K.; Jiskoot, W.; Hatzakis, N. S.; Nielsen, F. S.; Solberg, H.; Foderà, V.; Bartholdy, C.; Groenning, M. In Vitro and in Vivo Immunogenicity Assessment of Protein Aggregate Characteristics. *Int. J. Pharm.* **2023**, *631*, 122490. <https://doi.org/10.1016/j.ijpharm.2022.122490>.
- (11) Housmans, J. A. J.; Wu, G.; Schymkowitz, J.; Rousseau, F. A Guide to Studying Protein Aggregation. *FEBS J.* **2023**, *290* (3), 554–583. <https://doi.org/10.1111/febs.16312>.
- (12) Bemporad, F.; De Simone, A.; Chiti, F.; Dobson, C. M. Characterizing Intermolecular Interactions That Initiate Native-Like Protein Aggregation. *Biophys. J.* **2012**, *102* (11), 2595–2604. <https://doi.org/10.1016/j.bpj.2012.03.057>.
- (13) Žerovnik, E.; Stoka, V.; Mirtič, A.; Gunčar, G.; Grdadolnik, J.; Staniforth, R. A.; Turk, D.; Turk, V. Mechanisms of Amyloid Fibril Formation – Focus on Domain-swapping. *FEBS J.* **2011**, *278* (13), 2263–2282. <https://doi.org/10.1111/j.1742-4658.2011.08149.x>.
- (14) Sambashivan, S.; Liu, Y.; Sawaya, M. R.; Gingery, M.; Eisenberg, D. Amyloid-like Fibrils of Ribonuclease A with Three-Dimensional Domain-Swapped and Native-like Structure. *Nature* **2005**, *437* (7056), 266–269. <https://doi.org/10.1038/nature03916>.
- (15) Seetaloo, N.; Zacharopoulou, M.; Stephens, A. D.; Kaminski Schierle, G. S.; Phillips, J. J. Millisecond Hydrogen/Deuterium-Exchange Mass Spectrometry Approach to Correlate Local Structure and Aggregation in α -Synuclein. *Anal. Chem.* **2022**, *94* (48), 16711–16719. <https://doi.org/10.1021/acs.analchem.2c03183>.

- (16) Peterle, D.; Pontarollo, G.; Spada, S.; Brun, P.; Palazzi, L.; Sokolov, A. V.; Spolaore, B.; Polverino De Laureto, P.; Vasilyev, V. B.; Castagliuolo, I.; De Filippis, V. A Serine Protease Secreted from *Bacillus Subtilis* Cleaves Human Plasma Transthyretin to Generate an Amyloidogenic Fragment. *Commun. Biol.* **2020**, *3* (1), 764. <https://doi.org/10.1038/s42003-020-01493-0>.
- (17) Seuma, M.; Lehner, B.; Bolognesi, B. An Atlas of Amyloid Aggregation: The Impact of Substitutions, Insertions, Deletions and Truncations on Amyloid Beta Fibril Nucleation. *Nat. Commun.* **2022**, *13* (1), 7084. <https://doi.org/10.1038/s41467-022-34742-3>.
- (18) Fonda, B. D.; Kato, M.; Li, Y.; Murray, D. T. Cryo-EM and Solid State NMR Together Provide a More Comprehensive Structural Investigation of Protein Fibrils. *bioRxiv* **2024**, 2024.05.30.596698. <https://doi.org/10.1101/2024.05.30.596698>.
- (19) Mukherjee, S.; Sakunthala, A.; Gadhe, L.; Poudyal, M.; Sawner, A. S.; Kadu, P.; Maji, S. K. Liquid-Liquid Phase Separation of α -Synuclein: A New Mechanistic Insight for α -Synuclein Aggregation Associated with Parkinson's Disease Pathogenesis. *J. Mol. Biol.* **2023**, *435* (1), 167713. <https://doi.org/10.1016/j.jmb.2022.167713>.
- (20) Chen, Z.; Huai, Y.; Mao, W.; Wang, X.; Ru, K.; Qian, A.; Yang, H. Liquid–Liquid Phase Separation of Biomacromolecules and Its Roles in Metabolic Diseases. *Cells* **2022**, *11* (19), 3023. <https://doi.org/10.3390/cells11193023>.
- (21) De Groot, N. S.; Sabate, R.; Ventura, S. Amyloids in Bacterial Inclusion Bodies. *Trends Biochem. Sci.* **2009**, *34* (8), 408–416. <https://doi.org/10.1016/j.tibs.2009.03.009>.
- (22) Bhatwa, A.; Wang, W.; Hassan, Y. I.; Abraham, N.; Li, X.-Z.; Zhou, T. Challenges Associated With the Formation of Recombinant Protein Inclusion Bodies in *Escherichia Coli* and Strategies to Address Them for Industrial Applications. *Front. Bioeng. Biotechnol.* **2021**, *9*, 630551. <https://doi.org/10.3389/fbioe.2021.630551>.
- (23) Villaverde, A.; García-Fruitós, E.; Rinas, U.; Seras-Franzoso, J.; Kosoy, A.; Corchero, J.; Vazquez, E. Packaging Protein Drugs as Bacterial Inclusion Bodies for Therapeutic Applications. *Microb. Cell Factories* **2012**, *11* (1), 76. <https://doi.org/10.1186/1475-2859-11-76>.
- (24) Kopp, J.; Bayer, B.; Slouka, C.; Striedner, G.; Dürkop, M.; Spadiut, O. Fundamental Insights in Early-stage Inclusion Body Formation. *Microb. Biotechnol.* **2023**, *16* (5), 893–900. <https://doi.org/10.1111/1751-7915.14117>.
- (25) Gifre-Renom, L.; Seras-Franzoso, J.; Rafael, D.; Andrade, F.; Cano-Garrido, O.; Martinez-Trucharte, F.; Ugarte-Berzal, E.; Martens, E.; Boon, L.; Villaverde, A.; Opdenakker, G.; Schwartz, S.; Arís, A.; Garcia-Fruitós, E. The Biological Potential Hidden in Inclusion Bodies. *Pharmaceutics* **2020**, *12* (2), 157. <https://doi.org/10.3390/pharmaceutics12020157>.
- (26) Sabate, R.; De Groot, N. S.; Ventura, S. Protein Folding and Aggregation in Bacteria. *Cell. Mol. Life Sci.* **2010**, *67* (16), 2695–2715. <https://doi.org/10.1007/s00018-010-0344-4>.
- (27) Wasmer, C.; Benkemoun, L.; Sabaté, R.; Steinmetz, M. O.; Couлары-Salin, B.; Wang, L.; Riek, R.; Saupe, S. J.; Meier, B. H. Solid-State NMR Spectroscopy Reveals That *E. Coli* Inclusion Bodies of HET-s(218–289) Are Amyloids. *Angew. Chem. Int. Ed.* **2009**, *48* (26), 4858–4860. <https://doi.org/10.1002/anie.200806100>.
- (28) Chura-Chambi, R. M.; Farah, C. S.; Morganti, L. Human Growth Hormone Inclusion Bodies Present Native-like Secondary and Tertiary Structures Which Can Be Preserved by Mild

