

# Predicting and Measuring Molecular Mechanisms of Protein Aggregation

by

Heather Primmer

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, 2011

©Heather Primmer 2011

## **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

Protein aggregation is a hallmark of a number of neurodegenerative disorders including Alzheimer's Disease, Huntington's Disease, and Amyotrophic Lateral Sclerosis. Despite the common occurrence of protein aggregation in disease, the fundamental mechanisms controlling the propensity of a protein to aggregate are not well understood. Over the past decade, one of the most significant advancements in the field of understanding protein aggregation has been the development of several aggregation prediction algorithms. In this study, two separate approaches were used to investigate the detailed molecular mechanisms of protein aggregation. First, a thorough investigation that compared nine protein aggregation prediction techniques was performed. Protein aggregation propensity calculations were performed on wild type and mutant sequences of three diverse proteins including Superoxide Dismutase (SOD), human Acylphosphatase (AcP), and the amyloid beta peptide ( $A\beta_{42}$ ). This study presents the first wide-scale comparison of such a large number of prediction algorithms, and additionally provides new information on the ability of the algorithms to successfully predict the experimentally observed aggregation of several mutations of diverse proteins. The algorithms were predominantly developed based on a set of known amyloid-forming proteins and peptides, however, are quite diverse in the way they were designed and the proteins on which they were tested. Interestingly, significant variation was observed when predicting the aggregation propensity of identical sequences by multiple techniques, indicating that the algorithms do not possess a consensus on the primary factors that govern aggregation. Further analyses compared predicted and observed aggregation data for several mutants of the test proteins. The aggregation prediction algorithms predominantly demonstrated poor to moderate correlations with observed aggregation, and the strongest correlations occurred in instances where the test data was used in the development of the algorithms. The general lack of ability of the algorithms to predict the aggregation patterns of more than one test protein suggests that aggregation may be a much more specific process than it is generally attributed to be in that there may be inherently different

properties modulating the aggregation mechanisms of different proteins towards varying aggregate structures.

The second component of this project was to experimentally examine the role of salt in influencing protein aggregation as a method to elucidate the specific molecular mechanisms controlling protein aggregation pathways. The ALS-causing SOD1 mutation, A4V, in both the oxidized and reduced apo form, was used as a model protein. The role of NaCl and Na<sub>2</sub>SO<sub>4</sub> in mediating protein aggregation was studied using several techniques. While oxidized apo A4V showed very little evidence of aggregation even in the presence of salt, for reduced apo, aggregates readily formed and were promoted by the addition of salt. This finding correlated with the increasing kosmotropic nature of the salt as described by the Hofmeister series. The aggregates formed in the presence of salt contained intermolecular disulphide bonds and demonstrated ANS and ThT binding, indicating aggregates are likely to be largely hydrophobic and possess beta-sheet morphology. Salt promotes protein aggregation in two ways: 1) electrostatic interactions shield protein charges and reduce repulsion between proteins, and 2) specific interactions stabilize various aggregation-prone conformations of the protein. This work is evidence of the important role of salt in influencing protein aggregation and provides a framework for future studies into the complex effects of solution conditions in modulating protein aggregation pathways.

Both aspects of this study contribute greatly to furthering the understanding of the molecular mechanisms governing protein aggregation. This is of particular importance to neurodegenerative diseases, where uncovering the factors that modulate the formation of toxic aggregate species is important for disease treatment and prevention. The potential aggregation mechanisms of SOD1, and the contributions it may play in ALS pathogenesis, will be discussed throughout this study.

## Acknowledgements

I would first like to thank my supervisor, Dr. Elizabeth Meiering, for her leadership, support, guidance and encouragement. I have deeply appreciated many valuable conversations regarding the work presented in this thesis and possibly even more importantly those that had nothing to do with this work.

I would also like to thank the members of the Meiering lab for their help and support throughout my graduate work. It has been an absolute pleasure working with such an excellent team of people. I would specifically like to thank Helen Stubbs and Ming Tong for teaching me so much when I was first starting, and making all the hours spent in the lab so much more fun and Marty Smith for always being so excited to talk science.

Finally I would like to thank all of the incredible people in my life that have supported me for so many years. I thank my friends for giving me the opportunity to understand loving community, my parents for their endless love and believing that I can do anything, and my siblings for being my best friends and my biggest fans.

## Table of Contents

AUTHOR'S DECLARATION.....	ii
Abstract.....	iii
Acknowledgements.....	v
Table of Contents.....	vi
List of Figures.....	x
List of Tables.....	xiii
List of Abbreviations.....	xiv
Chapter 1 General Introduction.....	1
1.1 Protein Folding, Misfolding and Aggregation.....	1
1.2 Pathways and Products of Protein Aggregation.....	3
1.3 Modulating Factors in Protein Aggregation.....	7
1.3.1 Protein Stability and Structural Dynamics.....	7
1.3.2 Inherent Characteristics of the Polypeptide Sequence.....	8
1.3.3 Solution Conditions.....	9
1.4 Protein Aggregation and Disease.....	12
1.5 Amyotrophic Lateral Sclerosis.....	14
1.6 Cu,Zn Superoxide Dismutase and ALS.....	14
1.6.1 ALS Disease Durations are Specific for SOD1 Mutations.....	16
1.6.2 The Importance of Reduced Apo SOD1.....	17
1.6.3 SOD1 Aggregation and Toxicity in ALS.....	19
1.6.4 Pseudo Wild-type Superoxide Dismutase.....	22
1.7 Importance of Further Studies on the Mechanisms of Protein Aggregation.....	23
1.8 Research Objectives and Outline.....	24
Chapter 2 Predicting Protein Aggregation.....	25
2.1 Introduction.....	25
2.1.1 Chiti-Dobson Equation.....	27
2.1.2 Wang-Agar Equation.....	28
2.1.3 Zyggregator and Ztox.....	28
2.1.4 FoldAmyloid.....	29
2.1.5 PASTA.....	30
2.1.6 TANGO.....	31

2.1.7 WALTZ.....	32
2.1.8 Profile 3D.....	33
2.2 Methods.....	33
2.2.1 Quantifying predicted aggregation propensities .....	33
2.2.2 Normalization of Aggregation Prediction Data .....	44
2.2.3 Production of Hot-spot Maps.....	44
2.3 Results.....	45
2.3.1 Predicting Aggregation Propensity of Superoxide Dismutase Mutants.....	45
2.3.2 Predicting Aggregation Propensity of Human Acylphosphatase Mutants.....	49
2.3.3 Predicting Aggregation Propensities of Amyloid Beta Peptide Mutants.....	53
2.4 Discussion .....	55
2.4.1 Inconsistent Patterns in Predicted Aggregation Propensities between Algorithms .....	55
2.4.2 Overlapping prediction hot-spots using multiple algorithms.....	59
2.4.3 Current Barriers in Accurate Aggregation Prediction.....	61
2.4.4 Conclusion .....	64
Chapter 3 Correlating Predicted and Observed Aggregation.....	66
3.1 Introduction.....	66
3.1.1 Techniques for Measuring Protein Aggregation .....	68
3.2 Methods.....	70
3.2.1 Quantifying Observed Aggregation .....	70
3.2.2 Production of Correlation Plots and Statistical Analysis .....	71
3.3 Results.....	71
3.3.1 Human Acylphosphatase Correlations.....	71
3.3.2 Amyloid Beta Correlations .....	75
3.3.3 Superoxide Dismutase Correlations.....	78
3.4 Discussion .....	88
3.4.1 Aggregation Prediction Algorithms Do Not Accurately Predict Observed Aggregation for Diverse Proteins .....	88
3.4.2 Interpreting the Correlation Results between Predicted and Observed Aggregation.....	93
3.4.3 Difficulty in Predicting Protein Aggregation Outside of the Context of Algorithm Development .....	96
3.4.4 Predicted and Observed SOD1 Aggregation fails to determine ALS Disease Duration.....	96

3.4.5 Conclusions.....	102
Chapter 4 Probing the Role of Salt in Protein Aggregation Mechanisms.....	104
4.1 Introduction.....	104
4.1.1 The Role of Ionic Strength in Protein Aggregation .....	104
4.1.2 ANS Binding.....	104
4.1.3 Colloidal Stability and the Second Virial Coefficient.....	106
4.1.4 Monitoring Apo A4V Aggregation in the Presence of Salt .....	108
4.2 Methods.....	108
4.2.1 Protein Production and Purification .....	108
4.2.2 De-metallation of Holo SOD1 to Prepare Apo SOD1 .....	109
4.2.3 Disulphide Reduction.....	109
4.2.4 Iodoacetamide Modification SDS-PAGE Gels .....	110
4.2.5 Preparation of Small Volume Aggregation Trials .....	110
4.2.6 ANS Fluorescence to Monitor Protein Aggregation .....	111
4.2.7 Thioflavin T Fluorescence .....	111
4.2.8 Measuring Aggregate Sizes Using DLS .....	112
4.2.9 Measuring the Second Virial Coefficient Using Light Scattering .....	112
4.2.10 DNTB Assay for the Determination of TCEP Oxidation .....	113
4.3 Results.....	113
4.3.1 Apo Oxidized A4V Aggregation in Salt.....	113
4.3.2 Reduced Apo A4V Aggregation in 1 mM TCEP Buffer and Varying Salt Conditions.....	119
4.3.3 Reduced apo A4V Aggregation in 10 mM TCEP Buffer and Varying Salt Conditions.....	127
4.4 Discussion .....	140
4.4.1 Oxidized Apo A4V is only slightly Prone to Aggregate in Quiescent Conditions in a Range of Salt Conditions.....	140
4.4.2 Reduced apo A4V in the presence of 1 mM or 10 mM TCEP and Various Salt Conditions is Prone to re-oxidize and form Disulphide Cross-linked Aggregates with Amyloid-like characteristics .....	144
4.4.3 Conclusions.....	156
Chapter 5 Summary and Future Work.....	159
5.1 Summary and Conclusions.....	159
5.1.1 Comparison of Multiple Aggregation Prediction Algorithms.....	161

5.1.2 Correlating Predicted and Observed Protein Aggregation .....	162
5.1.3 Investigating the Role of Reduced Apo SOD1 in ALS.....	164
5.1.4 Probing the Role of Salt in Modulating Aggregation .....	165
5.2 Future Work .....	167
5.2.1 Improving Aggregation Prediction Algorithms .....	167
5.2.2 Further Experimental Investigations into the Specific Roles of Solution Conditions in Modulating Aggregation .....	170
Bibliography .....	172

## List of Figures

Figure 1.1 The energy landscape of protein folding.....	3
Figure 1.2 Simplified pathway of protein aggregation from the native state..	4
Figure 1.3 Interconnected pathways of protein aggregation..	5
Figure 1.4 Structural components of dimeric holo SOD1.....	16
Figure 1.5 The change in melting temperature induced by SOD1 mutations in different forms of the protein.....	19
Figure 1.6 Proposed mechanisms for mutant SOD1 toxicity in ALS.....	20
Figure 2.1 Calculation of energy scores using the PASTA algorithm..	31
Figure 2.2 Zyggregator output for pWT SOD1.....	36
Figure 2.3 FoldAmyloid prediction results for pWT SOD1.....	37
Figure 2.4 Example output of the PASTA algorithm.....	39
Figure 2.5 Output predicted aggregation from the TANGO server for pWT SOD1.....	40
Figure 2.6 Output data for aggregation prediction of pWT SOD1 using the Waltz server.....	42
Figure 2.7 Output data for the predicted aggregation propensity of pWT SOD1 using the Profile 3D method.....	43
Figure 2.8 Predicted aggregation propensities of SOD1 mutants display significant variability between algorithms for the same mutant.....	47
Figure 2.9 Aggregation-prone regions of pWT SOD1 predicted by seven different algorithms..	48
Figure 2.10 Predicted aggregation propensities of AcP mutants display significant variability between algorithms for the same mutant.....	49
Figure 2.11 Predicted aggregation-prone regions of AcP by seven different prediction algorithms..	52
Figure 2.12 Predicted aggregation propensity of amyloid beta mutants displays significant variability between algorithms for the same mutant.....	53
Figure 2.13 Predicted aggregation-prone regions of amyloid beta by seven different prediction algorithms.....	54
Figure 2.14 Aggregation-prone hotspots of three different proteins..	61
Figure 3.1 Chemical structure of Thioflavin T <sup>111</sup> .....	69
Figure 3.2 Correlation plots of observed versus predicted aggregation for 30 AcP mutants using nine different prediction algorithms.....	74

Figure 3.3 Correlation plots of observed aggregation <i>versus</i> predicted aggregation for A $\beta$ <sub>42</sub> mutants using nine different prediction algorithms.....	77
Figure 3.4 Correlation plots of observed aggregation <i>versus</i> predicted aggregation for SOD1 mutants using nine different prediction algorithms.....	80
Figure 3.5 The role of stability in aggregation propensity..	83
Figure 3.6 Relationship between observed aggregation for reduced apo SOD1 mutants, quantified using DLS, and ALS disease durations. ....	84
Figure 3.7 Correlation plots of ALS disease durations <i>versus</i> predicted aggregation scores for SOD1 mutants using nine different prediction algorithms (listed in the ordinate axis labels).....	86
Figure 3.8 Relationship between the change in melting temperature for reduced apo SOD1 mutants and the ALS disease duration characteristic for each mutant.....	87
Figure 3.9 Identification of outliers for reduced apo SOD1 correlations.....	100
Figure 3.10 Role of aggregate size in dictating ALS disease duration.....	101
Figure 4.1 ANS Fluorescence spectra in the presence of protein aggregates.....	106
Figure 4.2 Time-course of ANS Fluorescence for apo oxidized A4V samples in varying salt conditions and 20 mM Hepes, pH 7.4.....	114
Figure 4.3 DLS results for end point samples of oxidized apo A4V aggregation time-trials. ....	115
Figure 4.4 Debye plots for oxidized apo A4V in varying salt conditions.....	118
Figure 4.5 Summarized ANS fluorescence results for reduced apo A4V samples in 20 mM Hepes, 1 mM TCEP and various salt concentrations. ....	119
Figure 4.6 ANS Fluorescence for aggregation time-trials of reduced apo A4V samples in various salt conditions.....	120
Figure 4.7 Iodoacetamide-modified reduced apo A4V samples at various time points during aggregation time-courses in differing salt conditions and 1 mM TCEP, 20 mM Hepes, and 37°C...	121
Figure 4.8 Time-course of 1mM TCEP oxidation. The percent reduced TCEP on the y-axis is calculated from the ability of TCEP to reduce DTNB into two NTB chromophores. ....	122
Figure 4.9ThT Fluorescence for end point aggregation samples of reduced apo A4V with and without Na <sub>2</sub> SO <sub>4</sub> .....	124
Figure 4.10 DLS results for reduced apo A4V aggregation trials in 1 mM TCEP, 20 mM Hepes, and various concentrations of salt. ....	126

Figure 4.11 Relationship between ANS Fluorescence and DLS derived mean count rate at the end point of reduced apo A4V aggregation time-trials in 20 mM Hepes, 1 mM TCEP, and varying salt conditions. ....	127
Figure 4.12 Summary of ANS fluorescence results for reduced apo A4V aggregation trials. Samples were incubated in anaerobic conditions at 37°C.. ....	128
Figure 4.13 ANS Fluorescence time-course measurements for 1 mg/mL reduced apo A4V in 20 mM Hepes, 10 mM TCEP, pH 7.4, 37°C, and various salt conditions.....	129
Figure 4.14 SDS-PAGE gels of iodoacetamide-modified reduced apo A4V samples at various time points during aggregation time-courses in differing salt conditions and 10 mM TCEP, 20 mM Hepes, pH 7.4.....	131
Figure 4.15 Time-course of 10mM TCEP oxidation.....	133
Figure 4.16 ThT Fluorescence of reduced apo A4V aggregate samples.....	134
Figure 4.17 Average DLS result for of two independent reduced apo A4V aggregation trials in 10 mM TCEP, 20 mM Hepes, pH 7.4 and varying salt conditions.. ....	135
Figure 4.18 Relationship between ANS Fluorescence and DLS derived mean count rate at the end point of reduced apo A4V aggregation time-trials in 20 mM Hepes, 10 mM TCEP and varying salt conditions. ....	137
Figure 4.19 Debye plots for reduced apo A4V in a range of solution conditions.. ....	138
Figure 4.20 Possible roles of salts on promoting protein aggregation.. ....	149
Figure 4.21 Schematic of a potential aggregation mechanism of reduced apo A4V in salt conditions resulting in the formation of disulphide cross-linked, amyloid-like soluble aggregates.. ....	158
Figure 5.1 Overview of protein aggregation mechanisms.....	160

## List of Tables

Table 2.1 Summary of the variables involved in nine different aggregation prediction algorithms. ...	26
Table 2.2 Scales of hydrophobicity, $\beta$ -sheet propensity and charge for the 20 natural amino acids, reproduced from Chiti <i>et al.</i> <sup>8</sup> .....	35
Table 2.3 Normalized predicted aggregation propensity for wild type and twenty four SOD1 mutants. ....	46
Table 2.4 Normalized aggregation propensities for wild type and 30 AcP mutants using nine prediction algorithms.....	50
Table 2.5 Normalized aggregation propensities for wild type and 5 amyloid beta mutants using nine prediction algorithms.....	54
Table 2.6 Comparison of the methods used to test the validity of protein aggregation prediction techniques.....	57
Table 3.1 Predicted and Observed Aggregation Correlation Summary .....	81
Table 3.2 Correlation results for the sum of predicted aggregation propensity and mutant instability to observed aggregation propensity .....	83
Table 3.3 Correlation values between predicted aggregation and ALS disease duration.....	86
Table 4.1 Hydrodynamic diameters of particles in oxidized apo A4V samples at the end point of the ANS aggregation time-trial. ....	116
Table 4.2 Molecular weight and second virial coefficients of oxidized apo A4V in varying salt conditions determined by Debye plot analysis. ....	118
Table 4.3 Average hydrodynamic diameters of end point aggregate species for two independent time-courses of reduced apo A4V.....	127
Table 4.4 Average hydrodynamic diameter of end point aggregate species for two independent time-courses of reduced apo A4V aggregation in a range of salt conditions.....	136
Table 4.5 Molecular Weights and Second Virial Coefficients determined from Debye plot analysis of reduced apo A4V in varying solution conditions. ....	139
Table 5.1 Summary of potential variables that contribute to protein aggregation. ....	160

## List of Abbreviations

A $\beta$ .....	amyloid beta
AcP.....	human acylphosphatase
AFM.....	atomic force microscopy
ALS.....	amyotrophic lateral sclerosis
apo SOD1.....	metal free form of human copper zinc superoxide dismutase
ANS.....	8-Anilino-1-naphthalenesulfonate
B <sub>22</sub> .....	second virial coefficient
conA .....	concanavalin A
CCS.....	copper chaperone for superoxide dismutase
DLS.....	dynamic light scattering
DSC.....	differential scanning calorimetry
DTT.....	dithiothreitol
DTNB.....	5, 5-Dithiobis(2-nitrobenzoic acid)
EDTA.....	ethylenediaminetetraacetic acid
ER.....	endoplasmic reticulum
fALS.....	familial amyotrophic lateral sclerosis
Hepes.....	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
holo SOD1.....	fully metallated human Cu,Zn superoxide dismutase
MW.....	Molecular weight
NTB.....	2-nitro-5-thiobenzoate
PDB.....	Protein Data Bank
pI .....	isoelectric point
pWT.....	pseudo-wildtype human Cu/Zn superoxide dismutase
ROS .....	reactive oxygen species
sALS.....	sporadic amyotrophic lateral sclerosis
SDS-PAGE.....	sodium dodecyl sulphate polyacrylamide gel electrophoresis
SOD1.....	Cu,Zn superoxide dismutase
TCEP.....	tris(2-carboxymethyl) phosphine
ThT .....	Thioflavin T
T <sub>m</sub> .....	melting temperature
TMAO.....	Trimethylamine-N-oxide
WT .....	wild type

# Chapter 1

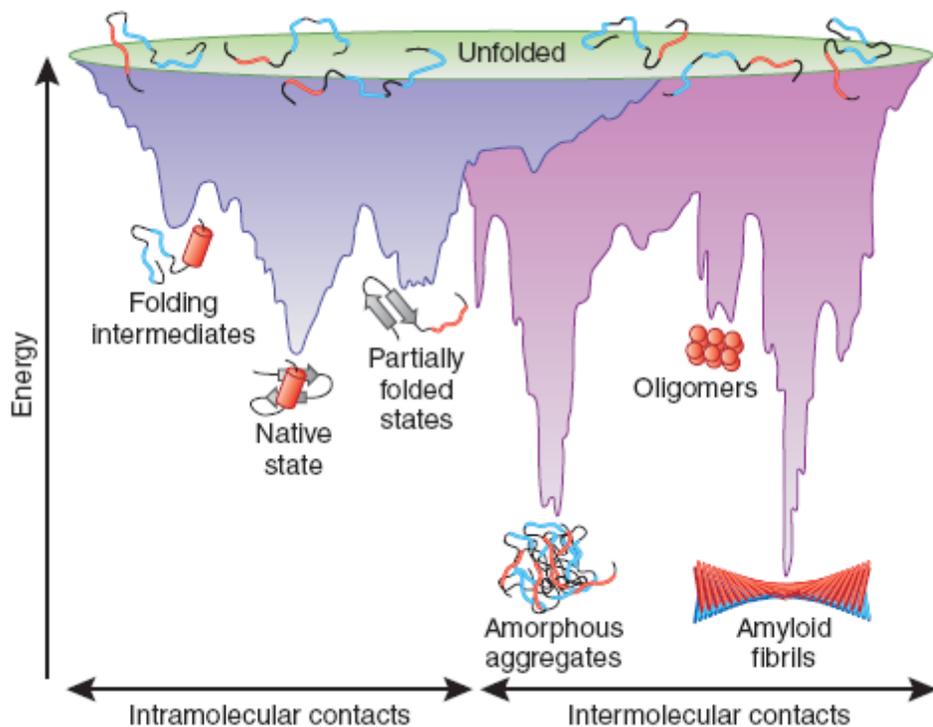
## General Introduction

### 1.1 Protein Folding, Misfolding and Aggregation

The ability of polypeptide sequences to fold into functional proteins is one of the most impressive phenomena in molecular biology. It requires extensive cellular precision to orchestrate the complex events that result in a mature protein product able to complete its role within a living machine. Inside a living cell, protein synthesis occurs on the ribosome, utilizing the information received by the mRNA<sup>1</sup>. Protein folding begins as the transcript is translated at the ribosome and continues to completion once the fully translated amino acid sequence is released from the ribosome<sup>2</sup>. Folding may be assisted by molecular chaperones, some of which bind non-specifically to the nascent polypeptide to prevent aggregation, and others that specifically guide later stages of the folding process, such as the formation of globular folds, the binding of cofactors, and the transition into multimeric forms<sup>3</sup>. Remarkably, however, most proteins are able to spontaneously fold into their native structure without the aid of chaperones. This phenomenon has led to the understanding that all of the information required for proper folding is contained in the amino acid sequence and that the native state has the most stable conformation under physiological conditions in order for the protein to naturally reach this structure<sup>4</sup>. Based on this observation, early ideas suggested that a protein followed a single pathway of folding events that eventually led to the desired structure. This theory was problematic, however, since it would take an impossible amount of time for a protein to find this particular conformation by systematically attempting all possible pathways<sup>3</sup>. This led to the concept of an energy landscape, developed by Leopold *et al.* in 1992<sup>5</sup>. This new framework suggested that protein folding is not just one pathway with a series of intermediates, but instead included a wide range of possible routes along a folding funnel that all led to the final, native structure. Figure 1.1 shows a visual depiction of a protein folding funnel. The left side of the funnel, coloured dark purple, represents the intramolecular contact pathway that results in the natively

folded protein. The rough edges on the sides of the funnel represent intermediate structures that are likely to be transiently populated during the folding process. The trough of the funnel represents the lowest energy, native state of the protein<sup>6</sup>.

Given the complexity of the protein folding process, it is expected that mistakes can and will occur. When a protein folds into a non-native form, this is termed “misfolding”. *In vivo*, a cell responds to these mistakes by using chaperones that rescue misfolded proteins by aiding in proper refolding. The formation of non-native structures from intermolecular interactions of protein monomer subunits is termed *aggregation*. When a misfolded protein persists, it often results in the exposure of hydrophobic areas that can lead to the development of protein aggregates through non-native protein-protein interactions<sup>3</sup>. Figure 1.1 includes aggregate structures in the protein folding funnel diagram. On the left-hand side of the funnel, in purple, intramolecular contacts that promote intermediate conformations leading to the native state dominate. On the right-hand side, in pink, intermolecular contacts result in funnelling toward amorphous and amyloid (*vide infra*) aggregate structures. Overlap of the two sides of the funnel represents the fact that similar factors can contribute to both protein folding and protein aggregation. Aggregation can be a result of “off-pathway” intermediates populated during folding or from a destabilized native state, resulting in partially folded states that tend toward intermolecular associations<sup>7</sup>. The types of aggregate structures that can result from intermolecular associations will be described in the next section.

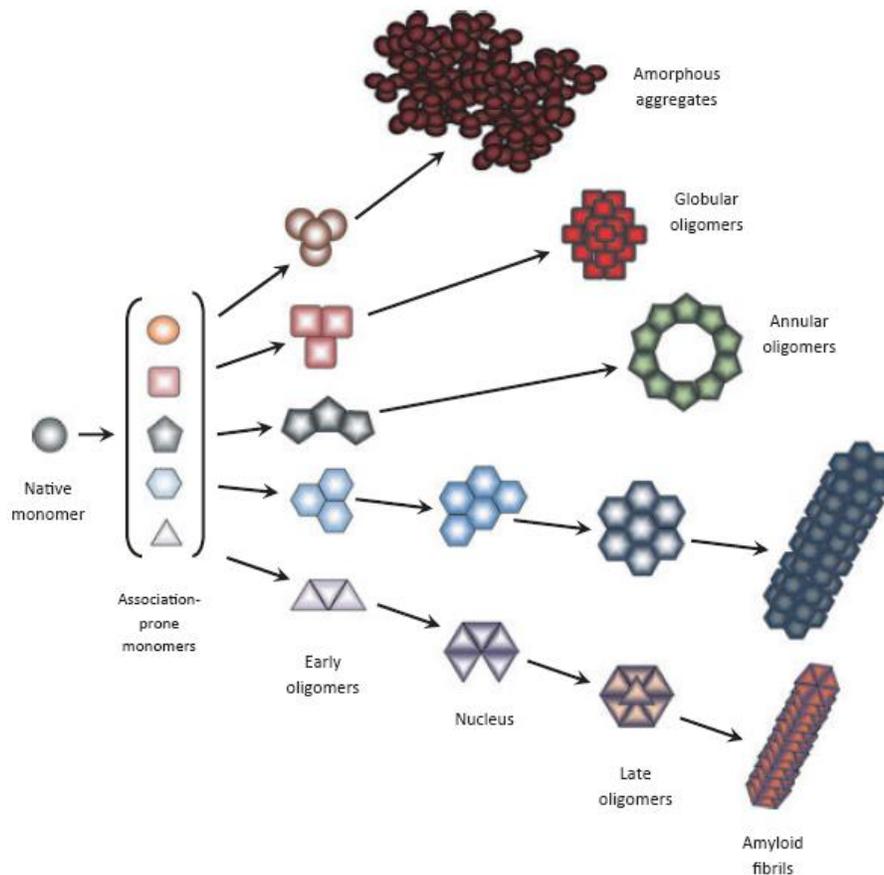


**Figure 1.1 The energy landscape of protein folding.** On the left side of the diagram, an unfolded polypeptide samples intramolecular contacts in order to achieve the lowest energy, native state. It may transiently populate folding intermediates on its way to a proper fold. On the right side of the diagram, an unfolded polypeptide, or a partially folded state, makes intermolecular contacts that lead to low energy aggregates such as amorphous aggregates or amyloid fibrils. Figure from Hartl & Hayer-Hartl, 2009<sup>7</sup>.

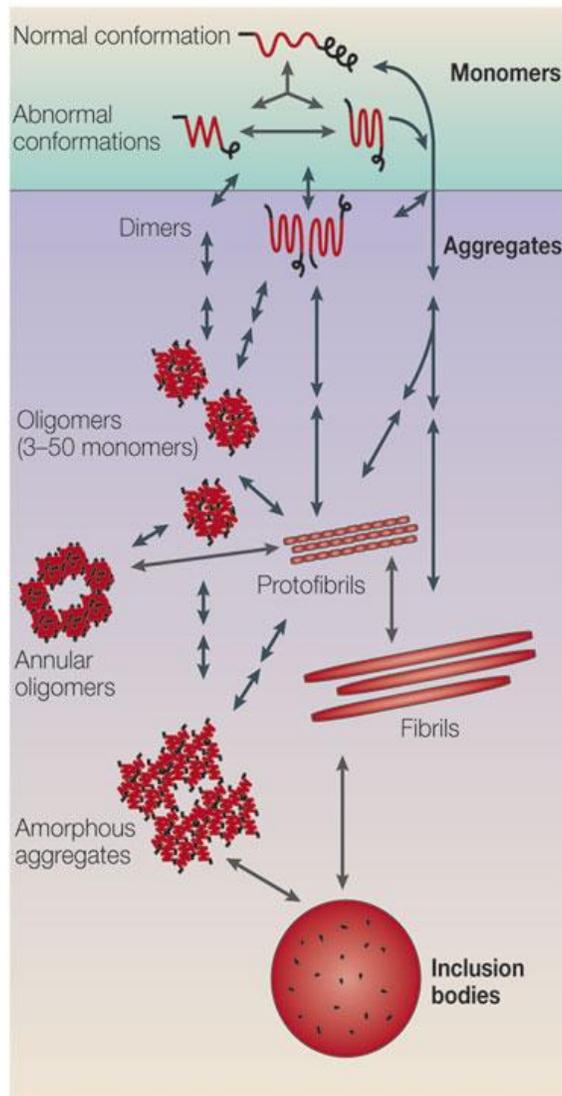
## 1.2 Pathways and Products of Protein Aggregation

Protein aggregation is the general terminology used to describe the process by which proteins form non-native, multimeric complexes of varied conformations. These aggregates can range from small soluble oligomers, to larger amorphous structures, or insoluble, well-structured fibrils. Growing evidence suggests that these morphologies have separate maturation pathways that depend on the environmental conditions, the degree of structure, and the inherent sequence characteristics of the protein<sup>8; 9; 10</sup>. Figure 1.2 shows a simplified scheme of how different pathways may result in varied aggregate types. Alterations in the native monomer can lead to the formation of early oligomers that can transform into a

range of aggregate structures<sup>11</sup>. Figure 1.3 shows another schematic representing the heterogeneity in aggregation pathways and the possible connections between mechanisms of formation of structurally varied aggregates. One of the most well characterized types of aggregates is amyloid, a fibrillar aggregate with extensive beta-sheet structure<sup>12</sup>. Interestingly, amyloid has been shown to form from both disease and non-disease related proteins<sup>13</sup>, and has further been suggested to be a possible conformation that could be reached by all proteins when exposed to harshly destabilizing conditions<sup>12; 14</sup>.



**Figure 1.2 Simplified pathway of protein aggregation from the native state.** Diverse aggregate structures can form through separate pathways as a result of the nature of the alterations to the native protein monomer. Figure from Uversky, 2010<sup>11</sup>.



Copyright © 2005 Nature Publishing Group  
 Nature Reviews | **Molecular Cell Biology**

**Figure 1.3 Interconnected pathways of protein aggregation.** Aggregate structure may form by similar mechanisms resulting in connected pathways of aggregate and the possibility of one type of aggregate to transform into another. Figure from Ross and Poirier, 2005<sup>15</sup>.

Variations in aggregate structures have been observed that depend on the protein in question and the environment in which the aggregates form. A current theory that attempts to explain differences in

aggregate morphology suggests that the variation between the pathway of structured fibril formation and amorphous aggregate formation depends on the degree of folding of the protein involved. Partially unfolded, highly flexible protein conformations result in the formation of structured fibres, whereas slight folding perturbations resulting in minimal protein unfolding tend to form amorphous aggregates<sup>10</sup>. A number of studies have validated this hypothesis. For example, it was shown that concanavalin A (ConA) has two separate aggregation pathways that depend on the pH of the reaction conditions. When the pH was far from the isoelectric point of ConA, large conformational changes in the protein resulted and induced amyloid formation. However at pH values closer to the isoelectric point, ConA experienced only slight conformational changes and formed amorphous aggregates<sup>16</sup>. A previous study on bovine  $\alpha$ -lactalbumin gave similar results<sup>17</sup>. When subjected to low pH, the protein adopted a molten-globule state with a high degree of flexibility, which resulted in the formation of amyloid. However, at neutral pH, partial unfolding due to metal interference resulted in the formation of amorphous aggregates which were attributed to arise from more rigid precursor species<sup>17</sup>. These results suggest that the morphology of aggregates is largely dependent on the degree of unfolding of a protein in solution. However, despite the seemingly viable explanation of the aforementioned studies, these results are not conclusively universal. A recent study on the dialysis-related amyloidosis protein  $\beta$ 2-microglobulin revealed similar fibrillar structures forming in a highly acidic environment (pH 2.5) and at neutral pH despite the extreme difference in time it takes for the aggregates to form<sup>18</sup>. This led to the suggestion that these two pathways for  $\beta$ 2-microglobulin aggregate formation may include similar intermediate structures. These results highlight the complexity of the mechanisms involved in governing the protein aggregation process.

Another interesting study examined the independence of fibrillization and oligomerization pathways<sup>19</sup>. A wide range of small molecule aggregation inhibitors were used to study the amyloid beta (A $\beta$ ) peptide that has been linked to Alzheimer's Disease. The inhibitory compounds could be divided into three classes: Class I inhibited oligomerization but not fibrillization, Class II inhibited both pathways, and Class III was selective for inhibiting fibrillization but not oligomerization. These findings confirmed that the pathway of formation of oligomers is distinct from the formation of fibrils<sup>19</sup>. This is an interesting

result considering contrary theories that suggest the formation of small soluble oligomers may be the nucleating event required for the formation of more structured fibrils<sup>20</sup>. Thus it seems that there are several types of aggregate pathways, some in which soluble oligomers may be an end-point on their own, and others in which oligomers may further convert into fibrillar structures<sup>20; 21</sup>. Considering the degree and variability in the literature regarding the range of potential aggregate structures that can be formed by proteins and the conditions required to produce these structures, no generic pathway or comprehensive view of protein aggregation has been developed. Instead, it appears that a single protein has the potential to form multiple aggregate structures by more than one pathway. These mechanisms may be greatly influenced by the nature of the protein in question and the conditions to which it is exposed.

### **1.3 Modulating Factors in Protein Aggregation**

The factors that modulate protein aggregation are not completely understood. Protein aggregation can occur from both an unfolded or partially folded state<sup>7</sup>. For an unfolded protein, the thermodynamic and kinetic properties that influence the pathway toward proper folding could be altered by the influence of many factors and thereby re-route the protein toward the formation of aggregates<sup>8; 22</sup>. Folded protein could also be directed to aggregate by factors that promote global unfolding or enhanced local fluctuations<sup>7</sup>. There are multiple variables that may play a role in controlling aggregation propensity, some of which are discussed below.

#### **1.3.1 Protein Stability and Structural Dynamics**

As mentioned above (Section 1.2), the extent of unfolding of a protein may play a significant role in its tendency to aggregate and the types of aggregate structures formed<sup>10</sup>. For example, the introduction of mutations that destabilize the conformation can promote fibril formation<sup>23</sup>. This has been demonstrated for several proteins, including lysozyme<sup>24</sup>, transthyretin<sup>25</sup>, immunoglobulin light chain<sup>26</sup>, and for soluble oligomer formation by superoxide dismutase<sup>27</sup>. Destabilization of the native state can result in partially folded intermediates that may expose aggregation-prone regions, resulting in intermolecular associations<sup>9; 23; 28</sup>. In addition to global changes in stability, local structural fluctuations may also promote

aggregation<sup>29</sup>. These fluctuations could result from a mutation that perturbs the dynamics of a particular region of the protein without having a large effect on the overall stability of the protein<sup>29</sup>. If the increased dynamics occurs in a region susceptible to causing aggregation, this could have major implications for the likelihood of aggregation of the whole protein. For these reasons, the stability and structural dynamics of a protein can have a large influence on its aggregation tendencies.

### **1.3.2 Inherent Characteristics of the Polypeptide Sequence**

There are several features of an amino acid sequence that may contribute to its propensity to associate into aggregate structures. The observation that stability alone cannot dictate the aggregation tendency of a protein led to the suggestion that the inherent characteristics of the amino acids within a protein sequence must play a role in dictating the tendency to aggregate. In particular, hydrophobicity, beta sheet propensity, and charge have been suggested to play substantial roles in aggregation propensity<sup>9</sup>. This was shown through mutational studies of an unfolded protein, where the substitution of a single residue determined the aggregation pattern of the entire protein based on the change in hydrophobicity, charge and beta-sheet propensity introduced by the mutation<sup>9</sup>. Interestingly, these features are also important for moderating correct protein folding, and so it has been suggested that similar forces contribute to both processes but that different key residues are involved in forming the initial contacts that drive the formation of either aggregates or natively folded protein<sup>30</sup>. The role of inherent residue properties in protein aggregation seems to depend on short consecutive stretches of amino acids that possess aggregation-prone characteristics. One interesting study inserted an amyloid-prone six amino acid peptide (hexapeptide) into a non-amyloid forming protein and showed that even though there were no global effects on stability, the insertion converted the protein into an amyloidogenic species<sup>31</sup>. This demonstrated that even short stretches (minimum 6 amino acids) were adequate to cause a protein to aggregate. It has further been suggested that the pattern of amino acids within a short peptide sequence is important in dictating the likelihood of amyloid formation<sup>32; 33</sup>. Another important intrinsic determinant is the presence of so-called “gatekeepers,” residues flanking aggregation-prone regions that act to prevent

aggregation, such as charged residues or proline<sup>30</sup>. It is apparent then, that protein aggregation is at least partially controlled by the properties of the specific residues found within a polypeptide sequence.

### 1.3.3 Solution Conditions

Experimental conditions can play an important role in governing protein aggregation *in vitro*. For example, sample temperature, pH, ionic strength, and surface effects have all been suggested to contribute to protein aggregation<sup>8; 22</sup>. These factors are capable of modulating the thermodynamic and kinetic properties of a protein in solution and thus can influence the rate of protein aggregation and the type of structure formed. The thermodynamic stability of a native protein is typically only ~5-15 kcal/mol greater than the unfolded state under physiological conditions<sup>34</sup>. For this reason, temperature can have drastic effects on the equilibrium ratios of folded and unfolded protein. Temperature-induced aggregation has been shown for a number of proteins including actin<sup>35</sup>, lysozyme<sup>36</sup>, beta-lactoglobulin<sup>37</sup>, and amyloid beta peptide<sup>38</sup>. Most of these studies observed aggregation as the temperature increased from well below to well beyond the melting temperature of the protein. The resulting general hypothesis was that the increased population of the partially unfolded protein state near the melting temperature promoted aggregation<sup>22</sup>.

The net charge of a protein can greatly affect proper folding and aggregation propensity<sup>22</sup>. High charge density, resulting from high or low pH, produces repulsive interactions and protein unfolding; however, this may not promote aggregation if the charge densities are high enough to also cause intermolecular repulsion. When the pH is closer to the isoelectric point (pI), the protein will possess both positive and negative charges and aggregation can become energetically favourable<sup>22</sup>. Both the overall charge effects and the role in affecting protein stability are ways in which pH can influence protein aggregation<sup>39</sup>.

Ions may bind to charged residues in a protein and can affect the stability, solubility, and electrostatic interactions of a protein<sup>22</sup>. When multivalent ions bind to a protein, they can increase the stability of the protein by bridging charged residues. Ions can also bind to peptide bonds due to the dipole

formed between the amino and carbonyl groups and can result in the destabilization of the protein's native state. At low concentrations, ions tend to associate with charged amino acids, resulting in charge shielding. This decreases electrostatic interactions between proteins<sup>22</sup>. The implications of ionic strength on aggregation are somewhat unclear. One study showed that, as the ionic strength in solution increased, the aggregate equilibrium shifted from higher oligomeric states to lower oligomeric states<sup>40</sup>. In contrast, a recent study utilized agitation along with high salt concentrations to form large, structured, amyloid aggregates<sup>41</sup>. Neither high ionic strength nor agitation on their own resulted in fibril formation, but the combination of these effects resulted in amyloid. These results seem to coincide with a more general observation that, at low pH, salt is often required to screen charges in order to promote aggregation<sup>42</sup>. Ionic effects also appear to depend on the strength of the ion being used. For example, a study on the  $\alpha$ -lactalbumin protein at a pH below its pI showed that aggregation was promoted by salts in the following order:  $\text{SO}_4^{-2} > \text{H}_2\text{PO}_4^- > \text{Cl}^-$ <sup>42</sup>. This was suggested to be due to preferential protein binding by  $\text{SO}_4^{-2}$  to block electrostatic repulsions between proteins, whereas  $\text{Cl}^-$  favoured hydration of  $\alpha$ -lactalbumin. The effects of these salts are consistent with the Hofmeister series, which ranks salts on their ability to stabilize or destabilize proteins. Salts that increase protein stability will decrease protein solubility and thus increase aggregation<sup>43</sup>. Similar results were observed for the yeast prion protein, where salts with increasing kosmotropic strength increased amyloid polymerization<sup>44</sup>. There are still some inconsistencies in the role of ionic strength in protein aggregation, however, as contrary results to those described above as to the role of kosmotropes have also been observed<sup>45</sup>. This is evidence of variable effects of salt on protein aggregation depending on the protein. It seems possible that salt could play different roles in different systems depending on the physicochemical properties of the protein involved, and the strength of the ion in question.

The type and size of the surface area a protein is exposed to can play an interesting role in the misfolding or aggregation of a protein. It has been suggested that the recruitment of proteins to a surface can cause local increases in concentration and affect the conformation of interacting proteins, resulting in enhanced aggregation. In particular, the presence of hydrophobic or charged surfaces could contribute to

the opening up of tightly folded native protein and thus increases the susceptibility of the protein to forming intermolecular interactions<sup>46</sup>. This is especially relevant when considering the biological surfaces within the cell, such as lipid membranes, which may catalyze protein aggregation and lead to cellular toxicity<sup>47; 48</sup>. A study involving the growth of insulin fibrils revealed that nucleation occurred quicker at hydrophobic polystyrene surfaces than in bulk solution, suggesting that aggregation mechanisms may depend on surface interactions<sup>49</sup>. Surface-catalyzed aggregation studies were also performed on SMA, a recombinant amyloidogenic light chain variable domain<sup>48</sup>. In these studies, fibrils formed readily on the negatively charged surface of mica but were not observed when the surface was modified to be non-polar or positively charged. This was explained by the presence of weak electrostatic interactions that recruit the SMA protein to the mica surface. Since the interactions were not very strong, it allowed for lateral mobility. Once the protein was recruited to the surface, the enhanced proximity of the protein monomers to each other allowed for hydrophobic interactions between proteins to favour conformational transitions that resulted in fibril growth. Two distinct mechanisms were observed for SMA aggregation on the mica surface, one in which unfolded monomers formed amorphous aggregates on the mica surface that further rearranged into fibrils branching from the amorphous structure. In the second, structured oligomers were recruited to the mica surface and associated to form twisted fibrils<sup>48</sup>. More recent studies employed  $\beta$ -2 microglobulin as a model for testing nucleation catalysis by nanoparticles. It was demonstrated that the surface area provided by the nanoparticles increased the nucleation rate and decreased lag-times in a manner dependent on the exposure and nature of the particle surface<sup>47</sup>. These studies indicate that both the degree of surface exposure and the nature of the surface in which the protein is exposed to can affect the aggregation pathway and the rate and the morphology of the aggregates being formed.

It is evident that the experimental conditions can drastically influence protein folding, misfolding, and aggregation of a protein. The conditions experienced by a protein *in vivo* are very different than that of a protein *in vitro*, and the variations that can occur in both, whether pH, temperature, ionic strength, surface exposure, or others, have the ability to alter the kinetic and thermodynamic properties of a protein and influence aggregation pathways. This is an important consideration when investigating the molecular

mechanisms controlling the aggregation process, and further understanding of the roles of these factors could serve to provide valuable information on the variables that govern protein aggregation.

#### **1.4 Protein Aggregation and Disease**

Protein aggregation has been linked to a large number of diseases including Alzheimer's disease, Creutzfeldt-Jacob disease, Type II diabetes, Parkinson's disease, Huntington's disease, and amyotrophic lateral sclerosis (ALS)<sup>29; 50; 51</sup>. The majority of aggregation-linked diseases can be divided into three major groups: neurodegenerative conditions, in which aggregation occurs in the brain, non-neuropathic, localised amyloidoses where aggregation occurs in a one specific tissue other than the brain, such as ALS, and non-neuropathic systemic amyloidoses where aggregates are present in multiple tissue types<sup>12</sup>. A general trend within protein aggregation disorders seems to be an increase in misfolded or unfolded proteins<sup>3</sup>. Protein misfolding has the potential to lead to a toxic gain of function where misfolded protein forms aggregates that can eventually lead to cell death.

The pathological mechanisms of protein aggregates are not well understood. Aggregates can possess a wide range of structural morphologies and growing evidence suggests that small, soluble oligomers are more toxic than large insoluble aggregates such as amyloid<sup>52</sup>. Patient data analysis for various neurodegenerative diseases, including Alzheimer's and Parkinson's, revealed that brain cells with large inclusion bodies were often healthier than the surrounding cells without noticeable fibril formation. Additionally, in some animal models of neurodegenerative diseases, symptoms appeared before insoluble aggregate formation occurred. Furthermore, clinical studies have confirmed the presence of insoluble inclusion bodies without disease symptoms<sup>53</sup>. This evidence suggests that the early stages in the aggregate formation pathway, such as protofibrils, small intermediate oligomers, or aggregate fragments and by-products, are the key instigators of neurodegeneration<sup>52; 53</sup>.

It is expected that there are common features between toxic aggregate species of different proteins considering the widespread occurrence of aggregation diseases. Work with antibodies has produced intriguing discoveries regarding this prospect by first identifying an antibody capable of

recognizing common structures of amyloids formed from a range of different proteins<sup>54</sup>, then by the isolation of an antibody that recognized pre-fibrillar amyloid oligomers of various proteins but not the native protein or the mature fibril<sup>55</sup>. It is thought that toxic oligomers may have common morphologies that can interact with similar cellular targets. Suggested mechanisms include: destabilization of membranes, interaction with neuronal receptors to initiate apoptosis, impairment of calcium currents as a result of membrane disruption, prevention of lysosome maturation due to accumulation of oligomers, autophagy dysfunction, oxidative damage, disruption of protein homeostasis and proteasome inhibition<sup>11</sup>. It is important to elucidate the common properties that may be involved in governing the toxicity of aggregates. A recent study from Campioni *et al.* compared the toxicity of two different types of soluble oligomers formed in different solution conditions by HypF-N<sup>56</sup>. The two conditions were as follows: A) 50 mM acetate buffer, 12% (v/v) trifluoroethanol (TFE), 2 mM DTT, pH 5.5 and B) 20 mM trifluoroacetic acid (TFA), 330 mM NaCl, pH 1.7. The oligomers formed in the two conditions were indistinguishable by atomic force microscopy (AFM) and Thioflavin T (ThT) fluorescence properties, but had very different toxic properties. When human neuroblastoma cells were exposed to each type of aggregate and cell viability was measured, the oligomers formed in condition A caused a significant decrease in cell viability, similar to that observed upon exposure to amyloid beta oligomers, a peptide linked to Alzheimer's disease. However, there was no effect on cell viability when exposed to the oligomers formed in condition B. Further investigations, using several techniques, into the structures of the aggregates formed under these conditions showed that the aggregates formed in condition B were tightly packed with buried hydrophobic regions, while those formed in condition A were less tightly packed and had exposed hydrophobic regions. In addition, microscopic imaging demonstrated that while both types of aggregate structures associated with the cells membrane, only those formed in condition A were able to penetrate the cell<sup>56</sup>. This study suggests that flexibility and exposed hydrophobic regions may contribute to the toxicity of oligomeric aggregates.

Another recent study compared the *in vivo* toxic effects of  $\alpha$ -synuclein oligomers and fibrils<sup>57</sup>. Mutants were constructed that favoured either the formation of structured fibrils or soluble, ring-like

oligomers with a hydrodynamic radius of approximately 100 nm. A rat model of synucleinopathies, involving the injection of a lentivirus vector containing the  $\alpha$ -synuclein variants into the rat's brain tissue, was used to assess toxicity. The mutants that favoured oligomer formation resulted in increased dopaminergic loss and neuronal death compared to wild-type and mutants favouring fibril formation. The neuronal cell death was coupled with increased influx of calcium, indicating damage to the cellular membrane. Further immunoblot studies demonstrated that oligomer-prone mutants bound strongly to membranes<sup>57</sup>. These studies confirm the role of oligomers in mediating toxicity and suggest that membrane interactions could be a key mechanism for aggregation-governed disease pathogenesis.

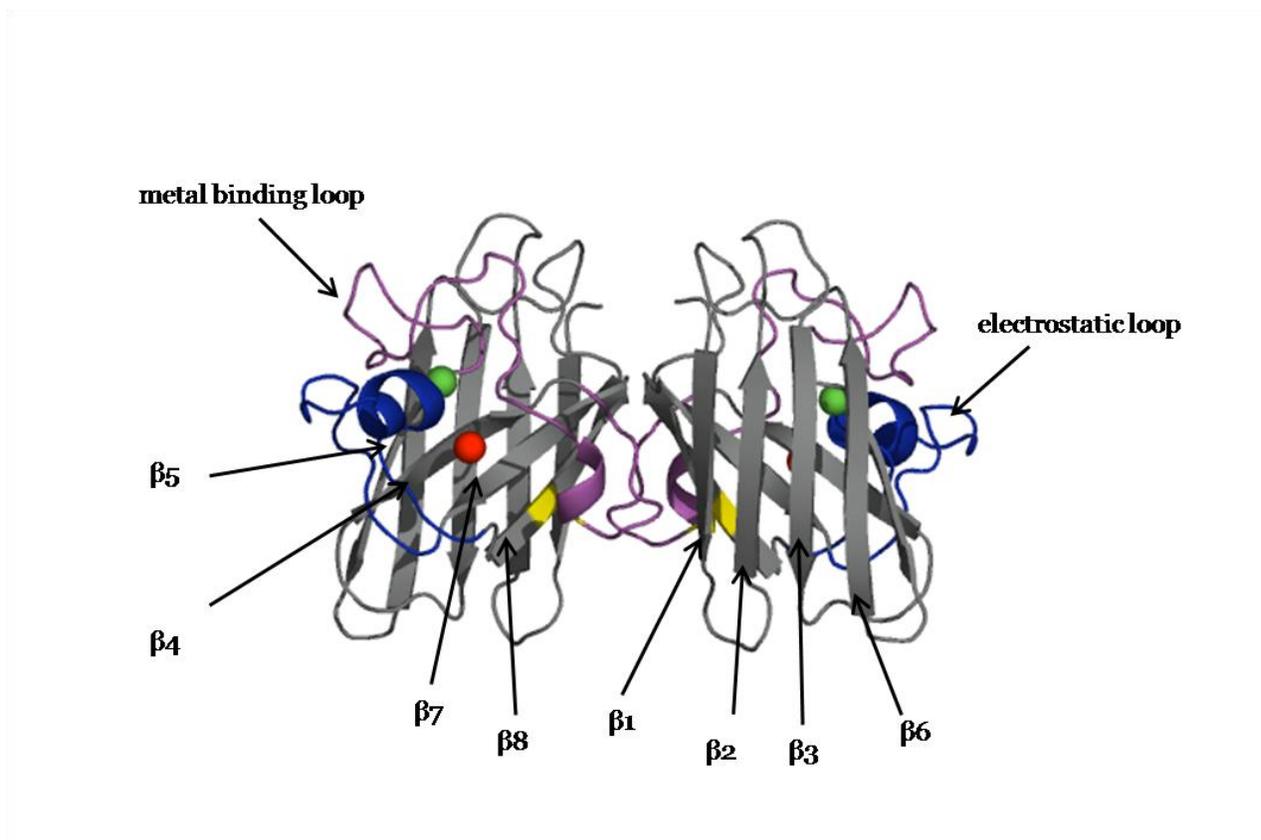
### **1.5 Amyotrophic Lateral Sclerosis**

Amyotrophic lateral sclerosis (ALS) is a debilitating neurodegenerative disease in which motor neuron degeneration results in muscle deterioration, the exaggeration of tendon reflexes, and spasticity. This commonly leads to respiratory failure and death within 2-5 years of disease onset. It is the most common adult onset motor degenerative disease. 90% of ALS cases are sporadic and 10% are familial. Within the 10% of cases that have hereditary links, about 20% of these patients possess mutations in the Cu, Zn Superoxide Dismutase (SOD1) protein, making SOD1 mutations the most common known causative agent of ALS. There is currently no cure for the disease, despite vigorous research efforts in the past decades<sup>58</sup>. ALS remains a fairly misunderstood disease that is likely triggered by a complex set of factors.

### **1.6 Cu,Zn Superoxide Dismutase and ALS**

Cu, Zn Superoxide Dismutase (SOD1) is a human homodimeric enzyme composed of two 153 amino acid monomers. SOD1 functions as a catalyst in the conversion of superoxide to hydrogen peroxide and oxygen, and for this reason it is an important enzyme in the control of reactive oxygen species (ROS) within a cell. Each monomer takes on a Greek key beta barrel conformation composed of eight antiparallel beta strands connected by a series of loops that bind one copper and one zinc ion each. Two of the loops are specifically involved in the binding of metals and are termed the "metal-binding

loop” and the “electrostatic loop” (see Figure 1.4). In addition to metallation, other post-translational modifications include the formation of a disulphide bond between cysteine 57 and cysteine 146, and dimerization to form the mature homodimeric protein<sup>59; 60</sup>. SOD1 is expressed in all tissues and is most concentrated in the cytoplasm of the cell. It is found in higher concentrations in the central nervous system (i.e. the brain and spinal cord) than in any other tissue<sup>51</sup>. More than 150 different SOD1 mutations have been characterized as contributing factors to the development of ALS (ALS Society of Canada, 2010). The mechanisms of SOD1 toxicity are still not completely understood, even more than 15 years after the discovery of the causative effect of mutant SOD1<sup>61</sup>. It was first hypothesized that mutations could result in a loss of function leading to increased oxidative stress in the motor neurons where SOD1 is particularly important in combating reactive oxygen species (ROS). However, mice models lacking the SOD1 gene were not found to develop symptoms of ALS, while mice models that had their endogenous SOD1 gene as well as a human ALS mutated SOD1 gene did develop ALS. This led to the understanding that SOD1 pathogenicity is dependent on a gain of toxic function rather than a loss of protective function<sup>51</sup>. The most accepted hypothesis is that the gain of toxic function is due to the formation of damaging protein aggregates<sup>62</sup>. It is assumed that SOD1 mutations influence physicochemical properties such as stability, hydrophobicity, aggregation propensity, susceptibility to post-translational modifications, loss of metals, and aberrant chemistry<sup>61</sup>. It has been suggested that the toxic effects of SOD1 mutants are dominantly a result of decreased stability and increased aggregation propensity<sup>61</sup>.



**Figure 1.4 Structural components of dimeric holo SOD1.** Beta strands 1-8 are labelled. Each monomer is composed of eight beta strands that together form a beta barrel. The metal binding loop shown in purple is also important to the dimer interface. The electrostatic loop is shown in blue. Cysteines 57 and 146 form an intra-molecular disulphide bond and are shown in yellow. Zinc is shown in green and copper is shown in red. This figure was produced using PyMol, with PDB coordinates taken from 1SOS<sup>63</sup>.

### 1.6.1 ALS Disease Durations are Specific for SOD1 Mutations

An interesting feature of the SOD1 mutations that lead to ALS is that they result in mutant-dependent disease durations. It was shown by Wang *et al.* in 2008 that the differences in disease durations for each mutant are statistically significant<sup>61</sup>. Table 1.1 lists data for several of the most common mutations, including the number of patients and the average disease duration associated with the mutation. It has been speculated that the tendency of each mutant to form aggregates may be what determines disease duration<sup>61</sup>.