- Solubilization for Refolding. *Microb. Cell Factories* **2022**, 21 (1), 164.
<https://doi.org/10.1186/s12934-022-01887-1>.
- (29) Oberg, K.; Chrunchy, B. A.; Wetzel, R.; Fink, A. L. Native-like Secondary Structure in Interleukin-1.β Inclusion Bodies by Attenuated Total Reflectance FTIR. *Biochemistry* **1994**, 33 (9), 2628–2634. <https://doi.org/10.1021/bi00175a035>.
 - (30) Naser, D.; Tarasca, M. V.; Siebeneichler, B.; Schaefer, A.; Deol, H. K.; Soule, T. G. B.; Almey, J.; Kelso, S.; Mishra, G. G.; Simon, H.; Meiering, E. M. High-Resolution NMR H/D Exchange of Human Superoxide Dismutase Inclusion Bodies Reveals Significant Native Features Despite Structural Heterogeneity. *Angew. Chem. Int. Ed.* **2022**, 61 (24), e202112645. <https://doi.org/10.1002/anie.202112645>.
 - (31) Talafová, K.; Hrabárová, E.; Chorvát, D.; Nahálka, J. Bacterial Inclusion Bodies as Potential Synthetic Devices for Pathogen Recognition and a Therapeutic Substance Release. *Microb. Cell Factories* **2013**, 12 (1), 16. <https://doi.org/10.1186/1475-2859-12-16>.
 - (32) Lipovsek, D. Adnectins: Engineered Target-Binding Protein Therapeutics. *Protein Eng. Des. Sel.* **2011**, 24 (1–2), 3–9. <https://doi.org/10.1093/protein/gzq097>.
 - (33) Chandler, P. G.; Buckle, A. M. Development and Differentiation in Monobodies Based on the Fibronectin Type 3 Domain. *Cells* **2020**, 9 (3), 610. <https://doi.org/10.3390/cells9030610>.
 - (34) Trainor, K.; Gingras, Z.; Shillingford, C.; Malakian, H.; Gosselin, M.; Lipovšek, D.; Meiering, E. M. Ensemble Modeling and Intracellular Aggregation of an Engineered Immunoglobulin-Like Domain. *J. Mol. Biol.* **2016**, 428 (6), 1365–1374. <https://doi.org/10.1016/j.jmb.2016.02.016>.
 - (35) Gee, E. P. S.; Yüksel, D.; Stultz, C. M.; Ingber, D. E. SLLISWD Sequence in the 10FNIII Domain Initiates Fibronectin Fibrillogenesis. *J. Biol. Chem.* **2013**, 288 (29), 21329–21340. <https://doi.org/10.1074/jbc.M113.462077>.
 - (36) Cota, E.; Clarke, J. Folding of Beta-Sandwich Proteins: Three-State Transition of a Fibronectin Type III Module. *Protein Sci. Publ. Protein Soc.* **2000**, 9 (1), 112–120.
 - (37) 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.* **2001**, 305 (5), 1185–1194. <https://doi.org/10.1006/jmbi.2000.4378>.
 - (38) Luo, J.; Teplyakov, A.; Obmolova, G.; Malia, T. J.; Chan, W.; Jacobs, S. A.; O’Neil, K. T.; Gilliland, G. L. N-terminal B-strand Swapping in a Consensus-derived Alternative Scaffold Driven by Stabilizing Hydrophobic Interactions. *Proteins Struct. Funct. Bioinforma.* **2014**, 82 (7), 1527–1533. <https://doi.org/10.1002/prot.24517>.
 - (39) Teplyakov, A.; Obmolova, G.; Malia, T. J.; Luo, J.; Jacobs, S. A.; Chan, W.; Domingo, D.; Baker, A.; O’Neil, K. T.; Gilliland, G. L. C-Terminal β-Strand Swapping in a Consensus-Derived Fibronectin Type III Scaffold. *Proteins Struct. Funct. Bioinforma.* **2014**, 82 (7), 1359–1369. <https://doi.org/10.1002/prot.24502>.
 - (40) Cheng, X.; Wang, K.; Zhao, Y.; Wang, K. Research Progress on Post-Translational Modification of Proteins and Cardiovascular Diseases. *Cell Death Discov.* **2023**, 9 (1), 275. <https://doi.org/10.1038/s41420-023-01560-5>.
 - (41) Gundry, R. L.; White, M. Y.; Murray, C. I.; Kane, L. A.; Fu, Q.; Stanley, B. A.; Van Eyk, J. E. Preparation of Proteins and Peptides for Mass Spectrometry Analysis in a Bottom-Up

- Proteomics Workflow. *Curr. Protoc. Mol. Biol.* **2010**, 90 (1).
<https://doi.org/10.1002/0471142727.mb1025s88>.
- (42) Antalis, T. M.; Bugge, T. H.; Wu, Q. Membrane-Anchored Serine Proteases in Health and Disease. In *Progress in Molecular Biology and Translational Science*; Elsevier, 2011; Vol. 99, pp 1–50. <https://doi.org/10.1016/B978-0-12-385504-6.00001-4>.
 - (43) Zheng, Z.; Zeng, Y.; Lai, K.; Liao, B.; Li, P.; Tan, C. S. H. Protein Painting for Structural and Binding Site Analysis via Intracellular Lysine Reactivity Profiling with O-Phthalaldehyde. *Chem. Sci.* **2024**. <https://doi.org/10.1039/D4SC00032C>.
 - (44) Trypsin as a Biochemical Tool for the Characterization of Ras-Related Protein Structure and Function. **2018**.
 - (45) Lambeth, T. R.; Julian, R. R. Proteolysis of Amyloid β by Lysosomal Enzymes as a Function of Fibril Morphology. *ACS Omega* **2021**, 6 (47), 31520–31527.
<https://doi.org/10.1021/acsomega.1c03915>.
 - (46) Shin, T. M.; Isas, J. M.; Hsieh, C.-L.; Kayed, R.; Glabe, C. G.; Langen, R.; Chen, J. Formation of Soluble Amyloid Oligomers and Amyloid Fibrils by the Multifunctional Protein Vitronectin. *Mol. Neurodegener.* **2008**, 3 (1), 16. <https://doi.org/10.1186/1750-1326-3-16>.
 - (47) Farrell, R. E. Chapter 2 - Creating a Ribonuclease-Free Environment. In *RNA Methodologies (Fifth Edition)*; Farrell, R. E., Ed.; Academic Press, 2017; pp 55–74.
<https://doi.org/10.1016/B978-0-12-804678-4.00002-6>.
 - (48) Sabaté, R.; Espargaró, A.; Saupe, S. J.; Ventura, S. Characterization of the Amyloid Bacterial Inclusion Bodies of the HET-s Fungal Prion. *Microb. Cell Factories* **2009**, 8 (1), 56.
<https://doi.org/10.1186/1475-2859-8-56>.
 - (49) Hilz, H.; Wieggers, U.; Adamietz, P. Stimulation of Proteinase K Action by Denaturing Agents: Application to the Isolation of Nucleic Acids and the Degradation of ‘Masked’ Proteins. *Eur. J. Biochem.* **1975**, 56 (1), 103–108. <https://doi.org/10.1111/j.1432-1033.1975.tb02211.x>.
 - (50) Karas, Michael.; Hillenkamp, Franz. Laser Desorption Ionization of Proteins with Molecular Masses Exceeding 10,000 Daltons. *Anal. Chem.* **1988**, 60 (20), 2299–2301.
<https://doi.org/10.1021/ac00171a028>.
 - (51) Whitehouse, C. M.; Dreyer, R. N.; Yamashita, Masamichi.; Fenn, J. B. Electrospray Interface for Liquid Chromatographs and Mass Spectrometers. *Anal. Chem.* **1985**, 57 (3), 675–679.
<https://doi.org/10.1021/ac00280a023>.
 - (52) Tuck, M.; Grélard, F.; Blanc, L.; Desbenoit, N. MALDI-MSI Towards Multimodal Imaging: Challenges and Perspectives. *Front. Chem.* **2022**, 10, 904688.
<https://doi.org/10.3389/fchem.2022.904688>.
 - (53) Gonnet, F.; Lemaître, G.; Waksman, G.; Tortajada, J. MALDI/MS Peptide Mass Fingerprinting for Proteome Analysis: Identification of Hydrophobic Proteins Attached to Eucaryote Keratinocyte Cytoplasmic Membrane Using Different Matrices in Concert. *Proteome Sci.* **2003**.
 - (54) Rubakhin, S. S.; Sweedler, J. V. A Mass Spectrometry Primer for Mass Spectrometry Imaging. In *Mass Spectrometry Imaging*; Rubakhin, S. S., Sweedler, J. V., Eds.; Methods in Molecular Biology; Humana Press: Totowa, NJ, 2010; Vol. 656, pp 21–49.
https://doi.org/10.1007/978-1-60761-746-4_2.

- (55) Glish, G. L.; Burinsky, D. J. Hybrid Mass Spectrometers for Tandem Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2008**, *19* (2), 161–172. <https://doi.org/10.1016/j.jasms.2007.11.013>.
- (56) Courouble, V. V.; Mann, M. D.; Griffin, P. R. Advances in Mass Spectrometry-Based Structural Proteomics: Development of HDX-MS and XL-MS Techniques from Recombinant Protein to Cellular Systems. *Isr. J. Chem.* **2023**, *63* (10–11), e202300084. <https://doi.org/10.1002/ijch.202300084>.
- (57) Espargaró, A.; Sabaté, R.; Ventura, S. Kinetic and Thermodynamic Stability of Bacterial Intracellular Aggregates. *FEBS Lett.* **2008**, *582* (25–26), 3669–3673. <https://doi.org/10.1016/j.febslet.2008.09.049>.
- (58) Schaefer, A.; Naser, D.; Siebeneichler, B.; Tarasca, M. V.; Meiering, E. M. Methodological Advances and Strategies for High Resolution Structure Determination of Cellular Protein Aggregates. *J. Biol. Chem.* **2022**, *298* (9), 102197. <https://doi.org/10.1016/j.jbc.2022.102197>.
- (59) Barth, A. Infrared Spectroscopy of Proteins. *Biochim. Biophys. Acta BBA - Bioenerg.* **2007**, *1767* (9), 1073–1101. <https://doi.org/10.1016/j.bbabo.2007.06.004>.
- (60) Lou, X.; Van Dongen, J. L. J.; Meijer, E. W. Generation of Csl Cluster Ions for Mass Calibration in Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2010**, *21* (7), 1223–1226. <https://doi.org/10.1016/j.jasms.2010.02.029>.
- (61) Trainor, K. Inclusion Body Formation by Mutants of the Tenth Human Fibronectin Type III Domain.
- (62) *Suppressing Posttranslational Gluconoylation of Heterologous Proteins by Metabolic Engineering of Escherichia coli.* <https://doi.org/10.1128/AEM.01790-07>.
- (63) Schweida, D.; Barraud, P.; Regl, C.; Loughlin, F. E.; Huber, C. G.; Cabrele, C.; Schubert, M. The NMR Signature of Gluconoylation: A Frequent N-Terminal Modification of Isotope-Labeled Proteins. *J. Biomol. NMR* **2019**, *73* (1–2), 71–79. <https://doi.org/10.1007/s10858-019-00228-6>.
- (64) Aebersold, R.; Agar, J. N.; Amster, I. J.; Baker, M. S.; Bertozzi, C. R.; Boja, E. S.; Costello, C. E.; Cravatt, B. F.; Fenselau, C.; Garcia, B. A.; Ge, Y.; Gunawardena, J.; Hendrickson, R. C.; Hergenrother, P. J.; Huber, C. G.; Ivanov, A. R.; Jensen, O. N.; Jewett, M. C.; Kelleher, N. L.; Kiessling, L. L.; Krogan, N. J.; Larsen, M. R.; Loo, J. A.; Ogorzalek Loo, R. R.; Lundberg, E.; MacCoss, M. J.; Mallick, P.; Mootha, V. K.; Mrksich, M.; Muir, T. W.; Patrie, S. M.; Pesavento, J. J.; Pitteri, S. J.; Rodriguez, H.; Saghatelian, A.; Sandoval, W.; Schlüter, H.; Sechi, S.; Slavoff, S. A.; Smith, L. M.; Snyder, M. P.; Thomas, P. M.; Uhlén, M.; Van Eyk, J. E.; Vidal, M.; Walt, D. R.; White, F. M.; Williams, E. R.; Wohlschläger, T.; Wysocki, V. H.; Yates, N. A.; Young, N. L.; Zhang, B. How Many Human Proteoforms Are There? *Nat. Chem. Biol.* **2018**, *14* (3), 206–214. <https://doi.org/10.1038/nchembio.2576>.
- (65) Geoghegan, K. F.; Dixon, H. B. F.; Rosner, P. J.; Hoth, L. R.; Lanzetti, A. J.; Borzilleri, K. A.; Marr, E. S.; Pezzullo, L. H.; Martin, L. B.; LeMotte, P. K.; McColl, A. S.; Kamath, A. V.; Stroth, J. G. Spontaneous γ -N-6-Phosphogluconoylation of a “His Tag” in Escherichia Coli: The Cause of Extra Mass of 258 or 178 Da in Fusion Proteins.
- (66) Chen, L.; Kashina, A. Post-Translational Modifications of the Protein Termini. *Front. Cell Dev. Biol.* **2021**, *9*, 719590. <https://doi.org/10.3389/fcell.2021.719590>.