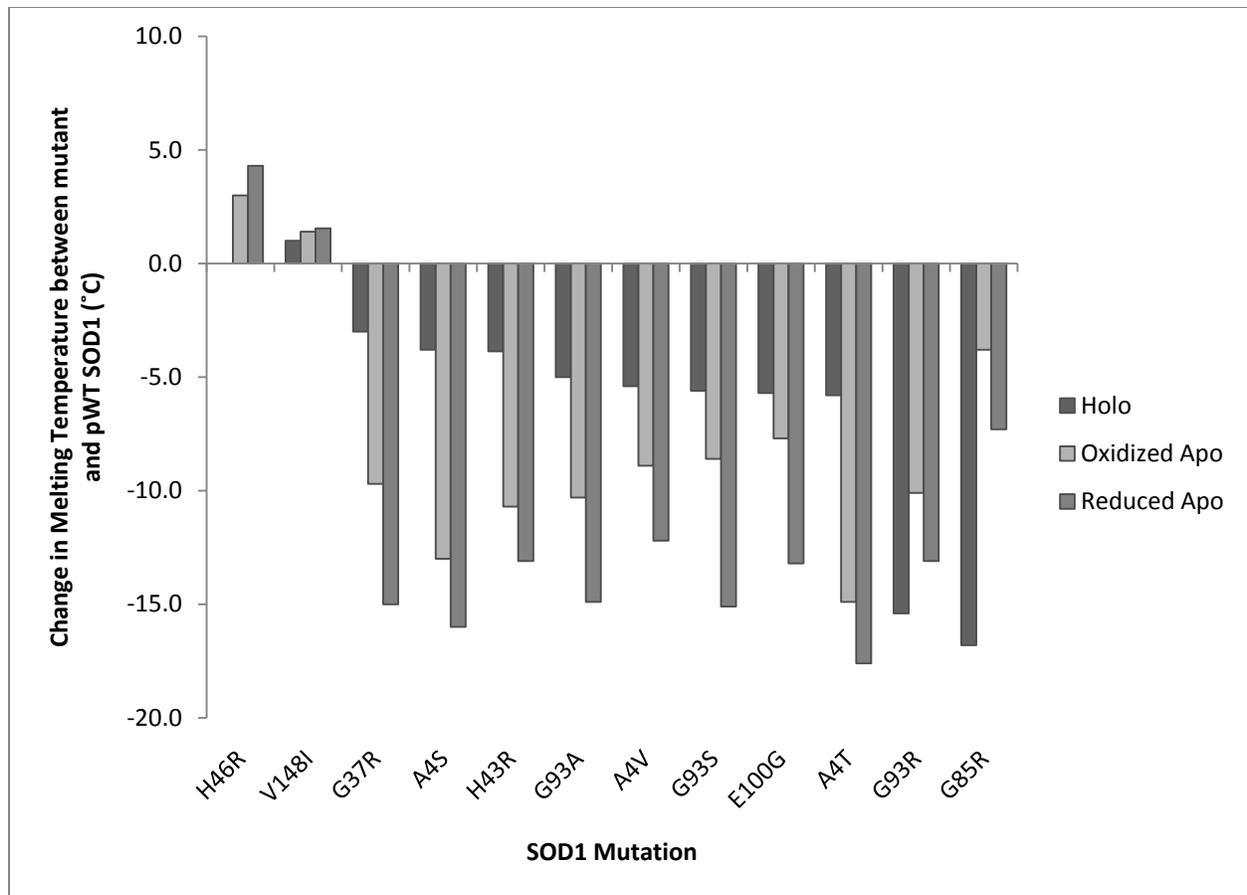
Table 1.1 Characteristic ALS disease durations are associated with SOD1 mutations.

<b>Mutation</b>	<b>Number of Patients</b>	<b>Disease Duration</b>
<b>A4T</b>	21	1.5
<b>A4V</b>	205	1.2
<b>G37R</b>	27	17
<b>L38V</b>	22	2.4
<b>G41D</b>	15	14.1
<b>G41S</b>	16	1
<b>H43R</b>	12	1.8
<b>H46R</b>	49	17.6
<b>L84F</b>	18	5.8
<b>L84V</b>	10	3.2
<b>G85R</b>	11	6
<b>D90A</b>	15	8
<b>G93A</b>	16	3.1
<b>G93C</b>	27	12.1
<b>G93S</b>	11	8
<b>G93R</b>	4	5.3
<b>E100G</b>	50	4.7
<b>D101N</b>	17	2.3
<b>S105L</b>	7	3.5
<b>I113T</b>	38	4.3
<b>L144F</b>	15	11.8
<b>V148G</b>	11	2.1
<b>V148I</b>	5	1.7
<b>I149T</b>	15	2.7

### 1.6.2 The Importance of Reduced Apo SOD1

Superoxide Dismutase goes through a series of post-translational modifications in order to achieve its native, mature state. These include metal-binding, disulphide formation and dimerization, resulting in at least 44 permutations that the polypeptide could adopt depending on its disulphide status, metal binding and multimer formation<sup>64</sup>. This begs the question as to whether any of these states are particularly vulnerable to disruption by mutation and could therefore be most relevant to ALS pathology.

Extensive study of the destabilizing effect of SOD1 mutations on both the holo and apo forms of the protein reveals that changes are more pronounced in the apo. *In vivo* and *in vitro* studies by Furukawa *et al.*, 2008, showed the importance of disulphide reduction in aggregate formation by comparing the effects of holo, apo-oxidized, and apo-reduced SOD1 in various mutant forms<sup>65</sup>. Specifically, they showed that the addition of zinc to apo-oxidized conditions reduced the amount of insoluble protein aggregates produced when incubated with agitation overnight. Conversely, when incubation occurred with the addition of a reducing agent, the fraction of insoluble aggregates formed increased. These results were supported by *in vivo* experiments that showed that the total fraction of aggregates formed in mouse neuroblastoma cell lines could be reduced by the over-expression of CCS, the copper-chaperone responsible for the addition of copper to SOD1 and speculated to be involved with disulphide bond formation. This study promoted the theory that fALS mutations have the greatest destabilization effects on the de-metallated, disulphide-reduced form of SOD1<sup>65</sup>. Additional *in vitro* experiments performed in the Meiering lab support these findings by showing that the changes in thermodynamic stability between wild-type and mutant SOD1s are generally largest in the reduced apo form of the protein. Figure 1.5 summarizes these data by showing the melting temperatures for a number of mutants studied in the Meiering lab. Interestingly, only ten out of the twelve mutants shown in this figure demonstrate protein destabilization. Two, however, cause an increase in thermal stability, which may implicate additional factors beyond stability as disease-causing properties of these mutants. For the ten mutants that destabilize the protein, eight have the most pronounced destabilization in the reduced apo form. These studies promote reduced apo SOD1 as an important form of the protein to study in order to determine whether the decreased stability of reduced apo mutants will also lead to an increased propensity to aggregate.

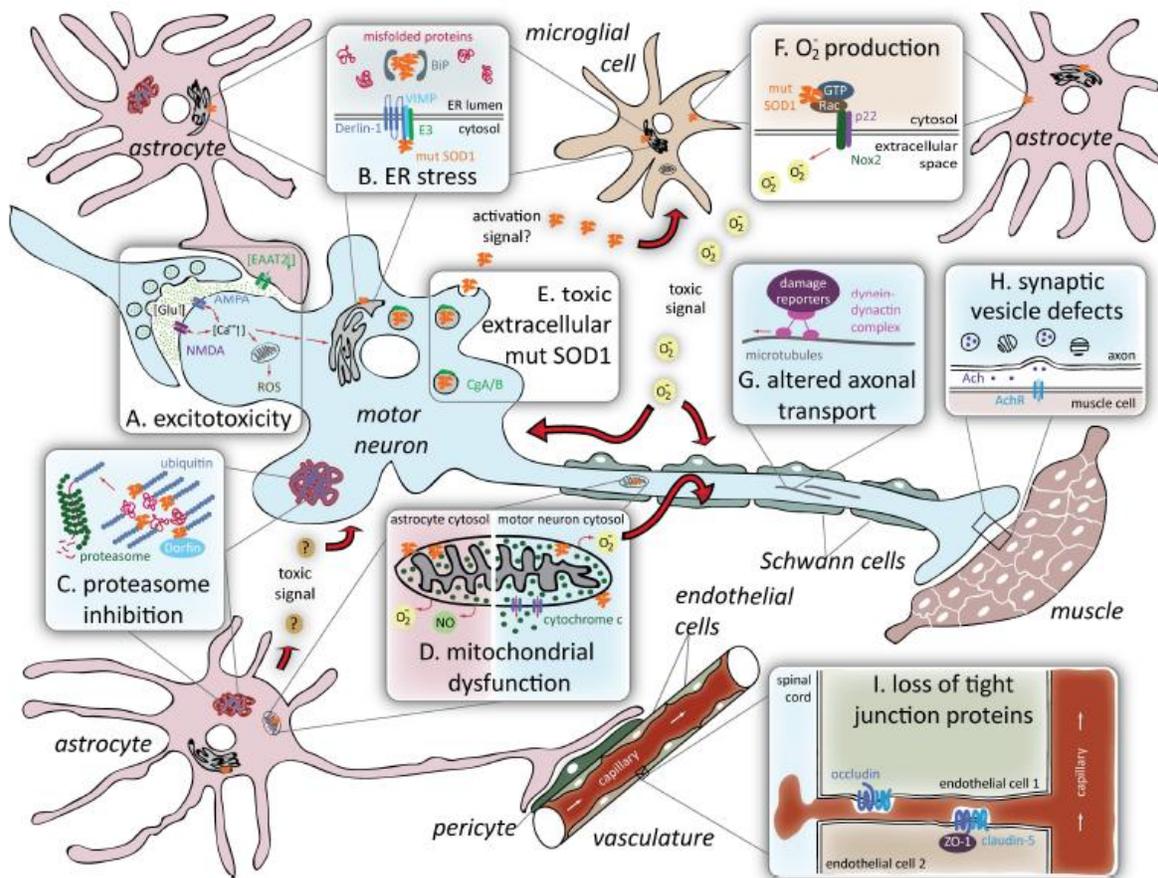


**Figure 1.5** The change in melting temperature induced by SOD1 mutations in different forms of the protein. The change in melting temperature was measured by DSC. All data was measured in the Meiering lab. In general, the largest changes in protein stability occur as a result of mutations in the reduced apo state.

### 1.6.3 SOD1 Aggregation and Toxicity in ALS

There is much debate surrounding the possible mechanisms by which mutant SOD1 exerts causes ALS. A review by Ilieva *et al.* in 2009 summarized nine potential mechanisms including: A) Glutamate excitotoxicity caused by excessive firing of motor neurons due to overstimulation by glutamate, resultant from the potential oxidative effects of mutant SOD1. B) Endoplasmic reticulum (ER) stress as a result of mutant SOD1 aggregates binding to a chaperone that regulates ER stress transducers or by inhibition of ER-associated degradation of improperly folded proteins. C) Proteasome inhibition due to overload with

misfolded protein. D) Mitochondrial dysfunction from association of mutant SOD1 with the mitochondrial membrane. E) Interaction of toxic extracellular mutant SOD1 with neurosecretory vesicles resulting in secretion and activation of microglia immune cells that drive neuronal death. F) Extracellular superoxide production from the interaction of mutant SOD1 with transcription factor Nox2. G) Altered axonal transport through mutant SOD1 interference with axonal cytoskeletal organization. H) Synaptic vesicle defects. I) Loss of tight junction proteins that maintain the blood-spinal cord barrier, because of damage to the vasculature by mutant SOD1 leading to leakage of toxic products. Figure 1.6 depicts these potential roles of SOD1. It is most likely that a combination of these possible mechanisms results in disease progression.



**Figure 1.6 Proposed mechanisms for mutant SOD1 toxicity in ALS.** Details described in text. Figure from Ilieva, 2009<sup>66</sup>.

It is important to further investigate the types of misfolded and aggregated protein that may be responsible for the disease mechanism in SOD1-linked ALS. As stated above in Section 1.4, soluble, pre-fibrillar oligomers are generally thought to be the cytotoxic species in aggregation-related diseases<sup>53</sup>. Patient data and mouse models have provided information as to the SOD1 species important in pathogenesis. Recent evidence based on the analysis of patient tissues using antibodies has suggested that protein inclusions contain misfolded SOD1 aggregates but not in the form of amyloid<sup>67</sup>. However, the role of insoluble inclusions is not fully understood, as not all transgenic ALS mice models display evidence of inclusion formation, even when ALS symptoms are observed<sup>51</sup>. Another study that confirms the role of oligomeric aggregates as opposed to larger insoluble inclusions in producing disease characteristics has found that small, soluble monomers, dimers, and trimers of the reduced apo form of SOD1 are present from the pre-symptomatic to final end stage of disease in murine models, suggesting that oligomers of SOD1 are a common cytotoxic species<sup>68</sup>. Although many *in vitro* studies have demonstrated amyloid formation by SOD1 and its variants<sup>65; 69; 70</sup>, the relevance to disease is questionable, both due to the lack of observance of amyloid in patient tissues<sup>67</sup> and because of the increasing evidence of soluble oligomer toxicity<sup>52; 53</sup>. Additionally, several studies have demonstrated the formation of non-amyloid aggregates by mutant SOD1<sup>27; 71</sup>. In general, there is considerable discrepancy between the formation of amyloid by SOD1 mutants and their toxicity. Potential aggregation pathways and toxic properties of SOD1 are discussed below.

Several specific mechanisms of SOD1 mutant aggregation and its potential toxic properties have been suggested. Much debate has surrounded the importance of intermolecular disulphide bonds to the mechanisms of SOD aggregate formation. There have been studies supporting both sides of the debate leading to general controversy as to the role of the free cysteines at positions 6 and 111, and the cysteines involved in the intramolecular disulphide at positions 57 and 146. Banci *et al.* 2007 have demonstrated that these two residues are necessary for the formation of amyloid at physiologically relevant conditions (37°C, pH 7.0 and a protein concentration of 100µM *in vitro*). They showed that when these residues are mutated to alanine and serine, respectively, Thioflavin T (ThT) binding, for the purpose of detecting

characteristic amyloid structure, is not observed for the apo form of SOD. However in the apo wild-type form of SOD, with two free cysteines, ThT fluorescence is observed within the first 25 hours<sup>72</sup>. In contrast, it has been shown that when all four cysteines in SOD1 are mutated aggregation is still observed in cultured cells<sup>73</sup>. Moreover, in mutants possessing free cysteines, inter-molecular disulphide bonds were shown to form only late in the progression of the disease in SOD fALS mice models<sup>74</sup>. Finally, Jonsson *et al.*, 2007, demonstrated that aggregates found in fALS mice models comprised SOD1 with reduced cysteine residues<sup>75</sup>. Thus, although there is still debate surrounding the importance of free cysteines, it seems reasonable to conclude that cysteines may play a role in aggregation formation at some point along the aggregate pathway but that the formation of disease-inducing aggregates is not dependent on intermolecular disulphide bond formation.

There are many other theories regarding possible avenues for SOD1 aggregation and toxicity. One such theory suggests that mutant SOD1 results in the formation of toxic protein aggregates by decreasing the repulsive net charge of the protein<sup>76</sup>. Conversely, it has been suggested that a decrease in net charge could be protective and result in a lengthened disease duration<sup>61</sup>. Other mechanistic ideas include mutant SOD1 destabilization causing interference in the interaction of SOD1 with the copper chaperone (CCS) and increasing the susceptibility of the protein to metal loss<sup>77</sup>, intra-molecular disulphide bond reduction resulting in local unfolding and fibril formation<sup>69</sup>, and increased monomerization resulting in exposure of the hydrophobic dimer interface<sup>78</sup>. The diversity in potential mechanisms suggests that there may be multiple pathways in which SOD1 aggregates and mediates disease.

#### **1.6.4 Pseudo Wild-type Superoxide Dismutase**

For the purpose of all SOD1 studies, a pseudo wild-type construct was used. In pseudo wild-type cysteines 6 and 111 are replaced with alanine and serine, respectively. This construct is often referred to as pWT SOD. The purpose of this pseudo wild-type background is two-fold. First, the removal of the two

free cysteines increases protein yield, and second it provides a form of the protein that can be analyzed thermodynamically. The wild-type protein cannot be reversibly unfolded for thermal unfolding studies, thus limiting the ability to calculate its thermodynamic properties. For this reason, the pWT SOD is commonly used for thermodynamic analyses. Studies have shown that pWT SOD behaves very similarly to wtSOD<sup>27; 63; 79</sup>, and this is a well established construct used within the Meiering and other labs. The majority of the former work done in the Meiering lab has been completed on pWT and mutants in the WT background. Thus, for consistency, the pWT background will be used for all experiments completed in this thesis.

## **1.7 Importance of Further Studies on the Mechanisms of Protein Aggregation**

Protein folding, misfolding and aggregation are fundamental to the understanding of molecular biology. Interestingly, the regions of a protein most important for protein folding are also the most prone to be involved in adverse aggregate formation<sup>80</sup>. Recent physicochemical studies have given some insight into protein folding and aggregation control, including the role of electrostatic and hydrophobic forces<sup>80</sup>; however mechanistic details remain elusive. Further analysis into the detailed mechanisms governing aggregation is required. Despite the extensive research in the past decade focussed on protein aggregation, there are still many unanswered questions. It has been suggested that protein maturation involves a crossroad in which a protein can fold properly, misfold, or not fold at all<sup>81</sup>. The factors governing the path a protein will take are not well understood and could have dramatic effects on the way we view the progression from polypeptide to protein and the potential for forming protein aggregates.

Elucidating the principles and mechanisms of aggregate formation is of importance to many fields, including: 1) contributions to further understanding in molecular biology, such as fundamental protein assembly and factors leading to protein misfolding and the intermolecular assembly into aggregates, 2) understanding the aggregation pathways that are attributed to a growing number of diseases, 3) application for protein-based materials and drugs including novel synthesis techniques and design templates<sup>41</sup>. Many studies have taken initial steps in understanding the variety of factors that may

have an impact on protein folding and aggregation pathways; however there is still a great need for significantly more work to be done outside of individual context-specific experiments to systematically address protein aggregation complexities on a wider scale. Thus, another major challenge within this field is to test the current understanding of the principles controlling aggregation and design further experiments that address the limits of our current understanding.

## **1.8 Research Objectives and Outline**

The Meiering lab is focused on developing a deeper understanding of the physical and chemical properties of ALS-linked SOD1 mutants and how these properties influence their aggregation pathway. The focus of this thesis is the investigation of the underlying principles that control the protein aggregation process. This is achieved by examining our current understanding of protein aggregation by systematically testing a series of protein aggregation prediction algorithms, which will be described in Chapter 2. This chapter will thoroughly explain the principles of nine different prediction algorithms and compare the output of the algorithms for a series of mutations of three different proteins, including Superoxide Dismutase. Chapter 3 extends this investigation by comparing predicted aggregation with observed aggregation and discusses our current understanding of the mechanisms of aggregation. This chapter also takes a detailed look at SOD1 correlations between observed aggregation, predicted aggregation, disease duration, and stability, and discusses the role of SOD1 aggregation in dictating ALS disease characteristics. Chapter 4 focuses on experimental data that investigates aggregation in different solution conditions as a method to further develop the role of ionic interactions and protein stability in contributing to protein aggregation. A range of techniques are used to investigate the differences in aggregation rate and aggregate structure for the aggregation-prone SOD1 mutant A4V when exposed to various salt types and concentrations. The overall objectives of this study are to increase understanding of the mechanisms of SOD1 aggregation and the role of SOD1 aggregation in mediating disease, and to take a broader look at protein aggregation - what we know and don't know about the principles involved in modulating these processes.

## Chapter 2

# Predicting Protein Aggregation

### 2.1 Introduction

Protein aggregation is the process by which intermolecular contacts between two or more molecules, generally of the same protein, result in the formation of aberrant, non-native structures<sup>3</sup>. The presence of aggregate structures in patients with a range of different neurodegenerative diseases has defined protein aggregation as a hallmark in the pathology of neurodegeneration<sup>82</sup>. Understanding protein aggregation is also key for aggregation prevention in pharmaceutical formulations of protein drugs<sup>8</sup>, and has applications for protein-based materials and drugs, including novel synthesis techniques and design templates for synthetic amyloid-like products<sup>83; 84</sup>. For these reasons, it is important to elucidate the factors that control aggregation processes. Currently, there is limited understanding of the detailed molecular mechanisms by which aggregate structures form, and the contributing roles of the driving factors of the rate of formation and morphology of the aggregate products. Over the past several years significant contributions to the understanding of protein aggregation have been made by the development of algorithms designed to predict the propensity of a polypeptide sequence to aggregate<sup>9; 33; 61; 85; 86; 87; 88; 89</sup>. These algorithms are based on experimental data sets and utilize properties of the protein sequence, such as patterns in hydrophobicity or probability of involvement in hydrogen bonding, as contributing variables that dictate the overall propensity of a protein to form aggregates and allow for the identification of aggregation-prone regions of a protein sequence. Each algorithm was developed based on different experimental data and tested on varying data sets. A summary table of the variables involved in each algorithm is given in Table 2.1.

With a continually increasing number of prediction algorithms it is important to test on a wider scale the ability of these algorithms to predict aggregation beyond the specific conditions in which they were developed. This will serve to further the understanding of the underlying principles of protein aggregation and to elucidate whether the variables involved in one confined type of aggregation may have

similarities to the principles governing other types of aggregation. By predicting protein aggregation and then testing the predictions, we can improve our understanding of the processes involved in this complex phenomenon and identify patterns that are involved in governing the overall aggregation process. This is important for identifying ways to interfere with protein aggregation which could guide the development of therapeutics for neurodegenerative diseases<sup>90</sup>.

**Table 2.1 Summary of the variables involved in nine different aggregation prediction algorithms.**

Algorithm Features	Chiti-Dobson <sup>9</sup>	Wang-Agar <sup>61</sup>	Zygg <sup>85</sup>	Ztox <sup>85</sup>	Fold Amyloid <sup>88</sup>	PASTA <sup>86</sup>	TANGO <sup>87</sup>	Waltz <sup>33</sup>	Profile 3D <sup>89</sup>
Sequence order-independent	x	x							
Sequence order-Dependent <sup>a</sup>			x	x	x	x	x	x	x
Aggregate structure-dependent <sup>b</sup>								x	x
Developed from amyloid aggregate datasets	x	x	x		x	x	x	x	x
Developed from non-amyloid aggregate datasets				x	x		x		
Sequence Hydrophobicity	x	x	x	x				x	
Sequence Charge	x	x	x	x				x	
Sequence Beta sheet propensity	x	x	x	x			x	x	
Potential for Hydrogen bonding and molecular contacts					x				
Energy minimization of self association						x			
Similarity to amyloid forming hexapeptides								x	x

<sup>a</sup>Sequence-order dependent means that the propensity of a certain residue to contribute to protein aggregation is based not only on the individual residue, but also on the properties of the surrounding residues in the protein sequence.

<sup>b</sup>Aggregate structure-dependent refers to the ability of portions of the polypeptide sequence to map onto known amyloid structures, and does not mean that the native structure of the protein in question is accounted for by the algorithm.

In this chapter, nine different prediction algorithms are used to predict the aggregation propensity of wild type and mutant sequences for three different proteins: human Cu, Zn Superoxide Dismutase (SOD1), human Acylphosphatase (AcP), and amyloid beta peptide ( $A\beta_{42}$ ). SOD1 was introduced in Section 1.6 and is particularly relevant because of its link to ALS. AcP is not a disease-linked protein, but has been extensively used in mutagenesis-based aggregation studies<sup>91; 92</sup> and provides an excellent dataset for comparison.  $A\beta_{42}$  is a short 42 amino acid peptide sequence that is implicated in the pathogenesis of Alzheimer's Disease<sup>93</sup>. These three proteins have diverse native states. SOD1 is a natively folded, 153 amino acid homodimer<sup>59</sup>. AcP is a relatively small, natively folded, 98 amino acid monomer<sup>94</sup>. However, in the conditions in which the aggregation of AcP mutants has been measured, the protein populated the unfolded state<sup>92</sup>. This is discussed further in Section 3.3.1.2.  $A\beta_{42}$  is a 42 amino acid peptide fragment without stable tertiary structure<sup>95</sup>. In organic solvents, or membrane-mimicking solution, the  $A\beta_{42}$  structure is composed of two alpha helices, while in water or aqueous solvents it appears to adopt beta sheet morphology. Structures for  $A\beta_{42}$  vary widely with differences in pH, concentration, or incubation time, indicating that it is generally unstable<sup>95</sup>. Performing prediction analysis on all three of these proteins and a series of mutants will allow for a thorough comparison of the performance of the algorithms for different types of proteins. In the next subsections nine different aggregation prediction techniques will be described. These algorithms were chosen based on the popularity of use and diversity in the strategies for algorithm design.

### 2.1.1 Chiti-Dobson Equation

The Chiti-Dobson method for predicting protein aggregation was published in 2003<sup>9</sup>. Its fundamental idea is that stability alone does not explain the pathogenic effects of several disease-linked proteins, and so there must be additional inherent factors, such as physicochemical properties and secondary structure propensities, in the amino acid sequence of a protein that influence its tendency to form protein aggregates. Based on mutational studies of human acylphosphatase (AcP) under conditions in which it was unfolded (pH 5.5, 25°C, 25% Trifluoroethanol (TFE))<sup>92</sup>, three main contributing variables

significant to amyloid formation by unfolded protein were suggested: hydrophobicity, charge, and beta-sheet propensity. Thioflavin T (ThT) fluorescence (a technique for quantifying amyloid formation, see Section 3.1.1) for a series of AcP mutants was correlated with each property separately to obtain empirical weighting coefficients, resulting in the following equation<sup>9</sup>:

$$\ln(v_{\text{mut}}/v_{\text{WT}}) = 0.633\Delta\text{Hydr.} + 0.198(\Delta\Delta G_{\text{coil-}\alpha} + \Delta\Delta G_{\text{coil-}\beta}) - 0.491\Delta\text{Charge} \quad (1)$$

$\ln(v_{\text{mut}}/v_{\text{WT}})$  represents the natural logarithm of the ratio of the rate of aggregation of mutant compared to wild type protein. The equation describes how mutation-dependent changes in hydrophobicity ( $\Delta\text{Hydr.}$ ), propensity to convert from  $\alpha$ -helical to  $\beta$ -sheet structure ( $\Delta\Delta G_{\text{coil-}\alpha} + \Delta\Delta G_{\text{coil-}\beta}$ ), and charge ( $\Delta\text{Charge}$ ), affect the rate of protein aggregation. Interestingly, although protein aggregation was measured at pH 5.5, the charge calculations were based on the charge adopted by each amino acid a neutral pH. The charge of an amino acid is dependent on the pH of the solution, and charge can have a strong influence on protein aggregation (see Section 1.3.3.). This discrepancy in the algorithm design may have implications in the accuracy of the predictions. This equation was developed and verified using experimental data of peptides and proteins that were unfolded in the conditions tested.

### 2.1.2 Wang-Agar Equation

The Wang-Agar equation<sup>61</sup> is based on the original equation proposed<sup>9</sup> by Chiti and Dobson. The equation was “re-calibrated” in 2008 by Wang and coworkers using updated ThT data for amyloid forming peptides and proteins, most of which were unfolded<sup>61</sup>. The updated equation used in their studies was:

$$\ln(v_{\text{mut}}/v_{\text{WT}}) = 0.82\Delta\text{Hydr.} + 0.52(\Delta\Delta G_{\text{coil-}\alpha} + \Delta\Delta G_{\text{coil-}\beta}) - 0.50\Delta\text{Charge} \quad (2)$$

### 2.1.3 Zyggregator and Ztox

The Zyggregator method for protein aggregation prediction was developed by Tartaglia *et al.* in 2008<sup>85</sup>. This algorithm is also based on the inherent properties of each amino acid that contribute to the tendency of a protein to aggregate. This is a sequence order-dependent algorithm. This means that the propensity of a certain residue to contribute to protein aggregation is based not only on the inherent

properties of the residue, but also on the properties of the surrounding residues in the protein sequence. The aggregation propensity of each residue is calculated based on a sliding window of 7 residues, including the three residues preceding and following a particular amino acid. Thus the score reported for each amino acid is an average of the scores for all seven residues in the sliding window centred on the residue being reported. Aggregation propensity is calculated based on the same three variables outlined by Chiti and Dobson<sup>9</sup>, and also takes into account the pattern of hydrophobicity of the sequences within the sliding window and the presence of gatekeeper residues (described in Section 1.3.2). The Zyggregator algorithm uses coefficients for the physical and chemical variables based on an experimental data set of amyloid forming peptides and unfolded proteins. The algorithm was tested using the AcP database developed by Chiti and Dobson, and with a range of unfolded proteins in varying experimental conditions<sup>85</sup>.

A second algorithm, termed Ztox, was additionally developed by Tartaglia *et al.*<sup>85</sup>, and is identical to the Zyggregator algorithm with the exception that the coefficients for the physical and chemical properties in the algorithm are instead based on a dataset of protofibrillar-forming polypeptide sequences<sup>85</sup>. The goal of this algorithm is to predict sequences most likely to form soluble oligomers, which are suggested to be particularly toxic<sup>52</sup>.

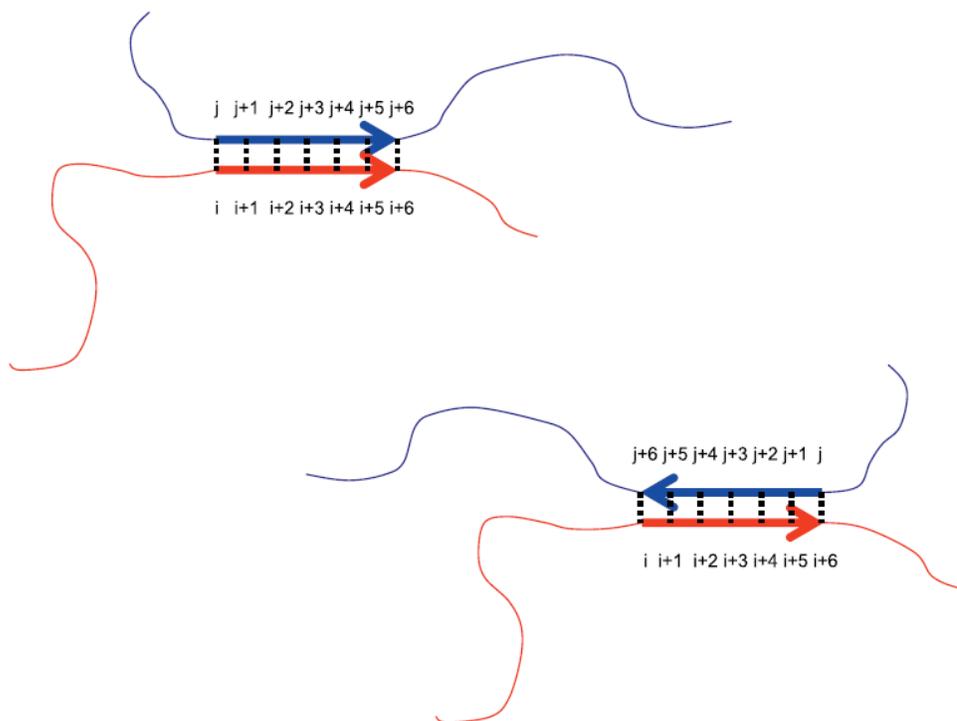
#### **2.1.4 FoldAmyloid**

The FoldAmyloid prediction algorithm was developed by Garbuzynskiy *et al.*<sup>88</sup>. The unique approach of this algorithm is based on the concept that amino acid sequences that have the potential to pack most tightly have the greatest likelihood of forming amyloid<sup>88</sup>. The ability to pack tightly is determined by the capability of the polypeptide sequence to form hydrogen bonds and molecular contacts. A dataset of 3769 native proteins was used to determine the average number of molecular contacts each amino acid makes within native structures and the probability of each amino acid to be involved in a hydrogen bond either as a proton donor or proton acceptor. Each amino acid within a sequence was given a score for each of these types of interactions. The average of these scores over a sliding window centered

on the residue in question determines the likelihood of contributing to amyloid formation by each amino acid within the sequence. A region of five or more positive scores in a row is considered to be aggregation-prone<sup>88</sup>. The algorithm was developed and tested using a dataset of amyloid-forming and non amyloid-forming peptides.

### **2.1.5 PASTA**

The PASTA prediction algorithm was introduced by Trovata *et al.*<sup>86</sup>. The purpose of this method is to determine the regions of an input amino acid sequence that are most likely to stabilize cross-beta fibrillation<sup>86</sup>. The algorithm uses pair-wise energy functions for residues facing each other when two identical sequences align in a  $\beta$ -sheet orientation. Pair potentials were derived from a set of 500 globular protein crystal structures and an energy score was given to every possible amino acid pairing in both parallel and anti-parallel conformation. For each residue in the input sequence a score is given for every possible pairing when the sequence is aligned with itself in the parallel and anti-parallel direction and the template strand is moved one residue at a time in the 3' direction (see Figure 2.1). The total score for each residue is the sum of all possible pairing values. The algorithm calculates an energy score over a sliding window of seven residues by averaging the total base pair potentials of each of the seven residues. The regions with the lowest scoring energy functions are considered to be the most prone to be involved in beta aggregation, since a lower score represents a lower free energy of interaction. This algorithm was tested by correctly predicting the amyloid-forming regions of the amyloid beta peptide,  $\alpha$ -synuclein, and PHF43 (a peptide fragment from the foetal form of human Tau)<sup>86</sup>.



**Figure 2.1 Calculation of energy scores using the PASTA algorithm.** The input sequence shown in red is aligned with an identical sequence in the parallel (top) or anti-parallel (bottom) direction. A score is assigned to each residue in the template sequence based on the knowledge based energies developed from a data set of 500 protein crystal structures. The template then shifts by one amino acid residue another score is assigned to each amino acid based on the new interaction. Once all possible pairings have been calculated, an aggregation score is given to each residue based on the sum of all the pairing scores assigned during the sequence alignment. Figure from Trovata *et al.*, 2006<sup>86</sup>.

### 2.1.6 TANGO

The TANGO algorithm for aggregation prediction was developed by Fernandez-Escamilla *et al.*<sup>87</sup>. This method uses a strictly computational, statistical mechanics approach. It is based on the concept that each amino acid residue can potentially be involved in one of the following conformations:  $\beta$ -turn,  $\alpha$ -helix,  $\beta$ -sheet, the folded conformation, or  $\beta$ -aggregate<sup>87</sup>. Each segment of the protein can populate these conformational states according to a Boltzmann distribution. The Boltzmann distribution requires that the probability of populating each state depends on the energy of that state. The TANGO algorithm uses

statistical and empirical methods to develop a partition function that determines the probability of populating each state for every residue in the sequence based on the energy calculations over a sliding window of seven amino acids. Here, the score given to each amino acid is the average probability of populating the beta-aggregate state of the seven residues centred on each particular residue. The algorithm identifies residues that have the highest likelihood of populating  $\beta$ -aggregates. Regions with several  $\beta$ -aggregate prone residues consecutively are considered to be hot-spot areas for aggregation. This validity of this method was verified using 179 peptide fragments from 21 different proteins<sup>87</sup>.

### **2.1.7 WALTZ**

The WALTZ method was developed in 2010 by Maurer-Stroh *et al.*<sup>33</sup>, and uses a position-specific scoring matrix to predict amyloid formation<sup>33</sup>. The algorithm is made up of three components. The first is a scoring matrix based on the location of amino acids in a six-residue stretch. A hexapeptide dataset consisting of 116 peptides that form amyloid and 103 that do not, was used to develop the scoring matrix by assigning each amino acid a score for every position within a hexapeptide. The score represents the probability of amyloid formation that exists when a given residue is in a certain position. The second component is a function that encompasses nineteen physical and chemical properties that have been implicated in favouring amyloid formation. The third component is a position-specific *pseudo-energy matrix* developed from structural modeling using amyloid backbone structures. In short, for this component, a known crystal structure for GNNQQNY was reduced to polyalanine. Then the alanines were replaced with all possible amino acid combinations and energy-optimized using FoldX. A position specific pseudo-energy database was created by assigning every residue a score for each position within the hexapeptide by averaging the calculated FoldX energies for each amino acid at each individual position in combination with every possible arrangement of all other amino acids at each of the other positions. These three components are summed in the final equation:

$$S_{\text{total}} = a_{\text{profile}}S_{\text{profile}} + a_{\text{physprop}}S_{\text{physprop}} + a_{\text{struct}}S_{\text{struct}} \quad (3)$$

where  $a$  represents the empirically determined weighting coefficients for each parameter and  $S$  represents the score for each parameter.  $S_{\text{profile}}$  is the score based on the location of the residue within the hexapeptide,  $S_{\text{physprop}}$  is the score based on the sum of nineteen physical and chemical properties, and  $S_{\text{struct}}$  is the score based on the pseudoenergy matrix from alignment to an amyloid crystal structure<sup>33</sup>. The algorithm was verified by predicting amyloid-forming six amino acid regions for series of functional amyloid-forming proteins and confirming experimentally whether these hexapeptides do in fact form amyloid.

### 2.1.8 Profile 3D

The Profile 3D prediction method was developed by Goldschmidt *et al.*<sup>89</sup>. It is an aggregate structure-based approach that utilizes the known crystal structure of the amyloid forming hexapeptide NNQQNY<sup>89</sup>. Six residue regions of the sequence in question are threaded onto the backbone of the crystal structure and comparative energies are calculated using the Rosetta Design<sup>96</sup> potential energy function. Segments that can form similar, self-complementary zipper structures achieve low energy scores and are considered capable of amyloid fibril formation. A threshold for high propensity for fibrillation was determined using a set of 16 hexapeptide zipper crystal structures and mapping them to the NNQQNY structure used in this method and calculating the comparative energy. Segments with calculated comparative energies below this threshold are considered highly aggregation-prone. The validity of the algorithm was verified by accurate predictions of the aggregation-prone regions within RNase A<sup>89</sup>.

## 2.2 Methods

### 2.2.1 Quantifying predicted aggregation propensities

#### 2.2.1.1 Chiti-Dobson

The Chiti-Dobson equation<sup>9</sup> is designed to specifically calculate the difference in aggregation propensity upon mutation, and so the only factors that contribute to the calculation are the changes in

chemical and physical properties between the WT residue and the mutated residue. The aggregation propensity calculations for each mutant were performed using equation 1 in Section 2.1.1, and parameters from Table 2.2 in Chiti *et al.*, 2003 (reproduced below). The propensity for alpha-coil was calculated using the online Agadir calculation (<http://agadir.crg.es/>) as described by Chiti *et al.*<sup>9</sup>. A sample calculation is shown below for the A4V mutation of SOD1:

$$\ln(v_{\text{mut}}/v_{\text{WT}}) = 0.633\Delta\text{Hydr.} + 0.198(\Delta\Delta G_{\text{coil-}\alpha} + \Delta\Delta G_{\text{coil-}\beta}) - 0.491\Delta\text{Charge}$$

$$\begin{aligned}\Delta\text{Hydr.} &= \text{Hydr. of alanine} - \text{Hydr. of valine} \\ &= (-0.39) - (-1.3) = 0.91\end{aligned}$$

$$(\Delta\Delta G_{\text{coil-}\alpha}) = RT \ln(P_{\alpha\text{WT}}/P_{\alpha\text{mut}})$$

\* $P_{\alpha\text{WT}}$  and  $P_{\alpha\text{mut}}$  calculated  
using Agadir online server

$$\begin{aligned}&= 0.008314 \text{ kJ mol}^{-1}\text{K}^{-1}(310\text{K})\ln(0.4/0.5) \\ &= -0.575\end{aligned}$$

$$\begin{aligned}(\Delta\Delta G_{\text{coil-}\beta}) &= \Delta G_{\text{coil-}\beta} \text{ for alanine} - \Delta G_{\text{coil-}\beta} \text{ for valine} \\ &= 0.47 - 0.13 = 0.60\end{aligned}$$

$$\begin{aligned}\Delta\text{Charge} &= 0 \text{ (Charge of alanine)} - 0 \text{ (Charge of valine)} \\ &= 0\end{aligned}$$

$$\begin{aligned}\ln(v_{\text{mut}}/v_{\text{WT}}) &= 0.633(0.91) + 0.198(0.58 + 0.60) - 0.491(0) \\ &= 1.38\end{aligned}$$

Since the aggregation score for A4V is greater than 0, this mutation is predicted to increase the propensity for aggregation relative to wild type. In this manner, aggregation propensity compared to wild type was calculated for all mutants using the Chiti-Dobson algorithm.

**Table 2.2 Scales of hydrophobicity,  $\beta$ -sheet propensity and charge for the 20 natural amino acids, reproduced from Chiti *et al.*<sup>9</sup>**

<b>Amino Acid Residue</b>	<b>Hydrophobicity (kcal mol<sup>-1</sup>)<sup>a</sup></b>	<b><math>\beta</math>-sheet propensity<sup>b</sup></b>	<b>Charge<sup>c</sup></b>
<b>Arg (R)</b>	3.95	0.35	1
<b>Lys (K)</b>	2.77	0.34	1
<b>Asp (D)</b>	3.81	0.72	-1
<b>Glu (E)</b>	2.91	0.35	-1
<b>Asn (N)</b>	1.91	0.4	0
<b>Gln (Q)</b>	1.3	0.34	0
<b>His (H)</b>	0.64 (2.87) <sup>d</sup>	0.37	0 (+1) <sup>d</sup>
<b>Ser (S)</b>	1.24	0.3	0
<b>Thr (T)</b>	1	0.06	0
<b>Tyr (Y)</b>	-1.47	0.11	0
<b>Gly (G)</b>	0	0.6	0
<b>Pro (P)</b>	-0.99	n.d.	0
<b>Cys (C)</b>	-0.25	0.25	0
<b>Ala (A)</b>	-0.39	0.47	0
<b>Trp (W)</b>	-2.13	0.24	0
<b>Met (M)</b>	-0.96	0.26	0
<b>Phe (F)</b>	-2.27	0.13	0
<b>Val (V)</b>	-1.3	0.13	0
<b>Ile (I)</b>	-1.82	0.1	0
<b>Leu (L)</b>	-1.82	0.32	0

<sup>a</sup> hydrophobicity values of the 20 amino acid residues at neutral pH based on the partition coefficients from water to octanol. The data are from column 6 of Table 4.8 in ref. 30.

<sup>b</sup>  $\beta$ -sheet propensities of the 20 amino acid residues normalized from 0 (high  $\beta$ -sheet propensity) to 1 (low  $\beta$ -sheet propensity). The data are from column 4 of Table 1 of ref. 29. The  $\beta$ -sheet propensity of proline is not reported due to the difficulty in determining it experimentally. The  $\beta$ -sheet propensity of glycine is from theoretical calculations.

<sup>c</sup> values of charge are at neutral pH.

<sup>d</sup> values in brackets are at a pH lower than 6.0, when the histamine residue is positively charged

### 2.2.1.2 Wang-Agar

The calculation of aggregation propensity using the Wang-Agar equation<sup>61</sup> is performed identically to that described for the Chiti-Dobson equation with the exception of different coefficients for the overall equation as described in Equation 3, Section 2.1.2.

### 2.2.1.3 Zyggregator and Ztox

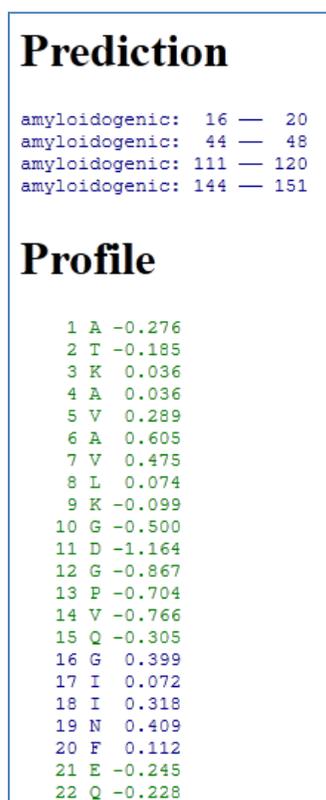
The calculation for aggregation propensity using the Zyggregator and Ztox algorithms<sup>85</sup> was performed using the online server at <http://www-vendruscolo.ch.cam.ac.uk/zyggregator.php>. The FASTA sequence for each mutant was input into the server, the pH was set to 7, and either Zagg or Ztox was selected depending on the prediction being performed. An example of the output is given in Figure 2.2. An aggregation propensity score is given to each individual residue. To calculate the total aggregation propensity for the protein, the Zyggregator scores greater than zero were summed (those below zero are not considered to be aggregation-prone and thus are not considered in the total propensity calculation). The difference between mutant and wild type protein was calculated by subtracting the value for the total propensity of wild type from the total propensity of mutant. Mutant sequences with a positive difference are considered to be more aggregation-prone than wild type, while those with a negative difference are considered to be less aggregation-prone than wild type (see Section 2.2.2).

```
Generic Amyloid Aggregation Propensity-----  
FILE: 34790438.inp      OPTIMAL SCORE: 0.561721      ERROR: 0.0287815  
OPTIMAL SCORE excluding termini:      0.700394  
#  
1 0.268412  
2 0.798985  
3 0.59632  
4 0.939979  
5 1.09464  
6 0.283272  
7 0.780749  
8 0.697121  
9 0.481675  
10 -0.727368  
11 -0.727368  
12 -0.910051  
13 -0.600223  
14 -0.433313  
15 -0.0166036
```

**Figure 2.2 Zyggregator output for pWT SOD1.** These data were generated using the online algorithm at <http://www-vendruscolo.ch.cam.ac.uk/zyggregator.php>. A sample of the output for residues 1-15 is shown here. Positive values indicate residues above the experimentally determined cut-off for being prone to aggregate while negative values correspond to residues below the cut-off.

#### 2.2.1.4 FoldAmyloid

Aggregation predictions were made with the FoldAmyloid algorithm<sup>88</sup> using the online server at <http://antares.protres.ru/fold-amyloid/oga.cgi>. Predictions were completed using the triple hybrid setting, which combined scores for molecular contacts, proton donors in hydrogen bonds, and proton acceptors in hydrogen bonds. Example output data is given in Figure 2.3. The predicted aggregation-prone regions are given under the "Prediction" heading. The profile lists the individual prediction scores for each residue. Scores above zero are considered to be aggregation-prone. Total aggregation sums were calculated identically to the method described for Zyggregator and Ztox.



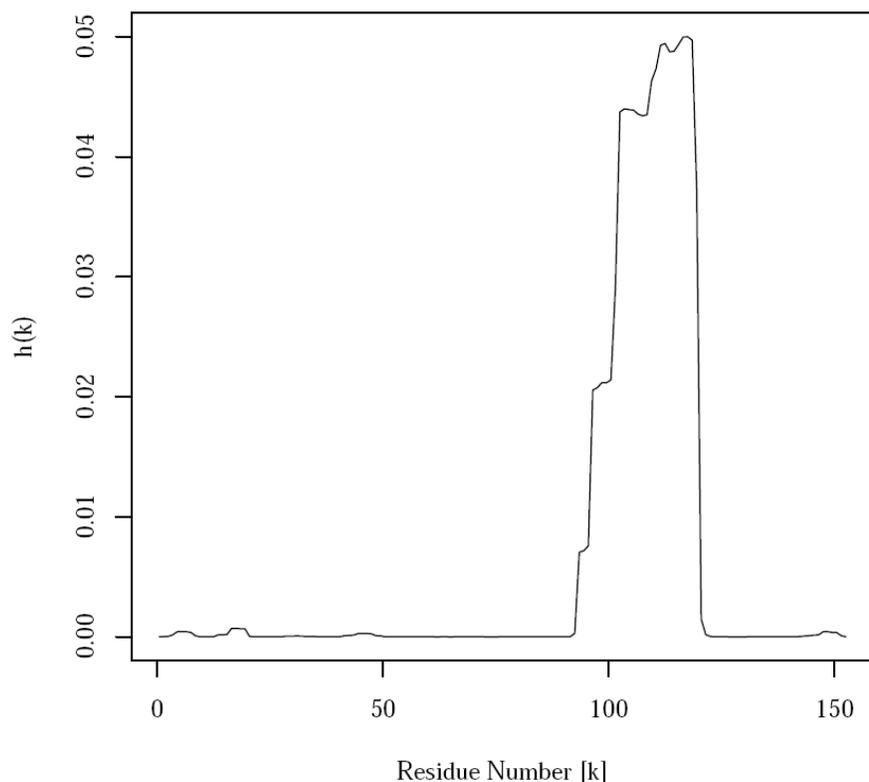
**Figure 2.3 FoldAmyloid prediction results for pWT SOD1.** This output was generated using the online algorithm at <http://antares.protres.ru/fold-amyloid/oga.cgi>. The pWT SOD1 sequence was used as the input data and the sample output shows the overall regions predicted to be aggregation-prone listed under the heading "Prediction", and the individual residue scores for residues 1-22. Negative scores correspond to residues not prone to aggregate and positive scores correspond to residues prone to aggregate. Regions coloured in blue (residues 16-20) are predicted to be amyloidogenic based on the presence of 5 consecutive significantly positive scores. Nonamyloidogenic regions are in green.

### 2.2.1.5 PASTA

Aggregation propensity calculations using the PASTA method<sup>86</sup> were performed using the online server at <http://protein.cribi.unipd.it/pasta/>. The output includes a visual profile of the aggregation-prone regions, and a list of the highest-scoring regions. For superoxide dismutase and amyloid beta predictions, the 10 highest scoring regions were used for comparison. However, for human acylphosphatase there was no difference in the 10 highest scoring regions for the mutants in question, and so the 20 highest scoring regions were calculated. The inclusion of more of the top scoring regions allowed for the observation of some slight differences between input sequences. Figure 2.4 shows an example of PASTA output. In order to quantify the prediction output, the sum of the pair-wise energy functions for the top aggregation-prone regions, termed the PASTA energy, was calculated for each input sequence. In this case, the number is negative, since the regions with the lowest energy scores are most aggregation-prone (See Section 2.1.5). For ease of comparison, this value was converted to its absolute value when used for comparison with other methods. The difference between the absolute overall aggregation values of mutant protein minus wild type protein was calculated for each mutant, and those with positive scores are predicted to be more aggregation-prone than wild type protein (see Section 2.2.2).

A

## Aggregation Profile



B

## PASTA OUTPUT

PASTA output					
pairing 1	PASTA energy	-6.939119	length 18	between segments 103-120 and 103-120	parallel
pairing 2	PASTA energy	-6.882765	length 24	between segments 97-120 and 97-120	parallel
pairing 3	PASTA energy	-6.625771	length 19	between segments 102-120 and 102-120	parallel
pairing 4	PASTA energy	-6.554064	length 27	between segments 94-120 and 94-120	parallel
pairing 5	PASTA energy	-6.403213	length 17	between segments 103-119 and 103-119	parallel
pairing 6	PASTA energy	-6.346859	length 23	between segments 97-119 and 97-119	parallel
pairing 7	PASTA energy	-6.112998	length 11	between segments 110-120 and 110-120	parallel
pairing 8	PASTA energy	-6.089864	length 18	between segments 102-119 and 102-119	parallel
pairing 9	PASTA energy	-6.018157	length 26	between segments 94-119 and 94-119	parallel
pairing 10	PASTA energy	-5.890441	length 9	between segments 112-120 and 112-120	parallel

*Beta-pairings known to appear in cross-beta fibrillar aggregates generally have PASTA energies  $\leq -4.0$ .*

**Figure 2.4 Example output of the PASTA algorithm.** A) The aggregation profile as calculated by the PASTA prediction server (<http://protein.bio.unipd.it/pasta/>) for pWT SOD1. The y-axis plots the value for  $h(k)$ , which is defined as the probability of a given residue to aggregate in an ordered  $\beta$ -structure. B) An example of the text output obtained from the PASTA server calculations of the top ten highest scoring regions for pWT SOD1.

### 2.2.1.6 TANGO

The TANGO algorithm<sup>87</sup> was used online at <http://tango.crg.es>. The amino acid sequences of the proteins involved in this study were input and the pH, ionic strength, and temperature were set to pH 7, 0.02 M and 310 K, respectively. An example of the output data is shown in Figure 2.5. The output displays the phase separation probabilities for each state. The total aggregation propensity for a given sequence is displayed at the top of the chart and labelled “AGG”. This value is the sum of the beta aggregation column. The total aggregation propensity of the wild type sequence was subtracted from the total propensity for each mutant sequence. A positive difference indicates the mutant is predicted to be more aggregation-prone than wild type and a negative difference indicates that the mutant is less aggregation-prone (see Section 2.2.2).

AGG	94.4612	AMYLO	178.168	TURN	129.825	HELIX	11.2416	HELAGG	0	BETA	338.713
Position	Amino acid	Beta sheet	Beta turn	Alpha helix	Beta aggregation						
01	A		0.3	0.0	0.000	0.000					
02	T		0.8	0.0	0.000	0.000					
03	K		1.1	0.0	0.000	0.000					
04	A		1.2	0.0	0.000	2.562					
05	V		8.2	0.0	0.000	2.562					
06	A		8.8	0.0	0.000	2.562					
07	V		9.6	0.0	0.000	2.562					
08	L		2.4	0.3	0.000	2.562					
09	K		1.3	2.4	0.000	0.000					
10	G		0.1	2.4	0.000	0.000					
11	D		0.0	2.4	0.000	0.000					
12	G		0.0	2.1	0.000	0.000					
13	P		0.0	0.0	0.000	0.000					
14	V		0.2	0.1	0.000	0.221					
15	Q		0.2	0.1	0.000	0.221					

**Figure 2.5 Output predicted aggregation from the TANGO server for pWT SOD1.** The prediction was made using the online algorithm at <http://tango.crg.es/>. These sample data show the individual probabilities for each residue and calculates the total aggregation score by summing all values for beta aggregation. The top left corner lists the total aggregation score next to the heading “AGG”.

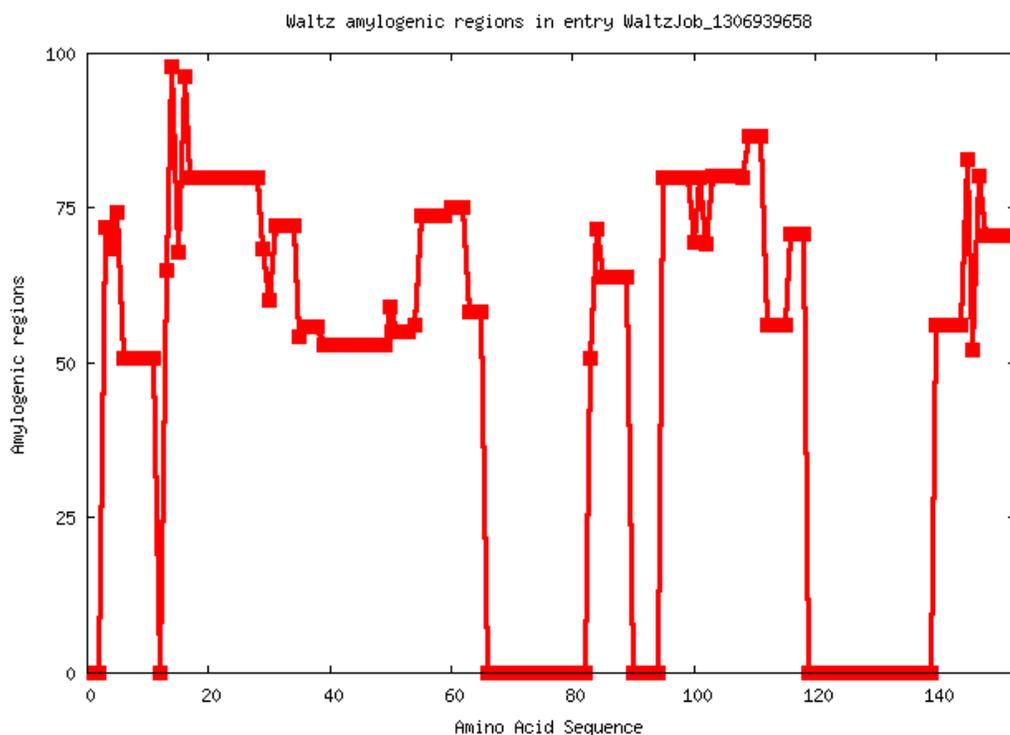
### 2.2.1.7 WALTZ

The Waltz method<sup>33</sup> for protein aggregation propensity prediction is available through an online server at <http://waltz.switchlab.org/>. There are several ways to adjust the output cut-off by selecting best overall performance, high sensitivity, high selectivity, or by choosing a custom cut-off. By decreasing the cut-off more data can be obtained about the aggregation tendencies for a greater portion of the protein. For the purposes of comparing a series of mutants, a cut-off of 50 was chosen to allow for the observation of more differences between mutants. An example of Waltz output is given in Figure 2.6. From this data, a total aggregation score for each protein sequence was determined by calculating the sum of the products of the lengths of each aggregation-prone region with the average score per residue within that region. The difference was calculated between each mutant and compared to wild type by subtracting the wild type score from the mutant score. In the cases where the difference was positive the mutant was predicted to be more prone to aggregate than wild type, while if the difference was negative, the mutant was predicted to be less aggregation-prone (see Section 2.2.2).

A

WaltzJob_1306939576		
Positions	Sequence	Average score per residue
3-11	KAVAVLKGD	57.7480488888889
13-65	PVQGIINFEQKESNGPVKVGSIKGLTEGLHGFHVHEFGDNTAGCTSAGPHFN	67.1294249245283
83-89	DLGNVTA	63.1151458571428
95-118	ADVSIEDSVISLSGSDHSIIGRTL	74.8745819583333
140-153	AGSRLACGVIGIAQ	65.6951745

B

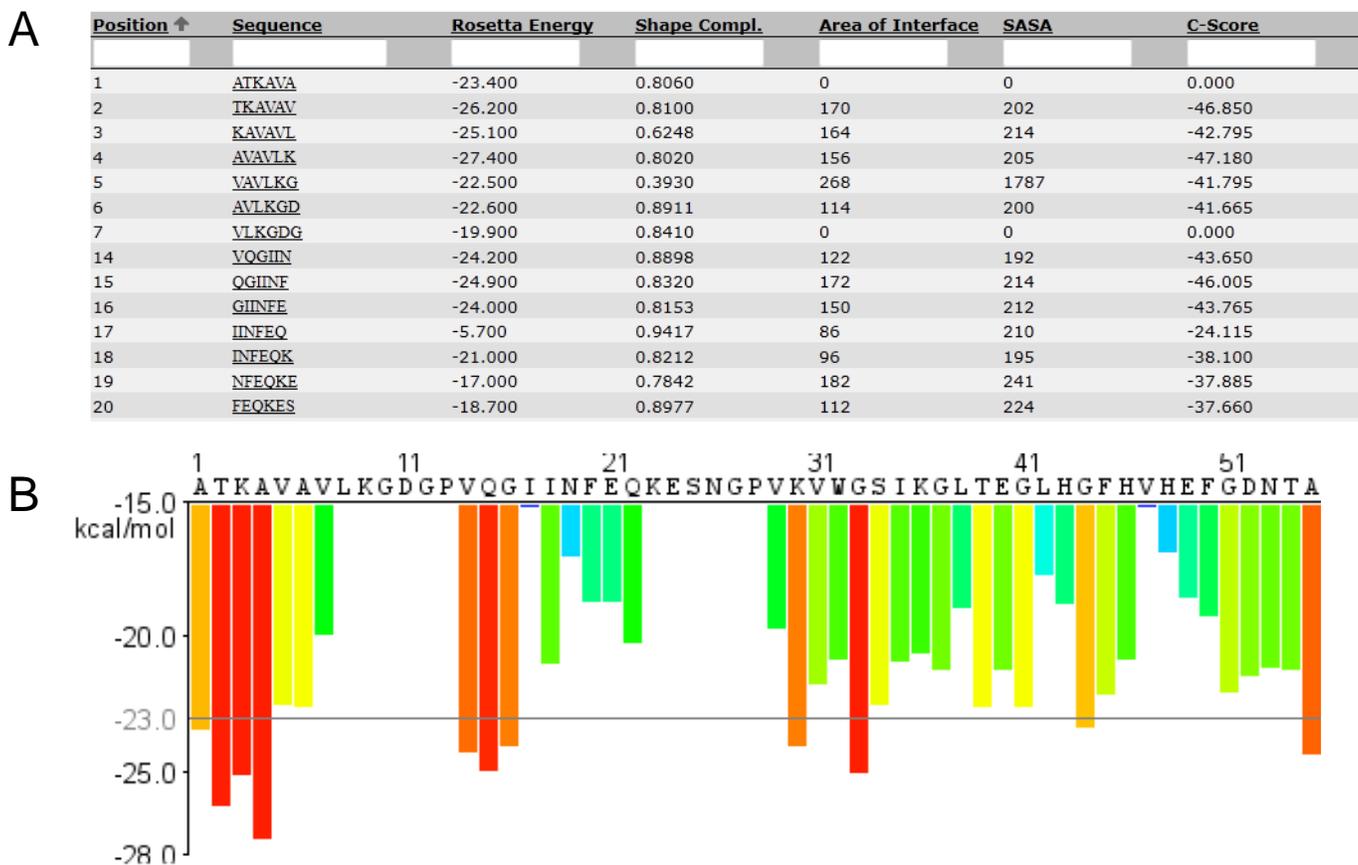


**Figure 2.6 Output data for aggregation prediction of pWT SOD1 using the Waltz server (<http://waltz.switchlab.org/>).** A) Detailed text output showing the positions, sequences, and score per residue for regions above the threshold of 50. B) Graphical output displaying the sequence regions predicted to be aggregation-prone.

### 2.2.1.8 Profile 3D

Profile 3D<sup>89</sup> aggregation prediction calculations were completed online at <http://services.mbi.ucla.edu/zipperdb/>. Figure 2.7 gives an example of the output. The threshold for aggregation was determined experimentally to be a Rosetta energy of -23.0 kcal/mol through the analysis

of predicted Rosetta energies and observed amyloid formation of a set of amyloid-forming hexapeptides. To determine a total aggregation propensity for each sequence, all values below -23.0 kcal/mol were summed, and the absolute value was compared to the absolute value of the wild type protein. A positive difference resulting from the subtraction of the wild type score from the mutant score indicated that the mutant was more prone to aggregate than wild type (see Section 2.2.2).



**Figure 2.7 Output data for the predicted aggregation propensity of pWT SOD1 using the Profile 3D method (available at <http://services.mbi.ucla.edu/zipperdb/>).** A) Numerical data displaying the calculated parameters for each hexapeptide region within the full amino acid sequence. Of particular importance is the Rosetta energy which describes how well the given peptide can map onto a known amyloid crystal structure; the lower the energy, the better the structural homology. B) A bar graph is displayed of the Rosetta energies for each hexapeptide (beginning at the indicated residue) region of the protein sequence. Bars coloured in orange and red have energies below the aggregation threshold (-23.0 kcal/mol) and thus are predicted to be highly aggregation-prone.

### 2.2.2 Normalization of Aggregation Prediction Data

In order to directly compare aggregation propensities predicted using different methods all data sets were normalized on a scale of 0 to 1. Each data set included all predicted aggregation propensities of mutations for one of the three test proteins, and the wild type protein, as predicted by one particular method. The difference between mutant and wild type was calculated for each mutant sequence giving a data set where wild type was zero and those with positive scores were predicted to aggregate more than wild type, while those with negative scores were predicted to be less aggregation-prone than wild type. The minimum and the maximum aggregation propensity values were determined for each data set, and the normalized values were calculated as:

$$\text{—————} \quad (3)$$

where  $N_i$  represents the normalized value for the mutant sequence,  $i$  is the initial aggregation propensity value for the mutant sequence,  $max$  is the value for the mutant in the dataset with the highest predicted aggregation and  $min$  is the value for the mutant in the dataset with the lowest predicted aggregation. In this way, each prediction algorithm was adjusted to the same scale.

### 2.2.3 Production of Hot-spot Maps

In order to visualize the most aggregation-prone regions of the input protein sequence, hot-spot maps were created. The individual residue scores for the wild type protein sequences were normalized over a range of 0 to 1 (as described in 2.2.2). Stacked column charts were prepared using Excel in which the data are displayed with the residue number along the x-axis and the cumulative normalized score for predicted aggregation propensity along the y-axis (for an example, see Figure 2.9). The columns are divided into the proportional contribution of each algorithm and colour coded to represent the methods used. The Chiti-Dobson and Wang-Agar calculations are not included in this comparison because these algorithms are designed to predict only the changes in aggregation upon mutation and are not designed to identify highly aggregation-prone regions of the protein.

## 2.3 Results

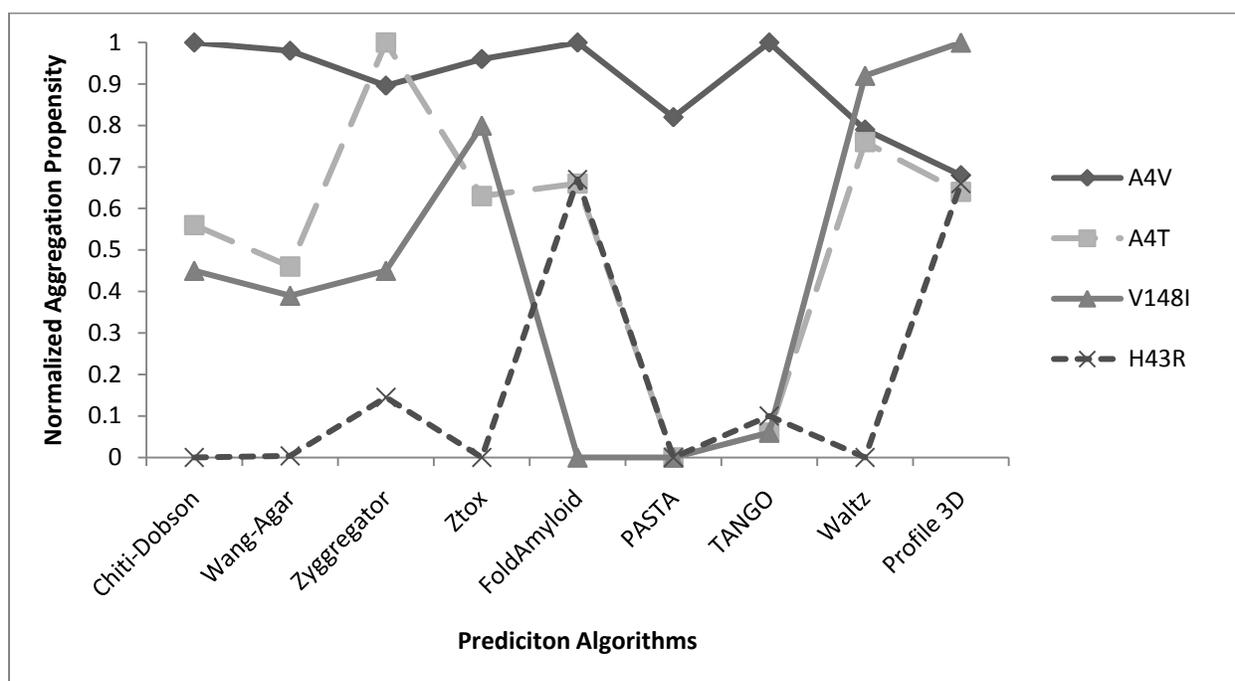
### 2.3.1 Predicting Aggregation Propensity of Superoxide Dismutase Mutants

All nine algorithms introduced in Section 2.1 were used to predict the aggregation propensity of pWT SOD1 and twenty-four of the most common SOD1 mutations found in fALS. Table 1.1 in Chapter 1 lists the mutants used for this analysis. Table 2.3 gives the normalized aggregation scores from highest to lowest for all mutants for each algorithm. The table is colour coded such that pWT SOD1 and any mutant with identical predicted propensity are bolded in black. Mutants with a higher predicted propensity to aggregate than pWT are coloured red and those with a lower predicted aggregation propensity are shown in green. Some of the mutants are consistently predicted to be more or less aggregation-prone than pWT by multiple algorithms. However, for most mutants, there are obvious discrepancies in the pattern of predicted aggregation scores by the nine different algorithms. To further illustrate this, Figure 2.8 shows a line graph of the predicted aggregation propensities for four different mutants using the nine algorithms. Among the mutants studied, A4V is most consistently predicted to be aggregation-prone, while A4T and V148I show significant variability, and H43R, with the exception of two algorithms, is primarily predicted not to aggregate.

**Table 2.3 Normalized predicted aggregation propensity for wild type and twenty four SOD1 mutants.**

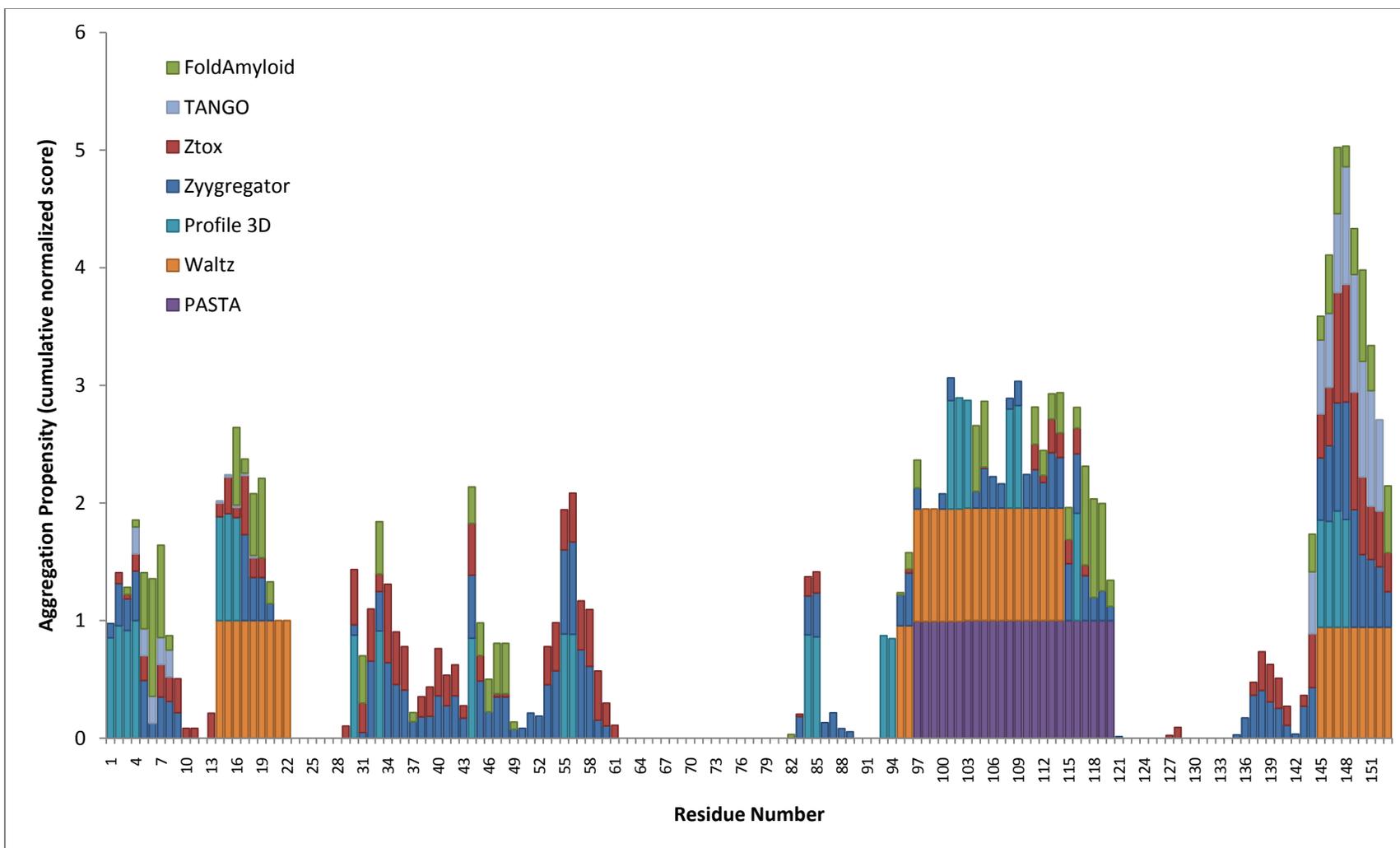
Chiti-Dobson		Wang-Agar		Zyggregator		Ztox		FoldAmyloid		PASTA		TANGO		Waltz		Profile 3D	
1.00	D90A	1.00	D90A	1.00	A4T	1.00	G37R	1.00	S105L	1.00	D101N	1.00	A4V	1.00	L144F	1.00	G41S
0.81	D101N	0.77	D101N	0.99	D101N	0.97	L144F	0.77	A4V	0.93	S105L	0.73	L144F	0.95	E100G	0.98	D101N
0.79	S105L	0.76	S105L	0.98	E100G	0.96	A4V	0.67	G37R	0.62	G93C	0.58	S105L	0.92	L38V	0.60	L144F
0.75	E100G	0.76	E100G	0.94	A4V	0.92	D101N	0.66	G93C	0.61	E100G	0.46	D101N	0.92	L84V	0.60	V148I
0.70	G93C	0.61	A4V	0.91	D90A	0.85	D90A	0.62	G85R	0.61	A4V	<b>0.45</b>	pWT	0.90	V148I	0.59	L38V
0.69	A4V	0.60	G93C	0.91	L144F	0.84	A4S	0.62	V148G	0.61	G93S	<b>0.45</b>	G93R	0.88	D90A	0.59	E100G
0.58	L144F	0.50	L144F	0.87	L84F	0.82	L38V	0.60	G93A	0.60	G93A	<b>0.45</b>	G85R	0.86	S105L	0.59	G41D
0.58	G41S	0.48	G93A	0.83	L38V	0.82	V148I	0.60	G93R	<b>0.59</b>	pWT	<b>0.45</b>	G37R	0.85	V148G	0.41	D90A
<b>0.56</b>	pWT	<b>0.47</b>	pWT	0.80	L84V	0.80	L84F	0.55	G41S	<b>0.59</b>	A4S	<b>0.45</b>	H43R	<b>0.83</b>	pWT	0.41	A4V
<b>0.56</b>	A4T	0.45	G41S	0.79	G93R	0.78	G93A	0.54	G41D	<b>0.59</b>	A4T	<b>0.45</b>	H46R	<b>0.83</b>	G93S	<b>0.39</b>	pWT
0.55	G93A	0.44	L38V	0.79	I149T	0.73	L84V	0.54	G93S	<b>0.59</b>	V148I	<b>0.45</b>	E100G	<b>0.83</b>	G93A	<b>0.39</b>	H43R
0.54	L38V	0.43	A4T	0.78	I113T	0.72	I149T	0.53	D90A	<b>0.59</b>	H43R	<b>0.45</b>	G93A	<b>0.83</b>	G93C	<b>0.39</b>	L84F
0.52	G93S	0.40	G93S	0.73	S105L	0.72	I113T	0.53	D101N	<b>0.59</b>	G93R	<b>0.45</b>	G93S	<b>0.83</b>	L84F	<b>0.39</b>	L84V
0.50	V148I	0.39	V148I	<b>0.70</b>	pWT	0.68	G85R	0.52	H46R	<b>0.59</b>	G85R	<b>0.45</b>	L38V	<b>0.83</b>	A4V	0.38	S105L
0.47	A4S	0.34	A4S	0.69	G93S	0.67	G93C	<b>0.51</b>	pWT	<b>0.59</b>	G37R	<b>0.45</b>	G41D	0.81	A4T	0.38	A4S
0.41	H46R	0.27	H46R	0.68	G93A	0.67	S105L	<b>0.51</b>	H43R	<b>0.59</b>	H46R	<b>0.45</b>	G41S	0.80	A4S	0.38	A4T
0.39	I113T	0.26	V148G	0.68	V148I	0.66	A4T	<b>0.51</b>	E100G	<b>0.59</b>	L38V	<b>0.45</b>	L84F	0.79	I113T	0.34	I149T
0.39	I149T	0.24	I113T	0.66	A4S	<b>0.65</b>	pWT	<b>0.51</b>	L84F	<b>0.59</b>	G41D	<b>0.45</b>	L84V	0.79	D101N	0.20	H46R
0.39	G85R	0.24	I149T	0.65	G41S	0.64	E100G	<b>0.51</b>	L38V	<b>0.59</b>	G41S	<b>0.45</b>	D90A	0.71	G85R	0.20	G93C
0.39	G37R	0.24	G85R	0.60	H46R	0.64	G93R	<b>0.51</b>	L84V	<b>0.59</b>	L84F	<b>0.45</b>	G93C	0.67	I149T	0.20	G93S
0.35	V148G	0.24	G37R	0.58	V148G	0.62	V148G	0.50	A4T	<b>0.59</b>	L84V	<b>0.45</b>	I113T	0.64	G93R	0.20	G93R
0.35	G93R	0.20	G93R	0.56	G85R	0.53	G93S	0.50	L144F	<b>0.59</b>	D90A	0.43	A4T	0.61	G37R	0.20	G93A
0.26	H43R	0.15	H43R	0.50	H43R	0.29	H46R	0.37	A4S	<b>0.59</b>	L144F	0.43	V148I	0.45	G41S	0.20	I113T
0.20	G41D	0.08	L84F	0.45	G41D	0.19	G41S	0.21	I113T	<b>0.59</b>	V148G	0.39	A4S	0.41	H46R	0.19	G37R
0.06	L84F	0.02	G41D	0.41	G37R	0.09	H43R	0.07	I149T	<b>0.59</b>	I149T	0.02	I149T	0.35	H43R	0.15	V148G
0.00	L84V	0.00	L84V	0.00	G93C	0.00	G41D	0.00	V148I	0.00	I113T	0.00	V148G	0.00	G41D	0.00	G85R

\*Mutants coloured red, green or black have aggregation propensity scores higher than, lower than, or equal to pWT SOD1, respectively.



**Figure 2.8 Predicted aggregation propensities of SOD1 mutants display significant variability between algorithms for the same mutant.** The predicted aggregation propensities for four ALS-linked SOD1 mutations were calculated using nine different prediction algorithms. There are considerable differences between prediction results for the same mutant.

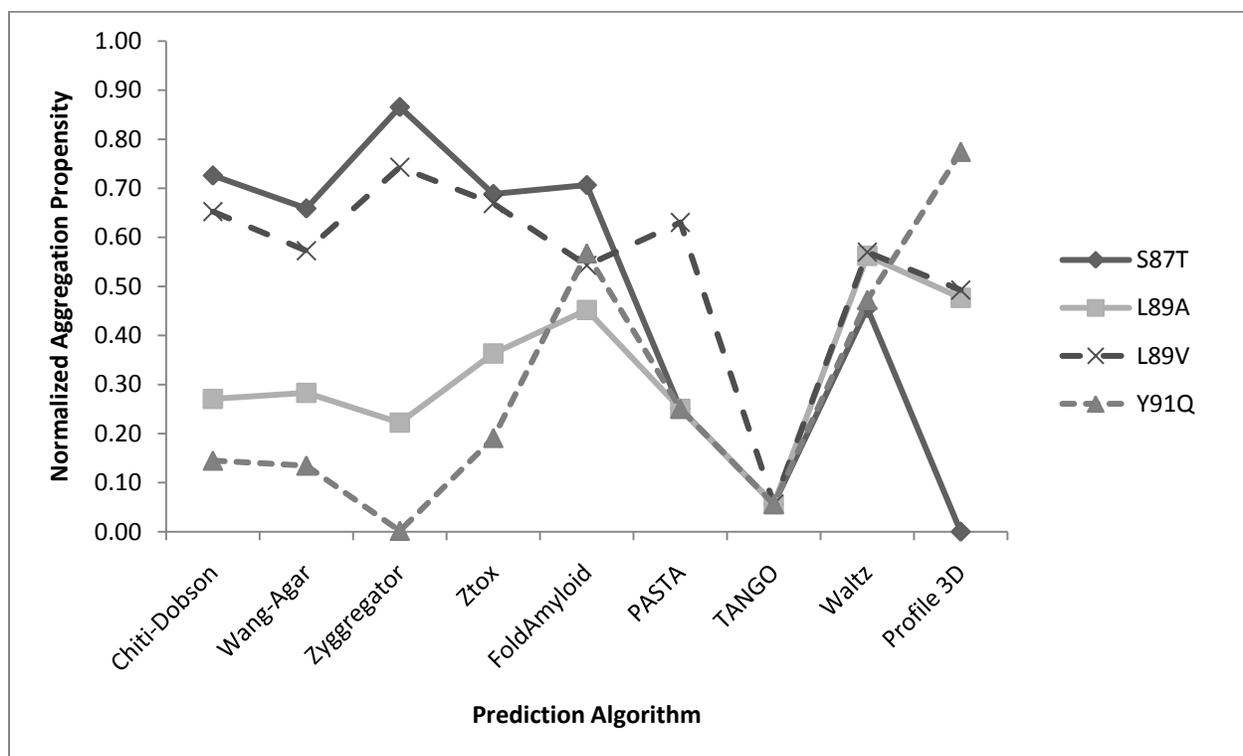
To compare the aggregation-prone regions predicted by the different algorithms, a map of the aggregation hot spots for pWT SOD1 is given in Figure 2.9. The graph is a stacked column representation in which the y-axis is the sum of the normalized scores for each prediction algorithm. There is some consistency in the regions determined to be particularly aggregation-prone. Region 145-148 is a hot-spot identified by six out of seven algorithms. Regions 16-18, 111-114 and 149-152 include areas predicted to be aggregation-prone by five out of seven algorithms, while several other regions are identified by three or four algorithms.



**Figure 2.9** Aggregation-prone regions of pWT SOD1 predicted by seven different algorithms. The normalized aggregation prediction scores for each residue are plotted as a stacked bar chart. The total height of each bar represents the sum of the aggregation scores from all 7 algorithms. Each bar is divided by colour (see legend) based on the contribution from each algorithm.

### 2.3.2 Predicting Aggregation Propensity of Human Acylphosphatase Mutants

The aggregation propensity of wild type and 30 mutants of human acylphosphatase (AcP) were calculated using the nine different prediction algorithms. The normalized predicted aggregation propensities are displayed in Table 2.4. Once again there is some consistency in certain mutations that consistently score higher or lower than WT. However, the overall pattern differs widely between algorithms. Figure 2.10 shows a line graph representing the predicted aggregation propensity for four mutations. All four mutants shown in this figure are predicted to have both high and low predicted aggregation propensities depending on the algorithm used, and so illustrate the large scatter between the various aggregation prediction techniques.



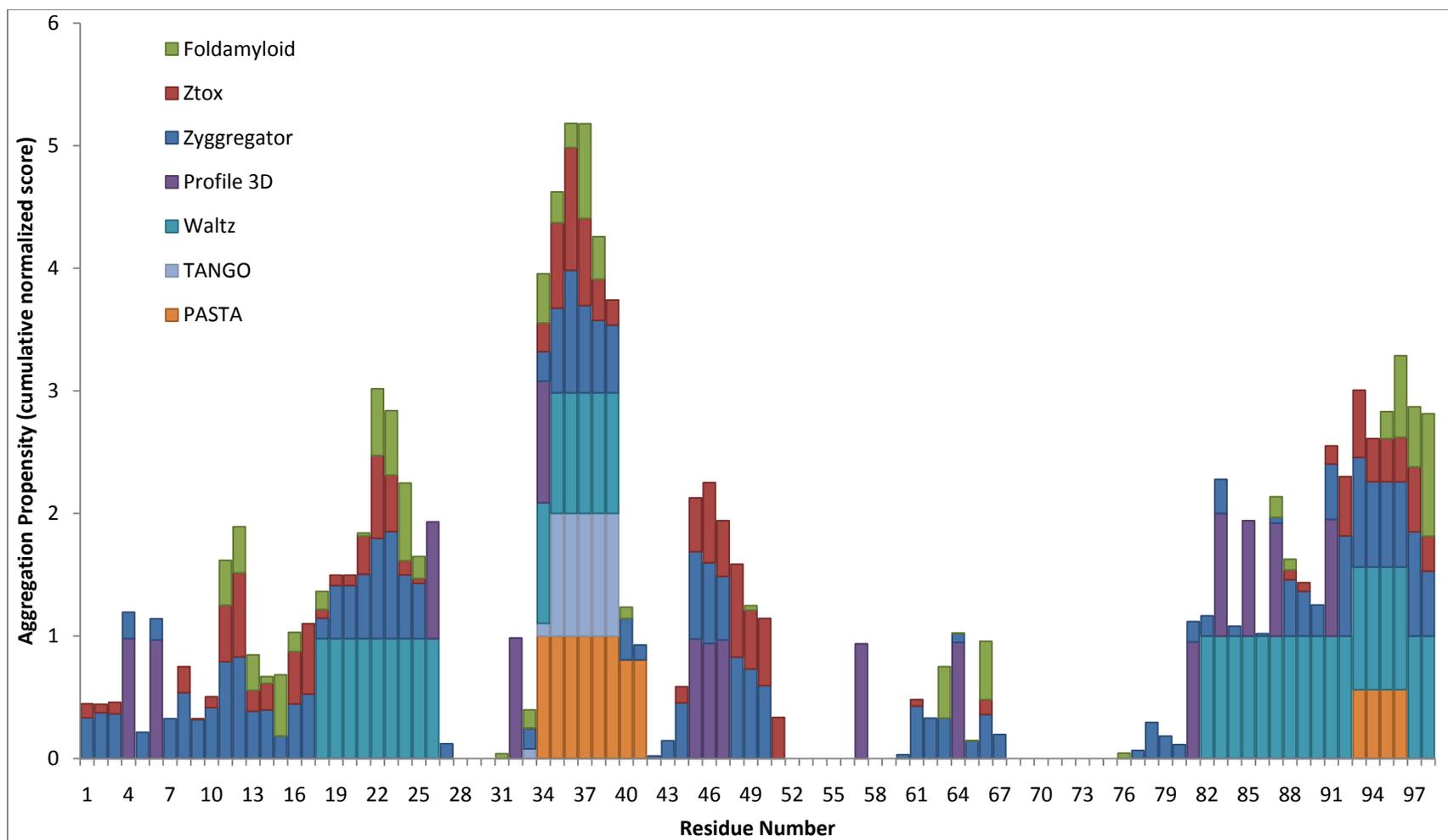
**Figure 2.10 Predicted aggregation propensities of AcP mutants display significant variability between algorithms for the same mutant.** The predicted aggregation propensities for the four AcP mutants show considerable differences between prediction results for the same mutant.

**Table 2.4 Normalized aggregation propensities for wild type and 30 AcP mutants using nine prediction algorithms.**

Wang-Agar		Chiti-Dobson		Zyggregator		Ztox		FoldAmyloid		PASTA		TANGO		Waltz		Profile 3D	
1.00	R23Q	1.00	R23Q	1.00	K88N	1.00	R23Q	1.00	S21R	1.00	E90H	1.00	E29K	1.00	S43E	1.00	K88N
0.98	R97Q	0.97	R97Q	0.91	K88Q	0.89	R97Q	0.94	S92H	0.63	L89V	0.94	E29Q	0.93	E55Q	0.94	Y25A
0.88	R77E	0.87	R77E	0.87	S92T	0.87	S92T	0.91	E90H	0.55	K88N	0.93	A30G	0.80	R23Q	0.94	F22L
0.88	R97E	0.87	R97E	0.87	S87T	0.86	S43E	0.84	S8H	0.51	S21R	0.84	E29R	0.59	R97E	0.94	Y98Q
0.82	K88Q	0.81	K88Q	0.86	R23Q	0.79	K88N	0.71	S87T	0.42	S92H	0.82	E29D	0.57	L89V	0.94	S43E
0.73	K88N	0.74	E29Q	0.83	R97Q	0.73	K88Q	0.70	K88Q	0.29	F94L	0.24	Y25A	0.56	L89A	0.77	Y91Q
0.69	E29Q	0.73	S87T	0.79	R77E	0.69	S87T	0.70	S92T	<b>0.25</b>	<b>WT</b>	0.13	S43E	0.55	K88Q	0.74	E90H
0.66	S87T	0.72	K88N	0.74	L89V	0.67	L89V	0.65	E29R	<b>0.25</b>	<b>V17A</b>	0.12	S92T	0.51	S92T	0.73	S92T
0.65	E55Q	0.68	E55Q	0.62	E29D	0.65	A30G	0.61	E29Q	<b>0.25</b>	<b>V20A</b>	0.11	R77E	0.50	R97Q	<b>0.71</b>	<b>WT</b>
0.60	S92T	0.65	L89V	0.62	A30G	0.64	R77E	0.60	V20L	<b>0.25</b>	<b>V20L</b>	0.11	R97E	0.50	E29K	<b>0.71</b>	<b>V17A</b>
0.57	L89V	0.64	S92T	0.59	R97E	0.62	E90H	0.60	F22L	<b>0.25</b>	<b>V20S</b>	<b>0.06</b>	<b>WT</b>	0.50	L89T	<b>0.71</b>	<b>V20A</b>
<b>0.55</b>	<b>WT</b>	<b>0.54</b>	<b>WT</b>	<b>0.58</b>	<b>WT</b>	0.59	R97E	0.59	F94L	<b>0.25</b>	<b>F22L</b>	<b>0.06</b>	<b>V17A</b>	0.49	F22L	<b>0.71</b>	<b>V20L</b>
0.51	V20L	0.53	A30G	0.56	S43E	<b>0.57</b>	<b>WT</b>	0.58	E29K	<b>0.25</b>	<b>Y25A</b>	<b>0.06</b>	<b>V20A</b>	<b>0.48</b>	<b>WT</b>	<b>0.71</b>	<b>V20S</b>
0.50	A30G	0.48	V20L	0.50	L89T	<b>0.57</b>	<b>E29D</b>	<b>0.57</b>	<b>WT</b>	<b>0.25</b>	<b>E29D</b>	<b>0.06</b>	<b>V20L</b>	<b>0.48</b>	<b>V17A</b>	<b>0.71</b>	<b>S21R</b>
0.42	F94L	0.48	E90H	0.45	E90H	0.43	L89T	<b>0.57</b>	<b>Y91Q</b>	<b>0.25</b>	<b>A30G</b>	<b>0.06</b>	<b>V20S</b>	<b>0.48</b>	<b>V20L</b>	<b>0.71</b>	<b>R23Q</b>
0.40	F22L	0.46	E29K	0.45	E29Q	0.42	Y25A	<b>0.57</b>	<b>S43E</b>	<b>0.25</b>	<b>S87T</b>	<b>0.06</b>	<b>F22L</b>	<b>0.48</b>	<b>V20A</b>	<b>0.71</b>	<b>E55Q</b>
0.40	E90H	0.44	L89T	0.42	V20L	0.41	V20L	<b>0.57</b>	<b>E55Q</b>	<b>0.25</b>	<b>L89A</b>	<b>0.06</b>	<b>S87T</b>	0.47	Y91Q	<b>0.71</b>	<b>R77E</b>
0.39	E29K	0.40	F94L	0.38	Y25A	0.37	V17A	0.55	E29D	<b>0.25</b>	<b>L89T</b>	<b>0.06</b>	<b>L89A</b>	0.47	F94L	<b>0.71</b>	<b>R97E</b>
0.37	S43E	0.38	F22L	0.37	E55Q	0.36	L89A	0.55	A30G	<b>0.25</b>	<b>Y91Q</b>	<b>0.06</b>	<b>L89T</b>	0.46	S87T	<b>0.71</b>	<b>R97Q</b>
0.34	E29D	0.35	S43E	0.34	Y98Q	0.34	F22L	0.55	R77E	<b>0.25</b>	<b>S92T</b>	<b>0.06</b>	<b>L89V</b>	0.45	E90H	0.70	E29K
0.32	L89T	0.32	E29D	0.32	F22L	0.33	S92H	0.54	L89V	<b>0.25</b>	<b>Y98Q</b>	<b>0.06</b>	<b>F94L</b>	0.45	V20S	0.50	K88Q
0.31	V20A	0.29	E29R	0.31	V17A	0.31	E29Q	0.53	K88N	<b>0.25</b>	<b>S8H</b>	<b>0.06</b>	<b>S8H</b>	0.44	Y98Q	0.49	L89V
0.28	L89A	0.27	L89A	0.30	F94L	0.29	S8H	0.45	L89A	<b>0.25</b>	<b>E29K</b>	<b>0.06</b>	<b>S21R</b>	0.44	K88N	0.48	L89A
0.26	V17A	0.27	V20A	0.26	S92H	0.28	Y98Q	0.45	L89T	<b>0.25</b>	<b>E29Q</b>	<b>0.06</b>	<b>K88N</b>	0.35	Y25A	0.48	L89T
0.24	Y25A	0.22	S8H	0.25	S8H	0.26	E55Q	0.40	R23Q	<b>0.25</b>	<b>E29R</b>	<b>0.06</b>	<b>K88Q</b>	0.35	E29D	0.47	E29D
0.23	E29R	0.22	S92H	0.22	L89A	0.25	F94L	0.38	V20A	<b>0.25</b>	<b>E55Q</b>	<b>0.06</b>	<b>S92H</b>	0.35	A30G	0.47	A30G
0.20	S8H	0.22	V20S	0.21	E29K	0.19	Y91Q	0.37	R97Q	<b>0.25</b>	<b>R77E</b>	<b>0.06</b>	<b>R97Q</b>	0.35	E29Q	0.47	F94L
0.20	S92H	0.20	V17A	0.21	E29R	0.14	V20S	0.32	V17A	<b>0.25</b>	<b>K88Q</b>	<b>0.06</b>	<b>Y91Q</b>	0.35	E29R	0.47	E29Q
0.20	V20S	0.18	Y25A	0.08	S21R	0.14	V20A	0.29	Y25A	<b>0.25</b>	<b>R97E</b>	<b>0.06</b>	<b>Y98Q</b>	0.32	R77E	0.47	E29R
0.13	Y91Q	0.14	Y91Q	0.05	V20S	0.05	E29K	0.28	V20S	<b>0.25</b>	<b>R97Q</b>	0.05	R23Q	0.30	S92H	0.24	S92H
0.10	Y98Q	0.10	Y98Q	0.00	Y91Q	0.05	E29R	0.15	R97E	0.23	R23Q	0.00	E55Q	0.02	S21R	0.23	S8H
0.00	S21R	0.00	S21R	0.00	V20A	0.00	S21R	0.00	Y98Q	0.00	S43E	0.00	E90H	0.00	S8H	0.00	S87T

\*Mutants coloured red, green or black have aggregation propensity scores higher than, lower than, or equal to pWT SOD1, respectively.

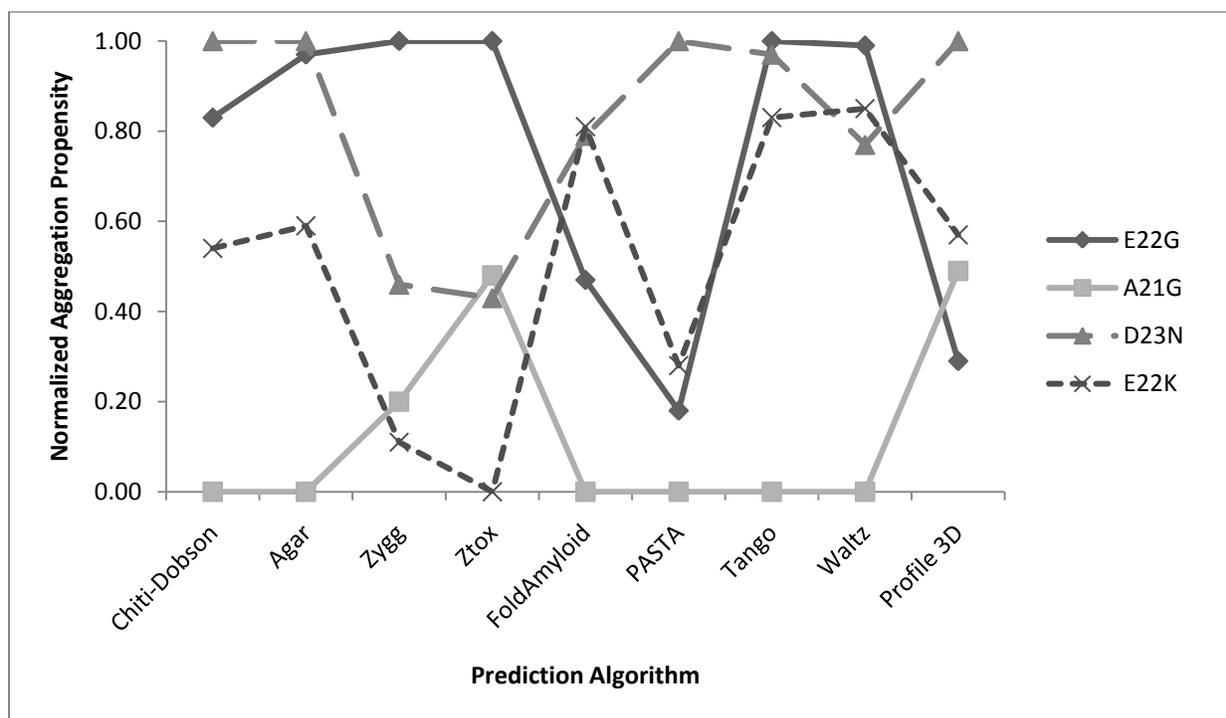
A comparison of the regions predicted to be most aggregation-prone was done by creating a hot-spot profile. As for SOD1 (Figure 2.9), the sum of the normalized aggregation scores for each residue as predicted by seven of the prediction algorithms (the Chiti-Dobson and Wang-Agar methods are not used, see Section 2.2.3) for wild type AcP was plotted against the residue number. Figure 2.11 displays the hot-spot profile as a stacked bar chart. Similarly to SOD1, the aggregation-prone regions for human acylphosphatase (AcP) are determined by identifying regions predicted by multiple algorithms as being prone to aggregate. Residue 34 is predicted as aggregation-prone by all seven algorithms and residues 35-39 are predicted to be prone to aggregation by six out of seven algorithms. Residues 95 and 96 are predicted to be aggregation-prone by five out of seven algorithms, and several others are identified by four algorithms.



**Figure 2.11 Predicted aggregation-prone regions of ACP by seven different prediction algorithms.** The normalized aggregation prediction scores for each residue are plotted as a stacked bar chart. The total height of each bar represents the sum of the aggregation scores from all 7 algorithms. Each bar is divided by colour based on the contribution from each algorithm.

### 2.3.3 Predicting Aggregation Propensities of Amyloid Beta Peptide Mutants

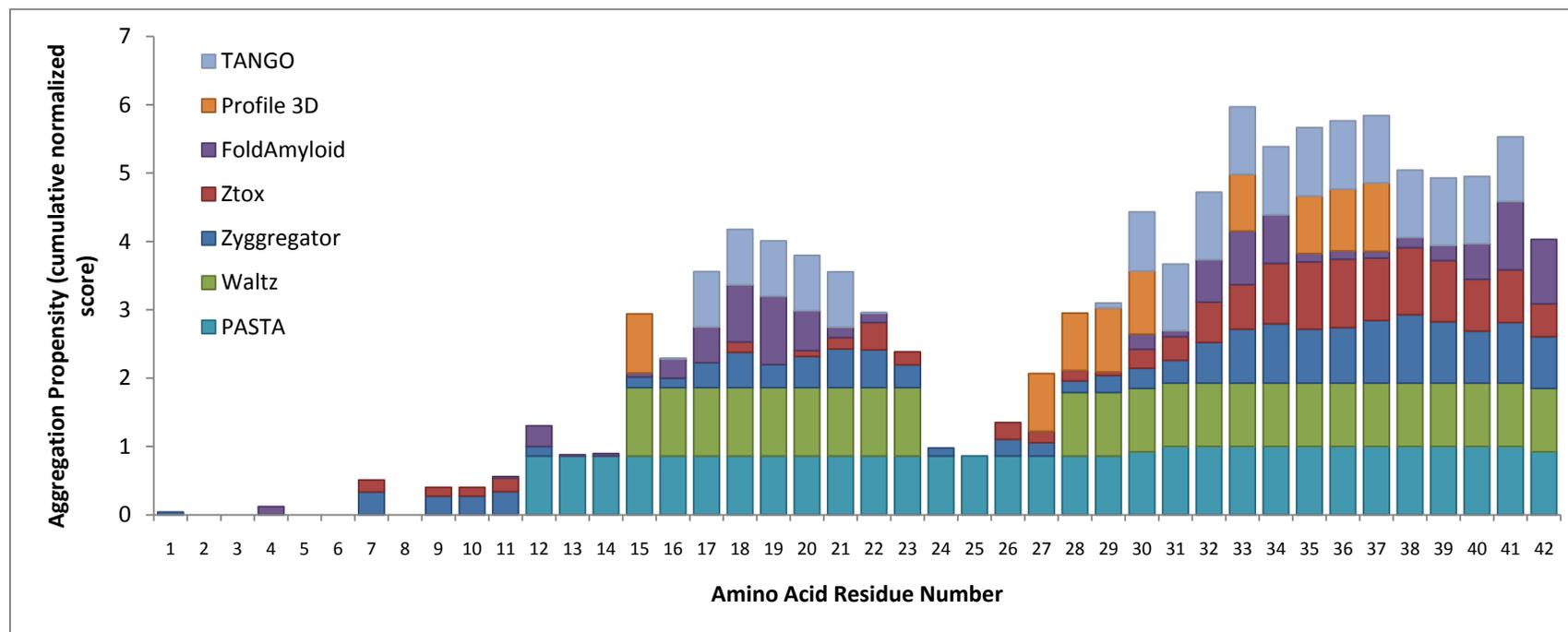
Aggregation propensities were calculated for WT and five amyloid beta peptide mutants. The normalized propensity results are given in Table 2.5. In general, WT is predicted to be the least, or second least, aggregation-prone sequence. However, the predicted aggregation propensities of the five mutations relative to WT vary widely. Figure 2.12 plots the aggregation propensity for four  $A\beta_{42}$  mutants for the nine prediction algorithms. E22G is consistently ranked as highly aggregation-prone by six of the nine algorithms whereas A21G has the lowest tendency to aggregate in six out of nine prediction techniques. D23N and E22K vary significantly in predicted propensity.



**Figure 2.12 Predicted aggregation propensity of amyloid beta mutants displays significant variability between algorithms for the same mutant.** The predicted aggregation propensities for four  $A\beta_{42}$  mutations were calculated using nine different prediction algorithms. There are considerable differences between prediction results for the same mutant.

**Table 2.5 Normalized aggregation propensities for wild type and 5 amyloid beta mutants using nine prediction algorithms.**

Chiti-Dobson		Wang-Agar		Zyggregator		Ztox		FoldAmyloid		PASTA		Tango		Waltz		Profile 3D	
1.00	D23N	1.00	D23N	1.00	E22G	1.00	E22G	1.00	E22Q	1.00	D23N	1.00	E22G	1.00	E22Q	1.00	D23N
0.83	E22G	0.97	E22G	0.73	E22Q	0.54	E22Q	0.81	E22K	0.31	E22Q	0.98	E22Q	0.99	E22G	0.57	E22K
0.68	E22Q	0.74	E22Q	0.46	D23N	0.48	A21G	0.79	D23N	0.28	E22K	0.97	D23N	0.85	E22K	0.50	E22Q
0.54	E22K	0.59	E22K	0.20	A21G	0.43	D23N	<b>0.76</b>	<b>WT</b>	0.18	E22G	0.83	E22K	0.77	D23N	0.49	A21G
<b>0.19</b>	<b>WT</b>	<b>0.17</b>	<b>WT</b>	0.11	E22K	<b>0.06</b>	<b>WT</b>	0.47	E22G	<b>0.16</b>	<b>WT</b>	<b>0.81</b>	<b>WT</b>	<b>0.44</b>	<b>WT</b>	0.29	E22G
0.00	A21G	0.00	A21G	<b>0.00</b>	<b>WT</b>	0.00	E22K	0.00	A21G	0.00	A21G	0.00	A21G	0.00	A21G	<b>0.00</b>	<b>WT</b>



**Figure 2.13 Predicted aggregation-prone regions of amyloid beta by seven different prediction algorithms.** The normalized aggregation prediction scores for each residue are plotted as a stacked bar chart. The total height of each bar represents the sum of the aggregation scores from all 7 algorithms. Each bar is divided by colour based on the contribution from each algorithm.

The regions predicted to be most aggregation-prone by each algorithm for WT A $\beta$ <sub>42</sub> are displayed in Figure 2.13. There is an obvious general consensus that the N-terminus of amyloid beta is not predicted to aggregate. Residues 33 and 35-37 are predicted to be aggregation-prone by all seven algorithms. Residues 18 and 20-22 are predicted as aggregation-prone by six out of seven algorithms. Several other residues are predicted to be aggregation-prone by 4 or 5 algorithms.

## **2.4 Discussion**

There are two main approaches in predicting protein aggregation. The first is the calculation of an overall aggregation propensity for a protein sequence. The second is the identification of the regions within a protein sequence that are most likely to be involved in aggregation. The algorithms used here were analyzed for both purposes and the results are discussed in the sections below. To date, this is the first broad comparison between such a large selection of algorithms. The ability of the algorithms to consistently predict a consensus in the overall propensity of a sequence to aggregate was quite weak. Based on the fact that the algorithms were developed by varying strategies and tested for the ability to predict different types of aggregation (see Section 2.1), the large inconsistencies in the overall predicted propensity for an identical sequence by multiple algorithms are not completely surprising. However, the extent of these differences, even for proteins that readily form amyloid, is greater than expected. This will be discussed in Section 2.4.1. In Section 2.4.2 a comparison of the ability of several algorithms to generate a consensus on the region within a protein that is most prone to aggregate is analyzed. Much greater success is observed by the algorithms for this technique than in predicting overall aggregation propensities. This may prove useful in identifying regions susceptible to promoting protein aggregation.

### **2.4.1 Inconsistent Patterns in Predicted Aggregation Propensities between Algorithms**

Extensive calculations of predicted aggregation propensity were performed using various techniques to look for consistency in the predictions made by multiple algorithms. It would be expected that if a certain mutant was predicted as highly aggregation-prone by multiple methods, it may enhance the likelihood that this sequence is prone to aggregate experimentally. However this is complicated by the

fact that different algorithms were designed to predict different types of protein aggregation (*vide infra*). Based on the role of aggregation in neurodegenerative disease pathology, the validation of predicted mutant aggregation propensity using multiple techniques could be particularly important to better understand the role of mutant aggregation in disease pathogenesis. To date, a comparison of this size between such a large number of algorithms and using a broad range of mutations had not been made. Somewhat surprisingly, there were significant deviations in the aggregation propensity scores for the same mutant of a certain protein by multiple prediction algorithms. This might suggest that the differences in design of these algorithms prevent consistent prediction scores for the same sequence. Table 2.6 compares the methods by which each algorithm was tested in order to confirm the validity of the algorithm. Eight out of nine algorithms (with the exception of Ztox) were verified by the formation of amyloid. Four of these algorithms (FoldAmyloid, TANGO, Waltz and Profile 3D) were verified based on amyloid formation by peptides. In the case of FoldAmyloid and PASTA, the algorithms were used to predict whether or not a short peptide sequence would form amyloid and this was tested experimentally. For Waltz and Profile 3D full length native protein sequences were used as input sequences to determine regions with high amyloidogenicity. Then peptides were acquired with the sequence of the regions highly predicted to aggregate and tested experimentally for amyloid formation. Based on the diversity in the prediction methods and the methods by which they were designed and tested, it might be expected that some of the algorithms would perform quite differently for the same sequence, while others may be more likely to give similar results because of commonalities in algorithm testing and design. This was additionally confirmed by performing several correlations between the prediction results for SDO1 by multiple algorithms. Most algorithms did not give statistically significant, strong correlations, when plotted against each other, confirming that the algorithms show scattered results in the predictions of identical sequences.

**Table 2.6 Comparison of the methods used to test the validity of protein aggregation prediction techniques.**

Method of Algorithm Verification	Prediction of Amyloid formation by Unfolded Protein	Prediction of Oligomer formation by Unfolded Protein	Prediction of Amyloid formation by hexapeptides	Prediction of amyloid-prone regions of Functional Amyloids	Prediction of amyloid-prone regions of Folded Protein
<b>Algorithms</b>	Chiti-Dobson Wang-Agar Zyggregator PASTA	Ztox	FoldAmyloid PASTA TANGO	Waltz	Profile 3D

Table 2.3 shows the numerical aggregation scores and the overall ranking for all SOD1 mutants studied. The pattern of highest to lowest aggregation propensity is different for every algorithm. In addition, the normalized scores vary widely for the same mutant, indicating that the degree of change in propensity as a result of a single mutation is quite different for different algorithms. For example the pWT SOD1 construct ranges from an aggregation score as low as 0.39 by Profile 3D to as high as 0.83 by WALTZ. Figure 2.8 further illustrates the diversity in aggregation predictions by demonstrating the range seen for four specific mutants between algorithms. It is interesting that while one mutation, A4V, scores somewhat consistently high, another, H43R, ranges quite significantly from 0 to 1. These discrepancies show there is a lack of consensus between the algorithms on which factors might govern the general principles contributing to protein aggregation.

To further examine the origin of the diverse range of prediction scores observed for different mutants, the algorithms were tested on more than one protein. Because many of the algorithms were developed using datasets of unfolded protein and short peptides, and many included aggregation data from AcP in unfolded conditions and A $\beta$ <sub>42</sub> (a short peptide with little secondary structure), both AcP and A $\beta$ <sub>42</sub> were chosen for additional analysis. An additional point for consideration is that most of the algorithms (see Table 2.1) were developed based on ThT data, a dye that binds specifically to amyloid. Since SOD1 is not a typical amyloid-forming protein, this may also contribute to the lack of consensus seen in the aggregation prediction results for SOD1. AcP forms ThT-binding aggregates at pH 5.5, while

A $\beta_{42}$  readily form amyloid aggregates at pH 7, and thus it is expected that these proteins may show more consistent aggregation predictions. Somewhat surprisingly, there are considerable differences in the predicted aggregation scores between different algorithms used for the analysis of the same mutant, and in the overall observed patterns of predicted propensities, for AcP mutants when calculated using the nine described techniques. Table 2.4 shows the propensity rankings for each AcP mutant by all algorithms, and while there are some mutants that consistently score at the top or the bottom of the list, there are many that vary widely. A similar pattern of variation is observed for the aggregation predictions of A $\beta_{42}$ . There may be a slight improvement in consistency of the ranking pattern, likely because of the short peptide is less complicated than a full protein and it may be easier to accurately predict aggregation; however, the range of aggregation scores is still quite high, with WT ranging from 0 to 0.81. Figure 2.12 displays a similar lack of consensus for the propensity of predicted aggregation for four A $\beta_{42}$  mutants between the nine algorithms despite the fact that this is a characteristic amyloid-forming peptide. As noted in Table 2.1, all algorithms except Ztox, were designed based on amyloid-forming proteins or peptides, and as described in Table 2.6, almost all algorithms were tested with amyloid forming peptides and unfolded proteins. Thus, the inconsistency in predictions for AcP and A $\beta_{42}$  is somewhat surprising; however it may again be an indication that the properties controlling the different types of aggregation used for the development of each algorithm are not identical.

The discrepancy in aggregation scores and patterns observed for mutations of three different proteins suggests an overall lack of consensus of the major contributing factors controlling the protein aggregation processes predicted by each of these algorithms. Table 2.1 lists the variables included in each algorithm and Table 2.6 summarize the types of aggregation that were predicted by each method. Some factors such as hydrophobicity, charge and beta-sheet propensity are included in multiple algorithms, while others, such as hydrogen bonding, are considered by only one of the algorithms. In addition, the role of sequence order or aggregate structure, as incorporated through different methods by some of the techniques, will also play a role in the outcome of the prediction. Thus, as has been stated previously, the large differences in predicted aggregation propensities for the same sequence between algorithms are

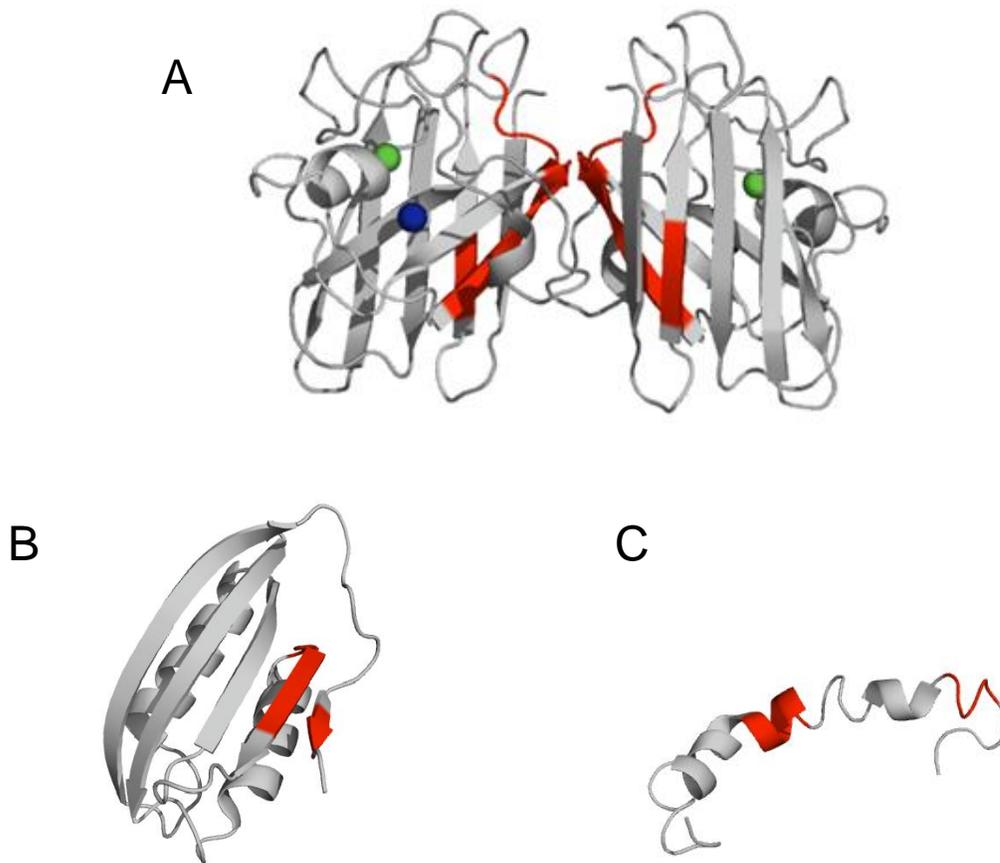
perhaps not surprising. The more fundamental question then is which of these algorithms, or which components within these various algorithms, are most important to aggregation propensity, and what are the differences in the variables that contribute to different types of aggregation.