- (67) Jürgen, B.; Breitenstein, A.; Urlacher, V.; Büttner, K.; Lin, H.; Hecker, M.; Schweder, T.; Neubauer, P. Quality Control of Inclusion Bodies in Escherichia Coli. *Microb. Cell Factories* **2010**, *9* (1), 41. <https://doi.org/10.1186/1475-2859-9-41>.
- (68) Kushnirov, V. V.; Dergalev, A. A.; Alexandrov, A. I. Proteinase K Resistant Cores of Prions and Amyloids. *Prion* **2020**, *14* (1), 11–19. <https://doi.org/10.1080/19336896.2019.1704612>.
- (69) Kyte, J.; Doolittle, R. F. A Simple Method for Displaying the Hydropathic Character of a Protein. *J. Mol. Biol.* **1982**, *157* (1), 105–132. [https://doi.org/10.1016/0022-2836\(82\)90515-0](https://doi.org/10.1016/0022-2836(82)90515-0).
- (70) Goldschmidt, L.; Teng, P. K.; Riek, R.; Eisenberg, D. Identifying the Amylome, Proteins Capable of Forming Amyloid-like Fibrils. *Proc. Natl. Acad. Sci.* **2010**, *107* (8), 3487–3492. <https://doi.org/10.1073/pnas.0915166107>.
- (71) 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.* **2004**, *22* (10), 1302–1306. <https://doi.org/10.1038/nbt1012>.
- (72) Schaefer, A. O. Structural Characterization and Engineering of Adnectin Inclusion Bodies.
- (73) Trainor, K.; Doyle, C. M.; Metcalfe-Roach, A.; Steckner, J.; Lipovšek, D.; Malakian, H.; Langley, D.; Krystek Jr., S. R.; Meiering, E. M. Design for Solubility May Reveal Induction of Amide Hydrogen/Deuterium Exchange by Protein Self-Association. *J. Mol. Biol.* **2022**, *434* (2), 167398. <https://doi.org/10.1016/j.jmb.2021.167398>.

Appendix

Appendix-Chapter 2

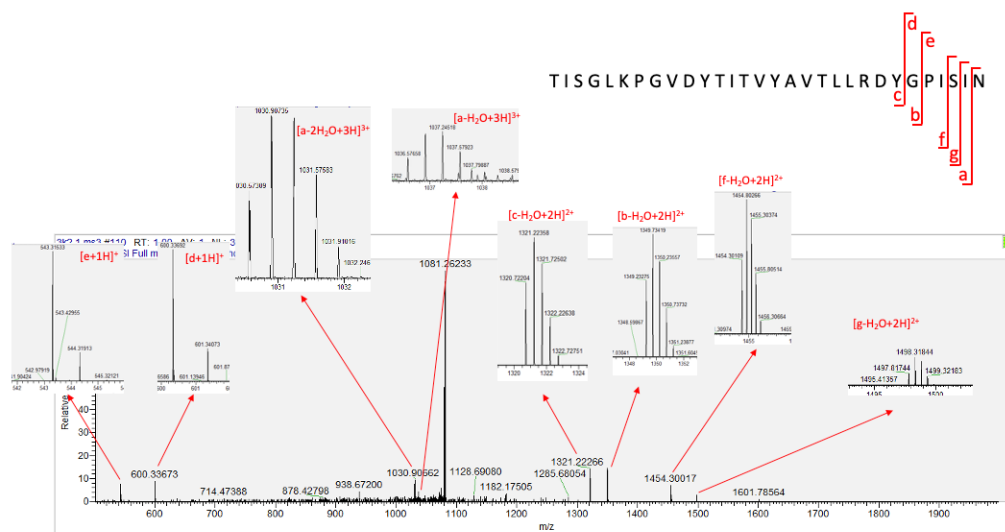


Figure A2.1: ESI-MS/MS for fragment identification. Data are shown for Parent IB PK digested fragment (T58-N87) ESI-MS/MS. The peak for the intact fragment is at 1081.0 ± 1.0 m/z, in the scan from 500.0 to 2000.0 m/z, NCE=10 ($z=3$). The peaks derived from the intact fragment are labelled on the primary sequence (top right) and on the spectra.