The context in which each algorithm was designed strongly influenced which factors were included in the algorithm. For instance, the Chiti-Dobson equation was based on ThT-fluorescence of aggregates formed from unfolded polypeptide sequences, and so the physical and chemical properties of the amino acids within the sequence are likely to play the predominant role in controlling aggregation. On the other hand, the Waltz algorithm was based on amyloid-forming hexapeptides, and so the sequence order of 6-amino acid long stretches within a protein are of particular importance to predicting whether these short regions may contribute to amyloid formation by the protein in question. This algorithm also treats each hexapeptide region as an individual unit available to aggregate, and does not consider its context within the entire protein structure. In these examples, and in the context-specific design of each algorithm, it may be true that the factors controlling one system are different than the factors controlling another system. For example, the features that cause an unfolded polypeptide sequence to aggregate may be different than the factors that promote amyloid formation of a short hexapeptide, or amorphous aggregation by a natively folded protein. The next chapter will investigate a comparison of predicted and observed aggregation to look at the ability of each algorithm to predict experimental *in vitro* aggregation of these three proteins.

#### **2.4.2 Overlapping prediction hot-spots using multiple algorithms**

An important component in predicting protein aggregation is the ability to identify which region of the protein is contributing to the tendency to aggregate. Seven out of nine algorithms were used to identify consensus regions of high aggregation propensity for the wild type proteins involved in this study. In this way the ‘hot-spots’ for aggregation can be predicted by looking for regions that are identified as potential aggregation-prone areas by multiple algorithms. There was much greater success identifying aggregation hot-spot regions by comparing prediction algorithm results than finding a

consensus in absolute overall aggregation propensity as described in the previous section. Figures 2.9, 2.11 and 2.13, give the hot-spot maps for the SOD1, AcP, and A $\beta$ <sub>42</sub>, respectively. For example the C-terminal of SOD1 is strongly predicted to aggregate according to the cumulative aggregation scores. Six out of seven algorithms predict this region as aggregation-prone. There are also several stretches with a cumulative score of zero, meaning that none of the algorithms predict this region to be aggregation-prone. Similar results were obtained for both AcP and A $\beta$ <sub>42</sub> where at least 5 algorithms predict the same region to be aggregation-prone and stretches exist where no algorithm predicts aggregation. This is important because it shows that despite the general inconsistencies in predicted overall probability of a polypeptide sequence to aggregate, there is some consensus in identifying the regions of the protein that are most likely to be involved in aggregation. Figure 2.14 shows crystal structures of native holo SOD1, folded human AcP, and A $\beta$ <sub>42</sub> with the adopted secondary structure when associated with a membrane<sup>97</sup>, and shows the aggregation hot-spot regions mapped onto the structures. For SOD1 and AcP the hot-spot regions are predominantly beta strands. This is not surprising since amyloid adopts a cross beta fibrillar structure. The fact that these seven algorithms are somewhat able to converge upon a few short sequences as the consensus regions for high aggregation propensity provides evidence for the potential ability of prediction techniques to locate vulnerable sections within a protein. For one of these segments, experimental results have previously confirmed the amyloidogenic nature of these regions. For example, the 36-40 amino acid segment of A $\beta$ <sub>42</sub> was predicted to be aggregation prone by Teng *et al.*<sup>98</sup>. This was verified experimentally by inserting this peptide sequence into RNase A and demonstrating that this resulted in amyloid formation by this RNase A construct. The consensus results from these experiments identify the C-terminus as being highly aggregation prone, with residues 35-37 being predicted as aggregation prone by all seven algorithms. Additional experimental results are required to confirm the amyloidogenicity of the other identified hot-spot regions. Further analysis to decipher whether there is increased consistency in the aggregation-prone regions predicted by algorithms that were developed based on similar data sets may also provide extended information as to whether certain regions are particularly important in one type of aggregation versus another.



**Figure 2.14 Aggregation-prone hotspots of three different proteins.** Portions in red indicate the protein regions with the greatest propensity to aggregate and are determined based on calculations of predicted aggregation propensities from seven different algorithms (see Table 2.1. All algorithms except for Chiti-Dobson and Wang-Agar were used for this comparison). A) Human Cu, Zn Superoxide Dismutase, with zinc shown in green and copper in blue (PDB 1SOS). B) Human Acylphosphatase (PDB 2VH7). C) Amyloid beta peptide (PDB 1Z0Q). The protein backbones are shown in ribbon representations generated using the program Pymol<sup>99</sup>.

### 2.4.3 Current Barriers in Accurate Aggregation Prediction

Accurately predicting protein aggregation remains an extremely difficult task. It is an important undertaking, however, because it provides an opportunity to test current understanding of what causes protein aggregation and what are the most important variables that affect this process. It may also allow

for the identification of proteins and mutants with a high risk of aggregating and provide a basis for the determination of the regions within a protein that are responsible for aggregation.

The heterogeneity of the aggregation process is a difficult problem when it comes to designing specific aggregate prediction tools, as it is possible that similar factors may be involved in controlling more than one aggregate pathway. Most of the current algorithms focus on the prediction of amyloid formation and were designed and tested based on amyloid forming proteins and peptides (See Table 2.6). However, amyloid is only one of many possible conformations that may result from protein aggregation. The molecular mechanisms resulting in protein aggregation are not well understood; however, it is clear that there are many potential pathways that can result in various aggregate morphologies (see Section 1.2). Some aggregate forms may be a component of a longer pathway, such as prefibrillar species that will further transform into mature amyloid fibrils, while others may be a separate endpoint in themselves, such as the formation of soluble oligomers<sup>11</sup>. Some algorithms intentionally try to maintain a narrow scope of prediction, such as Waltz, which specifically aims to predict only amyloid formation, whereas Tango aims to predict all forms of beta-sheet aggregation. These differences likely play a role in the inconsistencies observed for absolute aggregation scores for a given mutant. A major challenge, then, in the development of improved prediction algorithms is the ability to specify which type of aggregation is being predicted, and to find techniques to identify and measure the different types of aggregates being formed within an *in vitro* test system.

A second major difficulty in the effective design of aggregation prediction methods is the inclusion of native protein structure into the prediction method. All of the prediction algorithms included in this study do not take into account the native protein structure in the prediction calculation. This is important to consider because if the input sequence is a natively folded, extremely stable protein, than even if there is a highly aggregation-prone region buried in the core of the protein, it may not follow that the protein has an increased tendency to aggregate if it is uncommon for this region of the protein to be exposed<sup>100</sup>. A handful of groups have tried to address this issue by the incorporation of native structure, predominantly by the prediction of native structure for the given sequence, then the use of that structure in

the design of the prediction algorithm<sup>85</sup>. Another way to include some form of structure in aggregation prediction is to factor in protein stability. A measure of stability would inherently include a measure of the probability of a protein to persist in the unfolded or folded state, and this is expected to be an important contributing factor in protein aggregation<sup>23</sup>. This would assume, however, that a less stable protein is more likely to be unfolded, and that aggregation occurs from an unfolded state. Unfortunately, this is not always the case as aggregation is known to occur from both folded and unfolded states<sup>10; 29</sup>. Regardless, stability is known to correlate with aggregation rates and thus is an important factor in controlling aggregation<sup>65; 100; 101</sup>. The development of ways to more accurately represent protein structure in aggregation prediction tools should dramatically enhance propensity calculations.

Another question at the root of many of these prediction algorithms is whether or not a peptide region that is amyloidogenic as an individual peptide will cause the protein it is a part of to form amyloid. This is a fundamental issue that needs addressing, particularly considering the fact that several of the algorithms discussed here were developed based on and/or tested on peptide sequences. This is not an easy question to address as it depends on numerous other factors including the location of the peptide on the protein structure. If the region is buried inside the protein, it will likely not be a principal contributor to the aggregation tendency of the entire protein, whereas, if the amyloidogenic region is located on the surface of the protein it could play a predominant role<sup>10; 102</sup>. Thus, the assumption of the algorithms that are based on amyloid forming peptides (FoldAmyloid, TANGO, Waltz and Profile 3D) in identifying aggregation hot-spots is that a peptide sequence that forms amyloid *in vitro* will also promote aggregation when in the context of the native protein structure. There is some evidence to support this<sup>31; 98; 103</sup>, and other evidence that contests this notion<sup>100</sup>. More information is required to confirm or alter this hypothesis and this will be extremely important to the advancement of aggregation prediction technology.

Finally, the solution conditions are an additional contributing factor not well incorporated into most of the prediction algorithms. As discussed in 1.3.3, the temperature, ionic strength, pH and surface exposure can play significant roles in modulating the initial state and the aggregation patterns of a protein<sup>8; 22</sup>. These conditions can influence protein stability, charge, electrostatic and hydrophobic

interactions, and protein dynamics. Solution conditions can be used to direct the type of aggregate formed, and the rate of aggregation, of a protein in solution<sup>16</sup>. For these reasons, it is particularly important to consider the conditions in which a protein is in and how they will contribute to the state of the protein and the promotion of protein-protein interactions (discussed further in Chapter 4).

Given the complexity of the factors involved in controlling and mediating the aggregation process, it may not be possible to develop a prediction algorithm capable of predicting all types of protein aggregation. What may become a more promising and favourable research path is the development of specific algorithms for various types of aggregation pathways, allowing for the determination and differentiation of the aggregation patterns of different proteins. Thus instead of focusing on the general principles of protein aggregation it may be important shift the focus to very specific contexts of aggregation to look for the factors that influence a specific type of aggregation, then comparison of the factors between different aggregate types will provide information regarding how the contribution of different variables varies between the formation differing aggregate structures.

#### **2.4.4 Conclusion**

There is great diversity in current aggregation prediction techniques. Each method has potential strengths and weaknesses in its ability to predict the aggregation tendencies of protein sequences. Considering the variability in the components included in each algorithm, it is evident that the factors responsible for controlling the many pathways of aggregation are not well understood. Major differences were observed when comparing the prediction results by all nine methods for the same input sequence. This was true for all three proteins studied, suggesting that although each algorithm appears to be successful within the context it was generated, there are major limitations when attempting to apply the same principles over a broader range of proteins. Despite the inconsistencies in predicting overall aggregation tendencies, the algorithms do tend to converge in indentifying regions within a protein responsible for aggregation, which is a positive step in understanding the aggregation process. For these reasons, additional research is required to systematically test the variables contributing to aggregation and

determine whether there are differences in contributing factors depending on the type of aggregation observed.

## Chapter 3

### Correlating Predicted and Observed Aggregation

#### 3.1 Introduction

Prediction of protein aggregation is a potentially powerful tool that may further understanding of the mechanisms contributing to protein aggregation pathways. Comparison between predicted and observed aggregation provides a tangible way to test whether the variables included in the prediction algorithms control the experimental aggregation in the system being tested. Deciphering the detailed mechanisms controlling protein aggregation is particularly important in neurodegenerative disease research to elucidate the patterns contributing to the formation of toxic aggregates. A more comprehensive understanding of the mechanisms controlling protein aggregation pathways has potential to provide a basis for the development of much needed therapeutic techniques to interfere with and prevent aggregate formation.

A handful of studies have utilized protein aggregation prediction algorithms to compare predicted aggregation propensities of wild-type and mutant proteins with observed aggregation and/or toxicity for a variety of disease-linked proteins. Few, however, have taken a quantitative approach to calculate and correlate the predicted and observed aggregation values for several mutants of the same protein. A study by Luheshi *et al.*<sup>104</sup>, used a *Drosophila melanogaster* model for Alzheimer's disease with a range of amyloid beta mutant peptides and correlated the predicted aggregation propensity of these peptides to the lifespan of the flies<sup>104</sup>. In this case, the predicted aggregation propensity using both the Zyggregator and Ztox methods correlate well with the life span of the mutant flies. The correlation with Ztox, an algorithm designed specifically to predict the formation of oligomeric, pre-fibrillar structures, is stronger than the correlation with Zyggregator, suggesting an important role for pre-fibrillar aggregates in mediating aggregate toxicity<sup>104</sup>. However, this study did not directly correlate the prediction algorithms with observed aggregation, but instead assumed that fly death is dependent on amyloid beta aggregation and is a measure of the mutant aggregation tendencies. Another study investigated the effects of mutations on

the aggregation patterns of  $\beta$ 2-microglobulin and compared the results with a series of aggregation prediction techniques<sup>100</sup>. The lag and elongation times of fibril formation were measured using Thioflavin T (ThT) fluorescence for a series of point mutations. Experiments showed that mutations in only one short stretch of the sequence had the greatest effects on aggregation rates. This was in contrast to the results of three prediction methods, Zyggregator<sup>85</sup>, Tango<sup>87</sup> and Aggregscan<sup>105</sup>, which predicted two main regions of high amyloidogenic potential. The discrepancy between the experimental and predicted results was attributed to the role of native intramolecular interactions in the second region. Because this part of the protein is particularly important in forming the intramolecular contacts involved in the native folded protein, it was predicted to also be involved in protein aggregation. However, experimental results revealed this was not the case<sup>100</sup>.

These studies show some evidence for correlations between predicted and observed aggregation results, however, in general the limited number of these types of studies indicate a need for further investigation. To the best of our knowledge there has yet to be a quantitative comparison between predicted and observed aggregation for several algorithms using multiple, diverse, test proteins. A recent study performed a comparison of several prediction algorithms by correlating predicted aggregation propensities of several A $\beta$ <sub>42</sub> mutants with observed *in vivo* aggregation<sup>106</sup>. Aggregation was previously measured in an *E. coli* system using green fluorescent protein (GFP)<sup>107</sup>. An analysis of twelve algorithms demonstrated that most of the algorithms were able to accurately predict the effect of a mutation on aggregation propensity *in vivo*. Similar positive correlations were observed for the relationship between predicted aggregation propensities for HypN-F mutants and aggregation observed by the measuring of the ratio of soluble/insoluble protein in an *E. coli* system<sup>106</sup>. While these results are quite promising, they are also quite limited as they are solely based on the prediction of two short amyloid-forming peptides, and further studies on a more diverse set of proteins is required. Here, the three test proteins described in Chapter 2 will be used to test the correlation between predicted and observed aggregation *in vitro*. The results will show discrepancies in predicted and observed aggregation and highlight ideas surrounding the apparent specificity of aggregation predictions based on the context in which the algorithms were

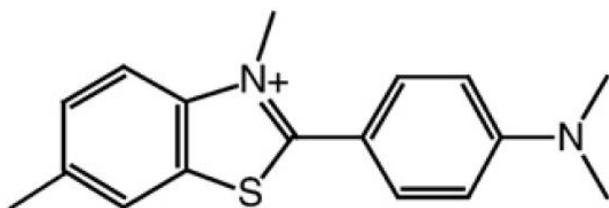
designed. Additionally, prediction methods will be used to investigate the nature of SOD1 aggregation in ALS disease pathogenesis.

### 3.1.1 Techniques for Measuring Protein Aggregation

Two methods will be used in this study to measure protein aggregation. Dynamic Light Scattering (DLS) utilizes scattered light to measure the amount and size of aggregated species in solution, while Thioflavin-T (ThT) is a fluorescent dye that can be utilized to quantify fibril formation in solution. DLS measurements for SOD1 mutants in the reduced apo form were acquired for this study in collaboration with Helen Stubbs in the Meiering lab<sup>27</sup>. ThT-fluorescence data from previously published work from other labs will be used for correlations between predicted and observed aggregation of AcP<sup>9</sup>; <sup>92</sup>, and A $\beta$ <sub>42</sub><sup>108; 109; 110</sup>. The aggregation prediction algorithms that will be used in this study were predominantly designed based on ThT data from amyloid-forming proteins (See Section 2.1). However, SOD1 is not a typical amyloid-forming protein and ThT data is not readily available for SOD1 mutants. Instead, reduced apo SOD1 has been shown to readily form soluble aggregates, which can be monitored by DLS<sup>27</sup>. For this reason, DLS data as opposed to ThT data will be used as the experimental results for observed SOD1 aggregation. These two methods for measuring aggregation are described in detail below.

Thioflavin T (ThT), shown in Figure 3.1, is a fluorescent dye that interacts with the beta-sheet structure of amyloid and amyloid-like aggregates<sup>111</sup>. It was first identified in 1959 as a histological amyloid marker<sup>112</sup>. Its fluorescence spectrum was further characterized in the 1980's and 1990's and found to exhibit a several-fold intensity fluorescence increase, an excitation maximum shift from 385nm to 450nm, and an emission maximum shift from 445nm to 482nm upon binding to amyloid fibres<sup>111</sup>. It was quickly determined that the shift and increase in fluorescence are linked to the beta-sheet morphology of amyloid fibres, allowing binding to occur with amyloids formed by any protein because the interaction is not dependent on the amino acid side chains but on the structural backbone interactions<sup>113</sup>. The specific molecular interactions of ThT with amyloid were far more elusive, and only recently have advancements been made to decipher the mechanism by which ThT interacts with amyloid fibres. It is suspected that in

solution the benzylamine and benzathiole rings of ThT freely rotate about their shared carbon-carbon bond (Fig. 3.1), thereby quenching the intrinsic fluorescence. However, upon interaction with a beta-sheet fibril, decreases in this rotation results in increased fluorescence. The most widely accepted model for ThT binding is that the side chains of amino acids involved in beta-sheet conformations form solvent-exposed channels that act as dye binding pockets<sup>113</sup>. Further investigation has suggested that ThT-interactions are predominantly mediated by aromatic side chains<sup>113</sup>. The interaction of ThT with beta sheet amyloid structures provides a useful tool to characterize the growth of amyloid fibrils. This is an extremely popular technique due to the relative simplicity of measuring the fluorescence increase. However, it is worth noting that because the specific mechanisms of ThT-amyloid interactions are not known, it is possible that ThT is not particularly specific to amyloid fibrils. Thus, interpretation of ThT data should be approached with a certain degree of caution. This chapter will utilize previously published aggregation rate data obtained from ThT-fluorescence experiments of several mutant proteins to explore the correlations between predicted and observed aggregation.



**Figure 3.1 Chemical structure of Thioflavin T<sup>111</sup>.**

Dynamic Light Scattering (DLS) is a useful technique for studying the size of macromolecules and their assemblies, both of which may be considered as light scattering particles. This method measures the amount of light scattered by the particles in a small volume of solution on a time scale of nanoseconds to seconds and directly correlates the intensity of light scattered to the motion of the particles<sup>114; 115</sup>. When an incident laser beam is directed at particles in solution this results in secondary wave formation, or in other words, scattered light<sup>115</sup>. A protein in solution will move randomly according to Brownian motion

and scatter light in all directions. When a detector is at a fixed angle and distance from the incident light beam relative to the sample, the scattered light received by detectors will fluctuate based on the motion of a particle in solution. The location of the particle may be such that the electromagnetic scattered light waves result in constructive or destructive interference<sup>116; 117</sup>. The change in the fluctuations in intensity of the scattered light received by the detector is fundamental to the ability of DLS to measure particle size. The decay times of the fluctuations in intensity can be related to the diffusion constant of the particle in solution<sup>116</sup>. The diffusion coefficient is dependent on the size and shape of the molecule. Therefore, the data received by the DLS detectors can give a great deal of information about the size and quantity of the particles present in solution<sup>114</sup>. During DLS measurement, the total acquisition time is divided into delay times ( $\tau$ ). An autocorrelation function is determined by plotting the averaged intensity measured during time  $\tau$  as a function of the time between  $\tau$ . From this, diffusion constants can be calculated and used to determine the hydrodynamic diameter of the species present in solution based on the Stokes-Einstein relationship<sup>116; 117</sup>. Importantly, light scattering intensity is dependent on the sixth power of the diameter, thus when only a small fraction of the percent mass of the sample is due to a high molecular weight species it can easily swamp the entire sample signal intensity<sup>116</sup>. The ability of this method to determine important information about particles in solution makes it to be a powerful tool in protein aggregation studies.

## **3.2 Methods**

### **3.2.1 Quantifying Observed Aggregation**

Two methods were utilized to quantify the overall observed aggregation of the proteins involved in this study. The first technique was Thioflavin T fluorescence. For AcP and A $\beta$ <sub>42</sub>, ThT data summarized and reported in Wang *et al.*, 2008 were used for comparison<sup>61</sup>. The rate of aggregation for mutants of these two proteins, as calculated from ThT fluorescence studies, were reported as  $\ln(v_{\text{mut}}/v_{\text{WT}})$ , as described in Section 2.1.1 and Chiti *et al.*<sup>9</sup>. The  $\ln(v_{\text{mut}}/v_{\text{WT}})$  values for all AcP and A $\beta$ <sub>42</sub> mutants were normalized over a scale of 0 to 1 based on the difference in observed aggregation rate between mutant and

wild type. Equation 3 in Chapter 2 was used for the normalization calculation according to the method described in Section 2.2.2. Thus observed aggregation was converted to a scale from 0 to 1, where 0 represents the lowest observed aggregation and 1 representing the highest observed aggregation of the data set. The second aggregation measurement technique was dynamic light scattering (DLS). The total light scattering intensity of particles >20 nm in diameter at the final time point in the aggregation time trials, as measured and described in Vassall *et al.*, 2011<sup>27</sup>, were used as a measure of overall observed aggregation for pWT SOD1 and 12 mutants. The originally published scale for the light scattering intensity measurements was from 0 to 100%, but this was adjusted to 0 to 1 by dividing each percentage by 100 for consistency with the normalized predicted aggregation results (see Section 2.2.2).

### **3.2.2 Production of Correlation Plots and Statistical Analysis**

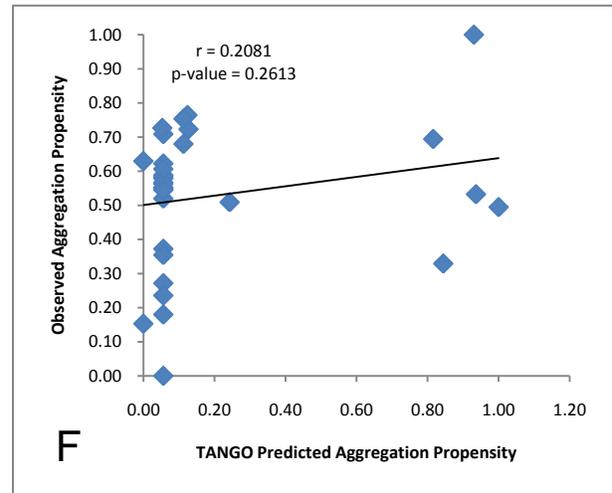
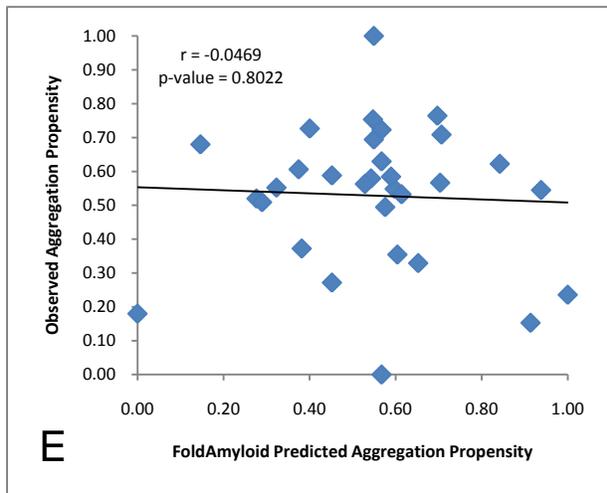
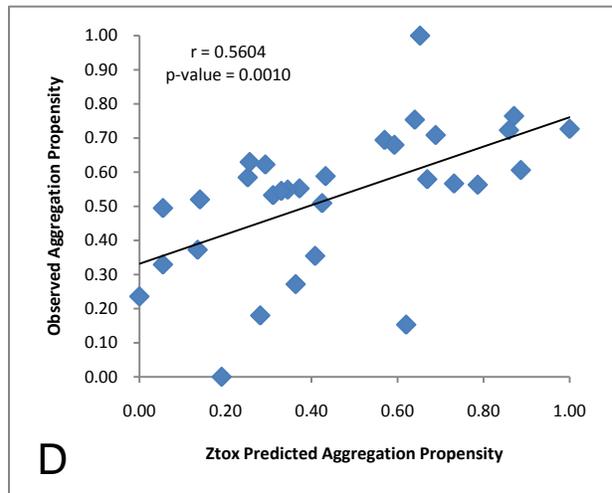
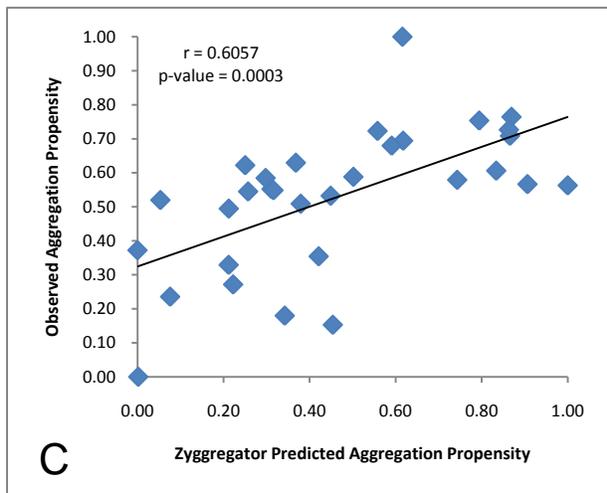
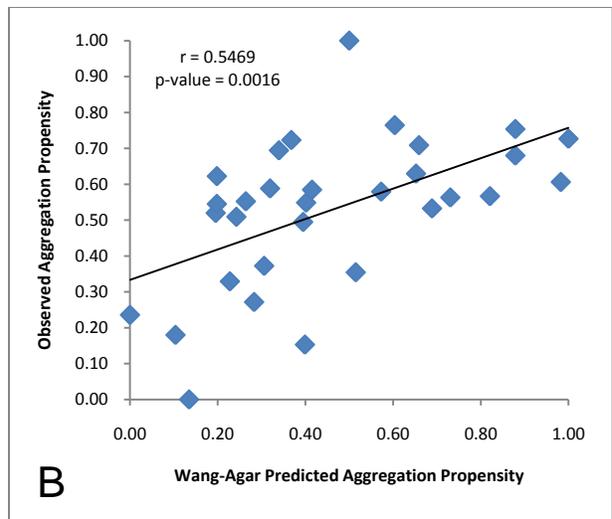
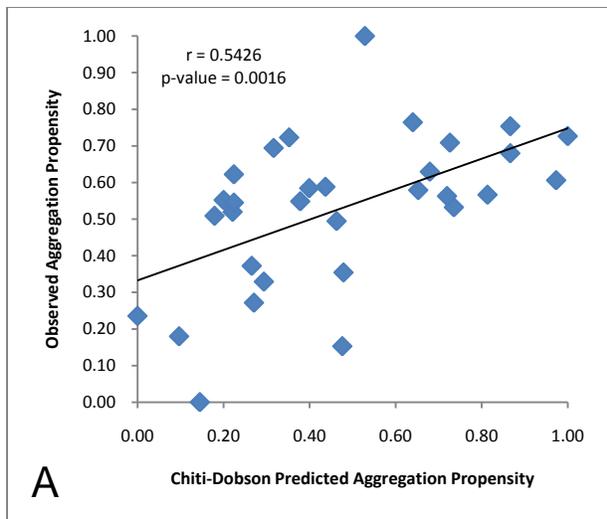
All correlation plots were prepared using Microsoft Excel 2007. Normalized observed and predicted aggregation values were compared using scatter plots. A line of best fit was obtained by linear regression analysis using Microsoft Excel, which also outputs a correlation coefficient value ( $r$ ) that describes of the linear dependence between two variables. A p-value, which reports the statistical significance of the linear correlation, was determined from the  $r$ -score and the degrees of freedom of the sample size using an online algorithm (<http://www.graphpad.com/quickcalcs/PValue1.cfm>). A confidence interval of 95% is typical required for determining statistical significance, which translates to a p-value < 0.05.

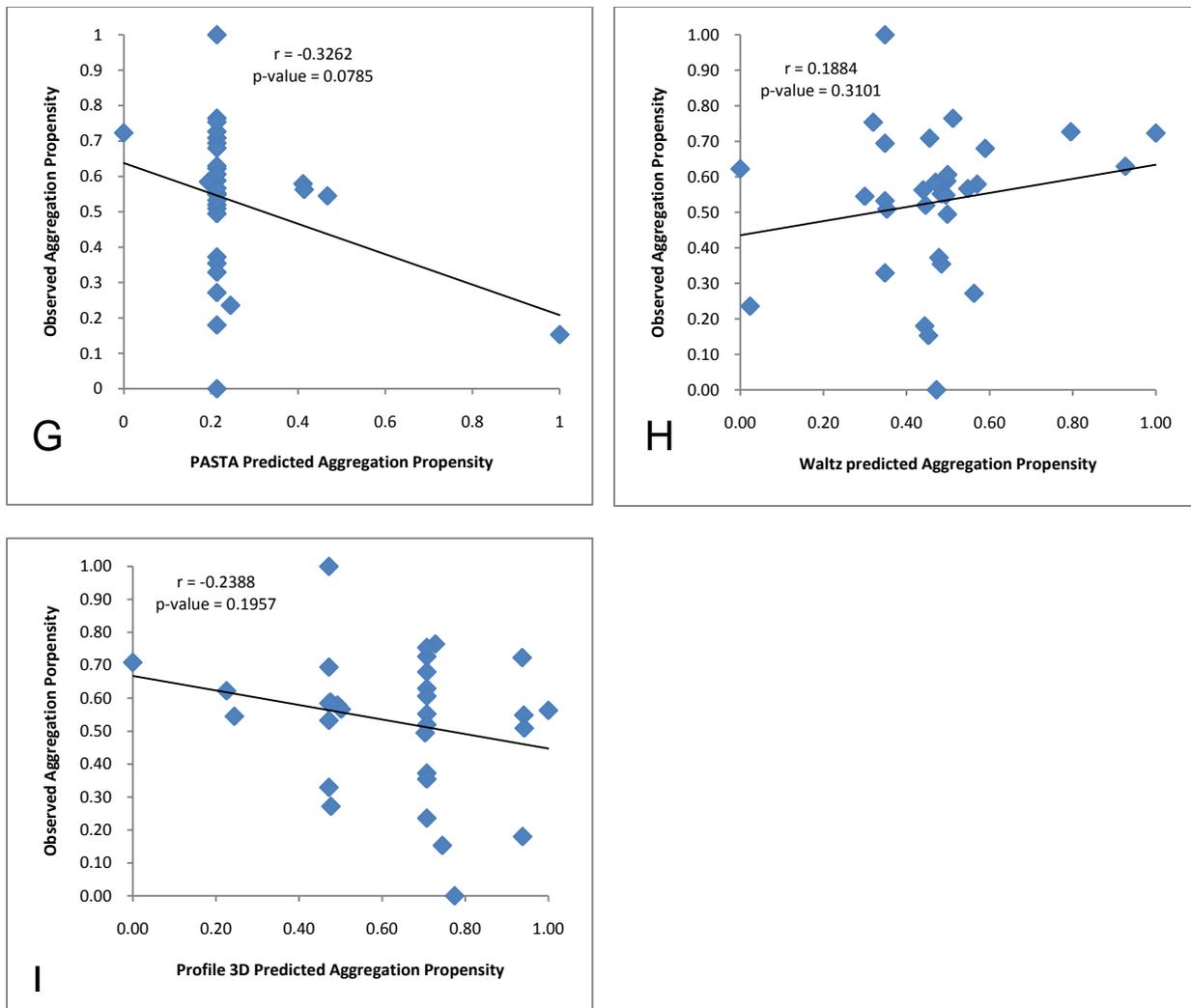
## **3.3 Results**

### **3.3.1 Human Acylphosphatase Correlations**

The predicted aggregation propensities for wild-type human acylphosphatase and 30 mutants (reported in Section 2.3 and listed in Table 2.4) were correlated with the observed aggregation for AcP as measured previously using ThT fluorescence<sup>91, 92</sup>. Figure 3.2 displays comparison plots for observed aggregation propensity *versus* predicted aggregation for all AcP mutants using the nine different algorithms. A summary of the  $r$ -scores and p-values for each algorithm is given in Table 3.1. The scores

that are considered to be statistically meaningful ( $p\text{-value} < 0.05$ ) are coloured red. Four out of nine algorithms produce statistically relevant results: Chiti-Dobson, Wang-Agar, Zyggregator and Ztox. However, even for these algorithms the correlations are not particularly strong, ranging from r-scores of 0.54 to 0.60. The other five algorithms do not result in statistically relevant correlations. Strikingly, three algorithms even result in negative (albeit not significant) correlations, where the mutations with lower predicted aggregation propensities have higher observed aggregation results.

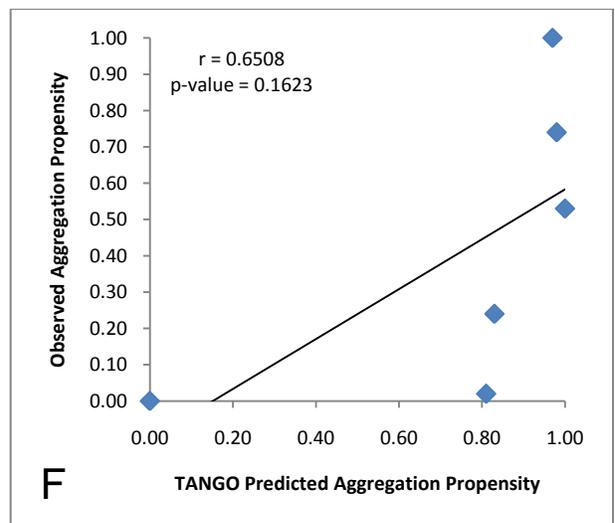
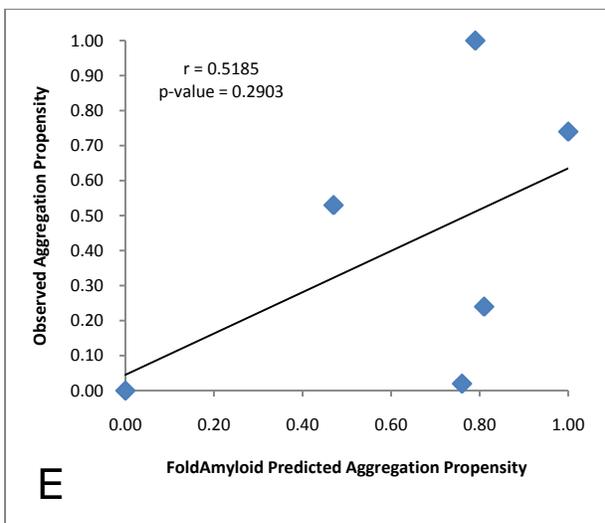
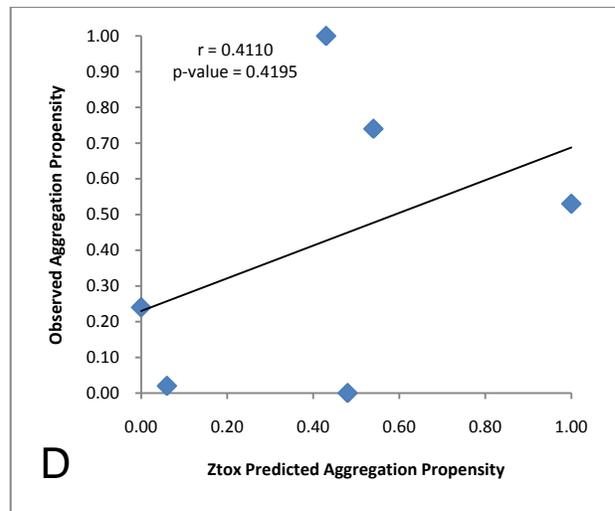
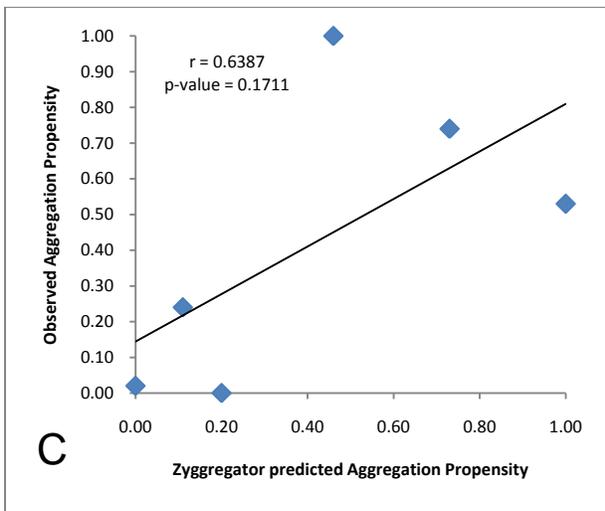
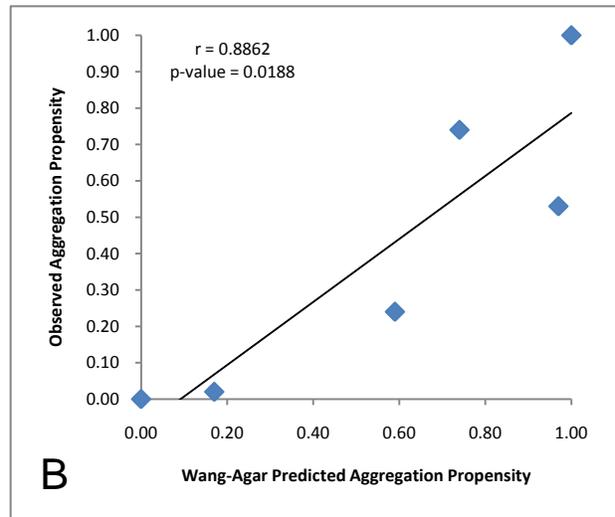
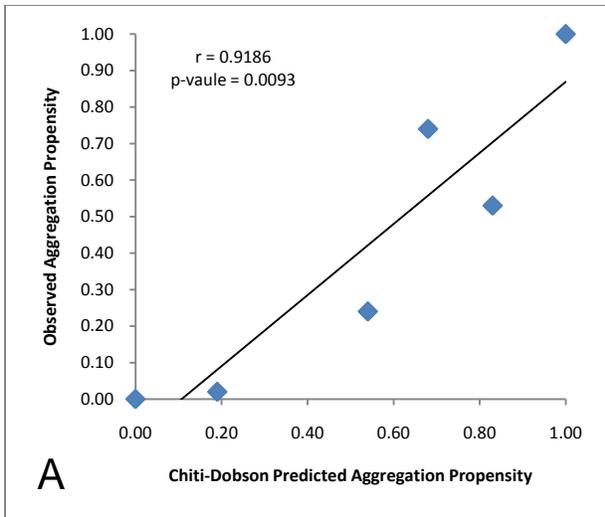


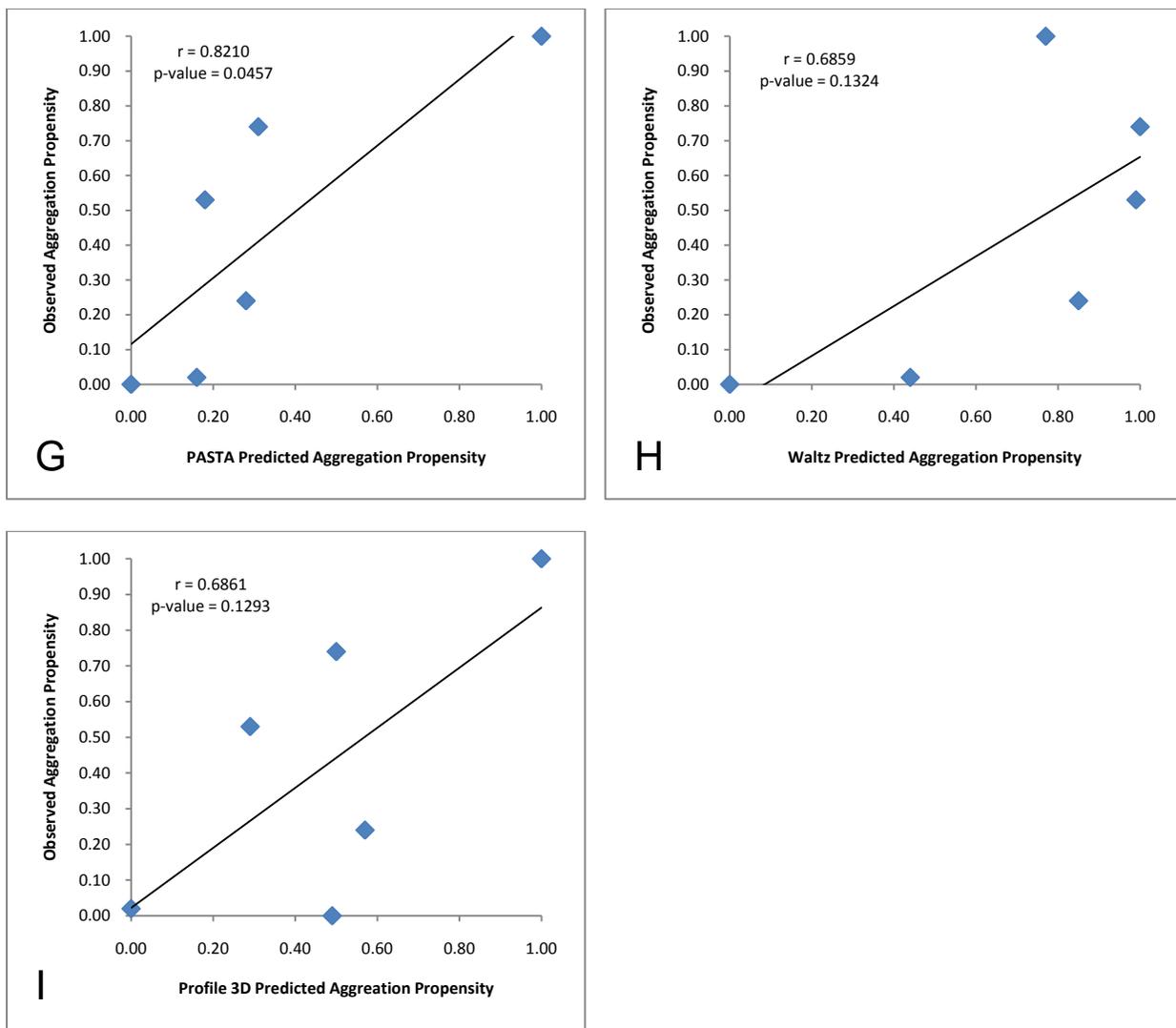


**Figure 3.2** Correlation plots of observed versus predicted aggregation for 30 AcP mutants using nine different prediction algorithms (listed in the ordinate axis labels). The data were fit to a straight line using linear regression.

### 3.3.2 Amyloid Beta Correlations

The predicted aggregation propensities for wild-type A $\beta$ <sub>42</sub> and five mutants linked to Alzheimer's Disease were calculated and are listed in Table 2.5. The relationship between predicted and observed aggregation was investigated using data reported previously on the aggregation rates of these mutants measured by ThT-fluorescence<sup>61; 108; 109; 110</sup>. Figure 3.3 shows the correlation plots between predicted and observed aggregation for A $\beta$ <sub>42</sub>. The correlations are summarized by r-scores and p-values in Table 3.1. Only three out of nine algorithms, including the Chiti-Dobson and Wang-Agar equations, and the PASTA method, result in statistically relevant correlations for A $\beta$ <sub>42</sub>. However, it should be noted that the extremely small sample size limits the likelihood of statistical relevance. The r-scores of the statistically significant correlations range from 0.82 to 0.92.

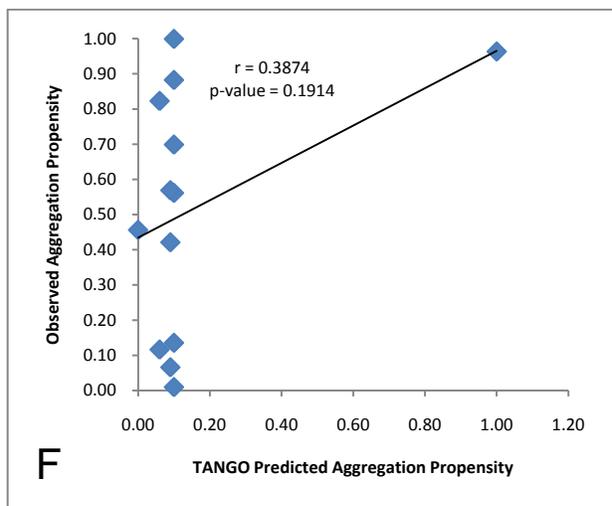
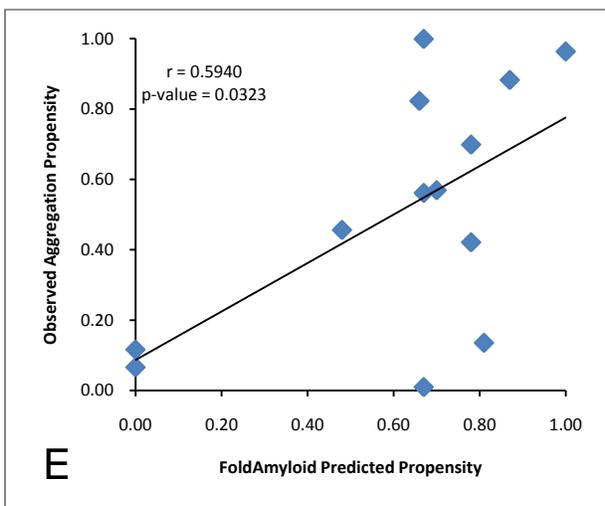
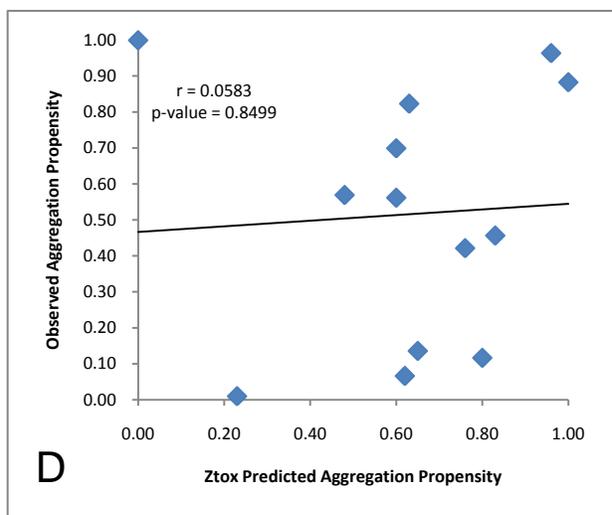
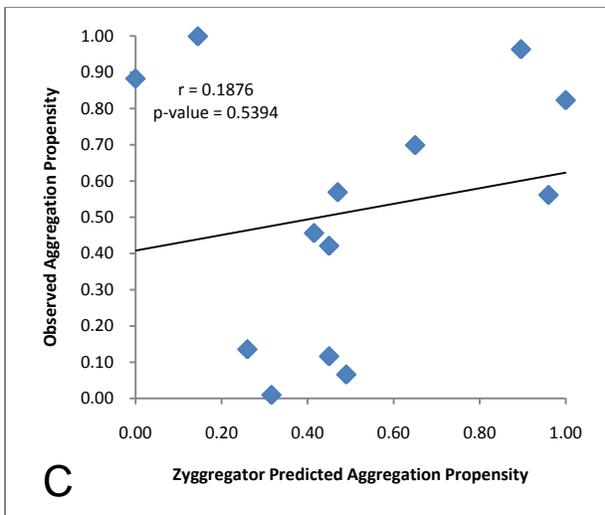
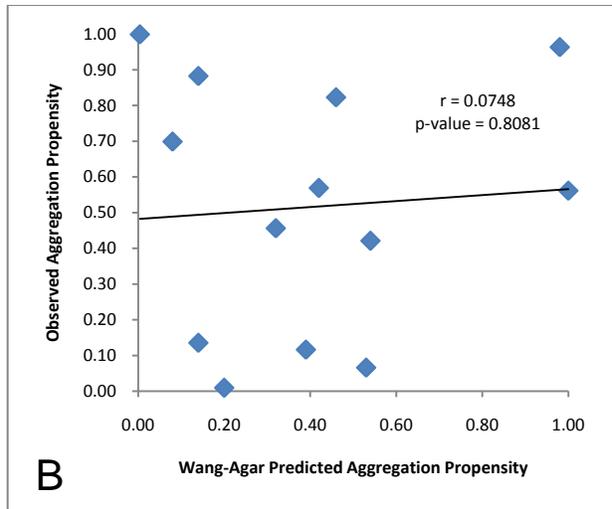
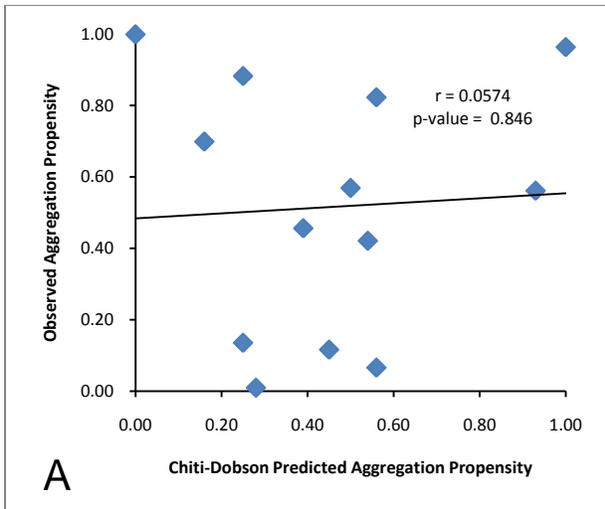


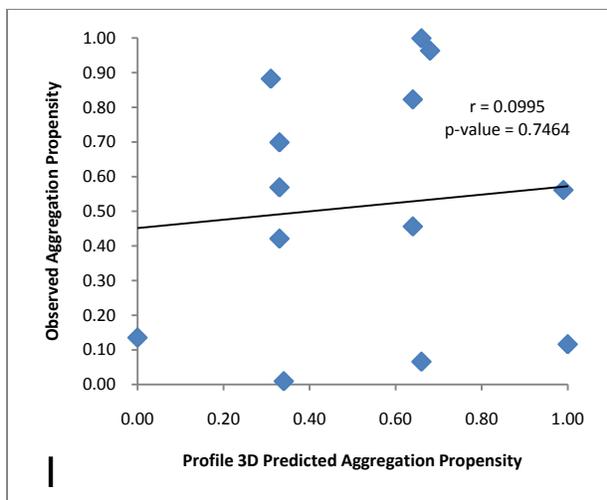
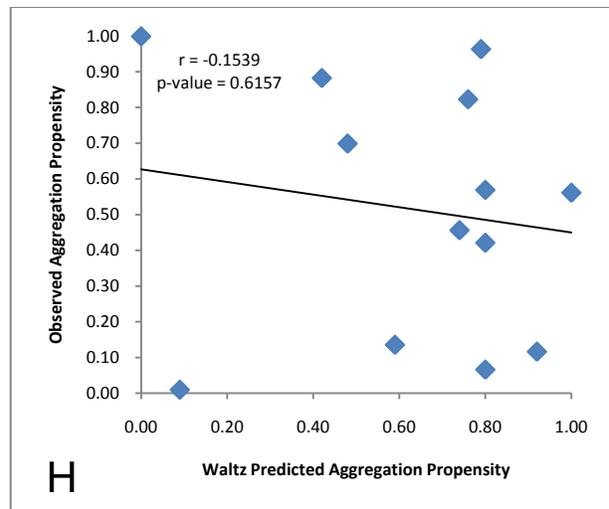
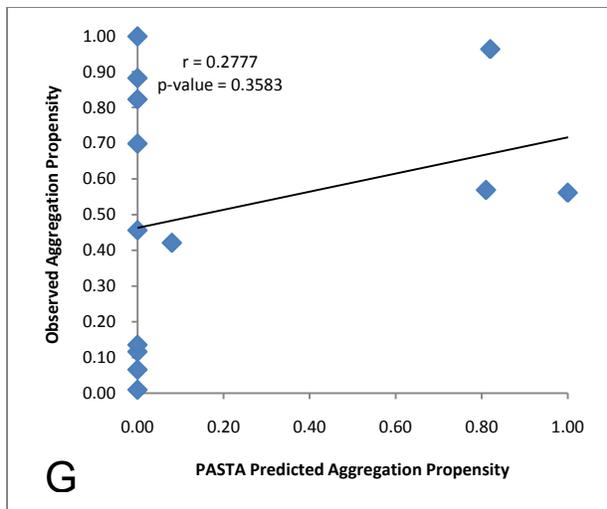


**Figure 3.3** Correlation plots of observed aggregation *versus* predicted aggregation for A $\beta$ <sub>42</sub> mutants using nine different prediction algorithms (listed in the ordinate axis labels). The data were fit to a straight line using linear regression.

### 3.3.3 Superoxide Dismutase Correlations

Aggregation propensity was measured previously in the Meiering lab for wild-type and 12 SOD1 mutants in the reduced apo form using DLS<sup>27</sup>. The observed aggregation propensity data were compared with the predicted aggregation propensities calculated in Chapter 2.3.1 and given in Table 2.3. The correlation plots are given in Figure 3.4. A summary of r-scores and p-values is given in Table 3.1. Only one of the nine algorithms, FoldAmyloid, results in a statistically relevant correlation. However, the r-score is not particularly high, 0.59, suggesting that some factors included in the algorithm are important to the aggregation observed here but additional variables not incorporated in this prediction technique also have a large influence on SOD1 aggregation.





**Figure 3.4** Correlation plots of observed aggregation *versus* predicted aggregation for SOD1 mutants using nine different prediction algorithms (listed in the ordinate axis labels). The data were fit to a straight line using linear regression.

**Table 3.1 Predicted and Observed Aggregation Correlation Summary**

	Chiti-Dobson	Wang-Agar	Zyggregator	Ztox	Fold Amyloid	TANGO	PASTA	Waltz	Profile 3D
<b>AcP r-score</b>	<b>0.5426</b>	<b>0.5469</b>	<b>0.6057</b>	<b>0.5604</b>	<b>-0.0469</b>	<b>0.2081</b>	<b>-0.3262</b>	<b>0.1884</b>	<b>-0.2388</b>
p-value	0.0016	0.0015	0.0003	0.0010	0.8022	0.2613	0.0785	0.3101	0.1957
<b>A<math>\beta</math><sub>42</sub> r-score</b>	<b>0.9186</b>	<b>0.8862</b>	<b>0.6387</b>	<b>0.4110</b>	<b>0.5185</b>	<b>0.6508</b>	<b>0.8210</b>	<b>0.6859</b>	<b>0.6861</b>
p-value	0.0093	0.0175	0.1711	0.4195	0.2903	0.1623	0.0457	0.1324	0.1293
<b>SOD1 r-score</b>	<b>0.0574</b>	<b>0.0748</b>	<b>0.1876</b>	<b>0.0583</b>	<b>0.5940</b>	<b>0.3874</b>	<b>0.2777</b>	<b>-0.1539</b>	<b>0.0995</b>
p-value	0.8522	0.8081	0.5394	0.8499	0.0323	0.1914	0.3583	0.6157	0.7464

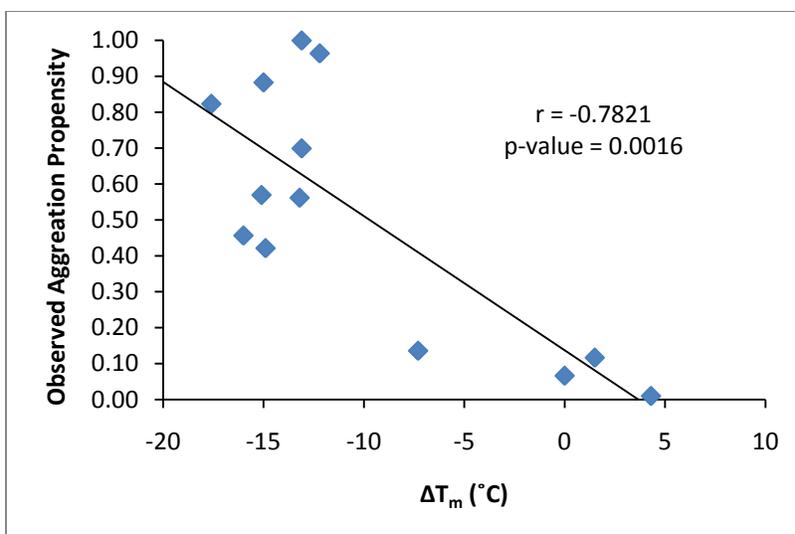
\*Values in red font are statistically significant ( $p < 0.05$ ) (See Methods, Section 3.2.2).

### 3.3.3.1 SOD1 Stability Correlated to Aggregation Propensity

Further investigations were conducted to determine factors important to SOD1 aggregation, as observed by dynamic light scattering. Considering the weak correlations described above between predicted and observed aggregation, it is important to look for variables that may be neglected by these algorithms but are significant contributors to protein aggregation. The stabilities of wild-type and mutant SOD1 proteins in the reduced apo form, as measured by Differential Scanning Calorimetry (DSC) in previously published data<sup>27</sup>, are correlated with observed aggregation measured by DLS. The results of this relationship are shown in Figure 3.5. The correlation between observed aggregation and change in melting temperature ( $\Delta T_m$ ) for each mutant compared to pWT in the reduced apo form gives a statistically significant r-score of -0.78 ( $p = 0.0016$ ), indicating that a decrease in stability causes an increase in aggregation. This correlation is stronger than any observed for the prediction algorithms with observed aggregation. A study by Wang *et al.* suggested that the sum of predicted aggregation and mutant destabilization for oxidized apo SOD1 mutants correlated with the disease duration of ALS patients with SOD1 mutations<sup>61</sup>. These investigators hypothesized that the incorporation of instability (the destabilization that occurs as a result of mutation) increases the accuracy of the aggregation prediction algorithms<sup>61</sup>.