Rep1	Rep2	Rep3	Rep4	unlab	15N	frag
2060.264	2058.771	2061.106	2092.647	2060.1962	2092.954	ITRDKSPQHHHHHHH
2964.76	2939.638	2941.928	2972.221	2940.408	2972.166	GLKPGVDITTVYALLTSLWLRKQSY
3067.62	3066.195	3068.694	3104.44	3066.6076	3104.3286	EYVATLTSKPGDKVYITVYALLTSLWLRKQSYRRYIT
3140.908	3140.086	3142.24	3140.576	3140.576	3139.5235	ISGLKPGVDITTVYALLTSLWLRKQSY
3243.515	3241.43	3243.718	3276.597	3241.726	3276.4855	ITSLKPGVDITTVYALLTSLWLRKQSY
3635.579	3634.824	3637.175	3678.78	3635.19	3677.8589	GVDPDLVEATLTSVLLTSLWLRKQSYRRY
3963.989	3963.349	3965.622	4011.194	3963.4738	4011.2215	ITPVLSSWLRKQSYRRYRITGETGNSPVQEF
4063.123	4005.538	4007.765	4055.167	4005.6411	4055.1279	GVDPDLVEATLTSVLLTSLWLRKQSYRRYIT
4105.19	4105.169	4107.453	4115.514	4105.631	4115.3979	IAITPVLSSWLRKQSYRRYRITGETGNSPVQEF
4216.521	4215.959	4218.213	4261.937	4215.8644	4261.5266	TPVNNYITATSLKPGDKVYITVYALLTSLWLRKQSY
4433.77	4433.174	4435.497	4485.872	4433.0109	4485.6213	EYVATLTSKPGDKVYITVYALLTSLWLRKQSYRRYITGETGNSPVQEF
4582.163	4581.987	4584.01	5046.117	5282.579	5046.0964	GLKPGVDITTVYALLTSLWLRKQSYRITRDKSPQHHHHHHH
5288.128	5282.928	5285.388	5350.438	5282.9157	5350.4162	ITSLKPGVDITTVYALLTSLWLRKQSYRITRDKSPQHHHHHHH
5372.651	5372.147	5374.422	5456.906	5372.0441	5456.6154	GVDPDLVEATLTSVLLTSLWLRKQSYRRYITGETGNSPVQEF
	6257.078	6259.843	6385.617	6277.0373	6375.6592	TPVNNYITATSLKPGDKVYITVYALLTSLWLRKQSYRRYITGETGNSPVQHHHHHHH
9570.339	9567.983		9679.941	9568.885	9677.0705	GVDPDLVEATLTSVLLTSLWLRKQSYRRYITGETGNSPVQHHHHHHH

Rep1	Rep2	Rep3	Rep4	unlab	15N	frag.
2059.547	2059.805	2061.009	2091.62	2060.1962	2092.954	YRTEDIKPSQHNNHH
2236.049	2238.128		2264.222	2238.415	2263.212	YRRTYTGEGNSPQVF
2398.441	2396.608		2628.624	2397.407	2625.821	TPSLLSWSARUKARYRIT
2667.745	2667.947		2700.923	2668.149	2701.896	APTPSLLSWSARUKARYRIT
2738.535	2738.796		2772.739	2739.282	2773.976	AAPTPLLSWSARUKARYRIT
2996.632	2996.89	2997.071	3030.829	2997.4555	3032.1584	GLKPDVDTTIVYARTLLRSDYPSIN
3065.949	3066.178	3066.335	3103.028	3066.676	3104.3283	EVAAATPSLLSWSARUKARYRIT
3196.512	3196.59	3197.159	3322.793	3197.6926	3324.421	ATPSTLSDYDTTIVYARTLLRSDYPSIN
3298.227	3298.373	3298.373	3335.064	3298.7973	3336.5185	ISGLGPDVDTTIVYARTLLRSDYPSIN
3546.141	3546.446	3546.586	3587.203	3547.038	3588.7299	GLKPDVDTTIVYARTLLRSDYPSINRYRT
	3634.303	3636.202	3677.044	3635.1904	3678.8669	GVSDPRLVEDAATPSLLSWSARUKARYR
3847.702		3850.132	3891.458	3848.3798	3893.0495	ISGLGPDVDTTIVYARTLLRSDYPSINRYRT
3961.851	3961.281		4000.428	3963.478	4001.1212	TPSLLSWSARUKARYRITGEGNSPQVF
4084.607		4007.005	4053.362	4051.641	4055.2739	GVSDPRLVEDAATPSLLSWSARUKARYRIT
4033.587		4035.889	4081.541	4034.5524	4083.1931	APTPSLLSWSARUKARYRITGEGNSPQVF
4104.998	4105.341	4107.439	4153.675	4105.631	4155.2643	AAPTPLLSWSARUKARYRITGEGNSPQVF
4272.502	4272.369	4274.411	4319.308	4272.931	4321.5593	TPKNVYATISGLKGVDTTIVYARTLLRSDYPSIN
4423.686	4423.07		4483.967	4423.0109	4485.622	YVAATPSLLSWSARUKARYRITGEGNSPQVF
5370.458	5373.995	5360.504	5303.641	5308.6286	5306.1291	GLKPDVDTTIVYARTLLRSDYPSINRYRTEDIKPSQHNNHH
5237.545	5238.06	5240.169	5305.793	5238.9598	5308.3531	ISGLGPDVDTTIVYARTLLRSDYPSINRYRTEDIKPSQHNNHH
5338.95	5339.374	5341.895	5408.145	5339.9704	5410.4489	ISGLGPDVDTTIVYARTLLRSDYPSINRYRTEDIKPSQHNNHH
5370.458	5371.031	5378.838	5433.999	5372.0441	5436.567	GVSDPRLVEDAATPSLLSWSARUKARYRITGEGNSPQVF
6311.07	6313.502		6391.48	6314.0921	6395.4004	TPKNVYATISGLKGVDTTIVYARTLLRSDYPSINRYRTEDIKPSQHNNHH
11443.645	11424.074	11429.227	11567.318	11423.851	11566.794	DVPRDVEDAATPSLLSWSARUKARYRITGEGNSPQVFTPKNVYATISGLKGVDTTIVYARTLLRSDYPSINRYRTEDIKPSQHNNHH