To investigate this possibility further, the destabilization of the SOD1 mutants in the reduced apo form was normalized (see Equation 3 in Chapter 2) over a scale of 0 to 1, with 1 being the largest

decrease in melting temperature when compared to pWT, and 0 being the smallest decrease (in this case a slight increase for H46R) in melting temperature compared to pWT. Analogous to the method of Wang *et al.*<sup>61</sup>, the normalized instability score was added to the (previously normalized) aggregation propensity score and again the summed values were normalized from 0 to 1, with 0 being the lowest cumulative score and 1 being the highest, and therefore predicted to have the greatest propensity to aggregate. Table 3.2 lists the r-scores for the relationship between the sum of predicted aggregation propensity and mutant instability and observed aggregation propensity. There is an obvious major increase in the positive correlations observed here compared to the predicted propensities on their own. This is a likely a result of the strong positive correlation between decrease in melting temperature and observed aggregation described above ( $r = 0.78$ ). Thus, the increase in r-score for the prediction algorithms summed with instability can be predominantly attributed to the correlation of instability to observed aggregation. Only one algorithm (TANGO) gives a stronger correlation when combined with instability ( $r = 0.83$ ) than instability on its own ( $r = 0.78$ ). However, this slight increase in correlation is likely due to the fact that most mutants were not predicted to aggregate more than pWT by the TANGO method (Fig. 3.4 F and Table 3.1) and so the correlation observed here is predominantly the correlation of the instability alone with observed aggregation. All but one of the correlations gave statistically significant p-values, also as a result of the strong correlation between instability and observed aggregation ( $p = 0.0016$ ). Thus, addition of predicted aggregation and measured stability does not improve correlation with observed aggregation.



**Figure 3.5 The role of stability in aggregation propensity.** Observed aggregation, as measured by DLS, for reduced apo SOD1 mutants is plotted against the change in melting temperature, observed using DSC, for each mutant in the reduced apo form.

**Table 3.2 Correlation results for the sum of predicted aggregation propensity and mutant instability to observed aggregation propensity.**

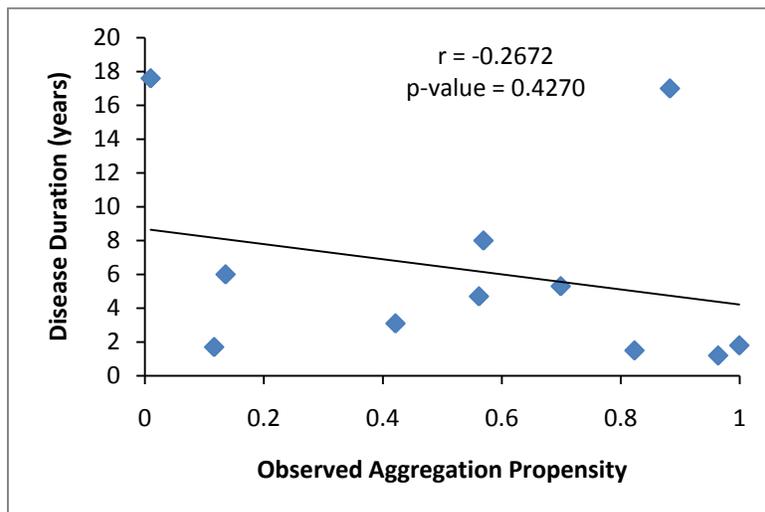
	Stability	Chiti-Dobson	Wang-Agar	Zyggregator	Ztox	Fold Amyloid	TANGO	PASTA	Waltz	Profile 3D
r-score	<b>0.7821</b>	<b>0.6048</b>	<b>0.5930</b>	<b>0.6367</b>	<b>0.5834</b>	<b>0.7775</b>	<b>0.8314</b>	<b>0.6416</b>	<b>0.4414</b>	<b>0.6935</b>
p-value	0.0016	0.0285	0.0327	0.0193	0.0363	0.0018	0.0004	0.0181	0.1311	0.0086

\*For each algorithm, the normalized predicted aggregation scores for each mutant were summed with the normalized instability scores and the new total was normalized and plotted against the observed aggregation results measured by DLS. A full description can be found in the text (Section 3.3.3.1).

### 3.3.3.2 SOD1 Correlations with ALS Disease Durations

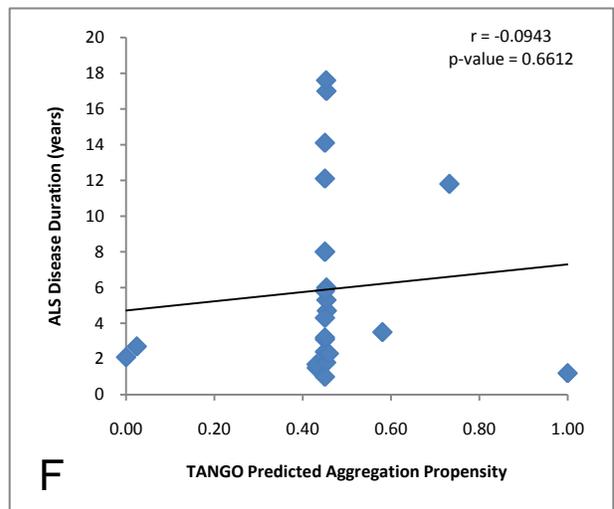
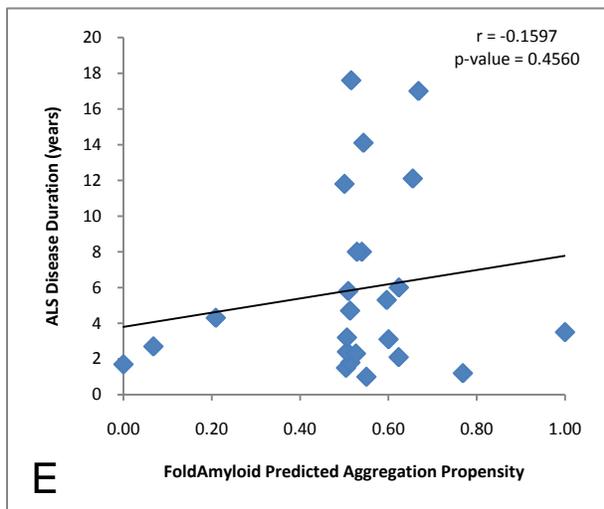
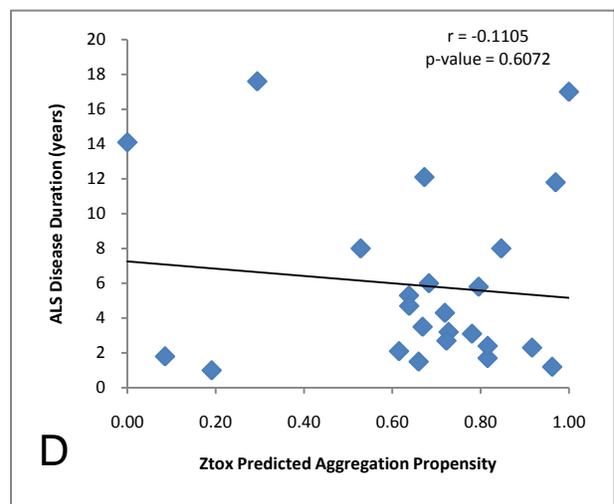
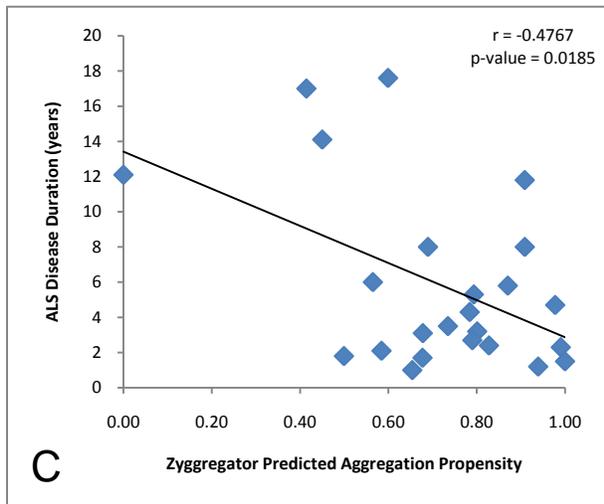
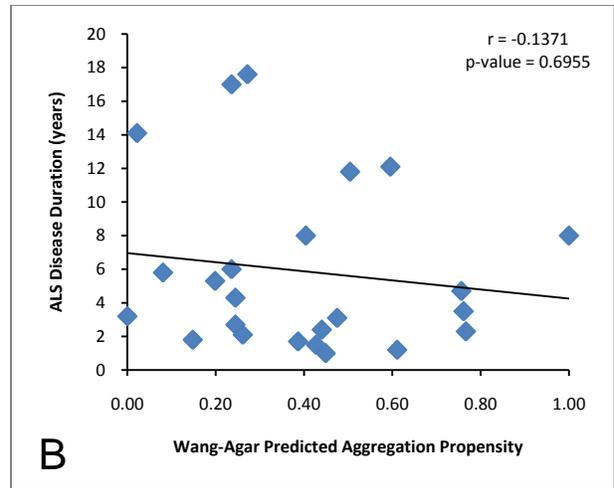
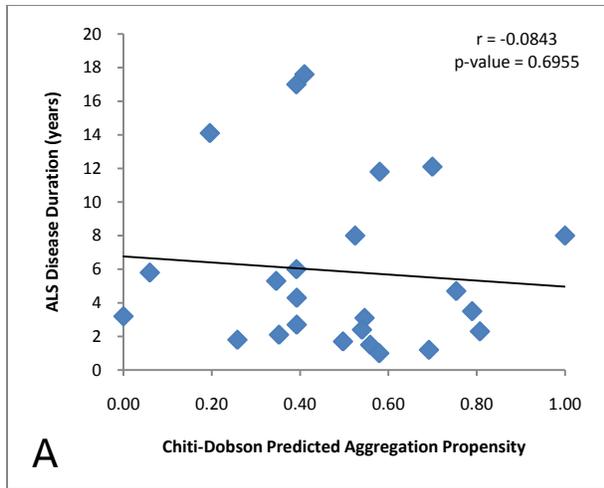
SOD1 mutations cause ALS with specific disease durations (see section 1.6.1). The relationship between disease-linked mutants and the characteristic time attributed to that mutant between onset and death was shown to be statically relevant by Wang *et al.*<sup>61</sup>. There is considerable evidence linking the pathogenesis of mutant SOD1 in fALS to protein aggregation<sup>62</sup>, and so it has been speculated that the propensity of a mutant to aggregate may dictate disease duration such that the mutants most prone to aggregate will cause the shortest disease durations<sup>61</sup>. To examine the role of reduced apo SOD1 aggregation in disease, the observed aggregation for 11 SOD1 mutations was plotted against the

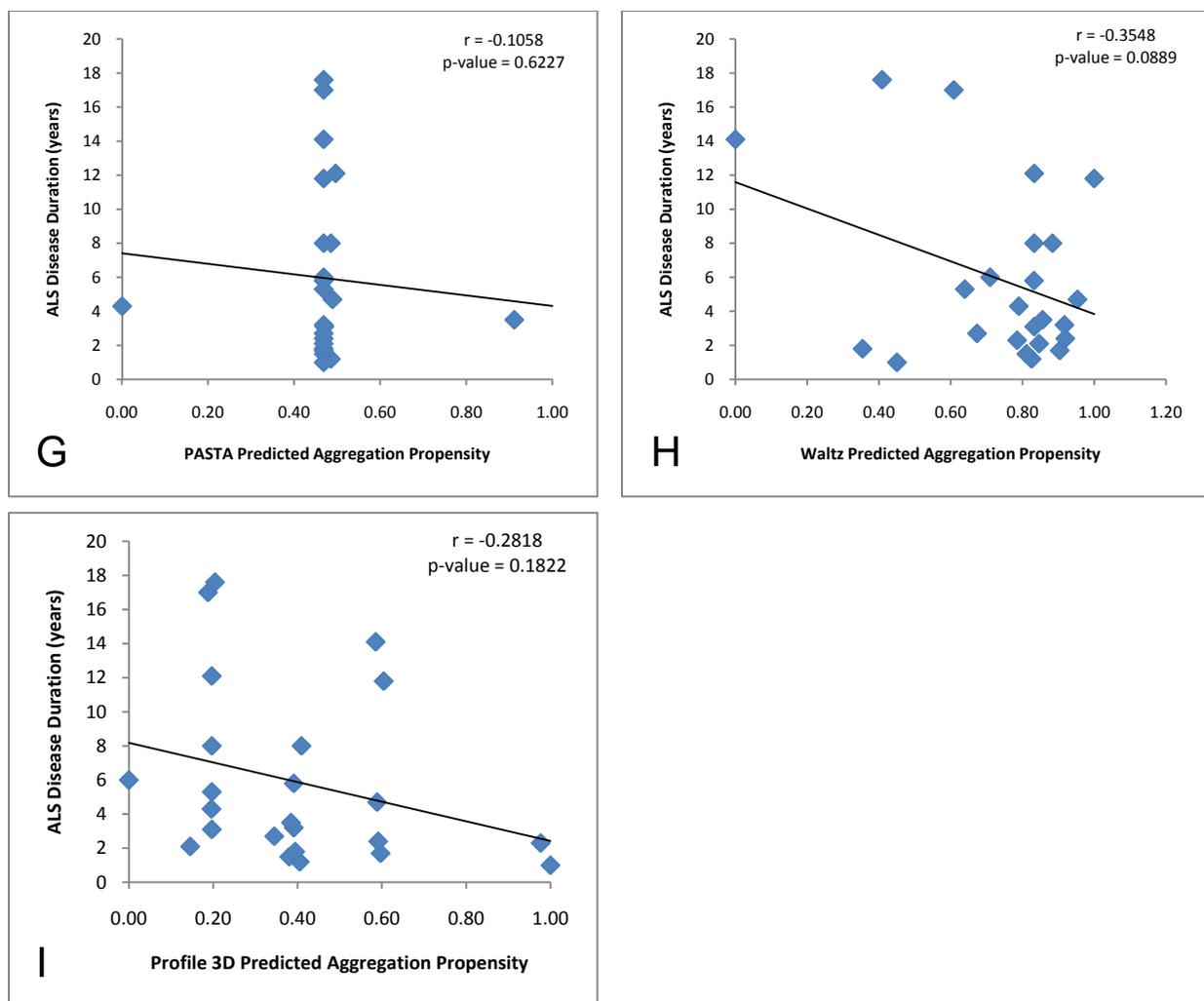
characteristic disease duration for each mutant. The results are shown in Figure 3.6. The correlation between observed aggregation and disease duration is not statistically relevant ( $r$ -score = -0.27,  $p$  = 0.43), however it appears there are significant outliers that greatly influence the overall trend. This will be discussed in 3.3.2.



**Figure 3.6 Relationship between observed aggregation for reduced apo SOD1 mutants, quantified using DLS, and ALS disease durations.**

In order to investigate whether the factors included in the prediction algorithms play a role in modulating ALS disease duration, the disease durations for each mutant were plotted against the predicted aggregation propensities. Twenty-four of the most common disease-linked SOD1 mutations were used for this investigation. Figure 3.7 shows the plots for ALS disease durations as a function of predicted aggregation propensity. Table 3.3 summarizes the correlations observed for these results. The correlations are in general quite weak. Two of the algorithms, FoldAmyloid and TANGO, actually resulted in a negative  $r$ -score, indicating that a higher predicted aggregation actually correlates with a longer disease duration, opposite to what might be expected, although not statistically significant. The only statistically significant correlation is that for Zyggregator, which has an  $r$ -score of 0.48 ( $p$  = 0.02). This is still a fairly weak correlation but may indicate that some of the factors included in the Zyggregator algorithm are related to how the mutant may impact ALS disease duration.





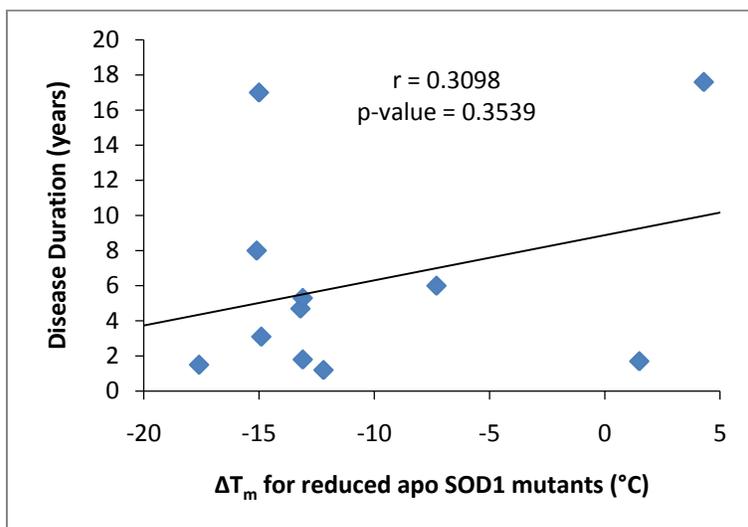
**Figure 3.7** Correlation plots of ALS disease durations *versus* predicted aggregation scores for SOD1 mutants using nine different prediction algorithms (listed in the ordinate axis labels). The data were fit to a straight line using linear regression.

**Table 3.3** Correlation values between predicted aggregation and ALS disease duration

	Chiti-Dobson	Wang-Agar	Zyggregator	Ztox	Fold Amyloid	TANGO	PASTA	Waltz	Profile 3D
r-score	0.0843	0.1371	0.4767	0.1105	-0.1597	-0.0943	0.1058	0.3548	0.2818
p-value	0.6955	0.5229	0.0185	0.6072	0.4560	0.6612	0.6227	0.0889	0.1822

\* Values in red font are statistically significant ( $p < 0.05$ ) (See Methods, Section 3.2.2).

One final correlation relevant to disease duration studies is that of the stability of reduced apo SOD1 protein with ALS disease durations. Figure 3.8 shows the relationship between the change in melting temperature for reduced apo SOD1 mutants compared to WT and the disease durations associated for each mutant with ALS patients. The resulting correlation is poor, with an r-score of only 0.31 ( $p = 0.35$ ).



**Figure 3.8 Relationship between the change in melting temperature for reduced apo SOD1 mutants and the ALS disease duration characteristic for each mutant.**

## 3.4 Discussion

### 3.4.1 Aggregation Prediction Algorithms Do Not Accurately Predict Observed Aggregation for Diverse Proteins

Testing the ability to predict protein aggregation provides significant information regarding how well the factors contributing to aggregation are understood. Recognizing the strengths and weaknesses of aggregation prediction algorithms provides a basis for further investigation into the complex processes involved in aggregation pathways. Section 1.3 discusses some of the potential contributing factors that have been shown to influence protein aggregation. Table 3.1 summarizes the correlation data for the relationship between predicted and observed aggregation for SOD1, AcP, and A $\beta_{42}$ . It is obvious that not all of the algorithms are able to accurately predict the measured aggregation for these three proteins and that there are very marked differences between the accuracy of the algorithms when comparing the three different proteins. There are various factors that may contribute to the poor correlations observed, which will be discussed in detail below, followed by interpretations of the correlation results.

#### 3.4.1.1 Differences in the Type of Aggregation being Monitored

The first issue to consider is the type of aggregation being measured for each of these proteins and how this is related to the design of the prediction algorithms. Both AcP and A $\beta_{42}$  aggregation were monitored using ThT-binding. As previously described (Section 3.1.1), ThT is a fluorescent dye that binds to the  $\beta$ -aggregate structure of amyloid fibrils resulting in an increase in intensity of the emission maximum and shift of the emission maximum from 445 nm to 482 nm<sup>111</sup>. Thus, ThT fluorescence is a direct measure of amyloid fibril formation. In contrast, SOD1 aggregation was measured by dynamic light scattering. This gives a measure of size but is not specific to the molecular structure of the aggregate. The reasoning behind this is that SOD1 does not typically form amyloid fibrils (see Section 1.6.2) and in the conditions used here it does not form amyloid, as confirmed by lack of ThT binding and by AFM studies revealing amorphous aggregates (Helen R Stubbs and Elizabeth M Meiering, unpublished data). The types of aggregates being observed by the DLS experiments are soluble and range in size from ~100

~1000 nm in hydrodynamic diameter<sup>27</sup>. Thus, it is important to consider whether the algorithms are designed to predict the type of aggregation being observed. Refer to Table 2.1 in Section 2.1 for a summary of the algorithms used in these studies. Four of the algorithms (Chiti-Dobson, Wang-Agar, Zyggregator and Ztox) utilized much of the same AcP and A $\beta$ <sub>42</sub> data used for the correlations shown here for the purpose of designing and training the algorithms. Thus, AcP and A $\beta$ <sub>42</sub> would be expected to perform well. The other algorithms were also developed based primarily on ThT-data of other proteins and peptides. For these reasons, it would be expected that the algorithms should be capable, at least to some degree, of accurately predicting aggregation for AcP and A $\beta$ <sub>42</sub>. The correlation results between predicted and observed aggregation for these two proteins does indicate some success by some of the prediction algorithms, however significantly weaker than expected. This will be discussed further in 3.4.2. Much less is known about how well the algorithms might be expected to perform in predicting non-amyloid aggregation as current development and use of these algorithms has been almost solely focussed on amyloid. SOD1 has the weakest correlations between predicted and observed aggregation (Table 3.1) compared to the other two proteins studied. This may be a direct indication that because the algorithms were designed based on amyloid formation, they are incapable of accurately predicting the soluble oligomer formation by reduced apo SOD1. This possibility is discussed in greater detail in 3.4.2.

#### 3.4.1.2 Variations in the Initial Structures of the Test Proteins

The state of a protein in solution may have a significant influence of the susceptibility of the protein to aggregation and the type of aggregate structures that might form. It has been suggested that proteins can aggregate from the natively folded state by local fluctuations that expose hydrophobic regions, i.e. without the requirement of crossing large energy barriers such as from the folded to the unfolded state<sup>29</sup>. The presence of mutations may enhance unfolding fluctuations as a result of the destabilization, which could further promote aggregate formation<sup>10</sup>. The lack of structure in fully unfolded protein may contribute to the ability to form extensive intermolecular associations, such as those present in amyloid<sup>10</sup>. Therefore, another particularly relevant difference between the proteins used in this study

that may play an important role in dictating the mechanism of protein aggregation is the initial state that the protein is in. As described in Section 2.1, SOD1, AcP and A $\beta$ <sub>42</sub>, represent three different types of protein structures. SOD1 is a natively folded, metal-binding, dimer<sup>59</sup>. For the purpose of these studies, SOD1 aggregation was analyzed in its reduced apo form, lacking metals and with a reduced disulphide bond. In this form, SOD1 exists as a marginally stable monomer<sup>27; 65; 118</sup>. The fact that the reduced apo monomer is only marginally stable suggests that it will have increased structural fluctuations or may exist in a partially folded state. In addition, the destabilization effects of several SOD1 mutants result in significant population of the unfolded state of the reduced apo protein at physiological conditions<sup>27</sup>. The experimental conditions used to measure reduced apo mutant SOD1 aggregation were pH 7.4 and 37°C in quiescent conditions. Some of the mutants observed in this study have melting temperatures at or below 37°C, and thus will populate both the folded and the unfolded state<sup>27</sup>. In contrast AcP is unfolded at the conditions used for experimental measurements<sup>9</sup> (see Section 2.1), and A $\beta$ <sub>42</sub> is a peptide lacking stable tertiary structure with the capability of adopting both alpha helical or beta sheet structure depending on the solution conditions<sup>95</sup> (see Section 2.1). In the solution conditions in which the aggregation measurements are taking place it is likely to preferentially adopt a beta sheet conformation<sup>95</sup>. For these reasons it is expected that AcP and A $\beta$ <sub>42</sub> may possess enhanced exposure of regions prone to aggregate, as opposed to the potentially folded protein, SOD1, in which aggregation-prone regions may remain buried in the folded state or only sometimes accessible depending on the stability and population dynamics of the mutant. Thus native state structure is more strongly favoured in the experimental conditions of SOD1 than for either AcP or A $\beta$ <sub>42</sub>. For these reasons, aggregation mechanisms are likely to differ between SOD1, AcP, and A $\beta$ <sub>42</sub><sup>119</sup>. This is evidenced by the differing structures formed in these two systems; amyloid formation by AcP and A $\beta$ <sub>42</sub>, and soluble oligomer formation by SOD1. The formation of amyloid *vs.* non-amyloid aggregates may be a direct result of the degree of structure of the initial protein. Several studies suggest that a higher degree of structure corresponds with the formation of amorphous aggregates, while less-structured, more unfolded, proteins or peptides are more likely to form amyloid<sup>10; 120</sup>.

In one such study, the variable domain of a recombinant amyloidogenic light chain, SMA, was used for investigating the role of structure in dictating aggregate morphology. Although at pH 4 and 6 secondary structure was maintained, there were significant alterations in the tertiary structure of SMA, resulting in the fast formation of amorphous aggregates. At pH 3, however, SMA was relatively globally unfolded with decreased secondary and tertiary structure, but remained compact and slow formation of amyloid fibrils was observed<sup>120</sup>. In addition,  $\beta$ 2-microglobulin mutational studies that measured the effects of mutations on the secondary structure (as monitored by circular dichroism) and aggregation rates (as monitored by ThT fluorescence) suggested that variations in the amount of ordered structure affect the aggregation pathway and final morphology of the amyloid being formed<sup>121</sup>. These examples provide evidence that varying mechanisms are involved in controlling different types of aggregation, and that the structure of the initial protein can greatly influence the structure of the aggregates formed. Starting structures remain an un-addressed complication in protein aggregation prediction algorithms and could play a significant role in the success of the algorithms.

#### 3.4.1.3 Variations in the Size of the Test Proteins

A recent study by Ramshini *et al.*<sup>122</sup> compared the sequence length for a sample of aggregation-prone, disease-linked proteins, and found striking evidence that the length of the polypeptide sequence may play an important role in dictating aggregation patterns. Proteins with longer sequences have a greater tendency to form non-amyloid protein deposits in disease, while proteins with shorter sequences (<250 amino acids) more often form amyloid deposits in patients<sup>122</sup>. The differing sizes of SOD1, AcP, and A $\beta$ <sub>42</sub> may also impact the aggregation of these proteins. The SOD1 monomer is 153 residues in length, while AcP is 98 and A $\beta$ <sub>42</sub> is 42. Although these differences are not particularly drastic, they could be significant. The SOD1 monomer is three times the length, and AcP twice the length, of A $\beta$ <sub>42</sub>. If only a small section of the protein is prone to aggregate, having a smaller sequence may increase aggregation propensity by preventing interference from non aggregation prone regions of the protein. Thus, protein

size is an important factor to consider when investigating the relationship between aggregation prediction algorithms and measured protein aggregation.

#### 3.4.1.4 Differences in Solution Conditions of Aggregation Samples

Solution conditions can influence many characteristics of a protein in solution including its stability, charge, and solubility, which can all affect aggregation (see Section 1.3.3). Therefore, a third important factor to consider when investigating the relationship between predicted and observed aggregation is the solution conditions for each aggregation experiment. The experimental conditions for aggregate formation of AcP were pH 5.5, 298 K and an ionic strength of 50 mM<sup>9; 92</sup>. A $\beta$ <sub>42</sub> aggregation was measured at pH 7.4 or 7.5, 298 K or 310 K and an ionic strength of 100 mM, 150 mM, or 155 mM<sup>108; 109; 110</sup>. The SOD1 aggregation data was collected at pH 7.4, 310 K and an ionic strength of 1 mM<sup>27</sup>. The differing solution conditions for the three proteins in question present additional complications when it comes to predicting protein aggregation. It can be argued that the relative difference between the aggregation tendency of various mutants remains constant regardless of solution conditions, as long as the conditions are constant. If this were the case, the predictions should not be affected by the differing solution conditions, as long as all data for the same protein are collected under the same conditions. However, this may not hold true if the conditions were altered in such a way that aggregation occurred by a different mechanism. For example, comparative differences in the aggregation patterns of mutant protein at a low pH versus a neutral pH may not be the same if the protein is unfolded in one context and folded in the other, and this could result in the formation of different aggregate structures<sup>16</sup>. Furthermore, the solution conditions are not always the same for all data collected for the same protein, as described for A $\beta$ <sub>42</sub>. Furthermore, many of the algorithms were developed based on aggregation data that was not all collected at the same experimental conditions, which also could have important implications to the success of the algorithms.

### 3.4.2 Interpreting the Correlation Results between Predicted and Observed Aggregation

With these complications in mind, general principles can nevertheless be deduced from the correlation data between predicted and observed aggregation for the three proteins in question using the nine different prediction techniques.

#### 3.4.2.1 Moderate Correlations Observed between Predicted and Observed Aggregation for AcP

The results for AcP will be considered first. Four out of nine algorithms (Chiti-Dobson, Wang-Agar, Zyggregator and Ztox) give statistically relevant positive correlations with r-scores between 0.54 and 0.61. These are not particularly convincing correlation coefficients, which is surprising considering AcP is unfolded in the conditions tested, and readily forms amyloid. Additionally, AcP data was used in the design of the Chiti-Dobson and Wang-Agar algorithms. The difference between the correlation given here for the Chiti-Dobson algorithm and that from the original paper<sup>9</sup> arises due to the fact that the original study only included the mutations found in primary sequence regions 16-31 and 87-98, which are considered to be particularly important for aggregation. The analysis shown here includes data for a larger set of AcP mutations, including mutations within and outside these regions. Thus the discrepancy results from a larger, less selective, dataset used in the correlations described here. Interestingly, a more inclusive dataset results in a weaker correlation of 0.54 (Table 3.1) compared to 0.76<sup>9</sup>. This may indicate that the physicochemical properties included in the Chiti-Dobson equation are important for determining the extent of aggregation in regions experimentally identified as being aggregation prone, but may not be the determining factors for whether or not these regions will cause a protein to aggregate. This algorithm may be better suited for the prediction of mutational effects in regions that have already been demonstrated to be involved in aggregation, while for a complete sequence analysis, this method may not be as accurate.

The best overall correlation between predicted and observed aggregation for AcP comes from Zyggregator, with an r-score of 0.61. Again, this weak relationship indicates that some significant factors are captured by the algorithm, while missing others. Interestingly, the five poorest performing algorithms for AcP (TANGO, Waltz, FoldAmyloid, PASTA and Profile 3D) are based on, or tested with,

hexapeptide fragments (six amino acid stretches based on amyloid forming proteins, see Section 2.1). This may demonstrate the limitations of studying peptide fragments as means for accurately predicting aggregation of longer protein sequences. Short sections predicted to be involved in aggregation based on amyloid formation by hexapeptides may not result in aggregation in the context of a larger protein sequence<sup>100</sup>. In general the aggregation prediction methods have limited success in predicting AcP aggregation. The four algorithms that gave positive, statistically relevant results focus on the role of inherent physical and chemical properties of the amino acid sequence. This indicates that these properties do strongly influence aggregation of unfolded AcP, however the moderate correlations observed suggest these are not the only factors contributing to the aggregation process.

#### 3.4.2.2 Partial Success in Correlating Predicted and Observed Aggregation for A $\beta$ <sub>42</sub>

Overall, the prediction algorithms are the most successful at accurately predicting ThT-monitored A $\beta$ <sub>42</sub> aggregation, similar to the successful correlations previously observed for the relationship between predicted aggregation using many algorithms with *in vivo* aggregation of A $\beta$ <sub>42</sub> in *E.coli*<sup>106</sup>. In the case of the work described here, the sample size is quite small, making it more difficult to interpret the accuracy of the results. With an increased number of mutational studies the correlation results could become much more or less significant. Interpretations of the presented results then must be approached with caution. Three out of nine algorithms, including Chiti-Dobson, Wang-Agar and PASTA, give statistically significant correlations with r-scores between 0.82 and 0.92. The Chiti-Dobson equation performs the best, as expected since A $\beta$ <sub>42</sub> data was used as a validation dataset when the equation was designed. The Wang-Agar equation follows closely behind since it is the same formula as the Chiti-Dobson method with updated coefficients. The third best algorithm was PASTA, with an r-score of 0.82 ( $p = 0.05$ ). This algorithm calculates the pair-wise energy function between two identical sequences when they align in a beta-sheet conformation; it was verified by correctly predicting the most amyloidogenic region of A $\beta$ <sub>42</sub>. The rest of the algorithms did not reveal statistically significant correlations, however, none resulted in exceptionally poor correlations. With a greater sample size more of the algorithms may have statistical

relevance. The better ability of the algorithms to accurately predict the aggregation observed here is likely due to the fact that A $\beta$ <sub>42</sub> is only a peptide and does not contain stable higher order protein structure. It also readily forms amyloid at neutral pH and low ionic strength. Thus, the dominant factors expected to contribute to its aggregation are the inherent properties of the amino acids within the sequence. These factors are well captured by several of the algorithms, including the Chit-Dobson and Wang-Agar equations, which give the strongest correlation results.

### 3.4.2.3 Poor Correlations for Predicted and Observed Reduced Apo SOD1 Aggregation

From Table 3.1 it is clear that the prediction algorithms do a poor job at predicting DLS-monitored aggregation of the reduced apo form of SOD1. Only one algorithm, FoldAmyloid, results in a statistically significant correlation; however, with an r-score of only 0.59, it is not a particularly strong correlation. Yet, the FoldAmyloid algorithm does better than the others, suggesting that hydrogen bonding and molecular contacts, the major components of this algorithm (described in Section 2.1.4), play some role in the formation of soluble aggregates by reduced apo SOD1 mutants. TANGO and PASTA have the next best r-scores, but at values of 0.39 and 0.28, neither of these algorithms is capable of accurately predicting SOD1 aggregation. The rest of the algorithms display very poor relationships, ranging from 0.19 for Zyggregator to -0.16 for Waltz. The general poor performance of the algorithms at predicting reduced apo SOD1 aggregation may be dependent on several factors. As mentioned above, the aggregation is being monitored by DLS and is not amyloid. Considering that these algorithms were primarily developed based on amyloid forming peptides and proteins, this may be evidence that amyloid formation and soluble oligomer formation are not governed by the same contributing factors. There are many possible pathways and products of aggregation (see Section 1.2), and so there must be determining factors that alter the course towards the formation of one type of aggregate or another. It is possible that completely different factors are responsible for modulating the different pathways. However, it may be more likely that similar factors, but with different levels of weighting, contribute to these processes. If the latter were true, algorithms may be able to be redesigned for increased suitability in predicting other types

of aggregation, beyond the context in which they were designed, by retraining with aggregation data from non amyloid-forming systems. Additionally, the fact that reduced apo SOD1 is marginally stable and thus exists in a folded, partially folded, or unfolded conformation, sets it apart from the proteins and peptides used to design and train the algorithms in question. The overall design of these algorithms may be less suitable for the prediction of aggregation by folded proteins.

### **3.4.3 Difficulty in Predicting Protein Aggregation Outside of the Context of Algorithm Development**

Thus, it appears that while the aggregation prediction algorithms are able to successfully predict aggregation within the context in which they were developed, none are able to accurately predict aggregation of a wider set of protein types. While some, such as the Chiti-Dobson and Wang-Agar equations perform reasonably well for unfolded AcP and the short A $\beta$ <sub>42</sub> peptide, they fail when it comes to the non-amyloid forming, potentially folded monomer, SOD1. Others that were developed to successfully identify aggregation-prone regions<sup>33; 89</sup>, are unable to predict overall aggregation propensity. There is still much unknown about of protein aggregation pathways. The present work clearly shows that significant ground has been made in the understanding some of the principles controlling aggregation within a confined context (such as amyloid formation by unfolded proteins<sup>9; 61; 85; 86</sup>). Continued research will allow for further understanding of the roles of each potential variable in dictating aggregation propensity and further deciphering of whether some factors may have differing roles in modulating the formation of different types of aggregate structures. The seemingly narrow scope of the aggregation prediction algorithms may require a re-think in field regarding the way protein algorithms are developed and used, and the types of protein aggregation they are capable of predicting.

### **3.4.4 Predicted and Observed SOD1 Aggregation fails to determine ALS Disease Duration**

An intriguing characteristic regarding the causative role of mutant SOD1 in ALS is the characteristic disease durations for different mutations (see Section 1.6.1 for details). It has been suggested that mutations introduced into SOD1 will influence the aggregation propensity of the protein

and that the aggregation propensity will control the disease duration<sup>61</sup>. To further investigate the possibility of this role of SOD1 mutant aggregation in dictating disease, the observed aggregation results for reduced apo SOD1 mutants were plotted against the disease durations of those mutants (Figure 3.6). The r-score of this relationship was only 0.27 ( $p = 0.43$ ), indicating no significant correlation. There are several ways to interpret this result with respect to the role of SOD1 in disease. The first is that reduced apo SOD1 is not the common denominator in ALS disease pathogenesis<sup>27</sup>. Section 1.6.2 describes the hypothesis that due to the destabilization that occurs as a result of mutations in the reduced apo form, this marginally stable form of SOD1 may be the most prone to aggregate and therefore may be the most toxic. Contrary to this hypothesis, the aggregation of reduced apo SOD1 mutants in quiescent, physiological solution conditions observed by DLS<sup>27</sup> does not clearly correlate with ALS disease duration. Importantly though, it does not mean that reduced apo is not playing a role in disease, or even that it is not a significant player. Instead, different mutants may be most toxic in different forms of the protein<sup>27</sup>. Thus, there could be different disease mechanisms for different mutants.

Another possible interpretation is that the type of aggregation measured by DLS may not be a toxic form that would contribute to disease. There is continued debate over what is the toxic species in protein aggregation diseases<sup>52</sup>. Increasing evidence suggests that soluble oligomers may be more toxic than insoluble amyloid<sup>52; 53</sup>, as discussed in Section 1.4. It seems that the species produced during these DLS experiments may be representative of soluble oligomers that could contribute to toxicity. However, currently the actual toxic aggregate species in ALS are not known and further investigation is required<sup>66</sup>. The weak correlation between reduced apo SOD1 aggregation and disease duration suggests that there is still much to investigate regarding the role of aggregation in modulating disease. It is evident that SOD1 mutations cause ALS (see Section 1.6), and it is likely that aggregation plays a role in the pathogenesis<sup>67</sup>. Therefore, it is important to continue pursuing the role of misfolded and aggregated SOD1 in all forms as it is apparent that complex mechanisms are involved in the syndrome of ALS.

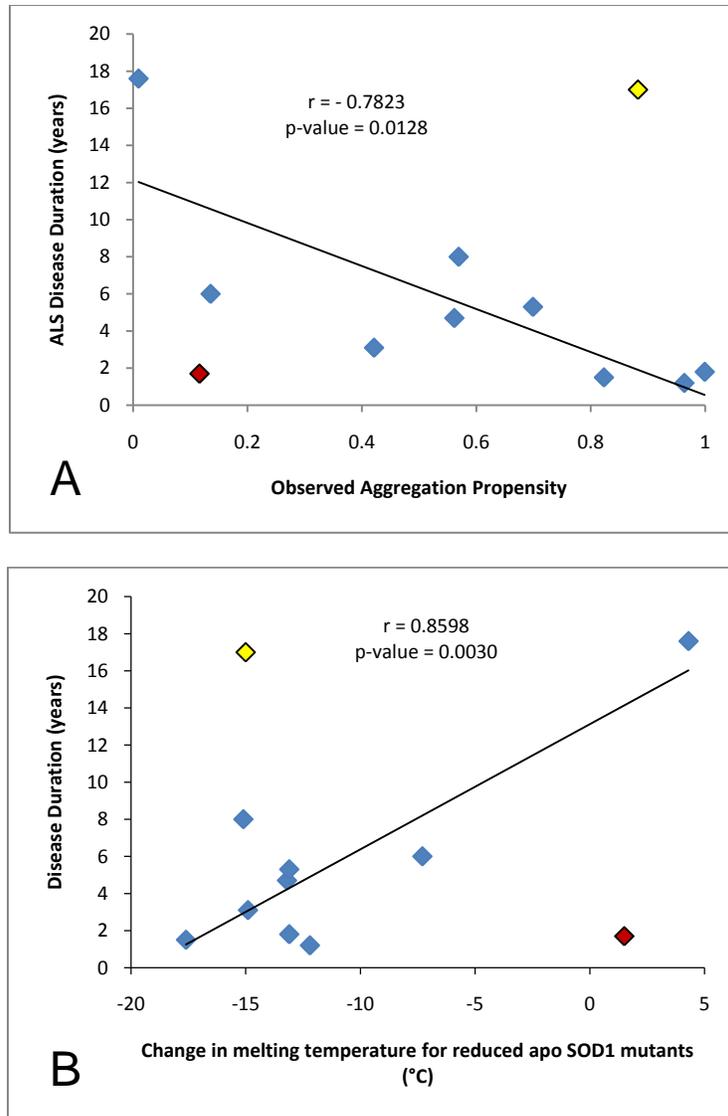
As correlations between experimental reduced apo SOD1 aggregation and ALS disease duration are not statistically relevant, further investigation of the variables contributing to disease duration were

carried out using the aggregation prediction algorithms. Twenty-four of the most common SOD1 mutations (see Table 1.1) were used for this analysis. The relationships between predicted aggregation and ALS disease duration were generally quite poor (Table 3.3). The only algorithm that resulted in a statistically relevant disease correlation was Zyggregator, with a correlation coefficient of 0.48 ( $p = 0.02$ ). This is not a particularly strong correlation, but it is not irrelevant either. It out-performs the Chiti-Dobson and the Wang-Agar methods, which suggest that disease properties are dependent on more than the physicochemical properties of the residue that is mutated. Zyggregator calculates physical and chemical properties over a sliding window, thereby incorporating effects of the sequence surrounding the mutation. It also includes a function for the pattern of the sequence (the pattern of hydrophobic residues can promote beta-sheet formation<sup>123</sup>), as well as the role of gate-keeper residues in preventing aggregation<sup>123</sup>; <sup>124</sup>. These four characteristics, taken together, appear to play some role in dictating ALS disease duration. These are by no means the whole story; however, they may provide a basis to build upon. All other algorithms do not result in statistically relevant correlations with disease duration; however, the primary function of these algorithms is to predict amyloid formation. The lack of correlation between predicted aggregation and disease duration may then be an additional indication that other forms of SOD1 aggregation, not amyloid, are involved in ALS, as discussed in section 1.6.3.

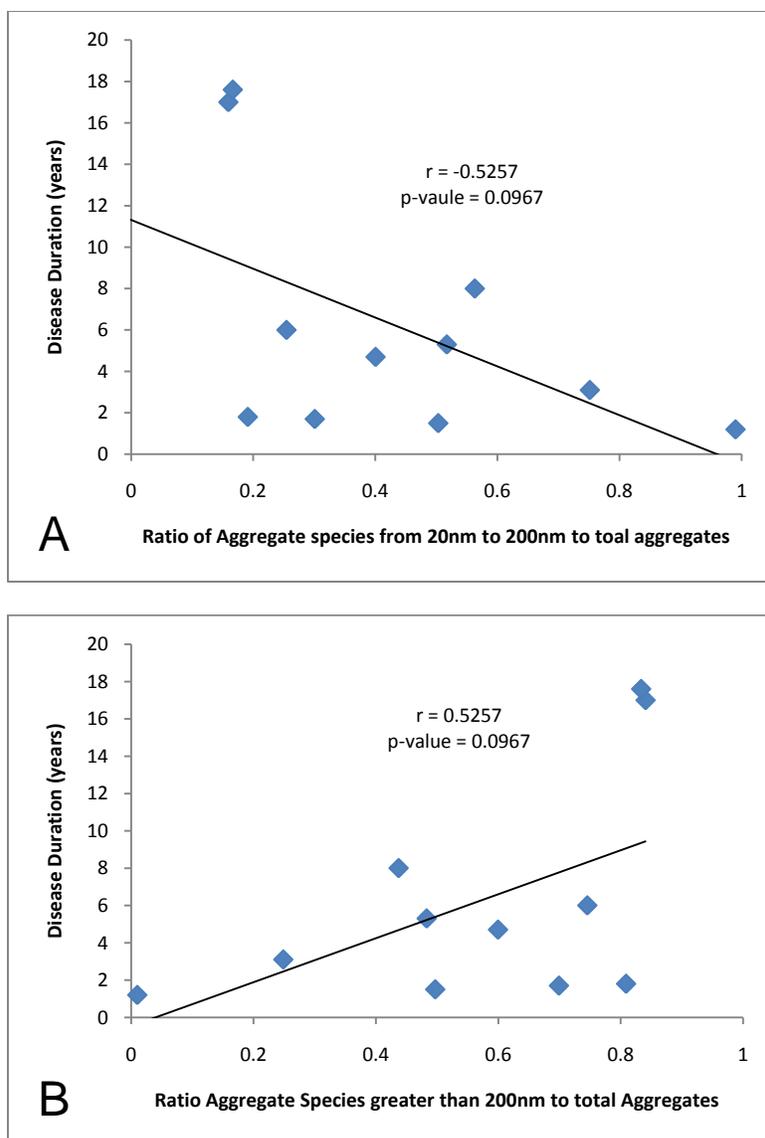
A final relationship that was investigated was reduced apo mutant SOD1 stability and disease duration. The correlation coefficient between the  $\Delta T_m$  of mutants and disease duration of ALS patients was only 0.31 ( $p = 0.35$ ). This is not statistically significant and thus does not strongly support that the decrease in stability of reduced apo SOD1 by ALS mutations directly causes the observed patient disease durations. This is important to consider because enhanced populations of misfolded species may not only trigger aggregation but may also have direct detrimental effects on cellular components as individual misfolded species<sup>66</sup>. The poor correlation does not necessarily mean that reduced apo SOD1 is an insignificant state in contributing to disease, but could indicate that a more complex set of variables, beyond stability alone, are involved in determining disease.

Interestingly, there are two main outlier mutants in this correlation, which when removed greatly strengthen the observed relationship. Figure 3.9 shows that when G37R and V148I are excluded, the correlation between observed aggregation and disease duration increases from an r-score of 0.27 ( $p = 0.43$ ) to 0.78 ( $p = 0.013$ ), while the correlation between the change in melting temperature and disease duration increases from an r-score of 0.31 ( $p = 0.35$ ) to 0.86 ( $p = 0.003$ ). It is not obvious why these two mutations might be such extreme outliers. Additional data will be needed in order to determine whether this trend continues with a larger data set. However, what might be implicit in these results is the identification of one of many potential mechanisms contributing to determining ALS disease duration: while most of the mutants here follow the trend of destabilization causing aggregation, and aggregation dictating disease duration, others, such as V148I and G37R, may have a different mechanism. In particular, G37R, though quite destabilized in the reduced apo form, has a very long disease duration. This could be a result of a differing degree of toxicity of the aggregates being formed.

Preliminary investigations based on further analysis of DLS data have hinted that the size of the aggregate structures may play a role in controlling disease. Figure 3.10 shows a comparison of the ratio of species with a hydrodynamic diameter between 20 and 200 nm to total aggregation (>20 nm) with the ALS disease durations. The same relationship is shown for the ratio of species formed with a hydrodynamic diameter greater than 200 nm to the total aggregation. The trends suggest that the greater the proportion of smaller species, the shorter the disease duration. Although there are not enough data to give a convincing, statistically reliable correlation, these results may be an indication that the size of aggregates plays a role in toxicity. This is similar to the proposed idea that small soluble oligomers are more toxic than large, insoluble amyloid species<sup>57</sup>. Interestingly, for G37R, 84% of the total scattered light is from aggregates larger than 200nm, which supports the notion of larger species being less toxic. On the other hand, V148I, shows very little aggregation in the reduced apo form, and is even slightly stabilized. This may indicate that V148I exerts its toxic effects in a different form of the protein.



**Figure 3.9 Identification of outliers for reduced apo SOD1 correlations.** A) The relationship between the observed aggregation of SOD1 mutants and ALS disease duration. B) The relationship between the change in melting temperature as a result of mutations in reduced apo SOD1 and disease duration. V148I is shown in red and G37R in yellow. Both mutants are excluded from the linear regression analysis and do not contribute to the given r-scores and p-values.



**Figure 3.10 Role of aggregate size in dictating ALS disease duration.** A) ALS disease duration *versus* the ratio of the percentage intensity of scattered light of aggregate species with hydrodynamic diameters between 20 and 200 nm to total aggregate species as monitored by DLS for 11 disease-causing SOD1 mutations (including G37R and V148I) in the reduced apo form. B) ALS disease duration vs. the ratio of the percentage of aggregate species with hydrodynamic diameters greater than 200 nm to total aggregate species as monitored by DLS for 11 disease-causing SOD1 mutations in the reduced apo form.

Unravelling the complex role of SOD1 mutations in modulating disease is a difficult task. Here we have seen hints at various factors that appear likely to be involved in controlling disease duration, including: protein stability, the pattern of physical and chemical properties of the amino acids in the protein sequence, and the average sizes of aggregate species formed. However, a complete mechanism has not been elucidated. This work highlights that reduced apo SOD1 is probably an important player in disease, since the instability and aggregation tendencies of this form of the protein can demonstrate significant correlation with disease duration when potential outliers are removed; however, scatter in this correlation and other data suggest that other forms of the mutant proteins may also play a role. Regardless, the inability of any of the correlations to fully explain ALS disease duration confirms the complexity of this disease. There is likely to be multiple effects of mutant SOD1s in various forms that work together to result in motor neuron toxicity<sup>27</sup>. Continued investigation is required to elucidate the principles governing the role of mutant SOD1 in ALS pathogenesis.

### **3.4.5 Conclusions**

Accurately predicting protein aggregation using computational methods is a complicated undertaking. The relationship between predicted and observed aggregation for three diverse proteins and a large set of mutants demonstrates that while moderate success can be achieved in predicting the aggregation propensity of various mutants for amyloid-forming proteins, there are still significant gaps in accurately understanding the underlying principles involved in diverse aggregation processes. Testing of widely used aggregation prediction algorithms, including the Chiti-Dobson equation<sup>9</sup>, Wang-Agar equation<sup>61</sup>, Zyggregator<sup>85</sup>, Ztox<sup>85</sup>, FoldAmyloid<sup>88</sup>, PASTA<sup>86</sup>, TANGO<sup>87</sup>, and Profile 3D<sup>89</sup>, demonstrates the limitations of applying methods developed in a specific test system to broader contexts. This points out that there are likely to be fundamental differences in the variables contributing to aggregation in different systems. It may be concluded, however, that although there may not be one set of global variables that governs all forms of aggregation, there could be an overlap in the factors governing more than one type of aggregation pathway. Thus, continued research should allow for the conceptualization of

multiple aggregation pathways, and the development of specific prediction tools for predicting certain types of aggregation. Gaining such understanding has important implications for many protein aggregation-linked diseases<sup>50</sup>, including the role of SOD1 in ALS<sup>62</sup>, and for practical aspects of formulating protein solutions for various pharmaceutical<sup>8</sup> or biotechnological<sup>84</sup> applications. By focusing future work on specializing prediction algorithms to predict aggregation within a narrow context, the comparison between algorithms will increase information regarding the differences in principles modulating different types of aggregation.

## Chapter 4

### Probing the Role of Salt in Protein Aggregation Mechanisms

#### 4.1 Introduction

Protein aggregation is a complex process in which multiple variables contribute to the pathways and products involved. Due to the diversity of the aggregate structures that can be formed, and the many factors that appear to modulate these processes, there is still a general lack of understanding of the principles of protein aggregation. In order to further probe the detailed mechanisms of aggregate formation, the role of ionic strength was investigated in detail here by monitoring the aggregation patterns of one of the most common ALS-causing SOD1 mutations, A4V, in the oxidized and reduced apo form, in a range of different salt solutions.

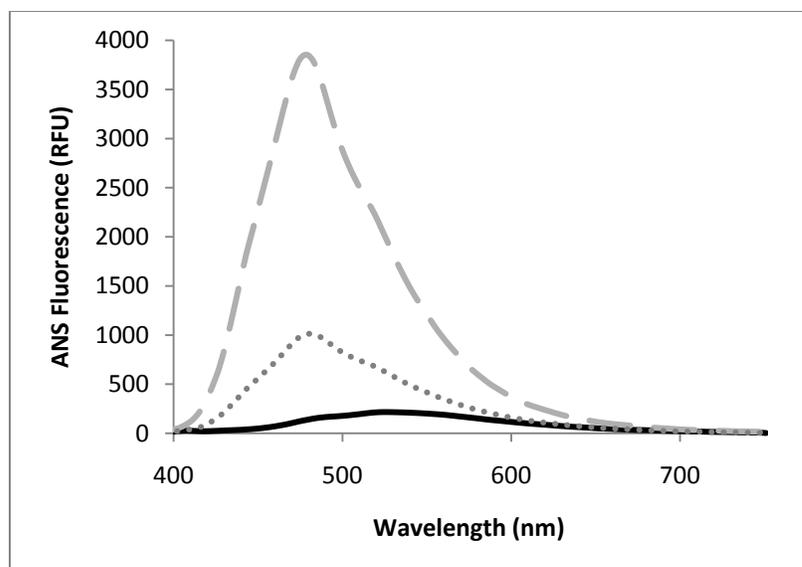
##### 4.1.1 The Role of Ionic Strength in Protein Aggregation

Solution conditions can have a very significant impact on protein aggregation (see Section 1.3.3). The details of how ions interact with proteins and affect intermolecular attractions are not well understood. Section 1.3.3 describes some of the previous work that has investigated the patterns of protein aggregation in various types and concentrations of salt. However, there does not appear to be one clear way in which all ions influence protein aggregation. In general it seems that ions can have electrostatic interactions, and stabilization or destabilization effects, that may result in aggregation promotion or inhibition depending on the state of the protein and the nature of the ion<sup>125</sup>.

##### 4.1.2 ANS Binding

In order to investigate the aggregation tendencies of A4V SOD1 in various salt conditions, 1-Anilino-8-naphthalene sulfonate (ANS), was used as a fluorescent monitor of aggregate formation. ANS is a dye that binds to exposed hydrophobic regions of a protein resulting in an increase in dye fluorescence intensity and a blue-shift of the fluorescence spectrum to a maximum of 475 nm (see Figure 2.1)<sup>126</sup>. This occurs because of the protein's ability to shield the dye from the polar solution environment<sup>126</sup>. ANS has often been used to characterize protein folding and it has been suggested that

ANS binding is a useful tool for detecting protein aggregation at early stages, with the capability of interacting with hydrophobic areas in fibrils and amorphous aggregates<sup>127</sup>. When the dye interacts with protein molecules there is a change in polarity and viscosity of the environment resulting in an increase in fluorescence intensity and shift in intensity maximum. Both hydrophobic and electrostatic interactions can contribute to ANS binding to protein<sup>127</sup>. There is still controversy as to what the specific binding modes of ANS to protein may be, and considerable evidence suggests that ANS likely has multiple mechanisms of binding<sup>128</sup>. ANS has been used very extensively in protein aggregation studies<sup>129</sup>. Recent investigations using ANS have suggested that it is an excellent probe for detecting toxic oligomers<sup>130</sup>. Several known amyloid forming peptides and proteins were monitored over time for ThT and ANS binding and for loss of cell viability when neuroblastoma cells were exposed to aggregates at various time points. ANS fluorescence, but not ThT fluorescence, correlated well with cell death, suggesting ANS is able to detect soluble oligomers that may be responsible for cell toxicity<sup>130</sup> (see Section 1.4 for more information about toxic protein aggregates). The ability of ANS to detect early aggregates makes it suitable for investigating differences in aggregation patterns as a result of varied ionic strength.



**Figure 4.1 ANS Fluorescence spectra in the presence of protein aggregates.** All samples contain 125  $\mu\text{M}$  ANS. The black line represents buffer, and the dotted and dashed grey lines represent aggregates formed by mutant A4V reduced apo SOD1. Both protein samples were incubated at 1 mg/mL in 20 mM Hepes, 1 mM TCEP, pH 7.4, and either 300 mM  $\text{Na}_2\text{SO}_4$  (dashed) or 300 mM NaCl (dotted), in anaerobic conditions at 37°C for approximately 400 hours.