Rep1	Rep2	Rep3	Rep4	Rep5	unlab	15N	frag.
2059.316	2060.671	2055.441	2092.077	2090.162	2092.954	WTEDKPSQHHHHH	
2639.48	2641.361	2640.288	2675.622	2640.098	2675.834	TPTSLRSWSKAKVARYIT	
2781.43	2783.486	2782.594	2815.959	2782.256	2819.977	ATPILSRSLRSKAKVARYIT	
2795.346	2797.005	2796.293	2834.255	2797.389	2833.966	EVRAATPSLRSLRSKAKVARYIT	
3030.785	3032.324	3030.841	3072.453	3040.408	3071.165	ATPILSRSLRSKAKVARYIT	
3165.758	3168.016	3166.959	3210.427	3166.096	3210.3674	EVRAATPSLRSLRSKAKVARYIT	
3241.029	3243.449	3241.271	3276.287	3241.758	3276.485	TGSLGQVDTTITVAATLRDGSIN	
3238.955		3320.24	3374.762	3239.862	3374.536	EVRAATPSLRSLRSKAKVARYIT	
3731.483	3736.344	3734.132	3784.131	3735.271	3784.905	ATPILSRSLRSKAKVARYIT	
3805.738	3808.476	3806.457	3857.506	3806.408	3857.1276	TPTSLRSWSKAKVARYITGETGNSPVQE	
4104.493	4107.523	4105.982	4161.413	4105.728	4161.3127	GVSDPDLVEAAATPSLRSLRSKAKVARYIT	
4147.631	4150.789	4149.123	4201.873	4148.659	4201.270	ATPILSRSLRSKAKVARYITGETGNSPVQE	
4255.058	4258.299	4256.819	4312.864	4256.394	4312.478	TPTSLRSWSKAKVARYITGETGNSPVQE	
4255.701		4257.198	4313.485	4255.8019	4313.394	ATPILSRSLRSKAKVARYITGETGNSPVQE	
4267.714		4269.123	4315.607	4268.896	4325.481	GVSDPDLVEAAATPSLRSLRSKAKVARYIT	
4322.08	4335.461	4333.567	4391.93	4335.493	4391.660	EVRAATPSLRSLRSKAKVARYITGETGNSPVQE	
4380.117	4382.988	4381.688	4438.981	4382.729	4438.904	GVSDPDLVEAAATPSLRSLRSKAKVARYITGETGNSPVQE	
5180.683	5184.674	5182.545	5248.652	5185.8107	5248.319	ISGLGQVDTTITVAATLRDGSINRTEKPSQHHHHH	
5194.05	5197.614	5195.419	5263.385	5196.386	5264.3288	GVSDPDLVEAAATPSLRSLRSKAKVARYITGETGNSPVQE	
5281.884	5285.823	5283.258	5350.819	5282.1597	5350.4162	TPKYNQVDTTITVAATLRDGSINRTEKPSQHHHHH	
5470.102	5474.763	5472.717	5541.607	5473.717	5541.607	TPKYNQVDTTITVAATLRDGSINRTEKPSQHHHHH	
6256.008	6260.543	6257.833	6275.782	6257.035	6275.452	TPKYNQVDTTITVAATLRDGSINRTEKPSQHHHHH	

Rep1	Rep2	Rep3	2004.104	2004.186	15N	frag.
2059.4969	2060.767	2060.104	2091.586	2092.954	IRTSGLSSRKARVYRRYTGSGNSPQVHHH	
2280.795	2281.728		2310.254	2281.431	2311.228	RYRRYTGSGNSPQVSGF
2716.346	2717.706	2717.243	2754.286	2717.415	2755.852	IAATPSSLRSSKARVYRRYTGSGNSPQVHHH
2858.4952	2859.784	2859.427	2898.425	2859.268	2899.975	AATPSSLRSSKARVYRRYTGSGNSPQVHHH
2939.486	2940.823		2971.004	2940.408	2972.166	GLGPGVDTYTVAVTLVRDPSGIN
3003.5237	3004.815	3004.451	3044.559	3004.413	3046.107	TVTLSSLRSSKARVYRRYTGSGNSPQVHHH
3145.276	3146.998	3146.403	3188.572	3146.705	3190.245	AAATPSSLRSSKARVYRRYTGSGNSPQVHHH
3214.384	3214.821	3214.731	3280.375	3243.731	3290.381	EVPAATPSSLRSSKARVYRRYTGSGNSPQVHHH
3259.9073	3261.551	3531.051	3578.668	3530.163	3580.639	EVATPSSLRSSKARVYRRYTGSGNSPQVHHH
3318.4239	3320.292	3919.714	3969.714	3920.388	3972.979	IRTSGLSSRKARVYRRYTGSGNSPQVHHH
4189.141	4183.347	4182.757	4420.182	4182.763	4421.337	GVDSPDLVREAAATPSSLRSSKARVYRRYTGSGNSPQVHHH
4214.905	4216.961		4299.878	4215.864	4261.356	TPVNNVYATLVRDPSGINNRTDKPSQVHHH
4445.7909	4447.443	4447.086	4505.586	4446.906	4507.514	EVPAATPSSLRSSKARVYRRYTGSGNSPQVHHH
4682.6575	4407.492	4470.004	4528.924	4470.033	4531.583	GVDSPDLVREAAATPSSLRSSKARVYRRYTGSGNSPQVHHH
4980.511	4982.408	4891.895	5004.362	4952.581	5040.069	GVDSPDLVTVAVTLVRDPSGINNRTDKPSQVHHH
5180.415	5192.399	5182.065	5245.823	5181.507	5248.135	ISGAKGVDTYTVATLVDGYSINRTDKPSQVHHH
5281.587	5283.767		5282.517	5282.517	5350.416	ITSGLPGRVDTYTVAVTLVRDPSGINNRTDKPSQVHHH
5384.1145	5386.031	5385.702	5454.512	5385.938	5458.482	GVDSPDLVREAAATPSSLRSSKARVYRRYTGSGNSPQVHHH
5065.5625	6056.714		6130.8513	6056.805	6233.136	PNNVYATLGLGPGVDTYTVAVTLVRDPSGINNRTDKPSQVHHH
6214.612	6216.882	6215.449	6290.897	6217.059	6292.489	IRSSGSLSSKARVYRRYTGSGNSPQVTVVNNVYATLVRDPSQVDTVTV
6255.445	6257.495		6315.452	6257.495	6315.452	TPVNNVYATLVRDPSGINNRTDKPSQVHHH
7171.1326	7173.299		7259.448	7172.929	7262.382	GSGNGPQVTPVNNVYATLVRDPSGINNRTDKPSQVHHH

58

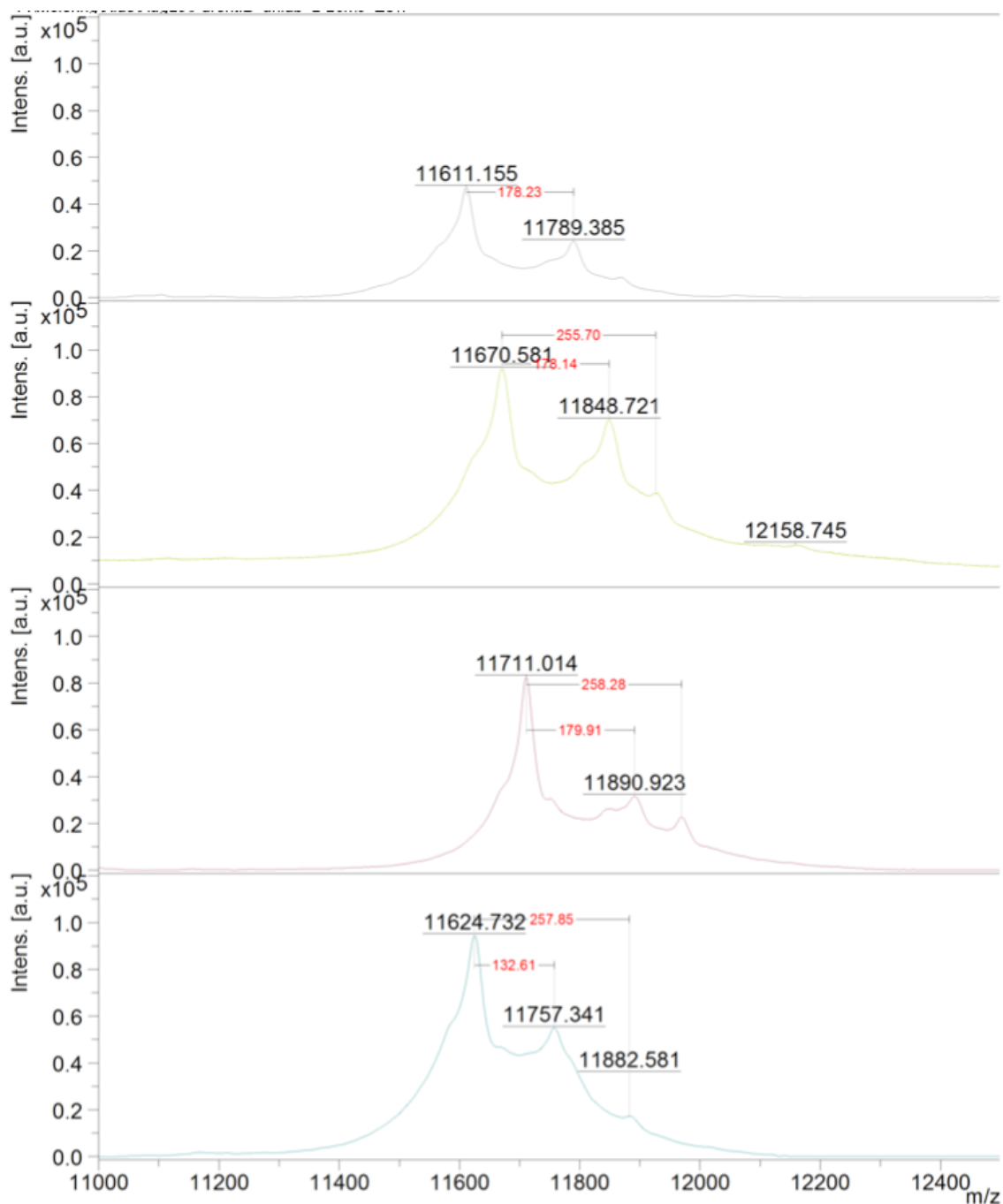


Figure A3.1: PTM analysis of Parent, V75R, RR, RRGR IBs. The 4 Adnectin variants are listed from top to bottom. Parent (gray), V75R (green), RR (purple), and RRGR (blue). RRGR IB shows no glucuronoylated species but shows significant full length Met species.

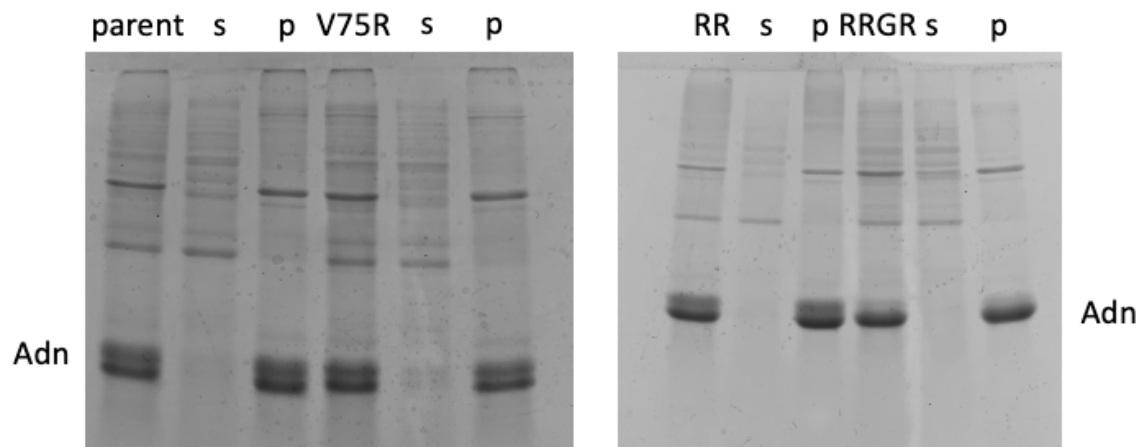


Figure A3.2: SDS-PAGE analysis of 4 Adnectin variant IBs. Parent, V75R, and RR showed 2 distinct bands, containing full-length Adnectin, gluconoylation and phosphogluconoylation. RRGR IB did not show clearly separated bands like the other 3 variants IBs, suggesting a faster aggregation process.