#### 4.1.3 Colloidal Stability and the Second Virial Coefficient

Protein aggregation is controlled by a complex set of factors that dictate whether dominant amino acid interactions will be intramolecular or intermolecular. Significant emphasis has been placed on the conformational stability of a protein as a dominant factor in dictating aggregation (see Section 1.3.1). A second important variable is the role the solution conditions play in favouring electrostatic interactions between the protein and the solvent or protein-protein interactions<sup>131</sup>. Colloidal stability is the strength of protein-protein interactions compared to the strength of protein-solvent interactions<sup>22</sup>. It is measured by the second virial coefficient ( $B_{22}$ ). When  $B_{22}$  is positive it indicates that protein-solvent interactions are stronger than protein-protein interactions. This results in dominating repulsive forces between protein molecules and aggregation is disfavoured. In the case where  $B_{22}$  is negative, protein-protein interactions are stronger than protein-solvent interactions, favouring aggregation<sup>22; 125</sup>. It has been suggested that non-native aggregation of folded protein involves two energy barriers<sup>125</sup>. The first is the free energy of

unfolding of the protein ( $\Delta G_{\text{unf}}$ ) and the second is the interaction energy of protein-protein interactions, represented by the second virial coefficient. In conditions where conformational stability is dominant over colloidal stability (large  $\Delta G_{\text{unf}}$  and a negative  $B_{22}$ ) protein unfolding becomes the rate-limiting step in aggregation. In this case, aggregation can be decreased by increasing  $\Delta G_{\text{unf}}$ . In solution conditions that have high colloidal stability ( $B_{22}$  is a large, positive value), the colloidal stability is dominant over conformational stability and the interaction between proteins becomes the rate-limiting step in aggregation. In this case aggregation can be reduced by altering the solution conditions to increase repulsive interactions, thus increasing  $B_{22}$ <sup>125</sup>. For these reasons, colloidal stability is an important, but often overlooked, factor to consider when analyzing protein aggregation in varying solution conditions.

It is worth evaluating why certain solvent conditions may increase or decrease the second virial coefficient as a way of deciphering some of the complexities of the overall mechanisms of protein aggregation. A study of the colloidal stability of lysozyme showed that increasing the concentration of NaCl resulted in a decrease in  $B_{22}$ , which in turn increased protein aggregation<sup>132</sup>. This effect was more pronounced for denatured than native state lysozyme. Two potential explanations were suggested: first that NaCl promoted the formation of hydrophobic clusters between proteins, and second that charge shielding reduced electrostatic repulsion and promoted protein-protein association<sup>132</sup>. Another study sought to correlate the second virial coefficient with aggregation rates for lysozyme and a monoclonal antibody (mAb1)<sup>133</sup>. Aggregation was monitored in several different salts and concentrations. For mAb1 in acidic conditions,  $B_{22}$  decreased as the strength of the chaotropic salt increased according to the Hofmeister series. This also corresponded with an increase in the rate of aggregation<sup>133</sup>, and followed the same pattern of aggregation measured previously<sup>45</sup>. The interesting result from this study is the positive relationship between chaotropic strength and protein aggregation. This is contrary to the typical interpretation of the Hofmeister series in which ions that act as stronger protein stabilizers (weaker chaotropes, therefore stronger kosmotropes) typically result in decreased protein solubility and increased aggregation (see Section 1.3.3)<sup>43</sup>. For mAb1, the role of the chaotropes in promoting aggregation was rationalized based on the idea that chaotropes bind weakly to protein and instead pair with each other,

allowing for the promotion of protein-protein interactions<sup>45</sup>. However, while these interactions are important for mAb1 aggregation in acidic conditions and low ionic strength, contrary results have been observed for other proteins in other studies<sup>42; 44</sup>, indicate the promotion of aggregation by chaotropes is not a universal principle. One study looked at the effect of several salts on seven different proteins and concluded that the effect of the salts generally followed the Hofmeister series, in that the stronger the kosmotrope, the larger the decrease in  $B_{22}$ <sup>134</sup>. It also demonstrated that at low salt concentrations (< 0.5 M) protein interactions could be attractive or repulsive, depending on the protein<sup>134</sup>. The diversity in these results further emphasizes the complexity of factors that can modulate protein-protein and protein-solvent interactions. In light of these results and the ability of the second virial coefficient to correlate with observed aggregation, the examination of colloidal stability is important for increased understanding of the variables that contribute to protein aggregation, and was investigated here.

#### **4.1.4 Monitoring Apo A4V Aggregation in the Presence of Salt**

The SOD1 mutation, A4V, was selected as a model protein for facilitating further investigations of the role of salt in protein aggregation. Reduced apo A4V aggregates readily in physiological solution conditions<sup>27</sup> and so provides an ideal system for monitoring the effects of changes in solution conditions on protein aggregation. Salts with differing kosmotropic strengths ( $\text{Na}_2\text{SO}_4$ , NaCl) were employed for these investigations. The aggregation patterns for both oxidized apo and reduced apo were monitored extensively in varying concentrations of these salts.

## **4.2 Methods**

### **4.2.1 Protein Production and Purification**

This project employed well established methods in the Meiering lab in order to produce holo SOD1<sup>135; 136</sup>. Briefly, plasmids containing the A4V SOD1 gene were transformed into SOD<sup>-/-</sup> *E.coli* cells (cells lacking all bacterial forms of SOD). Osmotic shock was used to release SOD1 from the periplasmic space. Purification was completed by heat-treating the osmotic solution while copper-charging the protein, then using an ammonium sulphate gradient on a hydrophobic interaction column to separate the

purified holo protein<sup>117; 136; 137</sup>. Purified protein samples were concentrated using a centricon device with a 10,000 Da molecular weight cut-off, cellulose membrane. Final concentrations were determined prior to storage by measuring absorbance at 280 nm and using an extinction coefficient of 5,400 M<sup>-1</sup>cm<sup>-1</sup> per monomer<sup>138</sup>. Samples were filtered with a 0.2 µm filter, flash frozen in liquid nitrogen, and stored at -80°C.

#### **4.2.2 De-metallation of Holo SOD1 to Prepare Apo SOD1**

In order to remove the metals from the holo protein, a well established protocol in which a series of dialyses with EDTA in acidic conditions, followed by dialyses in sodium chloride to remove EDTA and finally in water to remove salt, was used<sup>139</sup>. In brief, holo SOD1 at a concentration of approximately 0.5 mg/mL was prepared in 3500 Da molecular weight cut-off dialysis tubing. Four exchanges against 100 mM EDTA, 50 mM sodium acetate, pH 3.8 over an average time period of 36 hours was followed by four exchanges of similar length in 100 mM NaCl, 50 mM sodium acetate, pH 3.8. The final step was to exchange the protein into water by four exchanges over a 36 hour period. The protein was then concentrated using a centricon device, filtered with a 0.2 µm filter, flash frozen in liquid nitrogen, and stored at -80°C.

#### **4.2.3 Disulphide Reduction**

In order to study SOD1 in its most immature form it is necessary to reduce the cysteines at position 57 and 146 in order to eliminate the intramolecular disulphide bond formed at this site. This was accomplished through a protocol developed by K. Vassall in the Meiering lab<sup>27</sup>. First, in order to expose the disulphide bond which is located in the hydrophobic core of the protein, the protein must be denatured. The disulphide-intact apo protein was placed in a solution of 2 M guanidinium chloride, 20 mM Hepes, pH 7.8 that had been degassed for 30 minutes prior to the addition of protein. The sample was incubated at room temperature for 30 minutes to allow for complete unfolding. Next, the charged reducing agent, tris (2-carboxymethyl) phosphine (TCEP) hydrochloride was added to a final concentration of 10 mM, resulting in the reduction reaction of the disulphide bond. The mixture was

incubated at 37°C in an anaerobic environment (sealed dessicator) for 1 hour. Finally, the reduced protein mixture was exchanged into a 1 mM or 10 mM TCEP, 20 mM HEPES, pH 7.8 buffer using nanocsep centrifuge (molecular weight cut-off of 3000 Da) tubes by performing successive dilutions and concentrations. The disulphide status of the final protein was later confirmed by Differential Scanning Calorimetry (DSC) and SDS-PAGE<sup>27</sup>.

#### **4.2.4 Iodoacetamide Modification SDS-PAGE Gels**

To verify the disulphide status of reduced apo SOD1 samples a procedure involving iodoacetamide modification and separation by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) was used<sup>27</sup>. Disulphide-reduced apo SOD1 has a more expanded structure than disulphide-oxidized apo SOD1, and runs slightly slower than the more compact oxidized form on a denaturing gel, allowing the two forms of the protein to be distinguished. To prevent the re-oxidation of reduced SOD1 that can occur on the gel, the free cysteines were modified with iodoacetamide, as has been described previously<sup>64</sup>. Protein samples were precipitated by incubation on ice for 20 minutes with equal volume of 20% Trichloroacetic Acid (TCA). The samples were then pelleted by centrifugation and the supernatant was discarded. Acetone was used to wash the pellet by re-suspension in a small volume of acetone. Centrifugation was again used to pellet the sample and the supernatant was discarded. The pellet was then dried under vacuum and dissolved in buffer containing 50 mM Hepes, 2.5% SDS, 1 mM bathocuprione disulfonate, and 100 mM iodoacetamide, pH 7.2. The samples were incubated for 1 hour in an anaerobic environment to allow for modification of all free cysteines. After 1 hour the samples were diluted 1:1 in 2X SDS-PAGE loading buffer, loaded onto a 15% SDS-PAGE gel, and run at a constant voltage of 140 V for approximately 2 hours.

#### **4.2.5 Preparation of Small Volume Aggregation Trials**

Oxidized protein stock or freshly reduced SOD1 (filtered and confirmed with DLS as monodisperse for monomer species) was diluted to a final protein concentration of 1 mg/mL using the appropriate stock buffers to result in 140 µL samples with the following buffer conditions: 20 mM Hepes,

1 or 10 mM TCEP (for reduced samples only), and 0 mM, 150 mM or 300 mM of the appropriate salt (NaCl, Na<sub>2</sub>SO<sub>4</sub> or NaH<sub>2</sub>PO<sub>4</sub>), and pH 7.4. The 140 μL samples were prepared in conical glass vials (03-344-153, Wheaton Science Products) and lids were wrapped with parafilm. During the aggregation trials, all samples were kept at 37°C in anaerobic conditions.

#### **4.2.6 ANS Fluorescence to Monitor Protein Aggregation**

The fluorescence of 1-Anilino-8-naphthalene sulfonate (ANS) was monitored daily during aggregation time-trials using a Thermo Scientific NanoDrop 3300 Fluorometer. 1.5 μL of protein aggregate sample was combined with 0.5 μL of 500 μM ANS on the sample platform of the surface of the NanoDrop 3300, for a final concentration of 0.75 mg/mL protein and 125 μM ANS. After a period of 45 seconds a fluorescence measurement was taken by excitation with UV light at a wavelength of 365 nm ± 10 nm and an emission spectrum was recorded between 400 and 650 nm. The maximum fluorescence of ANS when bound to protein was approximately 475 nm, and so the fluorescence at 475 nm was recorded. The measurement was repeated 3 more times for the same sample, and two samples were measured for every vial at each time point, giving a total of 8 readings. The average fluorescence at 475 nm was plotted against time to monitor the time-course of aggregation for oxidized or reduced apo A4V in each of the solution conditions used.

#### **4.2.7 Thioflavin T Fluorescence**

ThT fluorescence was measured for the end point samples of several aggregation trials to evaluate the presence of amyloid-like aggregates. Protein samples were combined 1:1 with buffer containing 50 μM ThT and 100 mM Glycine, pH 9 and immediately a fluorescence measurement was taken using a Flouorolog fluorometer. The excitation wavelength was 445 nm and the emission spectrum was measured between 455 and 600 nm. The average of five spectra is reported as the ThT spectrum for each protein sample.

#### 4.2.8 Measuring Aggregate Sizes Using DLS

Dynamic Light Scattering was used at the beginning of reduced apo aggregation trials to confirm that the filtered reduced samples were monodisperse for monomer and there were no aggregate species present that could seed aggregation. In addition, at the end of the aggregation time-trials DLS was used to analyze the size of the aggregates formed. Light scattering measurements were made using the Malvern Zetasizer Nano-ZS. A clean, de-dusted, 45  $\mu$ L, quartz cuvette was used to measure the particle sizes present in each aggregate sample. Measurements were taken at 37°C and three sets of five measurements each were averaged to obtain the size distributions of each sample. The intensity plots were used to compare the size of the aggregated species.

#### 4.2.9 Measuring the Second Virial Coefficient Using Light Scattering

Light scattering was also used to determine the second virial coefficients for each of the solution conditions used in the aggregation trials. Debye's light scattering equation (Equation 1) can be used to obtain the second virial coefficient ( $B_{22}$ ) from the scattering intensity measured over a range of protein concentrations.

$$\frac{Kc}{R_{\theta}} = \frac{1}{M} + 2B_{22}c \quad (1)$$

$R_{\theta}$  is the Rayleigh ratio, which is the ratio of scattered light intensity at angle  $\theta$  to incident light intensity.  $K$  is an optical constant,  $c$  is the protein concentration and  $M$  is the molecular weight of the protein<sup>131</sup>. Light scattering was used to measure the scattering intensity which was used to calculate the Rayleigh ratio. A plot with  $Kc/R_{\theta}$  on the y-axis and concentration on the x-axis is called a Debye plot. The linear relationship between these variables allows for the determination of the molecular weight of the protein (1/y-intercept) and the second virial coefficient (slope/2). The Malvern software provides a method for molecular weight and  $B_{22}$  determination by Debye analysis, which was used to analyze a series of light scattering measurements that were taken for freshly prepared protein samples (prior to any detectable aggregation).

#### **4.2.10 DNTB Assay for the Determination of TCEP Oxidation**

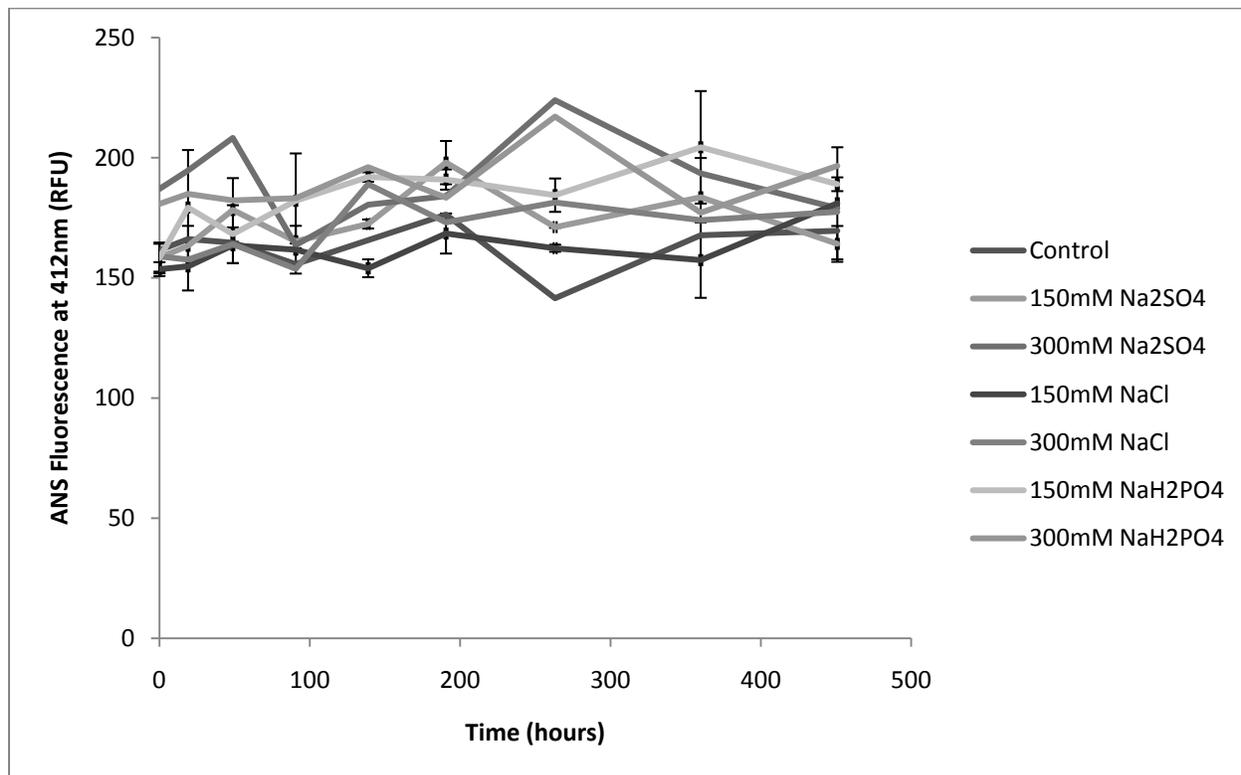
In order to monitor the extent of oxidation of the reducing agent, TCEP, used in the reduced apo A4V aggregation trials, an assay based on 5, 5'-dithiobis (2-nitrobenzoic acid) (DNTB) previously described by Han and Han<sup>140</sup>, was used. DTNB is composed of two 2-nitro-5-thiobenzoate (NTB) units linked by a disulphide bond. In the presence of TCEP, DNTB is reduced to form two NTB molecules. NTB is a chromophore with a maximum absorbance at 412 nm, where DTNB has negligible absorbance. By measuring the concentration of NTB produced, using the extinction coefficient  $14,150 \text{ M}^{-1}\text{cm}^{-1}$ , the concentration of reduced TCEP can be calculated as half of the concentration of NTB<sup>140</sup>. Original TCEP stock was made by weighing out solid TCEP-HCl and dissolving in water for a final concentration of 100 mM. Further dilutions were made with the appropriate buffer for a final concentration of 1 mM TCEP. Alternatively, neutral TCEP was stored as a stock solution at 500mM, according to manufacturer instructions, and was diluted into the appropriate buffer for a final buffer concentration of 10 mM TCEP. The assay was performed by combining 495  $\mu\text{L}$  of 100 mM TRIS, 100  $\mu\text{M}$  DTNB, pH 7.5, with 5  $\mu\text{L}$  of buffer containing 1 mM TCEP, or 499  $\mu\text{L}$  of TRIS/DTNB buffer with 1  $\mu\text{L}$  of 10 mM TCEP. An absorbance spectrum was measured from 600 to 200 nm and the absorbance at 412 nm was used to calculate the concentration of NTB. Based on the observed versus expected concentration of NTB the proportion of reduced TCEP to total TCEP was calculated.

### **4.3 Results**

#### **4.3.1 Apo Oxidized A4V Aggregation in Salt**

The aggregation patterns of oxidized apo A4V were monitored in a range of solution conditions using several methods for analysis. The solution conditions used were 20 mM Hepes, pH 7.4 and the following salt concentrations: 0 mM added salt, and 150 mM or 300 mM of sodium sulphate ( $\text{Na}_2\text{SO}_4$ ), sodium chloride (NaCl), or sodium phosphate ( $\text{NaH}_2\text{PO}_4$ ). The aggregation of these samples was measured over several weeks using ANS fluorescence to monitor aggregate formation. Figure 4.2 gives

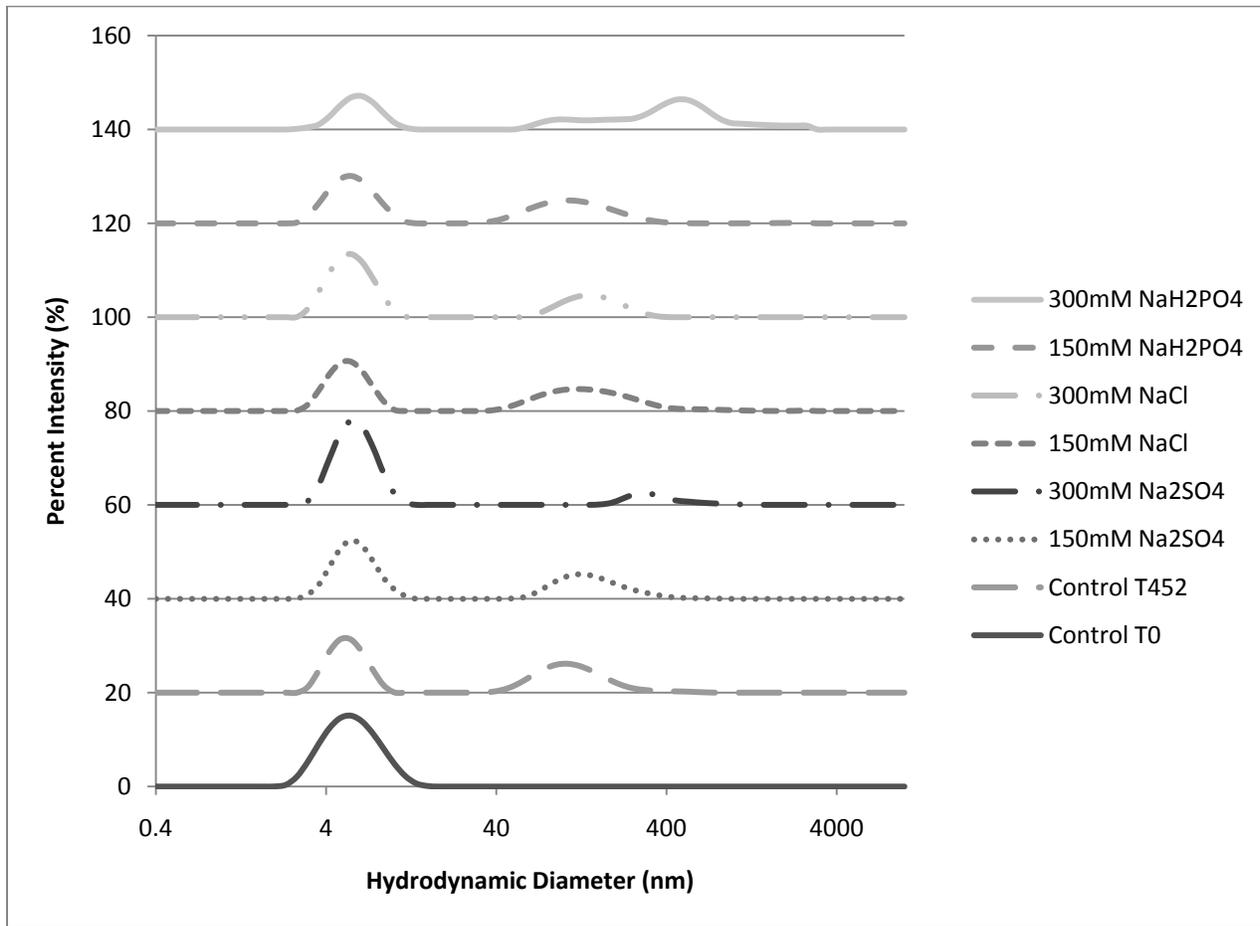
the summarized ANS results for oxidized apo A4V aggregation in all the solution conditions. There was very little difference in the ANS fluorescence over 450 hours for all samples.



**Figure 4.2 Time-course of ANS Fluorescence for apo oxidized A4V samples in varying salt conditions and 20 mM Hepes, pH 7.4. Samples were incubated at 37°C in anaerobic conditions. The control sample contained no added salt. Data presented is from 1 or 2 independent time-courses, and in the case of 2 time-courses, averaged values are plotted and standard deviation is represented by error bars shown in black.**

At the end of the time-course experiments, DLS was used to assess the size of the particles found in solution. Figure 4.3 shows the DLS results for the initial control sample and for all samples at the 452 hour end point, and Table 4.1 gives the corresponding hydrodynamic diameters of the particles in solution. These results show evidence for the presence of larger species by the end of the aggregation time-trials; however, the dominant species in all conditions is the dimer peak, at around 6 nm. Light scattering intensity is dependent on the sixth power of the diameter of the scattering particle<sup>116</sup>, thus when

only a small fraction of the percent mass of the sample is present as a high molecular weight species it can easily dominate the entire sample signal intensity. This indicates that even though two species are present in these samples, there is only very little of the larger molecular weight species.



**Figure 4.3 DLS results for end point samples of oxidized apo A4V aggregation time-trials.** 1 mg/mL oxidized apo A4V samples were incubated at 37°C, in 20 mM Hepes, pH 7.4, and various salt conditions. After 452 hours, 45  $\mu$ L aliquots were removed and DLS measurements were performed. All results shown are at the 452 hour time point with the exception of the control sample which is also shown at the initial time point.

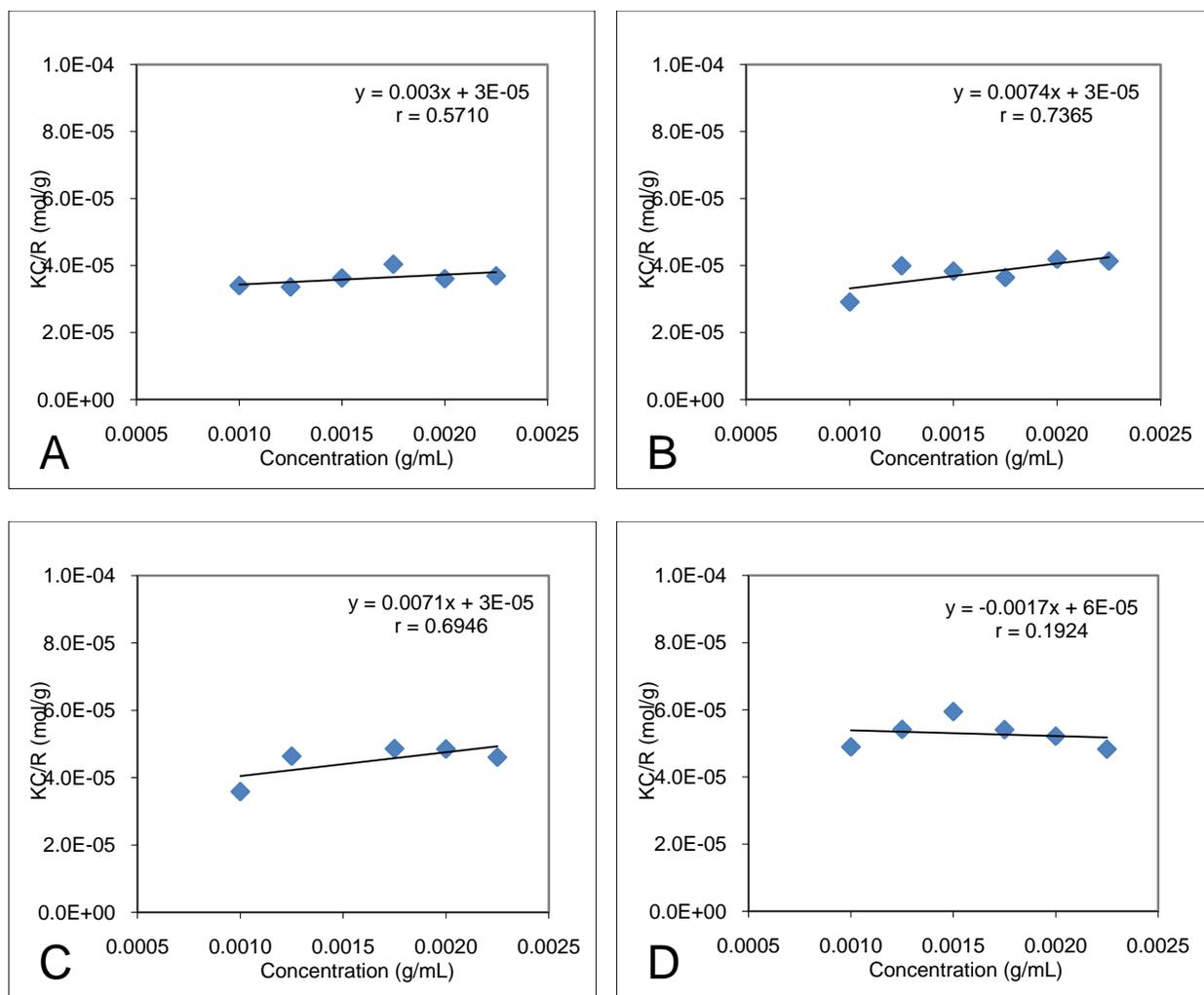
**Table 4.1 Average hydrodynamic diameters of particles in oxidized apo A4V samples at the end point of the ANS aggregation time-trial.**

	Control	Control	150 mM	300 mM	150 mM	300 mM	150 mM	300 mM
	T0	T452	Na <sub>2</sub> SO <sub>4</sub>	Na <sub>2</sub> SO <sub>4</sub>	NaCl	NaCl	NaH <sub>2</sub> PO <sub>4</sub>	NaH <sub>2</sub> PO <sub>4</sub>
<b>Hydrodynamic Diameter (nm)</b>	5.9	5.4	6.0	6.0	5.5	5.7	5.8	6.3
		117.0	155.9	373.1	172.9	153.4	125.9	519.8

The ANS and DLS results demonstrate limited aggregation of the apo oxidized A4V. To further investigate the role of the solution conditions in promoting protein-protein interactions, light scattering measurements were taken to create Debye plots for each type of salt at a concentration of 150 mM and in the control buffer. The plots are given in Figure 4.4 and the calculated molecular weights and second virial coefficients are given in Table 4.2. Three separate experiments were performed to obtain Debye plots for each solution condition, two at concentration ranges from 0.5 - 1.5 mg/mL and one at a range of 0.5 - 2.25 mg/mL. The data in the Debye plots show increased scatter at lower protein concentrations; this is not unexpected as this is reaching the lower limits for instrument performance. Accordingly, multiple samples were prepared for low protein concentrations and the data were averaged (Figure 4.4). In addition, for a few of the plots, major outliers were removed from the correlation based on the Q-test with a 90% confidence interval. In Figure 4.4, panels A, B, and C (Control, 150 mM NaCl, and 150 mM Na<sub>2</sub>SO<sub>4</sub>, respectively) give correlations with r-values above 0.5. The 150 mM NaH<sub>2</sub>PO<sub>4</sub> plot (Figure 4.4 D) gives an unconvincing result with an r-value of only 0.19. There are several reasons that may have contributed to the error in these experiments. Light scattering is an extremely sensitive technique and any non-homogenous distribution of particles in solution can have large impacts in the Debye plot calculations. If there was any contaminating dust in the solutions, this may have interfered with accurate results. Unfortunately, at the time when these experiments were performed the appropriate 0.02 μm filters typically used to remove dust and particulates from all buffers had been out of stock from the supplier for several months, and so samples were filtered instead with 0.2 μm filters which may have left some dust particles present in solution, potentially resulting in error in the light scattering measurements. Another

factor that could have been a source of error was the small volumes used for the preparation of these samples. Limited amounts of protein were available and thus it was typical for volumes of less than 100  $\mu\text{L}$  to be transferred during sample preparation. There is generally a greater chance of error when using small volumes which may also contribute to the weak linear relationship observed for the phosphate buffer plot.

Despite the possible sources of error, summary Table 4.2 shows consistency in the trend of the  $B_{22}$  values obtained for 3 out of 4 conditions, and an accurate molecular weight calculation for 2 of the conditions. Both the control and the 150 mM  $\text{Na}_2\text{SO}_4$  samples give molecular weights very close to the expected molecular weight of the dimer, 31.5 kDa. The NaCl and  $\text{NaH}_2\text{PO}_4$  plots gave higher and lower than expected molecular weights respectively, which may be attributed to experimental uncertainty in the experiment. The second virial coefficient values ( $B_{22}$ ) are quite similar for the NaCl, and  $\text{Na}_2\text{SO}_4$  samples, with values of 0.00371 and 0.00354  $\text{mL}\cdot\text{mol}/\text{g}$ , respectively, and are approximately 2.5-fold greater than the value obtained for the control (0.00148  $\text{mL}\cdot\text{mol}/\text{g}$ ) The positive sign of these values indicates that oxidized A4V has a greater tendency to form protein-solvent interactions than protein-protein interactions in these conditions. In phosphate, the sign of  $B_{22}$  is negative, however, due to the poor correlation obtained for these results, replicate experiments are required to confirm the validity of this result.



**Figure 4.4** Debye plots for oxidized apo A4V in varying salt conditions. All samples are in 20 mM HEPES at a pH of 7.4. Panels A-D represent the following salt conditions respectively: 0 mM salt, 150 mM NaCl, 150 mM Na<sub>2</sub>SO<sub>4</sub> and 150 mM NaH<sub>2</sub>PO<sub>4</sub>. The plots given are the average Kc/R<sub>0</sub> values for the concentrations in which more than one measurement was taken, and the measurements greater than 1.5 mg/mL are from a single experiment.

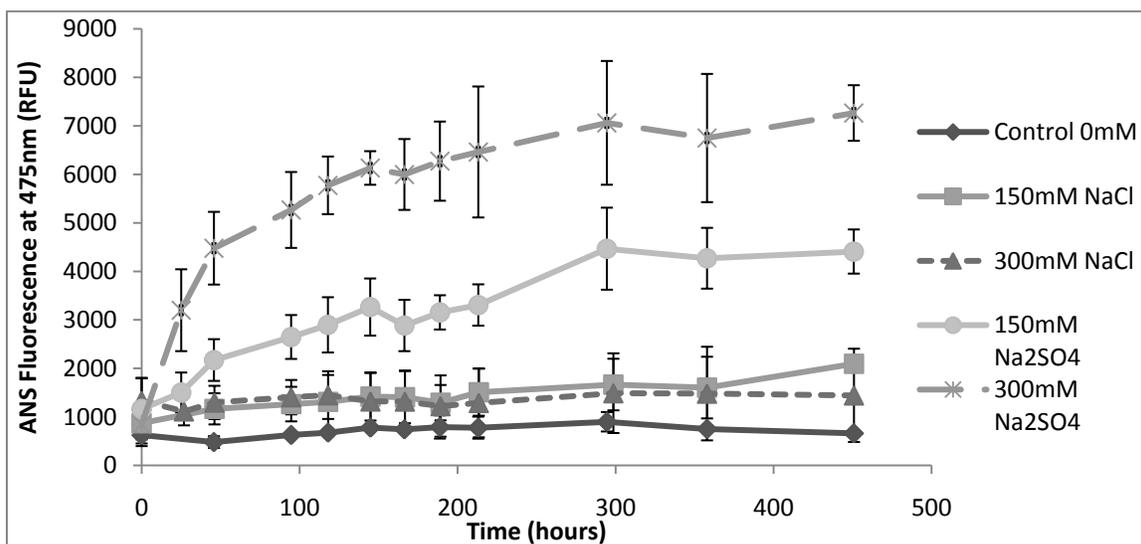
**Table 4.2** Molecular weight and second virial coefficients of oxidized apo A4V in varying salt conditions determined by Debye plot analysis.

	Control	150 mM NaCl	150 mM Na <sub>2</sub> SO <sub>4</sub>	150 mM NaH <sub>2</sub> PO <sub>4</sub>
<b>MW (g/mol)</b>	31,881	38,825	29,974	17,996
<b>B<sub>22</sub> (mL·mol/g)</b>	0.00148	0.00371	0.00354	-0.00084

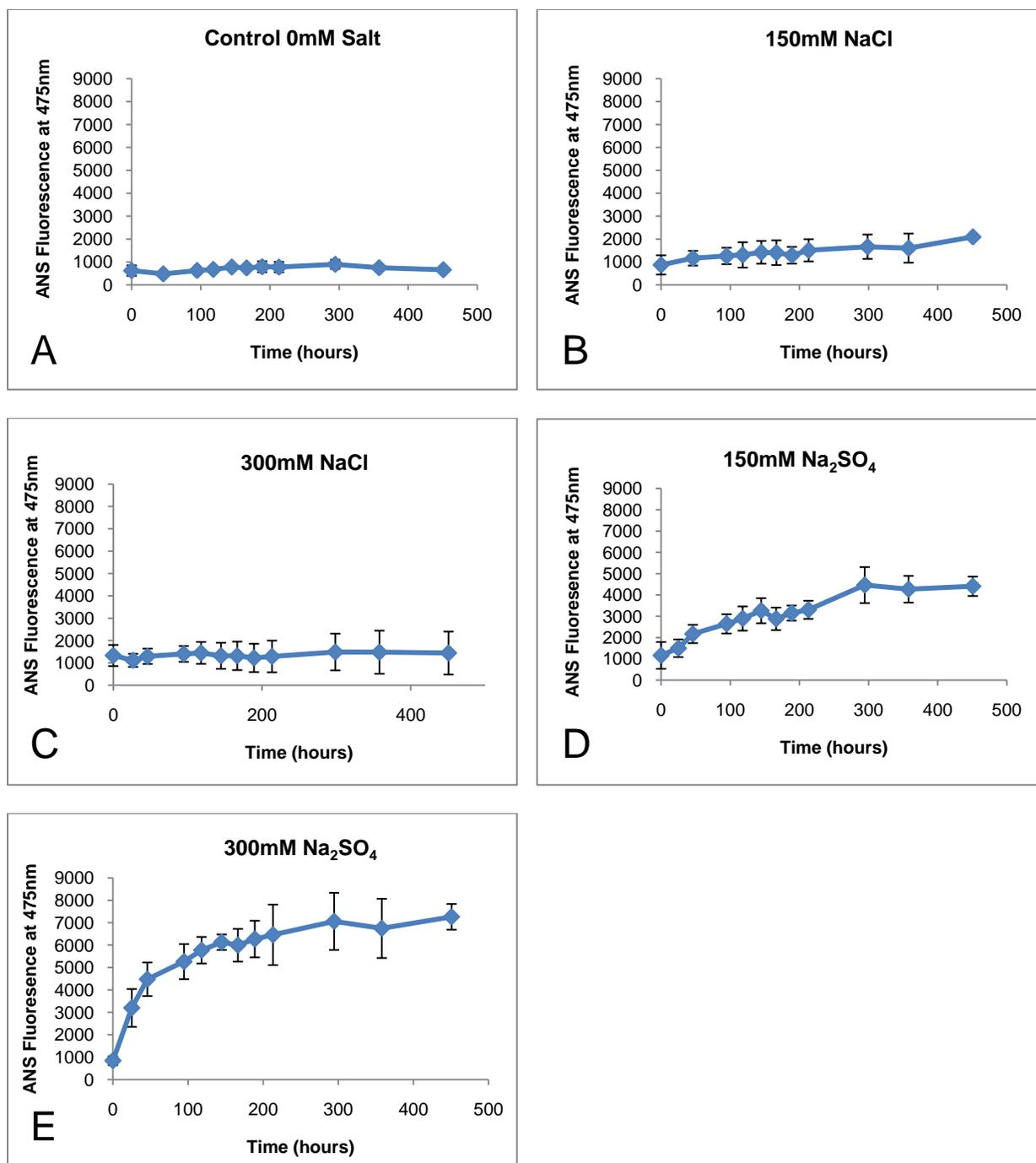
### 4.3.2 Reduced Apo A4V Aggregation in 1 mM TCEP Buffer and Varying Salt Conditions

#### 4.3.2.1 Time-course of ANS Fluorescence of Reduced apo A4V Aggregation in 1 mM TCEP Buffer and Varying Salt Conditions

Reduced apo A4V was prepared as described in Section 4.2.3 and aggregation time-course trials were observed using ANS fluorescence. Samples were monitored for an approximately 450 hour time-trial under anaerobic conditions at 37°C. Multiple independent time-courses were obtained giving a total of 2 to 4 replicates of each of the following solution conditions: 20 mM Hepes, 1 mM TCEP, pH 7.4 containing 0 mM, 150 mM, or 300 mM of NaCl or Na<sub>2</sub>SO<sub>4</sub>. Figure 4.5 compares the ANS results for all solution conditions and Figure 4.6 gives the individual ANS plots for each solution condition. The samples in Na<sub>2</sub>SO<sub>4</sub> exhibit the highest ANS fluorescence increase over time, while the NaCl samples have only moderate fluorescence increases. The 300 mM Na<sub>2</sub>SO<sub>4</sub> sample shows a larger ANS fluorescence increase than the 150 mM sulphate sample, while the two NaCl concentrations behave similarly to each other. The control sample, without added salt, has only a very slight overall increase in ANS fluorescence over the entire time-course.



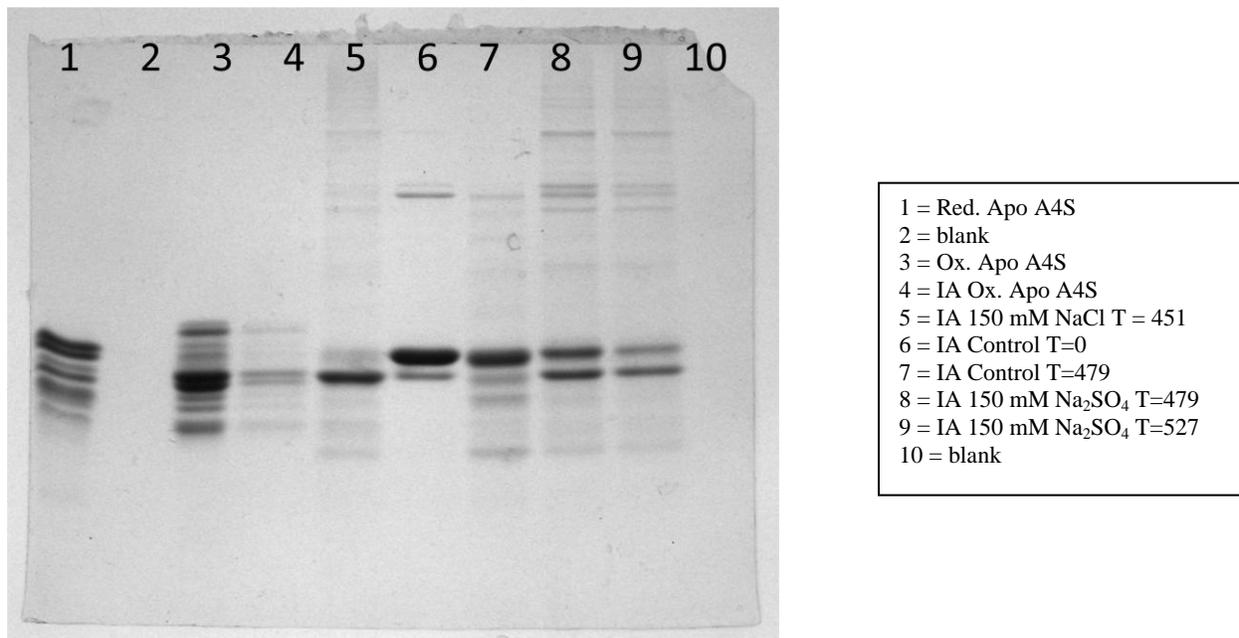
**Figure 4.5 Summarized ANS fluorescence results for reduced apo A4V samples in 20 mM Hepes, 1 mM TCEP and various salt concentrations as indicated in the legend.** Individual data points are averages for 2 or 3 independent time-courses, and error bars represent the standard deviation of the measurements.



**Figure 4.6 ANS Fluorescence for aggregation time-trials of reduced apo A4V samples in various salt conditions. All samples are in 1 mM TCEP, 20 mM Hepes, and incubated in anaerobic conditions at 37°C. Panels A-E represent the following salt samples respectively: 0 mM salt, 150 mM NaCl, 300 mM NaCl, 150 mM Na<sub>2</sub>SO<sub>4</sub>, and 300 mM Na<sub>2</sub>SO<sub>4</sub>. Individual data points are averages for 2 or 3 independent time-courses, and error bars represent the standard deviation of the measurements.**

#### 4.3.2.2 SDS PAGE gels and DTNB buffer Assays Reveal Reduced Apo A4V Oxidation

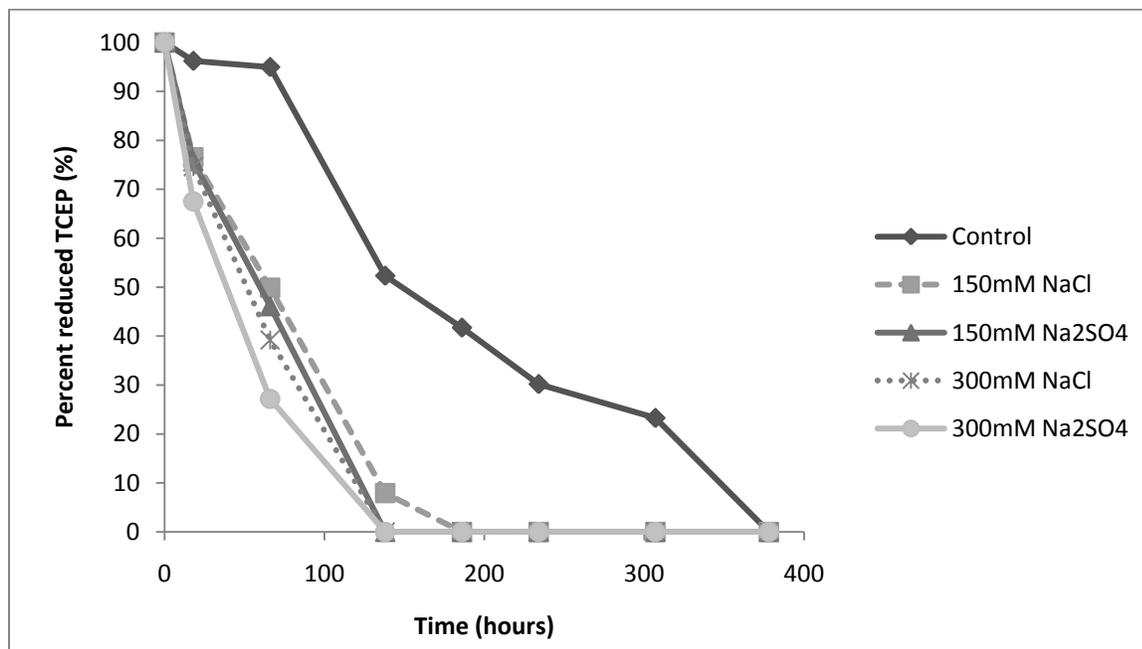
To determine whether the samples remained reduced throughout the time-course of the experiments, iodoacetamide-modified samples were run on an SDS PAGE gel (Figure 4.7). Some limited amounts of protein degradation are seen in a number of the A4V samples. This is sometimes observed for long duration time-courses. It is clear that although the protein is almost completely reduced at the starting time point (lane 6) it becomes substantially oxidized by the end of the time-trial (lanes 5, 8, and 9). Interestingly, the samples in salt seemed to oxidize more than the control sample, which remains significantly reduced (lane 7). In particular, the NaCl sample (lane 5) is almost completely oxidized by the 451 hour time point, and the Na<sub>2</sub>SO<sub>4</sub> samples (lanes 8 and 9) appear to be more than 50% oxidized. In addition, higher molecular weight bands are readily detectable in the sulphate samples, indicative of the formation of disulphide cross-linked species. There is some evidence of cross-linking, albeit less significant, in both the NaCl (lane 5) and control (lane 7) at the end of the aggregation time-course.



**Figure 4.7 Iodoacetamide-modified reduced apo A4V samples at various time points during aggregation time-courses in differing salt conditions and 1 mM TCEP, 20 mM HEPES, and 37°C.**

The lanes that have been modified by iodoacetamide have been denoted “IA.” The oxidized and reduced apo A4S control samples (lanes 1, 3 and 4) are significantly degraded, but nevertheless show where the oxidized and reduced bands should be. Time point of sample is listed in the given legend.

This oxidation of SOD1 was unexpected given that during previous experiments of reduced apo SOD1 aggregation in the absence of added salt, samples remained reduced over this time range<sup>27</sup>. The oxidation was therefore investigated further by monitoring the time-course of TCEP oxidation using the DTNB assay. Buffer samples were prepared identically to the buffers used in the reduced apo A4V aggregation trials. They were monitored without the presence of protein by incubation in the same conical glass vials used for the aggregation time-trials and were stored in an evacuated desiccator at 37°C. The time-course of the observed TCEP oxidation is shown in Figure 4.8. The solutions containing added salts become almost completely oxidized by approximately 150 hours while the control sample with no added salt becomes fully oxidized just prior to 400 hours. The oxidation of the TCEP likely plays a large role in the oxidation of the A4V samples observed by SDS PAGE (Figure 4.7). Even without salt present, the control buffer sample becomes oxidized more quickly than would be expected based on the stability of the TCEP molecule<sup>140; 141</sup>.

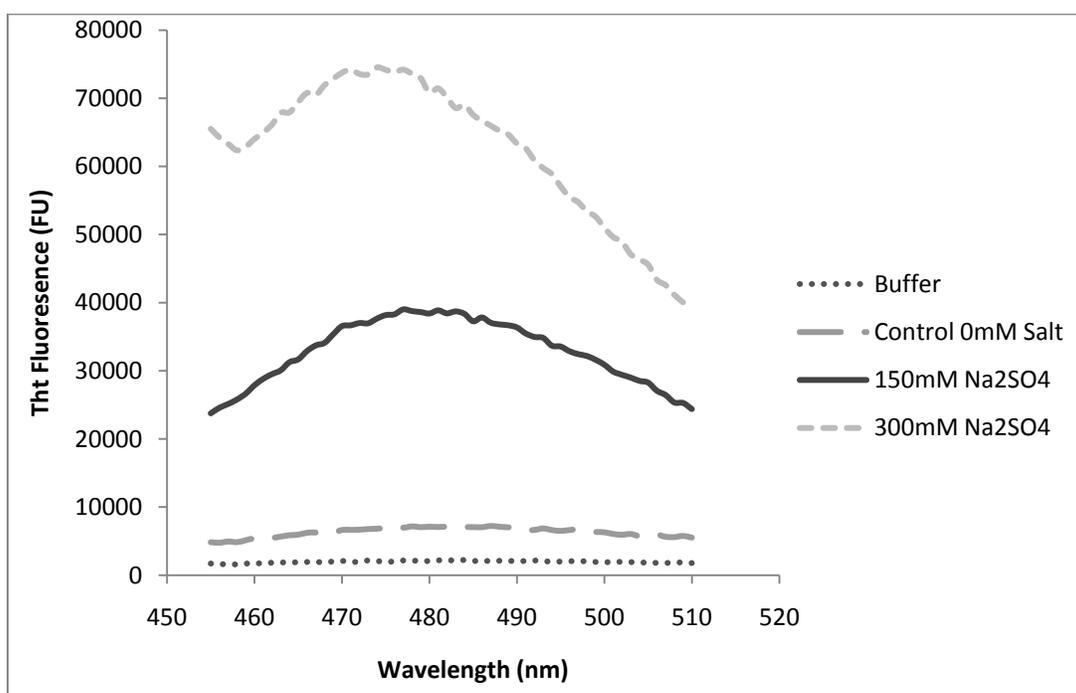


**Figure 4.8 Time-course of 1mM TCEP oxidation.** The percent reduced TCEP on the y-axis is calculated from the ability of TCEP to reduce DTNB into two NTB chromophores.

Previous studies have suggested that a 2 mM TCEP solution at pH 7.2, 25°C, in 50 mM 3-(N-morpholino)propanesulfonic acid (MOPS), 2 mM MgCl<sub>2</sub>, and 50 mM KCl, undergoes oxidation as a first order reaction and becomes approximately 10% oxidized in one week<sup>141</sup>. This corresponds to an oxidation rate of  $1.0 \times 10^{-5} \text{ min}^{-1}$ <sup>142</sup>. The half life of oxidation based on this value is 1155 hours, much longer than the rate of oxidation observed in the conditions used here. On the contrary, the reaction trend in Figure 4.8 appears to be closer to zero order kinetics than to first order. In the salt conditions, 1 mM TCEP is completely oxidized by approximately 150 hours, giving a zero order reaction rate of  $1.11 \times 10^{-4} \text{ mM/min}$ , while without salt complete oxidation occurs around the 380 hour mark, giving a reaction rate of  $4.39 \times 10^{-5} \text{ mM/min}$ . The rate of oxidation has also been measured previously in conditions similar to the conditions used in this experiment by Han and Han, 1994<sup>140</sup>. In this case, 2 mM TCEP in 50 mM Hepes and pH 6.8 or 8.2 became 14.8% or 13.6% oxidized respectively, over a three week period. This translates into zero order rate constants of  $9.79 \times 10^{-6} \text{ mM/min}$  at pH 6.8, and  $9.00 \times 10^{-6} \text{ mM/min}$  at pH 8.2. Again, these values are considerably smaller than the rates observed here at pH 7.4, 37°C. The kinetics of the reaction monitored by Han and Han<sup>140</sup> followed more closely to first order than zero order, which would result in first order rate constants of  $5.30 \times 10^{-6} \text{ min}^{-1}$  at pH 6.8 and  $4.83 \times 10^{-6} \text{ min}^{-1}$  at pH 8.2. These translate into half-lives of 2180 hours and 2392 hours, respectively. Therefore, the fact that the TCEP becomes fully oxidized within 150 hours for salt buffers or 400 hours for the control buffer is quite surprising. The observation that TCEP becomes oxidized so quickly likely plays an important role in the aggregation pathway of reduced apo A4V under these conditions (discussed further in Section 4.4.2.3).

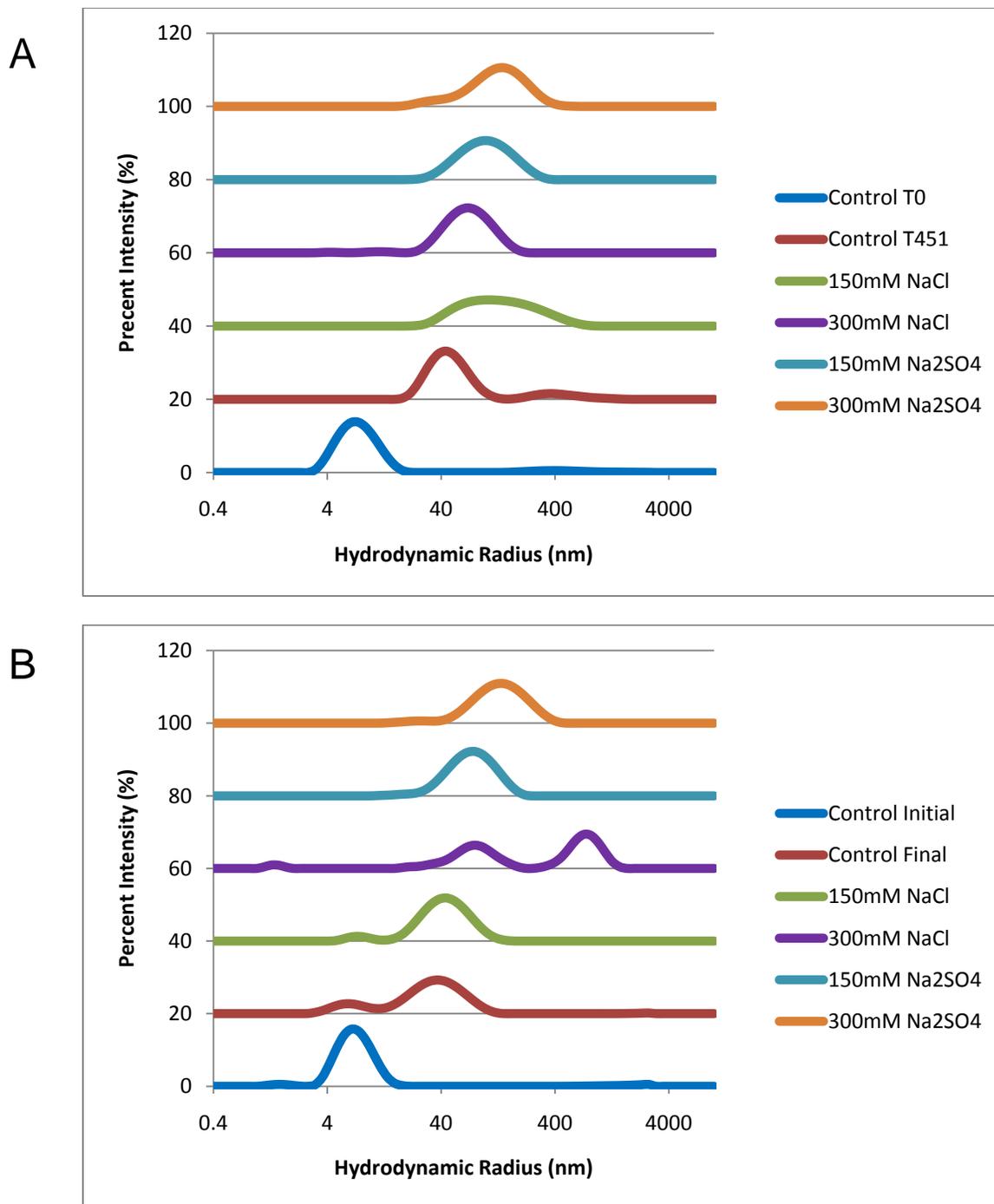
#### 4.3.2.3 ThT Fluorescence and DLS Measurements of End-point Samples of Reduced Apo A4V Time-courses

Two final experiments were performed to characterize the nature of the aggregates formed in these time-courses. To test whether the cross-linked aggregates, seen on the SDS PAGE gel (Figure 4.7), have amyloid-like characteristics, ThT fluorescence measurements were performed on the aggregates formed in the control and the sulphate samples. Unfortunately, there was not enough sample volume left from the chloride samples at the end of the time-course to perform ThT analysis. Figure 4.9 shows the ThT results from a single experiment with one sample for each condition. The increased ThT fluorescence suggests that the aggregates formed in the presence of sulphate have amyloid-like characteristics. The samples with no added salt have very little ThT fluorescence and thus do not contain amyloid-like aggregates. Although these results are intriguing, replicate samples are required to confirm their validity.



**Figure 4.9** ThT Fluorescence for end point aggregation samples of reduced apo A4V with and without Na<sub>2</sub>SO<sub>4</sub>. Samples were incubated at 1 mg/mL, 20 mM Hepes, pH 7.4, 37°C in an evacuated dessicator for approximately 475 hours and diluted 1:1 in ThT buffer (50µM ThT, 100 mM Glycine, pH 9) immediately prior to fluorescence measurement.