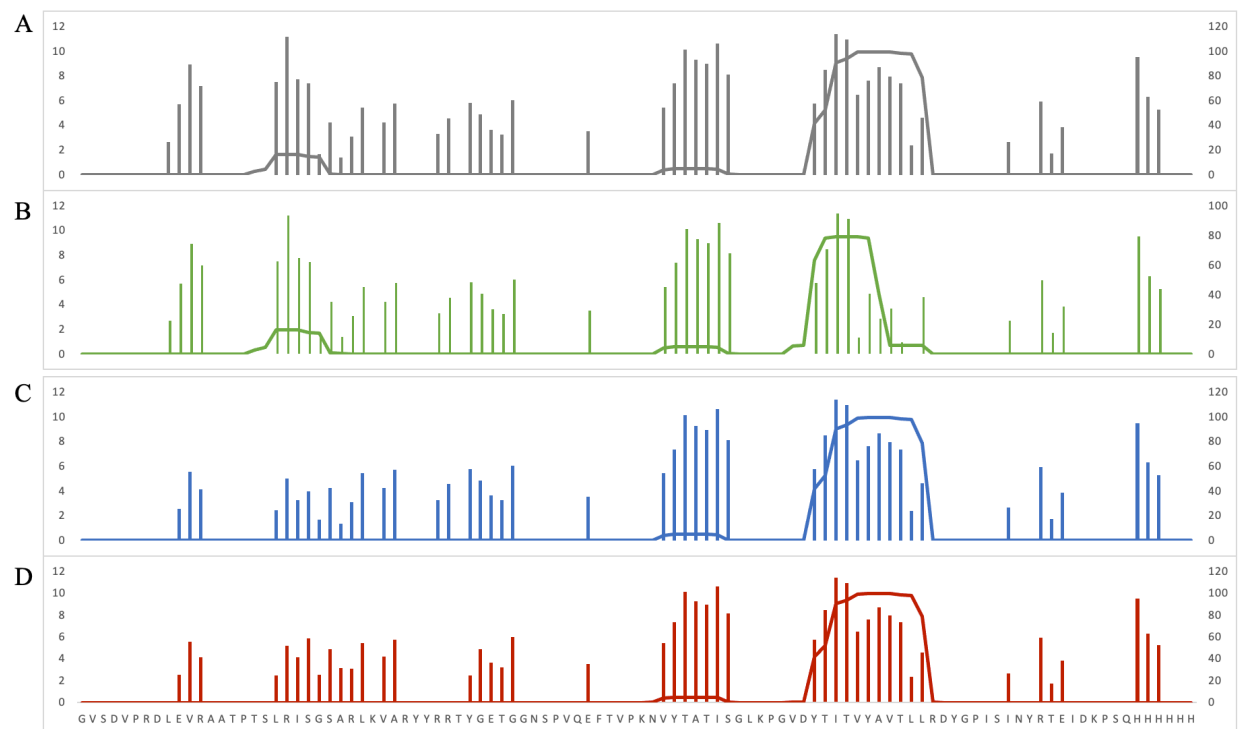


Figure A3.3: ZipperDB and Tango aggregation propensity prediction. The ZipperDB (bars, left scale)⁷⁰ and Tango (line, right scale)⁷¹ profile of Adnectin variants. A) Parent (black), B) V75R (green), C) RR (blue), and D) RRGR (red).

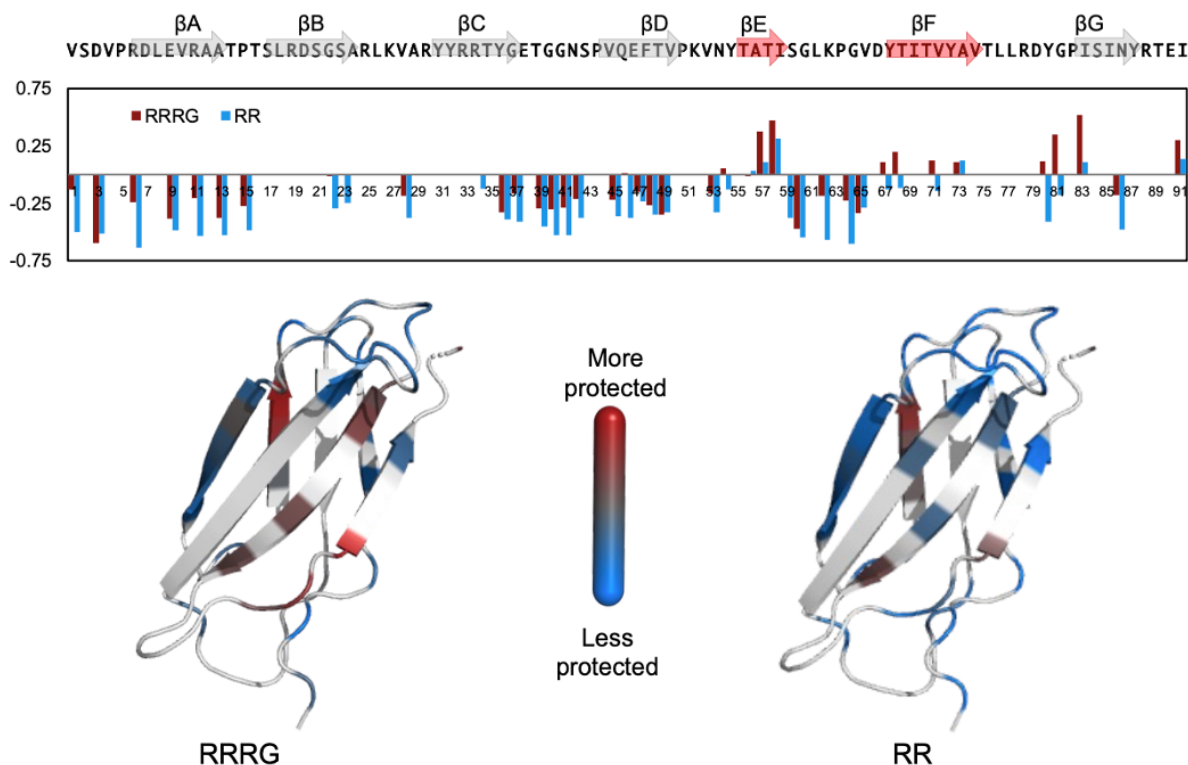


Figure A3.4: qHDX protection of RR and RRRG IB compared to Parent IB. Some residues on the βB , βC loop, and βG loop were unassigned. Their relative protection against Parent IB is presented in the bar graph and mapped on the ribbon structure. Blue represents the less protected region; red represents the more protected region⁷².

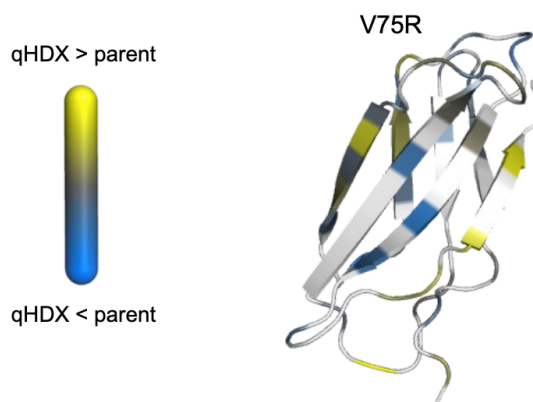


Figure A3.5: qHDX protection of V75R IB compared to Parent IB. Some residues on the βB , βC loop, and βG loop were unassigned. Relative protection against Parent IB is mapped on the ribbon structure. Blue represents the less protected region; yellow represents the more protected region⁷².