A second test to investigate the nature of the aggregates formed during these experiments was to use DLS to assess the size of the aggregates. Figure 4.10 shows the averaged DLS results of the end point samples for two independent time-courses in 20 mM Hepes, 1 mM TCEP and varying salt conditions. In all conditions the monomer peak disappears and the intensity is dominated by aggregated species. Table 4.3 gives the size of the dominant species in each solution for each trial. The species formed in the salt conditions have larger hydrodynamic diameters than those formed in the control buffer. The aggregates formed by the end of the time-trial in the control buffer are around 40 or 48 nm, whereas the dominant species formed in the salt samples ranges from 47 to 163 nm. The 300 mM NaCl sample in panel B shows two different sizes of aggregates being formed, one approximately 750 nm in size. These results suggest that salt induces the formation of aggregate species with greater hydrodynamic diameters than in the control buffer. A comparison was performed between the derived mean count rate, which reports on the total intensity of scattered light, for each end point sample using DLS and the final ANS fluorescence for the same sample. The correlation plots for two independent experiments are given in Figure 4.11. The  $r$ -values are both 0.98, with high statistical significance. This demonstrates a strong relationship between the amount of scattered light and the ANS fluorescence. Since total scattered light intensity is a common measure of total aggregation, it is likely that the ANS fluorescence increase results from increased protein aggregation.



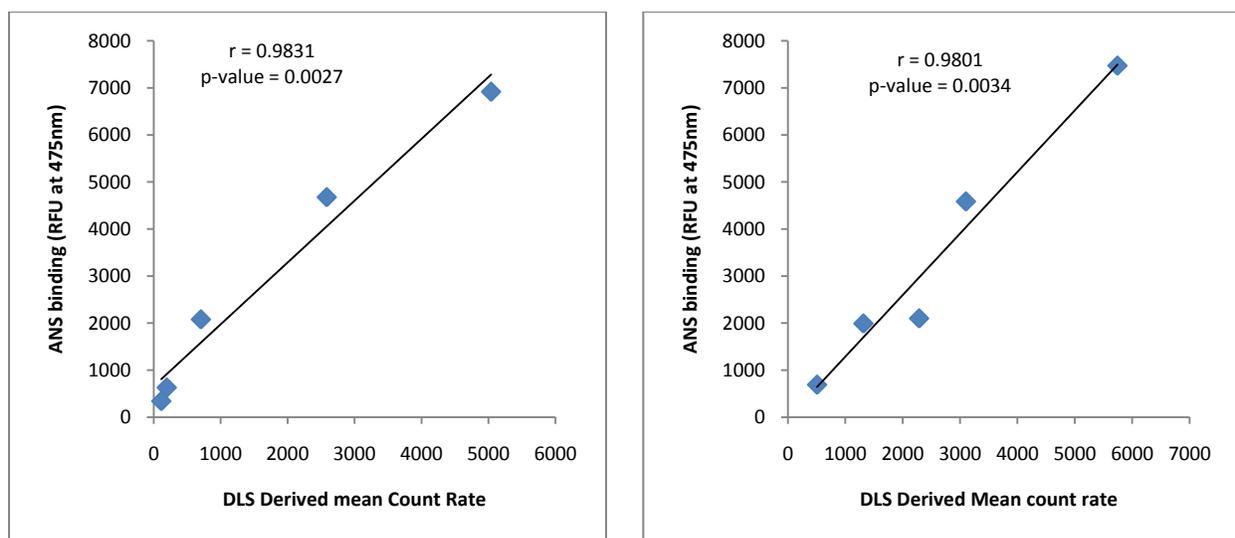
**Figure 4.10 DLS results for reduced apo A4V aggregation trials in 1 mM TCEP, 20 mM Hepes, and various concentrations of salt.** Measurements are given for the initial time point of the control sample and for all samples at the end point. Samples were incubated for approximately 450 hours at 37°C in an anaerobic environment prior to DLS measurements. Panels A and B represent two independent experiments.

**Table 4.3 Average hydrodynamic diameters of end point aggregate species for two independent time-courses of reduced apo A4V.**

		Initial Control	Final Control	150 mM NaCl	300 mM NaCl	150 mM Na <sub>2</sub> SO <sub>4</sub>	300 mM Na <sub>2</sub> SO <sub>4</sub>
Average Hydrodynamic Diameter (nm)	Trial 1	7.7	48.6	163.4	75.8	109.6	141.7
	Trial 2	7.3	40.7	47.94	753.17 91.35	80.1	145.0
	Average <sup>†</sup>	7.5 ± 0.3	44.7 ± 5.6	105.7 ± 81.6	83.6 ± 11.0	94.9 ± 20.9	143.3 ± 2.3

\*All samples excluding the initial control are at the end point of the aggregation time-course experiments. For Trial 1 this was 451 hours and for Trial 2 480 hours. All samples were at 1 mg/mL protein, 1 mM TCEP, 20 mM Hepes, pH 7.4, and 37°C. Each diameter value for the individual trials is an average of three consecutive measurements of the same sample.

<sup>†</sup>The average was determined for the hydrodynamic diameter values of trials 1 and 2 in each condition. In the case where one trial showed more than one species, the average is reported of the species from each trial with the most similar diameter size.



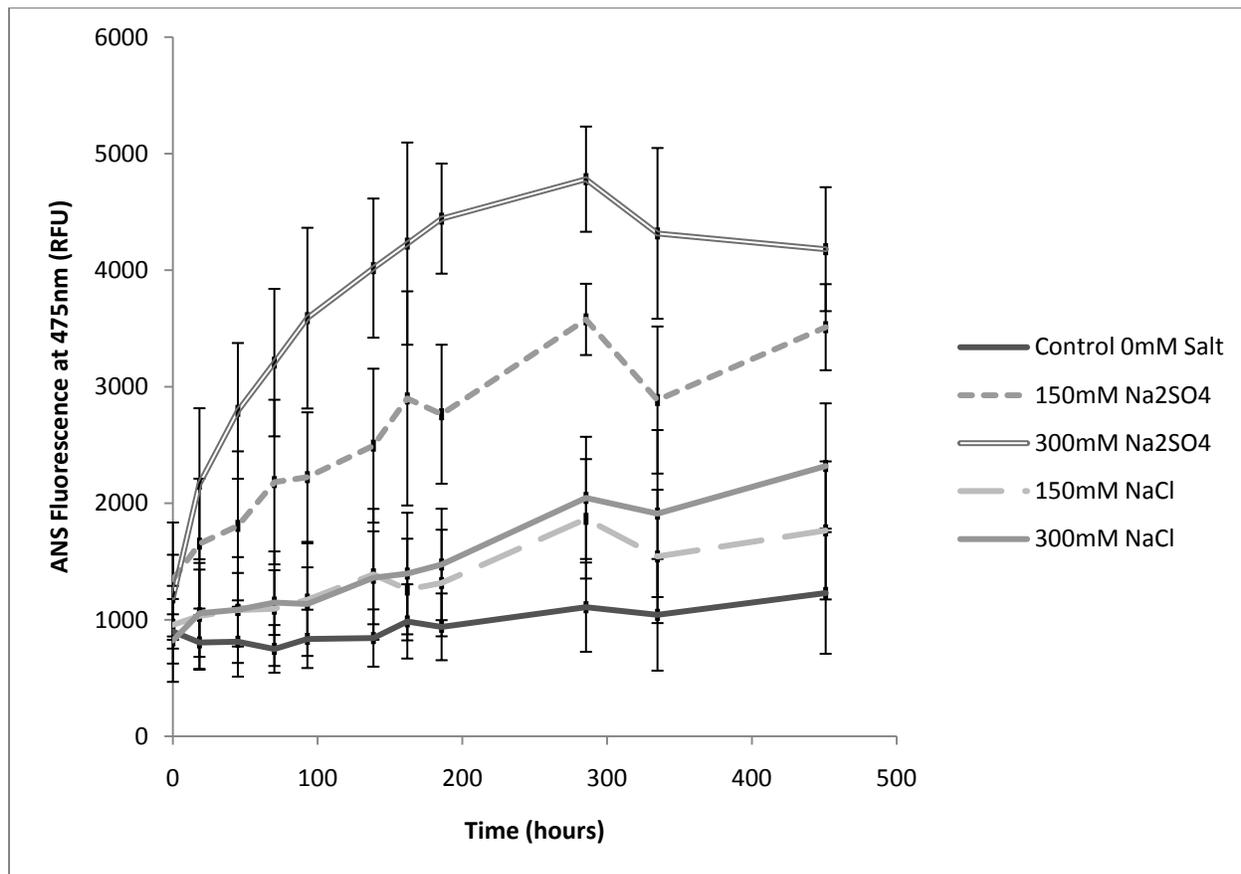
**Figure 4.11 Relationship between ANS Fluorescence and DLS derived mean count rate at the end point of reduced apo A4V aggregation time-trials in 20 mM Hepes, 1 mM TCEP, and varying salt conditions.**

### 4.3.3 Reduced apo A4V Aggregation in 10 mM TCEP Buffer and Varying Salt Conditions

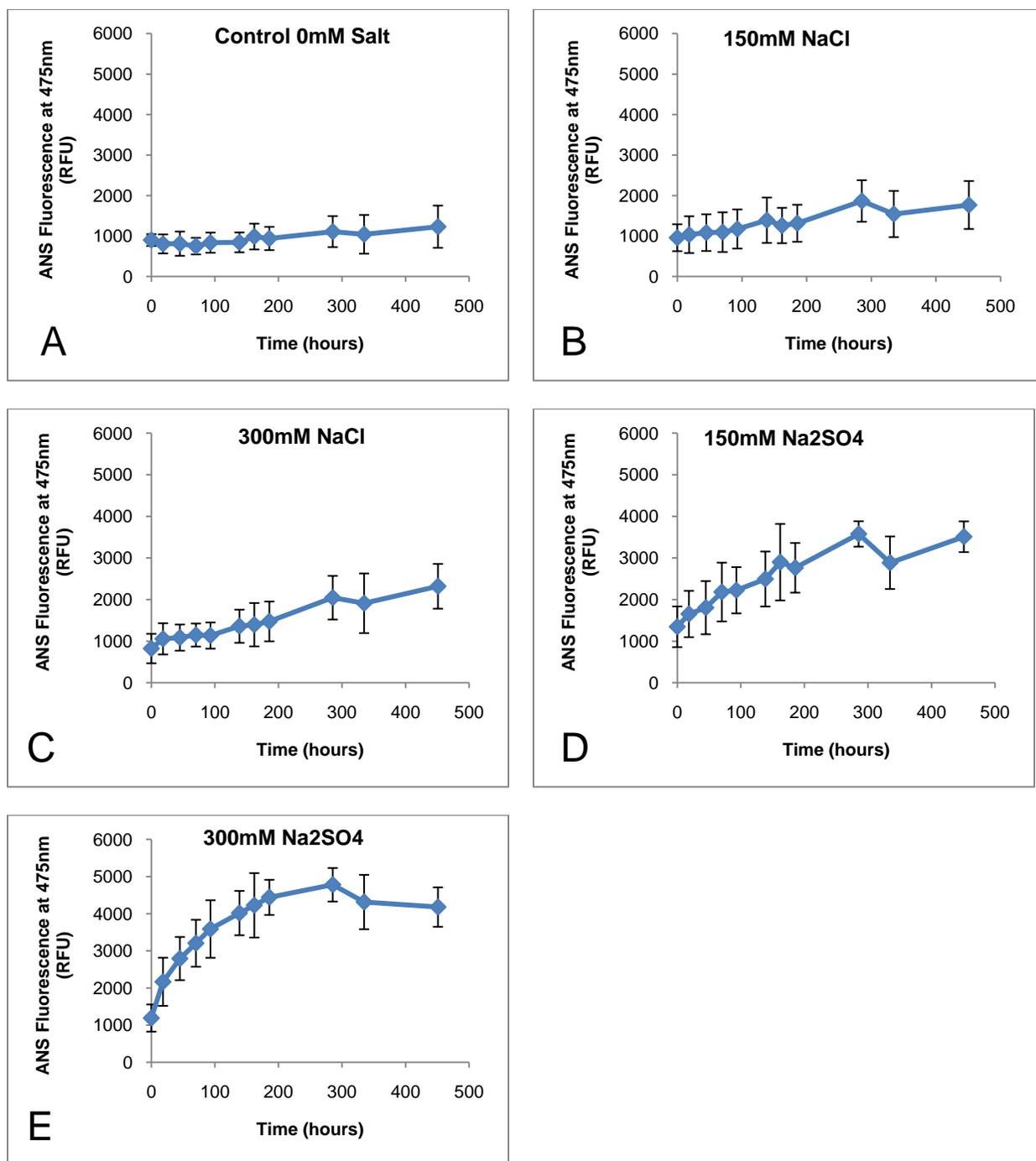
#### 4.3.3.1 Time-course of ANS Fluorescence of reduced apo A4V Aggregation in 10 mM TCEP Buffer and Varying Salt Conditions

To address the issue of TCEP oxidation during the course of the aggregation experiments in 1 mM TCEP, identical experiments were performed in 10 mM TCEP buffers, in attempt to keep the free

thiols of apo A4V reduced during the time-course of the experiments. Time-trials were carried out over approximately 450 hours by incubation in an anaerobic environment at 37°C as before. Samples were prepared containing 1 mg/mL protein, 20 mM Hepes, 10 mM TCEP, pH 7.4 and 0 mM salt or 150 mM or 300 mM NaCl or Na<sub>2</sub>SO<sub>4</sub>. The comparative ANS results for all of the solution conditions are given in Figure 4.12. It is evident that again Na<sub>2</sub>SO<sub>4</sub> has the greatest influence on aggregation, with 300 mM Na<sub>2</sub>SO<sub>4</sub> resulting in the highest ANS fluorescence, followed by 150 mM Na<sub>2</sub>SO<sub>4</sub>, 300 mM NaCl, 150 mM NaCl and the control sample with no added salt. Figure 4.13 shows the ANS results for the individual conditions to further illustrate the trend in aggregation and reproducibility of the ANS fluorescence time-courses.



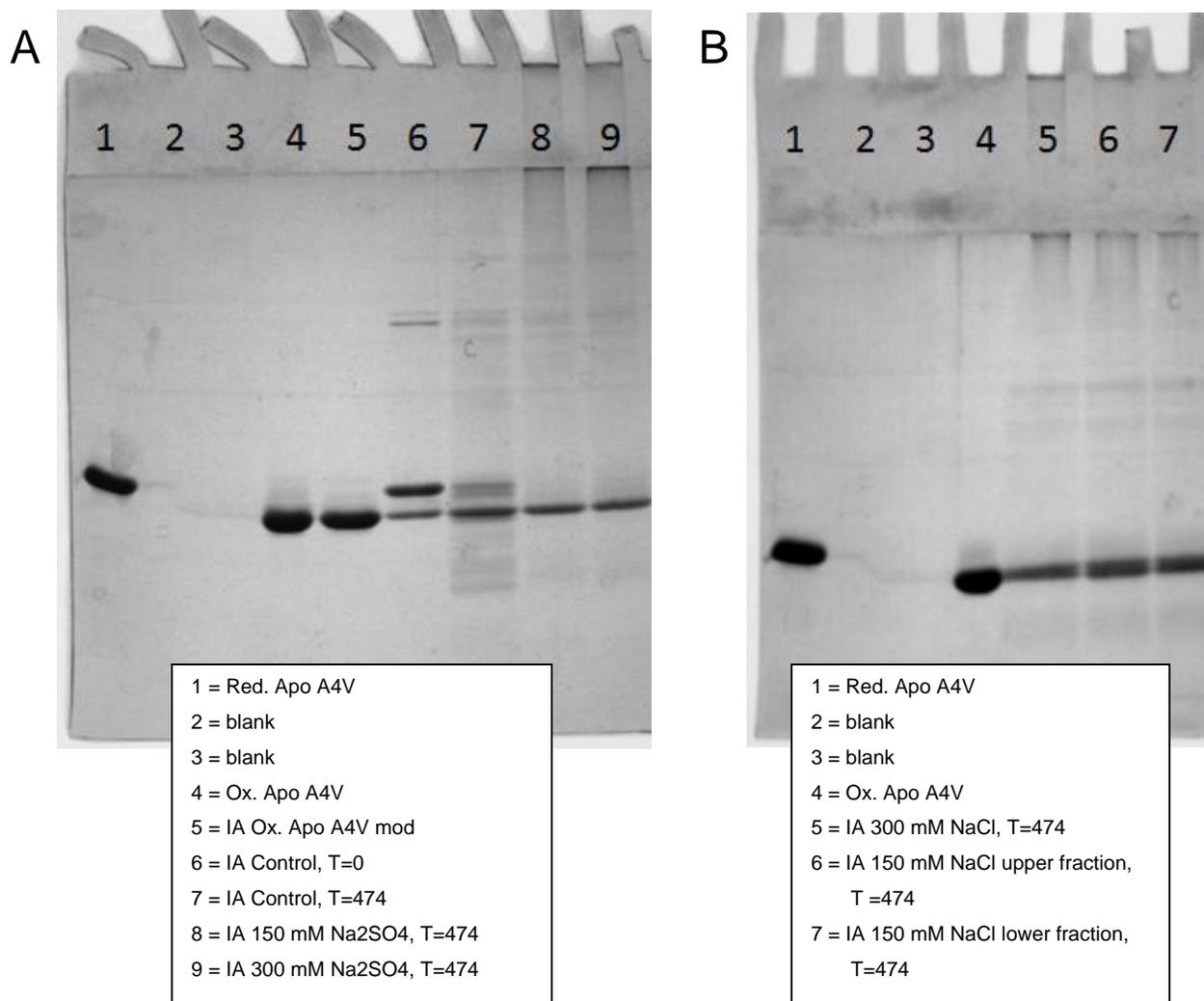
**Figure 4.12 Summary of ANS fluorescence results for reduced apo A4V aggregation trials.** Samples were incubated in anaerobic conditions at 37°C. All samples contain 1 mg/mL protein, 20 mM Hepes, 10 mM TCEP, pH of 7.4. Each data point shown is the average from two independent experiments (with some conditions replicated within each experiment). Standard deviations are shown as error bars.



**Figure 4.13 ANS Fluorescence time-course measurements for 1 mg/mL reduced apo A4V in 20 mM Hepes, 10 mM TCEP, pH 7.4, 37°C, and various salt conditions. A-E represent the following conditions respectively: 0 mM salt, 150 mM NaCl, 300 mM NaCl, 150 mM Na<sub>2</sub>SO<sub>4</sub>, and 300 mM Na<sub>2</sub>SO<sub>4</sub>. Each data point shown is the average from two independent experiments (with some conditions replicated within each experiment). Standard deviations are shown as error bars.**

#### 4.3.3.2 SDS PAGE gels and DTNB buffer Assays Reveal Reduced Apo A4V Oxidation

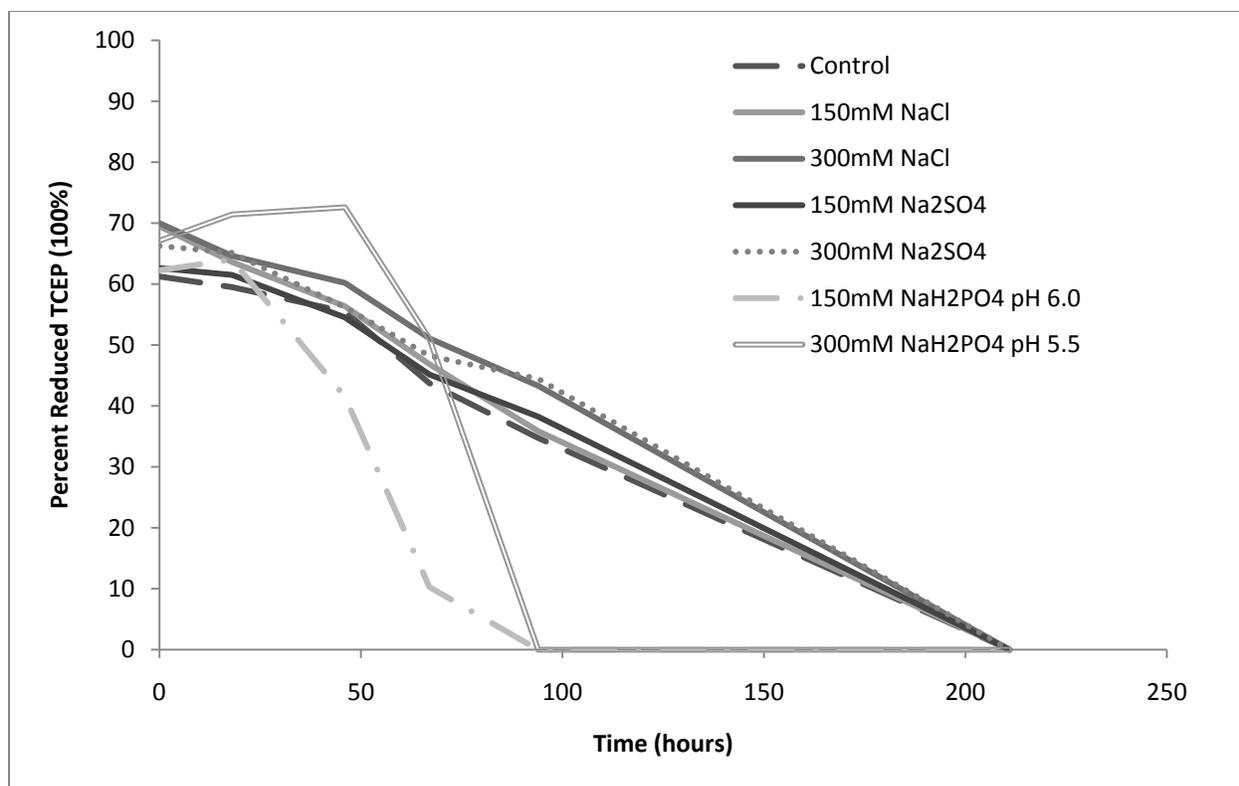
The reduced apo A4V aggregation samples were analyzed to determine the oxidation state of the protein at the end points of the aggregation time-courses, using iodoacetamide modification of free thiols, followed by SDS-PAGE, as before (Section 4.3.2.2). Figure 4.14 shows the gel results for all salt samples at the end of the aggregation time-course. Lane 6 in Panel A shows the initial time point of the control sample is mostly reduced. After 474 hours, lane 7 demonstrates that the control sample is mostly oxidized, with a small fraction reduced and some evidence of cross-linking in the higher molecular weight region. The sulphate samples in lane 8 (150 mM) and lane 9 (300 mM) show full oxidation by the end of 474 hours. Strong bands are found in the top and bottom of the stacking gel indicating that large cross-linked species that are too big to enter the gel are present in the sulphate samples. The chloride samples in Panel B lanes 5 (300 mM), 6 (upper fraction of 150 mM) and 7 (lower fraction of 150 mM) also show complete oxidation at the end of the 474 hour time-trial. The upper and lower fractions of the 150 mM sample were divided by pipetting the top 50  $\mu$ L and lower 50  $\mu$ L separately from the  $\sim$ 100  $\mu$ L aggregate sample left in the glass vial at the end of the time-course. The purpose of this was to see if any of the aggregates had settled in the bottom of the vial. The similarity of the two samples on the gel demonstrates that there was no major difference between the top and bottom fractions. Significant cross-linking is also observed for the NaCl samples. Protein is apparent in the stacking gel, in particular in the 300 mM NaCl sample well. These gels imply that reduced A4V protein was more readily oxidized in the salt solutions resulting in increased cross-linked species.



**Figure 4.14 SDS-PAGE gels of iodoacetamide-modified reduced apo A4V samples at various time points during aggregation time-courses in differing salt conditions and 10 mM TCEP, 20 mM Hepes, pH 7.4.** Panels A and B represent different samples from the same time-course experiment. The samples that have been modified by iodoacetamide have been denoted “IA.” In Panel B, the 150 mM NaCl samples in lanes 6 and 7 represent the upper and lower fractions removed from the sample vial at the final time point of the aggregation time-course.

A TCEP oxidation test was performed using the DTNB assay to determine how long the 10 mM TCEP buffers remain reduced. 140  $\mu$ L samples were prepared and incubated in the identical glass vials that were used for the aggregation time-course experiments. Figure 4.15 shows the DTNB assay results. Of particular interest is that the TCEP was not 100% reduced initially. The TCEP used for this experiment

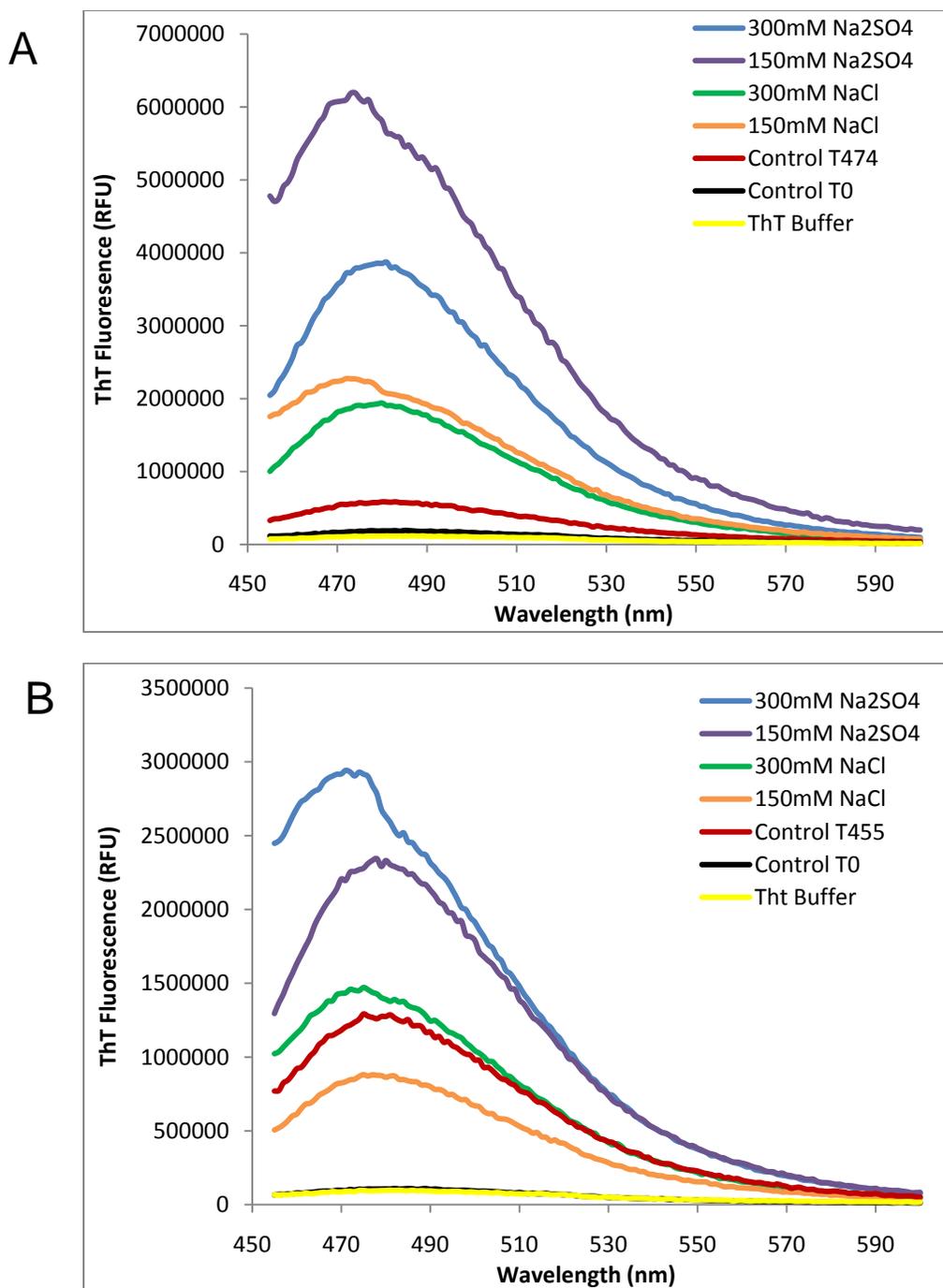
was a different compound than for the 1 mM TCEP trials. In the 1 mM TCEP trials, TCEP-HCl, which is available in solid form, was dissolved with water and diluted into the appropriate buffer. For the 10 mM TCEP trials, neutral TCEP was required since TCEP-HCl would greatly decrease the sample pH. Neutral TCEP is obtained from the supplier as a powder with instructions to reconstitute in 1 mL of water to obtain a 500 mM solution. These instructions were followed and the 500 mM solution was stored at -20°C. It is surprising that the initial measurement of freshly prepared buffer was only 60-70% reduced, and that although there was 6-7 times the amount of reduced TCEP in these solutions compared to the 1 mM TCEP buffer, the samples were completely oxidized by approximately 210 hours. This is only about 50 hours longer than for complete oxidation of the 1 mM samples (see Figure 4.8). Surprisingly, the 10 mM TCEP control sample seems to behave the same as the salt samples and actually becomes fully oxidized, much more quickly than the 1 mM TCEP sample. Thus, the inherent properties of the TCEP-HCl and TCEP-neutral appear to differ. It is unclear what the reasons behind this are. Once again, TCEP oxidation appears to be a zero order reaction, in this case, with a reaction rate of  $5.56 \times 10^{-4}$  mM/min for all samples at pH 7.8. The phosphate samples at pH 6 and 5.5 became oxidized by the 94 hour time point, giving a zero order reaction rate of  $1.24 \times 10^{-3}$  mM/min. This is consistent with previous findings that phosphate accelerates TCEP oxidation, potentially by forming an unstable TCEP-phosphate complex that enhances oxidation<sup>140</sup>. Additionally, in 50mM Hepes buffer, it was shown that a decrease in the solution pH increases the rate of TCEP oxidation<sup>140</sup>. Thus both the presence of phosphate and the decrease in pH may contribute the faster rate of TCEP oxidation than is observed for the control and salt conditions at pH 7.8. Given that previous results in similar conditions to those for the control, NaCl, and Na<sub>2</sub>SO<sub>4</sub> solutions used in this experiment give a zero order reaction rate of  $9.00 \times 10^{-6}$  mM/min<sup>140</sup> (discussed in Section 4.3.2.2), the reaction rates observed here are shockingly high. This probably has important implications for the observed aggregation behaviour of reduced apo A4V by allowing for the oxidation of free thiols to form intermolecular disulphide cross-linked species.



**Figure 4.15 Time-course of 10mM TCEP oxidation.** Buffers were prepared identically to the reduced apo A4V aggregation trials and monitored over time using the DTNB assay for the percent of TCEP remaining reduced. The initial TCEP concentration was 10 mM. All samples are in 20 mM Hepes, pH 7.8 except for the  $\text{NaH}_2\text{PO}_4$  samples which are at the pH listed.

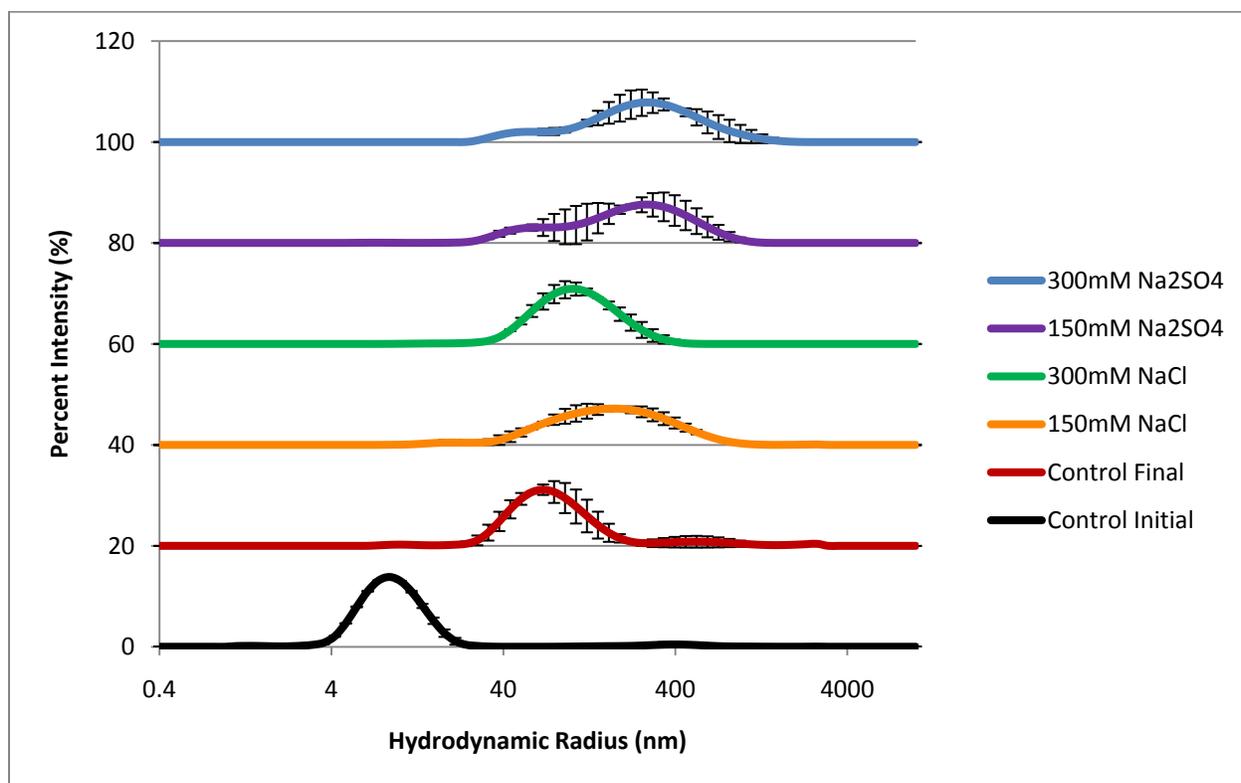
#### 4.3.3.3 ThT Fluorescence and DLS Measurements of End-point Samples of Reduced Apo A4V Time-courses

In order to gather more information about the aggregate species formed during these experiments, the end points of the aggregation time-trials were analyzed using ThT fluorescence and DLS. ThT fluorescence data for the end points of two different time-trials are shown in Figure 4.16. While there is some variability in the exact fluorescence readings for the two time-courses, both show the  $\text{Na}_2\text{SO}_4$  samples as having the highest ThT fluorescence, followed by NaCl, and then control (no added salt) samples. The ThT fluorescence is indicative of amyloid-like aggregation in the end point samples, which is promoted by the presence of salts, especially  $\text{Na}_2\text{SO}_4$ .



**Figure 4.16 ThT Fluorescence of reduced apo A4V aggregate samples.** All samples were in 20 mM Hepes, 10 mM TCEP, pH 7.4. The ThT fluorescence was corrected for protein concentration by dividing ThT Fluorescence by the measured protein concentration of each sample. Panels A and B represent two independent time-course experiments. All salt samples were measured at the end of the aggregation time-course which was 474 hours in panel A and 455 hours in panel B. The control sample (no added salt) was measured at both the initial and final time points of the reaction.

Dynamic Light Scattering (DLS) measurements were performed on samples from each of the solution conditions at the end of two independent trials (Figure 4.17). Table 4.4 summarizes the acquired data. Fairly high reproducibility can be noted by the shape of the curves and the relatively small error bars in Figure 4.17. In all end point aggregate samples the monomer peak has disappeared and a larger species with an average diameter of 80-370 nm appears. The salt samples seem to promote the formation of larger species than in the control sample, with the largest hydrodynamic diameter found in the 300 mM Na<sub>2</sub>SO<sub>4</sub> sample, as was also observed for the aggregation time-trials in 1 mM TCEP buffers. However, the greatest variations between replicate samples are also seen in the sulphate samples which may indicate that these samples are prone to form heterogeneous aggregate species of varying sizes.



**Figure 4.17 Average DLS result for of two independent reduced apo A4V aggregation trials in 10 mM TCEP, 20 mM Hepes, pH 7.4 and varying salt conditions.** The average percent intensity is plotted against the size of the species in solution. Error bars represent the variations in measurements from the two independent time-courses. All samples are from the end point of the aggregation trials with the exception of the initial control sample in black.

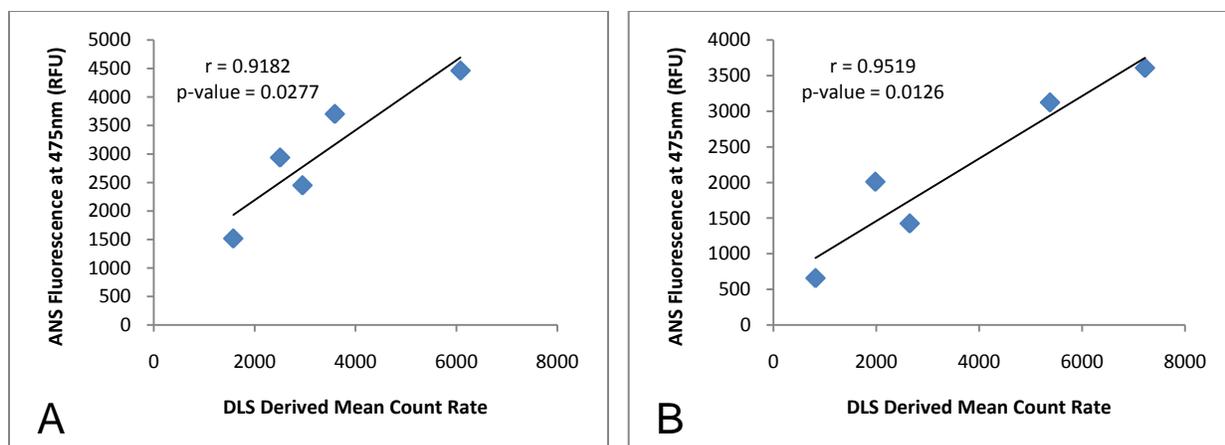
**Table 4.4 Average hydrodynamic diameter of end point aggregate species for two independent time-courses of reduced apo A4V aggregation in a range of salt conditions\*.**

		<b>Initial Control</b>	<b>Final Control</b>	<b>150 mM NaCl</b>	<b>300 mM NaCl</b>	<b>150 mM Na<sub>2</sub>SO<sub>4</sub></b>	<b>300 mM Na<sub>2</sub>SO<sub>4</sub></b>
<b>Average Hydrodynamic Diameter (nm)</b>	Trial 1	9.8	69.9 588.8	262.4	109.9	363.5 53.1	279.2
	Trial 2	9.2	87.5	164.1	129.2	209.4	461.2 70.8
	Average <sup>†</sup>	9.5 ± 0.4	78.7 ± 12.4	213.2 ± 69.5	119.6 ± 13.7	286.4 ± 109.0	370.2 ± 128.7

\*All samples excluding the initial control are at the end point of the aggregation time-course experiments. For trial 1 this was 474 hours and for Trial 2 450 hours. Each diameter-value for the individual trials is an average of three consecutive measurements of the same sample.

<sup>†</sup>The average was determined for the hydrodynamic diameter values of trials 1 and 2 in each condition. In the case where one trial showed more than one species, the average is reported of the species from each trial with the most dominant signal intensity.

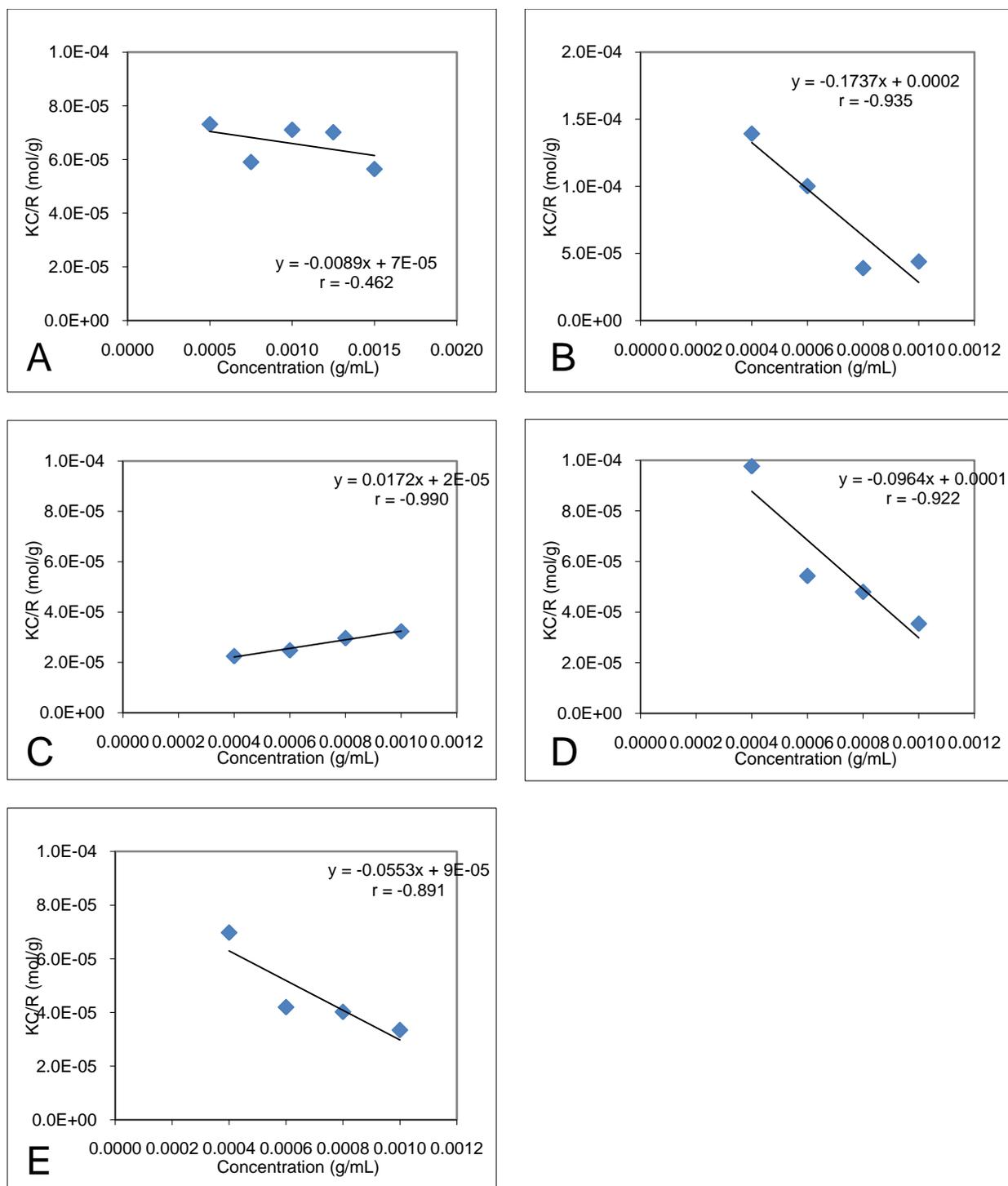
The derived mean count rates for the final aggregate samples were measured and correlated with the final ANS fluorescence measurements for each solution condition for two separate trials. Figure 4.18 gives the two correlation plots. For both replicate experiments there is a statistically relevant strong correlation ( $r = 0.92$  and  $0.95$ ) between the DLS count rate and final ANS fluorescence measurement. This indicates that the increase in ANS fluorescence is a result of an increase in aggregate species as monitored by the enhanced light scatter in the DLS measurements.



**Figure 4.18 Relationship between ANS Fluorescence and DLS derived mean count rate at the end point of reduced apo A4V aggregation time-trials in 20 mM Hepes, 10 mM TCEP and varying salt conditions.**

#### 4.3.3.4 Debye Plot Analysis of Reduced Apo A4V in Varying Salt Conditions

To aid in the understanding of the roles the salts are playing in controlling the aggregation pathways observed here, light scattering measurements were taken at various protein concentrations in each of the salt types to create Debye plots. This allowed for the determination of the second virial coefficient ( $B_{22}$ ) in 0 mM salt and 150 mM salt buffers in the background of 10 mM TCEP, 20 mM Hepes, and pH 7.4. The plots are given in Figure 4.13, and Table 4.4 summarizes the results. Due to limited time and available protein, the Debye plots were completed only over a small protein concentration range and were not repeated. For most plots the correlation coefficient is close to 0.9, indicating a strong correlation. Intriguingly, the value of  $B_{22}$  in  $\text{Na}_2\text{SO}_4$  is positive, indicating protein-solvent interactions are favoured over protein-protein interactions, which is contrary to the aggregation results observed for reduced apo A4V in these conditions. This is an unexpected result and requires replicate experiments to test the validity of this observation. In all other conditions the overall trend for  $B_{22}$  is negative, signifying that protein-protein interactions are favoured in these conditions.



**Figure 4.19 Debye plots for reduced apo A4V in a range of solution conditions.** All samples were in 20 mM Hepes, 10 mM TCEP, and pH 7.4. A-E represent the addition of the following salts respectively: 0 mM salt, 150 mM NaCl, 150 mM Na<sub>2</sub>SO<sub>4</sub>, 150 mM NaH<sub>2</sub>PO<sub>4</sub> and 150 mM TMAO.

**Table 4.5 Molecular Weights and Second Virial Coefficients determined from Debye plot analysis of reduced apo A4V in varying solution conditions.**

	Control	150 mM NaCl	150 mM N <sub>2</sub> SO <sub>4</sub>	150 mM NaH <sub>2</sub> PO <sub>4</sub>	150 mM TMAO
<b>Molecular Weight (Da)</b>	13,351	4,947	65,619	7,918	11,754
<b>B<sub>22</sub> (mL·mol/g)</b>	-0.0045	-0.0868	0.0086	-0.0482	-0.0277

In addition to the sulphate and chloride buffers that were used for the aggregation time-trial experiments, Debye plots were also made for reduced apo A4V in 150 mM NaH<sub>2</sub>PO<sub>4</sub> and Trimethylamine-N-oxide (TMAO). TMAO is an osmolyte found in the mammalian kidney and several deep sea animals, and serves to protect proteins from denaturation by urea present in the cellular environment<sup>143</sup>. It has been shown to be a universal protein stabilizer without significantly changing the native structure and activity of the protein<sup>144</sup>. Evidence has shown that in the presence of denaturant, TMAO interacts with N-H bonds in the protein backbone and restricts the conformational space sampled by the bond, thus promoting more restricted structural fluctuations<sup>145; 146</sup>. TMAO was used in the Debye plot analysis as a measure of the role of protein stabilization in preventing protein-protein interactions. Salts are also known to stabilize proteins, however the addition of TMAO allows for the investigation of the role of stabilization without the impact of the ionic interactions which play a major role in salt effects.

Both the NaH<sub>2</sub>PO<sub>4</sub> and TMAO samples behave similarly to the NaCl sample and result in a negative slope in the Debye plots which corresponds to the favouring of protein-protein interactions over protein-solvent interactions. Table 4.5 summarizes the molecular weights and second virial coefficients determined from these plots. The expected molecular weight for the SOD1 monomer is 15.75 kDa. The control sample and the TMAO sample result in molecular weight values closest to the expected monomer value while all other conditions show significant differences. This indicates that replicate experiments are necessary to validate these results. However, in general, the negative B<sub>22</sub> value suggests that in the reduced apo form, A4V preferentially forms protein-protein interactions rather than protein-solvent interactions.

## 4.4 Discussion

For the purpose of thoroughly investigating the aggregation patterns of the SOD1 mutant A4V, and to compare the effects of salt on different initial states of the same protein, two forms of the mutant were studied. The metal-free form of A4V with the disulphide bond intact, termed oxidized apo A4V, and the metal-free form of A4V with the disulphide bond reduced, termed reduced apo A4V, were investigated in detail. Both forms were monitored for aggregate formation in the presence and absence of added salts, and the results are discussed below

### 4.4.1 Oxidized Apo A4V is only slightly Prone to Aggregate in Quiescent Conditions in a Range of Salt Conditions

Limited investigations have been completed on the aggregation tendencies of oxidized apo A4V. In the studies described in Section 4.3.1, oxidized apo A4V was monitored in a range of salt conditions by ANS binding and revealed very limited evidence for aggregation. The ANS binding results (Figure 4.2) demonstrate limited fluorescence fluctuations are observed for all conditions, showing no convincing evidence of considerable aggregation. Further investigation was performed by dynamic light scattering of the samples after the completed time-course. Figure 4.3 shows that despite the lack of change in ANS fluorescence there does appear to be the formation of small amounts of larger species in the oxidized apo A4V samples in all solution conditions. The dimer peak at 6 nm remains the dominant species even after 450 hours. A second species between 177 and 155 nm arises in most solution conditions, and in 300 mM sulphate and 300 mM phosphate solutions, a species of approximately 373 nm or 575 nm, respectively, becomes the second most intense species. The fact that the 6 nm species remains the largest intensity peak even at the end of the time-course is evidence that the proportion of aggregated species in solution is very small. Light scattering intensity is dependent on the sixth power of the diameter of the scattering particle<sup>116</sup>, thus when only a small fraction of the percent mass of the sample is present as a high

molecular weight species it can easily dominate the entire sample signal intensity (see Section 4.3.1). The fact that both the small and large species are detectable simultaneously is evidence that only a tiny fraction of the sample is forming larger aggregate species and that the majority of the sample is not aggregated.

To investigate the role of the solution conditions in promoting protein-protein interactions, light scattering was used to create Debye plots to determine the second virial coefficient. The second virial coefficient ( $B_{22}$ ) gives information as to whether the protein in solution favours protein-solvent or protein-protein interactions. Table 4.2 gives the  $B_{22}$  results for oxidized apo A4V in 20 mM HEPES, pH 7.4 and 0 mM salt, or 150 mM NaCl,  $\text{Na}_2\text{SO}_4$  or  $\text{NaH}_2\text{PO}_4$ . For all conditions except for  $\text{NaH}_2\text{PO}_4$ , the  $B_{22}$  value is positive and varies between 0.00148 and 0.00371. This fits well with the observation that apo A4V is only slightly prone to aggregate in the oxidized state, as the positive  $B_{22}$  values indicate that the protein is more inclined towards protein-solvent interactions than protein-protein interactions. The negative value obtained for the  $\text{NaH}_2\text{PO}_4$  sample is acquired from a Debye plot with relatively high scatter and a low correlation coefficient (Figure 4.4 D,  $r = 0.19$ ). For this reason it is difficult to say whether this negative  $B_{22}$  is significant and repeat experiments are required to confirm the validity of this result.

In previous experiments comparing the effects of various salts on the second virial coefficient for seven different proteins it was commonly observed that NaCl had very little effect on  $B_{22}$ <sup>134</sup>. This is somewhat different than the results obtained here, where the value for  $B_{22}$  in NaCl (0.00371) compared to that of the control (0.00148) is about 2.5-fold higher. In addition, the previous study showed that in general the salt effects on a given protein followed the Hofmeister series, in that the stronger the kosmotrope, the greater the decrease in  $B_{22}$ <sup>134</sup>. In the results observed here, the  $B_{22}$  values observed in NaCl and  $\text{Na}_2\text{SO}_4$  are higher than in the control (approximately 2.5-fold), which is opposite of the expected trend. However there is significant scatter in these plots (Figure 4.4), and thus it is likely that the

$B_{22}$  values are not fully reliable. Repeat measurements are required to obtain more accurate values. Regardless, the trend in the sign of the  $B_{22}$  values is generally positive, indicating that oxidized apo A4V does not favour protein-protein interactions in the solutions conditions tested. Thus, even though salt has been suggested to increase the probability of forming protein-protein interactions<sup>134</sup>, the solubility and conformational stability of oxidized apo A4V prevent significant protein aggregation even in the presence of salt.

The general observations for oxidized apo A4V based on the studies here suggest that regardless of solution conditions it has relatively low propensity to aggregate. Oxidized apo A4V exists as a dimer with an apparent melting temperature of 50.7°C<sup>135</sup>. Thus, at the experimental conditions of 37°C, it is expected that the protein will remain predominantly as a folded dimer, and even in the presence of salt, aggregation does not become favoured. This does not necessarily mean that the oxidized apo state of SOD1 does not aggregate but instead that the conditions used here did not promote aggregation sufficiently enough to be observed by the ANS fluorescence method of monitoring protein aggregation. In these experiments aggregation time-courses were measured in quiescent, physiologically relevant conditions. Holo SOD1 has been shown to aggregate to low levels in similar solution conditions but at a concentration of 10 mg/mL<sup>71</sup>. Thus, the protein concentration may be one factor that could be altered to cause oxidized apo A4V to become more inclined to aggregate.

Various other experiments have demonstrated that oxidized apo SOD1 can be induced to aggregate *in vitro* under appropriate conditions. WT SOD1 (containing two free cysteines, C6 and C111) in the apo oxidized form was demonstrated to form heterogeneous aggregate samples that showed a steady ThT fluorescence increase for the first 150 hours of incubation in air at 37°C, pH 7 and 100µM (~1.6 mg/mL) protein concentration<sup>72</sup>. Follow-up studies showed that fALS mutants in the WT background were also prone to aggregation in the oxidized apo form in the same conditions<sup>147</sup>. The

conditions used in these studies were quite similar to those of the experiments described here except for the exposure to air. All samples monitored here were kept in capped vials in a sealed, evacuated, dessicator, and thus in an anaerobic environment. The aggregates formed under the air-exposed conditions formed intermolecular disulphide bonds and resulted in large, soluble, oligomeric aggregates that bound ThT<sup>72; 147</sup>. A key difference in these experiments is the use of the WT background *vs.* the pWT background used for the experiments described in Section 4.3.1. In this case, the free thiols at position 6 and 111 have been replaced with alanine and serine respectively (see Section 1.6.4). Thus by using the pWT construct and by incubation in an anaerobic environment, the A4V protein is not expected to form intermolecular disulphide bonds as this would require the reduction of the intramolecular disulphide followed by re-oxidation between the thiols of different monomers. In another previous study that used agitation to induce aggregation of 1 mg/mL samples of oxidized apo WT SOD1 in 50 mM MOPS, 0.1 M NaCl, 1 mM EDTA, pH 7, it was demonstrated that the aggregates bound ThT and contained intermolecular disulphide bonds<sup>69</sup>. Since the oxidized apo aggregates in the WT background were characterized by disulphide cross-linking in both of these previous examples, it is not surprising that similar aggregation is not observed here for oxidized apo A4V in the pWT background and in anaerobic conditions.

Oxidized apo A4V in the pWT background is not prone to aggregate in quiescent, physiologically relevant, anaerobic, solution conditions. Even in the presence of salts of varying kosmotropic strengths, no observable aggregation was measured by ANS fluorescence. Thus, although salts have the potential to promote aggregation (see Section 1.3.3), oxidized apo A4V in the solution conditions used here, resulted in very little aggregate formation and the prevalence of non-aggregated soluble protein dimers in all test conditions. This suggests that oxidized apo SOD1 has a relatively low propensity to form non-covalent

aggregates, but can be induced to aggregate quite readily by the oxidation of free thiols<sup>69; 72</sup>, which is additionally true for other proteins<sup>148</sup>.

#### **4.4.2 Reduced apo A4V in the presence of 1 mM or 10 mM TCEP and Various Salt Conditions is Prone to re-oxidize and form Disulphide Cross-linked Aggregates with Amyloid-like characteristics**

##### **4.4.2.1 Reduced apo A4V Favours Protein-Protein Interactions more than Protein-Solvent Interactions**

It has been previously suggested that reduced apo SOD1 may be more prone to aggregate than both the oxidized apo and holo forms of the protein<sup>65</sup> (See Section 1.6.2) . This hypothesis is supported by the difference in second virial coefficients of the oxidized and reduced form of apo A4V. In the oxidized apo form the  $B_{22}$  values in various buffer conditions are generally positive (see Table 4.2), indicating that the protein favours protein-solvent interactions over protein-protein interactions, and therefore it is colloidally stable. Initial inspection of the Debye plots for reduced apo A4V (Figure 4.19) shows obviously negative slopes indicating that the protein is colloidally destabilized and favours protein-protein interactions over protein-solvent interactions. The  $B_{22}$  values are summarized in Table 4.5. Interestingly, the only plot that shows a positive correlation for reduced apo A4V is that of the  $\text{Na}_2\text{SO}_4$  sample, which based on the aggregation results would be expected to have the strongest negative correlation because it appears to have the greatest tendency to aggregate (*vide infra*). However, these are preliminary results that have not been replicated and are based only on a few concentration points and thus require further validation.

To test the role of protein stability in influencing the second virial coefficient, a Debye plot was performed in a 150 mM TMAO buffer (Figure 4.19 E). TMAO is a stabilizing agent expected to prevent protein unfolding. Surprisingly, the TMAO does not appear to have a significant effect on promoting protein-solvent interactions, as represented by the negative  $B_{22}$  value. In fact, the slope of the Debye plot

is actually more negative in TMAO than in the control sample without salt, suggesting that TMAO may not aid in preventing reduced apo A4V aggregation, and may actually induce aggregation. TMAO can cause the native-like folding of naturally unfolded proteins<sup>149</sup>. If in the same way, TMAO stabilizes the unfolded state of reduced apo A4V, resulting in a more compact structure, this could be a potentially mechanism of promoting aggregation since partially-folded states are considered to be particularly aggregation prone<sup>7</sup>. In addition, other factors beyond stability, including factors that influence electrostatic and Van der Waals interactions, such as net charge and exposed hydrophobic groups, influence colloidal stability<sup>22</sup>, and are likely to play a role in determining the  $B_{22}$  value of reduced apo A4V in varying solution conditions. Regardless, based on the general trends observed in Debye plot analysis, it can be generally concluded that when the disulphide bond is reduced, apo A4V has a negative  $B_{22}$  value indicating an increased tendency to form intermolecular associations than in the oxidized apo form. Further investigation is required to confirm the role of salt in influencing the second virial coefficient of reduced apo in solution.

#### 4.4.2.2 Salt Promotes Reduced apo A4V Aggregate Formation with Increased ANS Fluorescence

The role of salt in the mechanisms of protein aggregation is not fully understood. Salts can have both electrostatic and stabilization effects on proteins that may promote or prevent aggregation depending on the protein in question and the state the protein is in<sup>125</sup>. The aggregation patterns of 1 mg/mL reduced apo A4V were monitored by ANS fluorescence in a range of salt conditions in a buffer containing 1 mM or 10 mM TCEP. Interestingly, the initial time points for all samples had higher ANS fluorescence signals in the reduced apo form than in the oxidized apo form. This is likely a result of increased exposure of hydrophobic groups in the monomeric, more expanded structure of the reduced form of the protein, which are more buried in the more compact, dimeric oxidized apo SOD1<sup>150</sup>. In addition, reduced apo

A4V is only marginally stable at 37°C, with an apparent melting temperature of 36°C<sup>27</sup>, indicating the population of unfolded monomer would be expected to be around half of the total protein concentration. The unfolded species may also contribute to the increase in ANS fluorescence by the exposure of hydrophobic residues. Based on the fact that at the conditions used for experimentation reduced apo A4V will be approximately half unfolded, salt may have complex effects on the aggregation of this protein by interactions with both the folded and unfolded states (*vide infra*).

Salts can interact with proteins and play a role in shielding charges, resulting in decreased repulsion between proteins and potentially increased aggregation. Monomeric SOD1 has a net negative charge at neutral pH<sup>76</sup>, and so it would be expected that under the experimental conditions (pH 7.4) charge shielding may promote aggregate. Ions could also be involved with the selective binding and stabilization of various states of the protein (*ie.* folded, unfolded, or partially folded). Stabilization of the folded state would result in less unfolded protein by shifting the equilibrium towards folded protein. Na<sub>2</sub>SO<sub>4</sub> has previously been shown to cause stabilization and structural compaction of a monomeric intermediate form of apo SOD1<sup>151</sup>. Additionally, stabilization of the unfolded state could cause partial refolding into a more compact state, similar to what has been observed for osmolytes<sup>149</sup>. The potential stabilization roles of salt could significantly affect the aggregation patterns of the protein. Investigation into the role of ions in reduced apo A4V aggregation was performed by monitoring the time-course of aggregation in the following salt conditions: 0 mM salt or 150 mM NaCl or Na<sub>2</sub>SO<sub>4</sub>.

Over the approximately 450 hour time-course of the experiments, the protein samples in 150 mM or 300 mM Na<sub>2</sub>SO<sub>4</sub> show large increases in ANS fluorescence over the first 50 hours, followed by a period of more gradual increases in ANS fluorescence for the rest of time-course of the experiment (see Figures 4.5 D,E and 4.12 D,E). The samples in 150 mM NaCl show a steady, gradual increase in ANS fluorescence over the entire 450 hours (Figures 4.5 B and 4.12 B). In 300 mM NaCl the average ANS

fluorescence remains fairly constant over the time-course in 1 mM TCEP (Fig.4.5 C), while showing a moderate increase in fluorescence over time in 10 mM TCEP (Fig. 4.12 C). In the control samples with no added salt, the 1 mM TCEP sample remains fairly constant in fluorescence while in the 10 mM TCEP sample there is a slight gradual increase in ANS fluorescence over time. In addition to differences in the overall pattern of ANS fluorescence for the different types of solution conditions, the initial ANS fluorescence also varies. In general, the salt samples seem to have slightly higher initial ANS fluorescence than the control samples. The difference is slightly more pronounced for the Na<sub>2</sub>SO<sub>4</sub> samples than the NaCl samples. This may be indicative of some initial aggregation occurring before the first time point of fluorescence is taken, or that the salt promotes increased hydrophobic exposure in the reduced apo form.

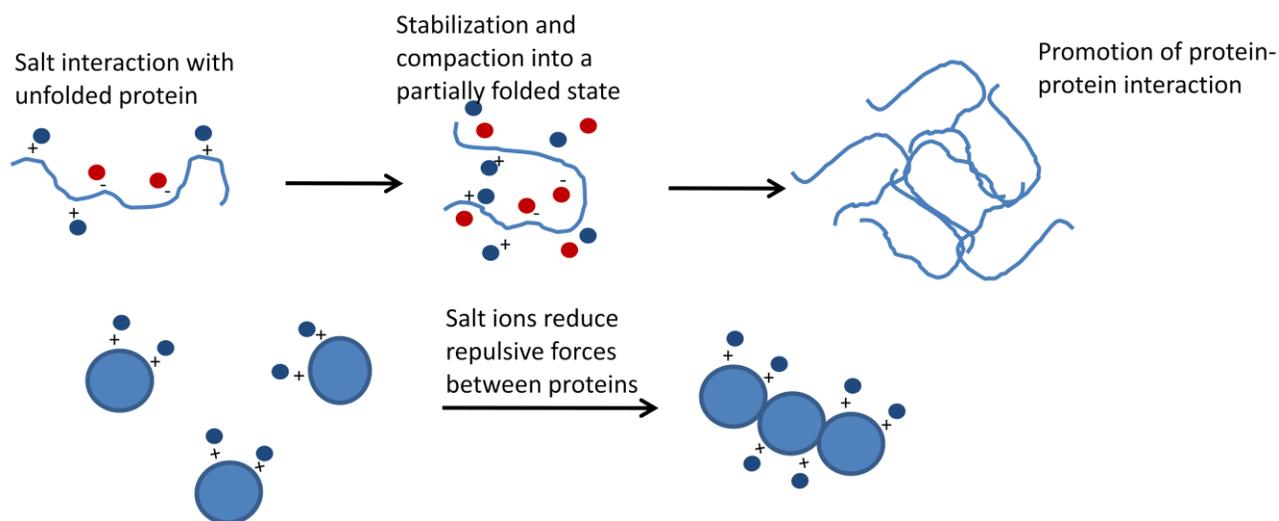
From the ANS results the overall aggregation pattern seems to follow the Hofmeister series in that the order of greatest to least aggregation is Na<sub>2</sub>SO<sub>4</sub> > NaCl > no salt. This is somewhat expected as this pattern has been observed for the role of salt in protein aggregation previously<sup>42; 44</sup> and additionally corresponds to the ThT results (see Section 4.4.2.4). The kosmotropic salts in the Hofmeister series have been suggested to decrease protein solubility and can lead to increased protein aggregation<sup>43</sup>. The ionic strength of the salts allows for charge shielding which may decrease the net repulsive forces between like-charged protein molecules. As mentioned above, a significant portion of the reduced apo A4V species will be unfolded. This may result in the exposure of aggregation prone regions, and in addition the exposure of charged regions that may cause electrostatic repulsions between monomers. The presence of salt could provide the ions necessary to shield these charges and allow for intermolecular association into aggregate structures. This would explain the trend observed for the salts used in this experiment. Sulphate has a higher ionic strength than chloride and thus would be better at shielding charges and promoting aggregation. ANS-binding aggregates appear faster in the sulphate samples than in the chloride samples,

and faster in chloride than without salt. This suggests that salt increases the rate of aggregation (see Section 4.4.2.4).

It is important to consider whether the aggregation observed here is occurring from the unfolded or folded state, or from a transiently populated intermediate state. From these experiments it is not possible to determine this information; however, it is relevant to consider the possibility of aggregation from each of these states. Kosmotropic salts have been shown to stabilize proteins<sup>43</sup>, and specifically, Na<sub>2</sub>SO<sub>4</sub> has been demonstrated to stabilize SOD<sup>151</sup>. If aggregation were to occur from the folded state than it may be expected that the stabilization effects of salt would correspond to an increase in aggregation from the folded state since the presence of salt would shift the equilibrium toward folded protein. However, additional effects of salts, such as charge-shielding as mentioned above, complicate data interpretation. The increase in exposed charges characteristic of the unfolded state could be shielded by the presence of ions, supporting that salts may enhance aggregation from the unfolded state. This behaviour has been confirmed by the necessity of high ionic strength in protein solutions in which aggregation is induced by low pH<sup>42</sup>. In addition, the presence of salt may influence the stability of potential intermediate states. For example, the monomer intermediate for oxidized apo SOD is selectively and markedly stabilized by Na<sub>2</sub>SO<sub>4</sub><sup>151</sup>. Additionally, the stabilization of a folding intermediate has been observed for the four helical protein Im7<sup>152</sup>. If this were the case, the stabilization of a partially folded intermediate may be particularly influential in promoting protein aggregation as it has been suggested that too much structure, or too little structure, may both inhibit aggregation<sup>119</sup>, and that partially folded intermediates are particularly aggregation prone<sup>7</sup>. Further experimentation is required to elucidate which scenario is leading to the aggregation observed for reduced apo A4V in these conditions.

The two predominant effects of the interactions between salt and protein are electrostatic interactions that shield net charge, and specific interactions resulting in protein stabilization (see Figure

4.20). Both of these factors can contribute to promoting protein aggregation. The fact that both NaCl and Na<sub>2</sub>SO<sub>4</sub> are observed to promote reduced apo A4V aggregation suggests that salt may be decreasing the negative repulsive charges between protein monomers and stabilizing partially folded protein structures in such a way to increase intermolecular reactions, resulting in protein aggregation.



**Figure 4.20 Possible roles of salts on promoting protein aggregation.** Two important roles of salt on aggregation include: A) Stabilization of an aggregation-prone state (whether folded, unfolded, or a partially folded intermediate), and B) electrostatic interactions resulting in charge shielding and reduced repulsive charges between monomers.

#### 4.4.2.3 Salt Promotes Protein Re-oxidation and Disulphide Cross-linking

The role of disulphide cross-linking in SOD aggregation has been a significant point of controversy among researchers. Some studies have suggested that free thiol oxidation is required for the formation of amyloid-like aggregates<sup>72</sup>; while in contrast, others have shown that aggregation still occurs when all four cysteines are mutated<sup>73</sup>. It has also been suggested that intermolecular disulphide formation in ALS mice models is only important in late stage disease, and it is not required to initiate aggregation<sup>74</sup>. Further investigation into the patterns of aggregation of reduced apo A4V reveal that in both 1 mM and

10 mM TCEP buffers (20 mM Hepes, pH 7.4) the protein became significantly re-oxidized and formed high molecular weight, cross-linked species during a 450 hour time-course with incubation at 37°C in an anaerobic environment. The SDS PAGE gel shown in Figure 4.7 clearly demonstrates that the reduced apo protein becomes at least partially oxidized throughout the time-course of the experiment in 1 mM TCEP buffers and that the salt conditions promote significant protein re-oxidation and the formation of covalent intermolecular disulphide bonds, particularly in the sulphate samples (Figure 4.7 lanes 8 and 9). Similarly, in the 10 mM TCEP buffer conditions, considerable cross-linking is represented by higher molecular weight bands seen in the salt conditions (Figure 4.14, panel A lane 8 and 9 and panel B lanes 5-7). The relative intensities of the bands for cross-linked species follows the order of  $\text{Na}_2\text{SO}_4 > \text{NaCl} > \text{no salt}$ , which is identical to the pattern observed for ANS fluorescence. This may indicate that the cross-linked species induced by the presence of salt are responsible for the increase in ANS fluorescence.

The salt samples demonstrate increased oxidation and disulphide linking in both the 1 mM and 10 mM TCEP buffer conditions. Interestingly, while in 1 mM TCEP buffers there is some remaining reduced protein at the end of the time-course in the control and sulphate samples (Figure 4.7 lanes 7-9), in the 10 mM TCEP buffers there is only a small fraction of reduced protein remaining in the control sample, and none in the sulphate samples (Figure 4.14 panel A lanes 7-9). This suggests that the samples are more easily oxidized in the 10 mM TCEP samples as opposed to the 1 mM TCEP samples. This is an unexpected result considering that the concentration of TCEP has been increased 10-fold. To investigate whether the TCEP buffers are remaining reduced during the time-course of the experiments, DTNB assays were performed for both the 1 mM and 10 mM TCEP samples. The buffer analysis test to determine the time-course of 1 mM TCEP oxidation (Figure 4.8) clearly demonstrates that the salt buffers become oxidized more quickly than the control buffer. Interestingly, in the 10 mM TCEP buffer (Figure 4.15), the control behaves similarly to the salt buffers at the same pH. In both cases, the TCEP becomes

oxidized more quickly than expected. The rate of TCEP oxidation in similar conditions (50 mM Hepes, pH 8.2) to what was used here was previously measured and gives a zero order rate constant of  $9.0 \times 10^{-6}$  mM/min. However, in the 1 mM TCEP samples the salt conditions result in a zero order rate constant of  $1.1 \times 10^{-4}$  mM/min, or  $4.39 \times 10^{-5}$  mM/min for the control sample. In 10 mM TCEP the zero order rate constant was approximately identical in all conditions at pH 7.8 with a value of  $5.6 \times 10^{-4}$  mM/min. Therefore in all experimental conditions the rate was significantly faster than previously measured (five-fold greater for the control in 1 mM TCEP and 12 to 60-fold greater in all other conditions). The fastest rate occurred in the 10 mM TCEP samples, which may be indicative of why these samples resulted in almost complete protein oxidation by the end of the time-course experiments as observed in the SDS PAGE gels (Figure 4.14). These results are quite surprising, first because the rate of TCEP oxidation is much higher than previously determined, and second because the 10 mM TCEP samples had higher rates of oxidation than the 1 mM TCEP samples.

There are several factors that may promote the oxidation of the TCEP in the conditions used here including the small sample volumes, large surface area exposure to the glass of the vials and the presence of salt. The glass vials used in these experiments hold volumes of 140  $\mu$ L. The conical shape results in a large solvent area to glass exposure. The presence of oxygen in the silicon oxide glass, and the potential for accumulation of charges on the glass surface<sup>153</sup>, may contribute to the redox chemistry and enhance TCEP oxidation. However, there do not seem to be any previous publications noting similar glass effects on the rate of oxidation of reducing agents. It is additionally unclear why the 10 mM TCEP solution would have a faster rate of oxidation than the 1 mM TCEP solution. As discussed in Section 4.3.3.2, two different forms of TCEP were used for the preparation of the 1 mM and 10 mM TCEP solutions. However, there is no obvious reason as to why this might influence the rate of oxidation. The 1 mM solution is prepared from a TCEP-HCl molecule, while the 10 mM solutions from neutral TCEP. Thus the

only difference is the presence of HCl, and it is not obvious whether the presence of this ionic acid might slow the oxidation reaction. Another point for consideration is the redox potential of the solution, which is determined by the ratio of oxidized to reduced TCEP. Since the 10 mM TCEP had a higher proportion of oxidized TCEP at the beginning of the time-course, its potential to reduce was lower. This likely contributed to the observation of the faster oxidation of reduced apo SOD1 in the 10 mM TCEP buffer compared to the 1 mM TCEP buffer, but it does not explain the usually high rate of TCEP oxidation in both of these buffers. Further investigation into this intriguing observation is required. The rate of TCEP oxidation is likely to have a significant impact on the formation of disulphide cross-linked aggregates as the reduced apo A4V becomes re-oxidized.

Disulphide cross-linked aggregates have been described previously for SOD1 in various forms. Banci *et al.*, in 2008, demonstrated that oxidized apo mutants in a WT background (containing free C6 and C111) formed disulphide-linked soluble aggregate structures by the oxidation of free thiols upon incubation in air in solution conditions of 37°C, pH 7, and 100 µM protein<sup>154</sup>. The aggregates showed increased ThT-fluorescence which was proportional to the percentage of aggregated species in solution. The oligomeric species were destroyed in the presence of the reducing agent dithiothreitol (DTT)<sup>154</sup>. In addition, it was shown that reduced apo WT SOD1 could form aggregate species with or without intermolecular disulphide bonds depending on the solution conditions<sup>69</sup>. In 1 M guanidine hydrochloride (Gdm-HCl), 10 mM TCEP, cross-linked aggregates were formed. However in 100 mM TCEP, aggregates formed that did not possess intermolecular disulphide bonds. Agitation was used to induce aggregation in both of these conditions<sup>69</sup>. This seems to suggest that destabilization by Gdm-HCl and reduction by 10 mM TCEP was adequate for exposing free thiols but that the reducing potential of 10 mM TCEP was not able to keep the thiols reduced and cross-linked aggregates were formed. It is possible that the Gdm-HCl also was involved in charge shielding which may have promoted aggregation. Similarly, in the quiescent

conditions of the experimental results for reduced apo A4V in varying salt conditions described here, intermolecular disulphides were also a characteristic trait of the aggregates. Salt may have promoted the formation of intermolecular disulphide formation by the reduction of repulsive electrostatic forces allowing for increased protein-protein interactions. Because the buffer did not remain a reducing environment, this allowed for the re-oxidation of the free thiols to result in the formation of intermolecular disulphide bonds. This may have occurred first through the aggregation of reduced apo A4V in a reducing environment (as A4V has previously been shown to aggregate in quiescent solutions in the reduced apo form without formation of disulfide bonds<sup>27</sup>) that situated the protein monomers in such a way that when the solution became an oxidizing environment free thiols were in close proximity to free thiols from other monomers as opposed to thiols on the same strand. The result then becomes the favoured formation of intermolecular instead of intramolecular disulphide bonds (see Figure 4.20). This potential mechanism will be discussed further in Section 4.4.3.

#### 4.4.2.4 Salt Promotes Larger Aggregates with Increased Amyloid-like characteristics

Salt could have multiple effects in modulating the aggregation patterns of proteins. Two possible effects of salt may be altering the rate of aggregate formation and moderating the morphology of the aggregates being formed. It is obvious from the ANS binding results that aggregates appear faster in salt conditions than in the absence of salt (see Section 4.4.2.2). In order to further characterize the structure of the aggregates formed by A4V in the varying salt conditions used here, a series of tests were designed to gather more information about the structure and size of the aggregates being formed. ThT-fluorescence was used for the investigation of the amyloid-like nature of the aggregates and Dynamic Light Scattering (DLS) was employed to gather information about the sizes of the aggregates. Preliminary ThT data of the sulphate samples in 1 mM TCEP buffers (Figure 4.9) demonstrates that the aggregate species formed in sulphate solutions are amyloid-like. The endpoint samples of 150 mM and 300 mM sulphate show that a

pronounced difference in ThT results when compared to the endpoint control sample. The control sample has a slight increase (approximately 3-fold) in ThT fluorescence compared to the buffer, while sulphate samples show a ThT fluorescence increase of 18-fold (150 mM sulphate) of 36-fold (300 mM sulphate). This indicates that aggregate species formed in the presence of sulphate possess the cross-beta fibrillar structure that interacts with ThT to cause fluorescence.

More extensive ThT binding investigations were completed for the samples in 10 mM TCEP and are displayed in Figure 4.16. From these results it is evident that the protein samples in sulphate display the strongest ThT-fluorescence, followed by the NaCl samples, and finally the control samples. All endpoint samples have an increase in ThT-fluorescence compared to the control sample at the initial time point, which behaves identically to the ThT buffer without protein added. These results are in agreement with both the SDS-PAGE and ANS-binding results in which aggregation is promoted by salt in accordance with the Hofmeister series. It appears that the salt causes an increase in disulphide cross-linked aggregate species which is associated with an increase in ANS and ThT fluorescence. The interaction with ThT is characteristic of amyloid-like aggregates and suggests that these cross-linked species possess beta-sheet morphology.

Dynamic Light Scattering results demonstrate differences in the hydrodynamic diameters for aggregates formed in varying solution conditions. Figures 4.9 and 4.16, and Tables 4.3 and 4.4 describe the summarized results for the endpoint DLS measurements for 1 mM TCEP and 10 mM TCEP buffer conditions, respectively. In both cases, the sizes of the aggregates are larger in the salt samples compared to the control sample. In general, the species size is larger in the 10 mM TCEP samples which may be a result of the faster TCEP oxidation described in Section 4.4.2.2, if disulphide oxidation promotes aggregate growth. In both sets of experiments the largest species was formed in the 300 mM sulphate buffers. Figures 4.10 and 4.17 demonstrate the correlation between total light scattering intensity and

ANS fluorescence. The high correlations (all  $r > 0.9$ ) indicate that a sample with high ANS fluorescence also has the greatest light scatter. This could be a result of two possibilities: 1) the aggregates interacting with ANS are large species (as evidenced by DLS) and larger species will scatter more light than smaller species, or 2) there are more aggregates forming in the samples with higher ANS fluorescence and the increased number of aggregated species results in more light scatter. It is possible that both of these possibilities are contributing factors and that the salt samples induce the formation of larger aggregated species, and more aggregated species, than the control sample. Overall, these results suggest that the presence of salt increases the tendency of reduced apo A4V to form increased numbers of large, soluble protein aggregates. Similar results have been observed for  $\alpha$ -amylase from *Bacillus halmapalus*, where the presence of salt caused an increase in aggregate size compared to aggregate formation in the absence of salt<sup>155</sup>.

These results suggest that the salt samples have an effect on the structure of the aggregates formed by reduced apo A4V. In the presence of salt, aggregates show increased amyloid-like characteristics and are larger sizes than in the absence of salt. However, in these experiments the amount of total protein in each sample was the same, but the amount of protein aggregates in each sample differed. Thus, the differences in ThT fluorescence could actually be a result of differing amounts of aggregates in solution and not necessarily that salt promotes aggregation in the form of amyloid-like structures more than in the absence of salt. In order to appropriately determine this, samples would need to be normalized for aggregate concentration instead of overall protein concentration. Similarly, the DLS results suggest that salt promotes the formation of larger aggregates, however it could also be argued that salt increases the rate of aggregate growth and that over time, all solutions may reach the same size aggregates, but that the salts might promote reaching the final structure faster. For these reasons, the complex roles of salts in the mechanisms of protein aggregation must be investigated further (see Section

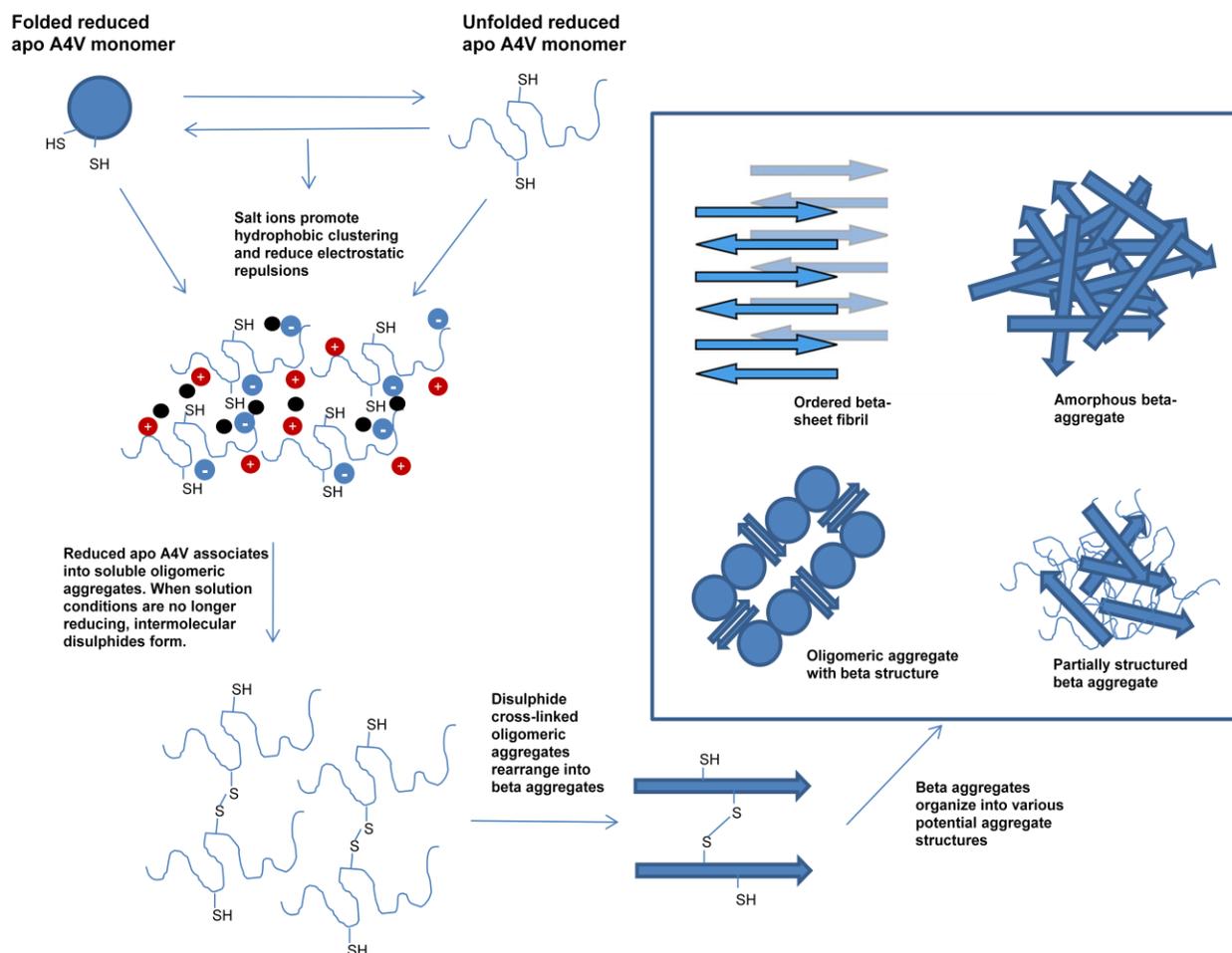
5.2.2) to define the implications of ionic strength on aggregate structure. Regardless, from the results obtained here, it can be concluded that for reduced apo A4V, salt promotes the appearance of larger, ThT-binding aggregates compared to samples with no added salt.

#### **4.4.3 Conclusions**

The roles of salt in the mechanism of protein aggregation are difficult to generalize as they can vary considerably from one protein to another. For the A4V aggregation experiments discussed in this chapter, increasing ionic strength was observed to play three major roles: 1) promote the rate of formation of ANS-binding aggregates, 2) promote the formation of disulphide cross-linked aggregates, and 3) promote the growth of protein aggregates into large, amyloid-like species. Both sulphate and chloride were extensively studied and all experimentation demonstrated that the order in which these ions promoted protein aggregation was: no salt < NaCl < Na<sub>2</sub>SO<sub>4</sub>, following the increasing kosmotropic nature described by the Hofmeister series. This confirms previous results that demonstrate similar aggregation patterns for the role of salt on the aggregation of  $\alpha$ -lactalbumin<sup>42</sup> and yeast prion protein<sup>44</sup>. The salt buffers are likely to play multiple significant roles in contributing to the enhancement of reduced apo SOD1 aggregation. It is difficult to separate the different roles of the salts, and the roles these properties might play in the aggregation results observed here. It is reasonable to suggest that both the stabilization and electrostatic effects of the salts could contribute to the observations. Kosmotropes are known to increase protein stability and enhance the salting out effect by decreasing protein-solvent interactions<sup>43</sup>. In this way, kosmotropes can promote the formation of hydrophobic clusters. Sulphate has a higher ionic strength than chloride, and since it is observed to have a greater influence on increasing protein aggregation of reduced apo A4V, this is consistent with a potentially important role of charge shielding on protein aggregation. These factors may work together to stabilize aggregation-prone conformations of reduced apo SOD1 (whether folded, unfolded, or partially folded) and promote protein-protein

interactions by decreasing solvent interactions, enhancing hydrophobic associations, and shielding repulsive charges. A schematic of what this aggregation pathway could look like is given in Figure 4.21.

Salts play a significant role in modulating the aggregation of protein. As seen in the results here, salts have the potential to cause increased protein aggregation in accordance with the Hofmeister series. Both ionic interactions and stability effects may contribute to the role of salt in protein aggregation (see Figure 4.20), which is an important consideration in improving understanding of the fundamental principles of protein aggregation. These studies provide a significant basis for improved understanding of the role of salt in modulating the pathways of protein aggregation. Further investigations will continue to increase understanding in the molecular pathways of protein aggregation.



**Figure 4.21 Schematic of a potential aggregation mechanism of reduced apo A4V in salt conditions resulting in the formation of disulphide cross-linked, amyloid-like soluble aggregates.** Unfolded, partially folded, or folded reduced apo A4V associates into soluble oligomeric aggregates. Salt ions (shown as black circles) prevent intermolecular repulsion between charged residues (shown as positive red circles and negative blue circles) and promote hydrophobic clustering. Oxidative solution conditions results in the formation of disulphide cross-linked species that further rearrange into aggregates with increased beta-sheet structure.

## Chapter 5

### Summary and Future Work

#### 5.1 Summary and Conclusions

This project set out to investigate the molecular mechanisms of protein aggregation. Two approaches were used for the purpose of achieving this goal. First, an investigation was completed that compared nine protein aggregation prediction techniques. The algorithms were used to predict aggregation for several mutations of three very different proteins. Predictions were compared for identical sequences to determine whether varying techniques would form a consensus on the mutants most prone to aggregate and the regions of a protein most likely to be involved in aggregation. Further investigations were completed by comparing predicted aggregation results with experimentally observed aggregation.

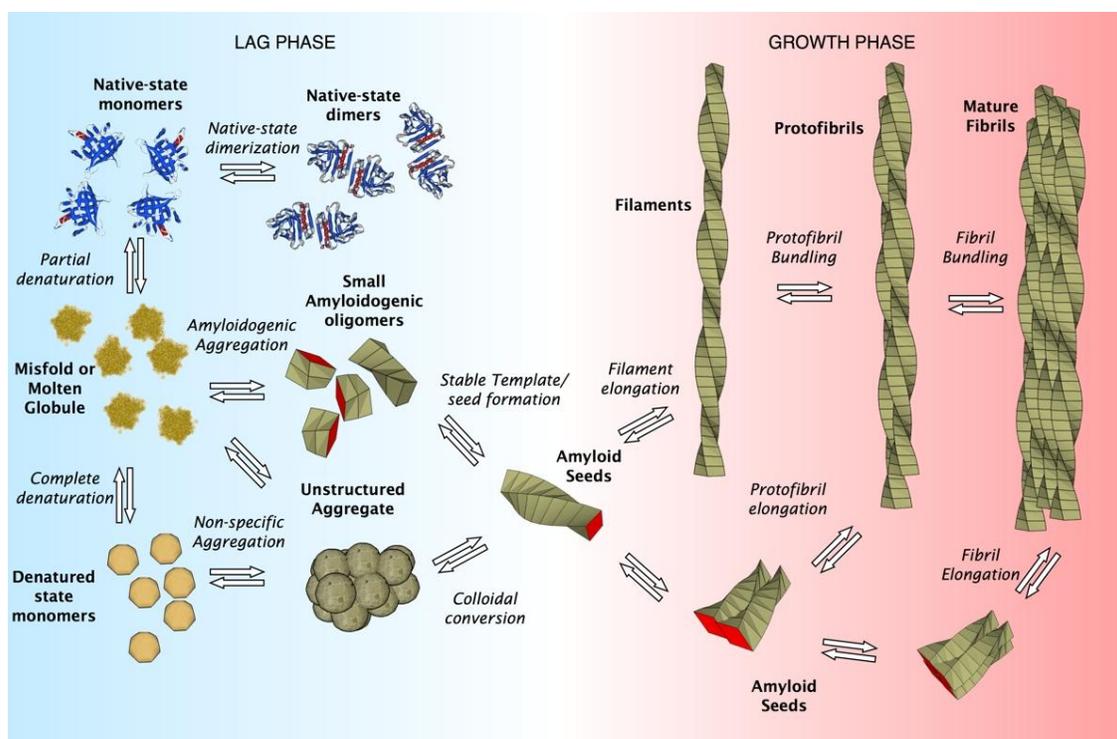
The second component of this project was to examine experimentally the role of salt in influencing protein aggregation. The ALS-causing SOD1 mutant, A4V, was used as a model protein for this study. The aggregation patterns of the oxidized and reduced apo form of this protein were monitored using a variety of techniques including ANS and ThT fluorescence, DLS, and SDS-PAGE gels. The two major roles of salt in modulating protein aggregation were examined in detail, including charge shielding which can decrease repulsive charges between proteins, and specific interactions which can stabilize aggregation-prone forms of the protein.

Table 5.1 provides a summary of potential variables discussed throughout this study that may influence protein aggregation, and the impact they can have on the protein. These factors will be further discussed in the summary sub-sections below. In addition, Figure 5.1 gives a summary scheme of the diversity of the potential pathways that could be involved in protein aggregation. Importantly, solution conditions could have significant effects on modulating these pathways.

**Table 5.1 Summary of potential variables that contribute to protein aggregation.**

Variable	Potential Impact on Protein
<b>Solution Conditions</b> (ie. Ionic strength, temperature, pH etc.) <sup>8; 22*</sup>	Charge, solubility, stability, electrostatic and hydrophobic interactions
<b>Inherent Properties of the amino acids in the protein sequence</b> <sup>9</sup>	Net charge, alpha helix or beta sheet propensity and hydrophobicity
<b>Native Structure</b> <sup>10</sup>	The degree of foldedness of the native protein, and the amount of alpha helix or beta sheet secondary structure
<b>Stability</b> <sup>28</sup>	Probability of populating folded and unfolded states
<b>Local fluctuations</b> <sup>29</sup>	Exposure of aggregation prone regions

\*The solution conditions have the potential to alter all additional variables listed in the row below.



**Figure 5.1 Overview of protein aggregation mechanisms.** Figure reproduced from Giurleo *et al.*, 2008<sup>156</sup>.

### 5.1.1 Comparison of Multiple Aggregation Prediction Algorithms

The following nine protein aggregation prediction algorithms were used to predict the aggregation propensity of many mutants of three different proteins: Chiti-Dobson equation<sup>9</sup>, Wang-Agar equation<sup>61</sup>, Zyggregator<sup>85</sup>, Ztox<sup>85</sup>, FoldAmyloid<sup>88</sup>, PASTA<sup>86</sup>, TANGO<sup>87</sup>, and Profile 3D<sup>89</sup>. These algorithms are quite diverse in the way in which they were designed and the proteins on which they were tested (see Tables 2.1 and 2.5). In general, most of the algorithms were developed and tested based on Thioflavin T (ThT) fluorescence data, which measures the rate of amyloid formation. Three proteins, including Superoxide Dismutase (SOD1), a 153 amino acid, homodimer<sup>59</sup>, human Acylphosphatase (AcP), a 98 amino acid, natively-folded monomer<sup>92</sup>, and the Amyloid Beta peptide (A $\beta$ <sub>42</sub>), a 42 amino acid peptide without stable tertiary structure<sup>95</sup>, were used as test proteins. Both SOD1 and A $\beta$ <sub>42</sub> are additionally interesting because they have been linked to neurodegenerative disorders<sup>50; 51</sup>. The aggregation prediction propensities were calculated for all three wild-type proteins and mutants, by all nine algorithms. This study presents the first wide-scale comparison of many prediction algorithms by comparing the predicted aggregation propensities of multiple mutants of three different proteins. Comparisons of predicted aggregation propensities for a given protein sequence revealed marked differences, even for A $\beta$ <sub>42</sub>, which readily forms amyloid. This led to the conclusion that the prediction algorithms do not form a consensus on the factors that control protein aggregation. This is not particularly surprising because of the diversity in the contexts in which the algorithms were designed. For example, the Chiti-Dobson equation was developed based solely on the role of the physical and chemical properties of single point mutations in causing amyloid formation of unfolded proteins<sup>9</sup>. In contrast, the FoldAmyloid method predicts aggregation based on the ability of consecutive residues in a sequence to form extensive molecular contacts including hydrogen bonds, and was not specifically designed to predict amyloid only, but may be relevant in predicting additional forms of aggregates<sup>88</sup>. Therefore, the

observation that the algorithms predict differing aggregation propensities for the same sequence may imply that the factors that modulate differing types of aggregation play different roles in varying types of aggregation. Figure 5.1 demonstrates a broad range of aggregation pathways. These pathways may be influenced differently by different factors and so algorithms developed based on one specific pathway may not be successful in predicting the aggregation of a different pathway. For example, an algorithm developed based on ThT binding by full-length proteins forming mature fibrils may result in the inclusion of variables that are not particularly relevant to the formation of branched protofibrils by short peptides.

In addition to predicting the overall propensities of aggregation for several protein sequences, the algorithms were used to examine aggregation-prone regions of the three wild type test proteins. Although the algorithms did not converge upon a consensus of the factors governing protein aggregation, as discussed above, they were much more successful at converging upon the regions within a sequence that are most prone to aggregate. Hot-spot maps were generated for SOD1, AcP and A $\beta$ <sub>42</sub>, and in all three proteins, regions in which positive aggregation propensities, as predicted by at least 5 out of the 7 algorithms, were identified (the Chiti-Dobson and Wang-Agar equations could not be used because they only predict overall aggregation propensities and do not identify aggregation-prone regions of a sequence). This is a promising result because it demonstrates the agreement between algorithms in locating aggregation-prone stretches within a protein sequence. In this manner, multiple algorithms may be used together to locate sections within a protein that are particularly vulnerable to protein aggregation. Further experimental validation of the success of using this approach to identify aggregation-prone regions is required.

### **5.1.2 Correlating Predicted and Observed Protein Aggregation**

Over the past decade several advances have been made in the field of understanding protein aggregation, including the development of aggregation prediction algorithms. In this study, many of these

algorithms were used to predict the aggregation propensity of a series of mutants of three different test proteins, and the results were correlated with previously measured experimental aggregation propensities<sup>9; 27; 92; 108; 109; 110</sup>. This provides the first broad-scale analysis that tests the success of the aggregation prediction techniques in correlating with observed aggregation. The results were somewhat surprising and are an important consideration for future advancement in refining the understanding of the detailed mechanisms that modulate protein aggregation pathways. In brief, the aggregation prediction algorithms predominantly demonstrated poor to moderate correlations with observed aggregation for three test proteins. The best correlations were those of the A $\beta$ <sub>42</sub> and its mutants, likely because A $\beta$ <sub>42</sub> readily forms amyloid experimentally and it is a short peptide without extensive native structure. The Chiti-Dobson (r = 0.92) and Wang-Agar (r = 0.89) equations gave the strongest correlations for predicted and observed aggregation propensity of this protein. It is not surprising that these correlations are high as the A $\beta$ <sub>42</sub> data used to make these correlations was used in the design of the algorithms. It is more surprising that so few of the algorithms were able to accurately predict the aggregation of A $\beta$ <sub>42</sub>, resulting in statistically insignificant correlations for six of the prediction techniques.

The second best correlations were observed for AcP. In the conditions in which the experimental data was collected (pH = 5.5), AcP was unfolded<sup>9</sup>. The aggregation data for AcP mutants was also used in the development of the Chiti-Dobson and Wang-Agar equations and was used to test the accuracy of the Zyggregator algorithm. Thus, it is expected that these algorithms would be able to accurately predict AcP aggregation. Statistically significant correlations, albeit not particularly strong (r-values ranging from 0.55 to 0.61), were obtained for these three algorithms and the Ztox algorithm, which is based on the Zyggregator method. Surprisingly, all other algorithms could not accurately predict amyloid formation by unfolded AcP. This may be attributed to the fact that most of the unsuccessful algorithms were developed

based on hexapeptide data and there the factors modulating the aggregation patterns of short peptides may differ than that of a larger protein sequence.

Finally, the algorithms gave the poorest correlations for SOD1. Only one algorithm, Zyggregator, gave a statistically relevant, moderate correlation ( $r = 0.59$ ). SOD1 aggregation was measured in the reduced apo state by Dynamic Light Scattering. This method does not monitor amyloid formation like ThT fluorescence, but instead measures light scatter as an indicator of total protein aggregation. Reduced apo SOD1 mutants form soluble oligomeric structures and not fibrillar amyloid<sup>27</sup>. The fact that the algorithms are incapable of predicting the aggregation observed here may reflect the specificity of these algorithms for the prediction of amyloid. That being said, the algorithms did not perform particularly well at predicting amyloid formation by A $\beta$ <sub>42</sub> and AcP. These results imply that, in general, the aggregation prediction algorithms are incapable of predicting aggregation outside of the individual contexts in which they were designed and tested (see Table 2.6). This provides further evidence of the specificity of the factors involved in different types of protein aggregation pathways (as discussed in Section 5.1.1).

### 5.1.3 Investigating the Role of Reduced Apo SOD1 in ALS

The instability of the reduced apo form of SOD1 has resulted in the suggestion that this form of the protein may be specifically important in causing ALS<sup>77</sup>. For the purpose of further investigating the potential role of reduced apo, several possible correlations were investigated between the properties of SOD1 mutants in the reduced apo form and the characteristic disease durations of ALS patients with SOD1 mutations (see Section 1.6.1). The aggregation of reduced apo SOD1 mutants, as measured by DLS<sup>27</sup>, was plotted against ALS disease durations, resulting in a poor correlation ( $r = 0.27$ ). There are multiple ways to interpret this result. First, it could be that reduced apo SOD1 aggregation is not important in disease. However, a more likely interpretation may be that reduced apo SOD1 aggregation is not the *only* important factor in disease. Instead, it is likely that different forms of the protein may be most

toxic depending on the mutant<sup>27</sup>. Quite intriguingly, there appears to be two main outliers (G37R and V148I) in the relationship between observed aggregation and disease duration which when removed markedly improve the correlation ( $r = 0.78$ ). This may indicate that some of the mutants involved in this study form aggregates in the reduced apo form that may be implicated in disease duration, while others are outside of this trend and likely exert their toxic effects in a different form of the protein.

#### **5.1.4 Probing the Role of Salt in Modulating Aggregation**

The role of salt in protein aggregation was investigated by monitoring the aggregation patterns of oxidized and reduced apo A4V in physiological relevant, quiescent solution conditions. Oxidized apo A4V showed very little tendency to aggregate regardless of the experimental solution conditions. This is in contrast to previous studies that showed that wild type and mutant oxidized apo SOD1 readily form large, disulphide-linked, ThT-binding, soluble oligomers<sup>72; 154</sup>. As opposed to the wild-type protein used for these former studies, the experiments described in Chapter 4 employ the A4V mutant in the pWT background (see Section 1.6.4) in which the free cysteines at positions 6 and 111 are replaced with alanine and serine, respectively. Additionally, former studies incubated protein samples in air-exposed conditions, while the current study incubated samples in sealed vials in anaerobic conditions. For these reasons, it is not surprising that the oxidized apo A4V in the pWT background did not form disulphide-linked aggregate species. This suggests that oxidized apo SOD1 has a relatively low propensity to form non-covalent aggregates, but can be induced to aggregate quite readily by the oxidation of free thiols<sup>69; 72</sup>. The presence of salt did not cause a large difference in the observed aggregation of oxidized apo A4V when compared to conditions without added salt. The only slight evidence for a role of salt comes from the DLS results which show an increase in hydrodynamic diameter of aggregated species in salt samples compared to the control sample. However, the dominant species contributing to the DLS light scattering

intensity in all solution conditions is the dimer peak, indicating only a very small fraction of the oxidized apo A4V protein is aggregating (see Section 4.3.1). This implies that although salt has been suggested to promote aggregation of proteins<sup>134</sup>, under the salt conditions used for these experiments, the conformational stability of oxidized apo A4V prevents the favouring of protein-protein interactions and thus minimal protein aggregation is observed.

In the reduced apo form of A4V, salt was shown to promote protein aggregation in correlation to the increasing kosmotropic nature of the salt. Previously, reduced apo A4V was shown to readily form soluble, oligomeric aggregates that were not disulphide-linked, in quiescent, physiologically relevant (pH 7.4, 37°C) solution conditions<sup>27</sup>. The experiments described in Sections 4.3.2 and 4.3.3 for reduced apo A4V employed the same solution conditions to those used previously except for incubation in small-volume glass vials, and the addition of NaCl or Na<sub>2</sub>SO<sub>4</sub> at a concentration of 150 or 300 mM. Interestingly, the aggregates formed in the salt conditions during ~450 hour time-trials with incubation in anaerobic conditions at 37°C, resulted in the formation of high molecular weight, disulphide-linked species. These aggregates had increased ANS and ThT fluorescence, and larger hydrodynamic diameters, when compared to aggregates formed in the control conditions without salt added. The formation of disulphide-linked aggregates was partially attributed to TCEP oxidation over the time-course of the experiment, which promoted protein re-oxidation in all solution conditions. Based on the cumulative results gained from the extensive studies performed here it was determined that the order in which salt promoted protein aggregation was: no salt < NaCl < Na<sub>2</sub>SO<sub>4</sub>. Previous investigations into the role of salt in the aggregation patterns of  $\alpha$ -lactalbumin<sup>42</sup> and yeast prion protein<sup>44</sup> followed the same pattern in which the increasing kosmotropic strength of the salt, according to the Hofmeister series, caused increased protein aggregation.

Salt can influence protein aggregation by two main mechanisms: 1) electrostatic interactions which shield protein charges and reduce the electrostatic repulsion between proteins, and 2) specific interactions that stabilize various forms of the protein<sup>22</sup>. In the reduced apo A4V system investigated here salt could potentially promote aggregation through both of these mechanisms. Monomeric SOD1 has a net negative charge at neutral pH<sup>76</sup>, and salt ions could be involved in shielding these charges to promote protein-protein interactions. Secondly, salt may influence the stability of the protein in such a way that aggregation-prone states are stabilized. Na<sub>2</sub>SO<sub>4</sub> has been shown to preferentially stabilize the monomeric intermediate during apo SOD1 unfolding, resulting in a compaction of the expanded structure<sup>151</sup>. If Na<sub>2</sub>SO<sub>4</sub> were able to interact with the unfolded state in a similar manner, this could result in the compaction of the unfolded protein into a partially folded protein<sup>7</sup>, which may be particularly prone to aggregate. Similarly, it has been demonstrated for the Im7 protein that Na<sub>2</sub>SO<sub>4</sub> can stabilize a folding intermediate state of the protein<sup>152</sup>. If a similar mechanism applies for reduced apo A4V, this could also result in the stabilization of aggregation-prone partially folded structure.

Thus the presence of salt can have important implications for protein-protein interactions in solution. For reduced apo A4V, Na<sub>2</sub>SO<sub>4</sub> strongly promotes the formation of disulphide-linked, soluble aggregate species with amyloid-like, beta-sheet structure. A schematic of the potential mechanism by which aggregates were formed in the presence of salt can be found in Figure 4.20.

## **5.2 Future Work**

### **5.2.1 Improving Aggregation Prediction Algorithms**

The correlations between observed and predicted aggregation discussed in Chapter 3 provide pertinent evidence that 1) current aggregation prediction techniques perform poorly in predicting observed protein aggregation for proteins outside of the specific context in which the algorithms were created, and 2) different types of protein aggregation pathways are controlled by differing variables. This

information highlights an important question in the study of the molecular mechanisms of protein aggregation, and that is whether there are general principles involved in modulating all types of aggregation. These results suggest that the variables that control aggregation may differ depending on the type of aggregation. Figure 5.1 highlights that there are many potential complex pathways of protein aggregation. Future work in the development of protein prediction algorithms may need to focus on narrowing the scope of prediction algorithms towards the accurate prediction of a very specific aggregation pathway. This would result in many algorithms, each specific to one type of aggregation (*ie.* amorphous, amyloid, oligomeric etc.). From this, the comparison of the variables important to each algorithm will provide insight into the specific factors involved in modulating different types of aggregation. It may be discovered that similar variables, but with difference empirical significance, may contribute to more than one type of aggregation pathway.

SOD1 could provide a practical test protein for the development of enhanced algorithms. The large number of mutants for this protein is advantageous for forming a test sample large enough to gain reliable results. A way to begin this type of algorithm design may be to take a previous algorithm (for example, the Chiti-Dobson equation) and re-calibrate the variables based on experimental aggregation results in differing solution conditions. SOD1 may be directed to form different types of aggregates in different conditions. For example the presence of  $\text{Na}_2\text{SO}_4$  promotes the formation of beta-sheet aggregates by reduced apo A4V that interact with ThT (see Chapter 4). Thus, several mutants could be monitored in high concentrations of  $\text{Na}_2\text{SO}_4$  to create a data base of aggregation rates for the formation of beta-sheet aggregates by reduced apo SOD1 mutants. This data could be used to re-train the Chiti-Dobson equation specifically for this type of aggregation. Alternatively, the DLS data that monitors the formation of soluble, oligomeric aggregates could additionally be used to re-train the equation within the context of this type of aggregation. Then, the comparison of the coefficients in each of these systems may provide

insight into the factors contributing to aggregation in each of these specific contexts. Furthermore, the algorithms could be used to generate the probability of a protein forming one type of aggregate over another.

There are several other significant factors known to have potential to modulate protein aggregation that are currently not well accounted for in the available prediction algorithms. Table 5.1 lists some of the important contributing factors. Potentially the most influential factor is the solution conditions. As noted, the solution conditions can have large effects on protein charge, solubility, and electrostatic and hydrophobic interactions<sup>8; 22</sup>. These factors can furthermore affect the protein structure, stability and local fluctuations that are important in influencing the potential intermolecular interactions that modulate protein aggregation (see Section 1.3). Currently, the only methods that incorporate effects of solution conditions are Zyggregator and Ztox<sup>85</sup> which allow for the input of pH, and TANGO<sup>87</sup>, which allows for the input of pH, ionic strength, and temperature. Because solution conditions can have such significant impacts on a protein and its aggregation tendencies, it is important that future work in the development of prediction algorithms address this issue.

Additionally, the native structure and stability of a protein controls the populations of folded and unfolded species that exist in solution. The degree of foldedness of a protein can modulate the type of aggregate structure formed<sup>10</sup>. For this reason, the structure of the native state, and the proportion of folded vs. unfolded species that exist in the experimental conditions, have significant influences on observed aggregation. This issue has only been addressed by the creators of Zyggregator, who have attempted various methods<sup>157</sup>, including the input of PDB files, or the prediction of structure from primary sequences using CamP<sup>85</sup>, for incorporating native structure into prediction techniques. However, this remains a relatively unaddressed issue by prediction methods and further research on this issue is necessary for increasing the effectiveness of aggregation prediction algorithms.

## 5.2.2 Further Experimental Investigations into the Specific Roles of Solution Conditions in Modulating Aggregation

The experiments described in Chapter 4 only scratch the surface of the types of experiments needed to further examine the specific roles solution conditions can play in modulating protein aggregation. It was observed that increasing the kosmotropic strength of the salts in a reduced apo A4V aggregation time course experiment results in increased aggregate formation. This implies a potential role of both stability and charge in controlling aggregation. Further experimentation using a broader range of salt concentrations, and incorporating other salts with varied kosmotropic strengths (such as  $\text{HPO}_4^{2-}$ ,  $\text{ClO}_3^-$ ,  $\text{SCN}^-$ ), would further verify the role of kosmotropic salts in reduced apo A4V aggregation. In addition, it will be important to perform similar experiments on various other mutants. By selecting mutants with differing properties than A4V, such as V148I which is slightly stabilized compared to pWT ( $T_m = 50^\circ\text{C}$ )<sup>27</sup>, and A4T, which is even more destabilized than A4V compared to pWT ( $T_m = 31^\circ\text{C}$ )<sup>27</sup>, in the reduced apo form, more information could be gathered as to how salt effects influence aggregation of SOD1 mutants with differing stabilities. Additionally, H43R, which has a melting temperature similar to A4V in the reduced apo form ( $T_m = 35^\circ\text{C}$ )<sup>27</sup>, but causes an overall reduction in net charge, would provide additional information about the potential charge-shielding effects of the salts (salt may have less of an effect in H43R aggregation since the overall net charge is already less due to the presence of positively charged arginine).

Another pertinent issue to address in the experiments of reduced apo SOD1 in the presence of salt is the re-oxidation of the protein during the time course. The reason reduced apo protein is becoming re-oxidized is likely due to the oxidation rate of the TCEP used in these studies. It is necessary that this is addressed in order to monitor the effects of salt on aggregates that are not disulphide linked. This is particularly important in the study of SOD1 because disulphide cross-linked aggregates may not be relevant to early stage ALS disease pathogenesis (see Section 1.6.3)<sup>67; 75</sup>.

Finally, to thoroughly address the role of the contributing factors to protein aggregation that are modulated by solution conditions it will be important to monitor aggregation in conditions beyond the addition of salt. By performing similar analyses while varying pH and temperature, and by the addition of crowding agents, considerable insight can be gained to provide a broader perspective on the effects of solution conditions on protein aggregation. It is likely that different conditions will result in the formation of different types of protein aggregates. This insight will be extremely valuable in advancing our current understanding of the molecular mechanisms that contribute to the many pathways protein aggregation.

## Bibliography

1. Lengyel, P. & Soll, D. (1969). Mechanism of protein biosynthesis. *Bacteriol Rev* **33**, 264-301.
2. Cabrita, L. D., Dobson, C. M. & Christodoulou, J. (2010). Protein folding on the ribosome. *Curr Opin Struct Biol* **20**, 33-45.
3. Dobson, C. M. (2004). Principles of protein folding, misfolding and aggregation. *Semin Cell Dev Biol* **15**, 3-16.
4. Anfinsen, C. B. (1962). Some observations on the basic principles of design in protein molecules. *Comp Biochem Physiol* **4**, 229-40.
5. Leopold, P. E., Montal, M. & Onuchic, J. N. (1992). Protein folding funnels: a kinetic approach to the sequence-structure relationship. *Proc Natl Acad Sci U S A* **89**, 8721-5.
6. Onuchic, J. N., Nymeyer, H., Garcia, A. E., Chahine, J. & Socci, N. D. (2000). The Energy Landscape Theory of Protein Folding: Insights into Folding Mechanisms and Scenarios. *Advances in Protein Chemistry* **53**, 87-152.
7. Hartl, F. U. & Hayer-Hartl, M. (2009). Converging concepts of protein folding in vitro and in vivo. *Nat Struct Mol Biol* **16**, 574-81.
8. Mahler, H. C., Friess, W., Grauschopf, U. & Kiese, S. (2009). Protein aggregation: pathways, induction factors and analysis. *J Pharm Sci* **98**, 2909-34.
9. Chiti, F., Stefani, M., Taddei, N., Ramponi, G. & Dobson, C. M. (2003). Rationalization of the effects of mutations on peptide and protein aggregation rates. *Nature* **424**, 805-8.
10. Uversky, V. N. & Fink, A. L. (2004). Conformational constraints for amyloid fibrillation: the importance of being unfolded. *Biochim Biophys Acta* **1698**, 131-53.
11. Uversky, V. N. (2010). Mysterious oligomerization of the amyloidogenic proteins. *Febs J* **277**, 2940-53.
12. Bellotti, V. & Chiti, F. (2008). Amyloidogenesis in its biological environment: challenging a fundamental issue in protein misfolding diseases. *Curr Opin Struct Biol* **18**, 771-9.
13. Stefani, M. & Dobson, C. M. (2003). Protein aggregation and aggregate toxicity: new insights into protein folding, misfolding diseases and biological evolution. *J Mol Med* **81**, 678-99.
14. Fandrich, M., Fletcher, M. A. & Dobson, C. M. (2001). Amyloid fibrils from muscle myoglobin. *Nature* **410**, 165-6.
15. Ross, C. A. & Poirier, M. A. (2005). Opinion: What is the role of protein aggregation in neurodegeneration? *Nat Rev Mol Cell Biol* **6**, 891-8.
16. Vetri, V., Canale, C., Relini, A., Librizzi, F., Militello, V., Gliozzi, A. & Leone, M. (2007). Amyloid fibrils formation and amorphous aggregation in concanavalin A. *Biophys Chem* **125**, 184-90.
17. Goers, J., Permyakov, S. E., Permyakov, E. A., Uversky, V. N. & Fink, A. L. (2002). Conformational prerequisites for alpha-lactalbumin fibrillation. *Biochemistry* **41**, 12546-51.
18. Platt, G. W. & Radford, S. E. (2009). Glimpses of the molecular mechanisms of beta2-microglobulin fibril formation in vitro: aggregation on a complex energy landscape. *FEBS Lett* **583**, 2623-9.
19. Necula, M., Kaye, R., Milton, S. & Glabe, C. G. (2007). Small molecule inhibitors of aggregation indicate that amyloid beta oligomerization and fibrillization pathways are independent and distinct. *J Biol Chem* **282**, 10311-24.
20. Lee, J., Culyba, E. K., Powers, E. T. & Kelly, J. W. (2011). Amyloid-beta forms fibrils by nucleated conformational conversion of oligomers. *Nat Chem Biol*.
21. Kodali, R. & Wetzel, R. (2007). Polymorphism in the intermediates and products of amyloid assembly. *Curr Opin Struct Biol* **17**, 48-57.

22. Chi, E. Y., Krishnan, S., Randolph, T. W. & Carpenter, J. F. (2003). Physical stability of proteins in aqueous solution: mechanism and driving forces in nonnative protein aggregation. *Pharm Res* **20**, 1325-36.
23. Kelly, J. W. (1998). The alternative conformations of amyloidogenic proteins and their multi-step assembly pathways. *Curr Opin Struct Biol* **8**, 101-6.
24. Dumoulin, M., Canet, D., Last, A. M., Pardon, E., Archer, D. B., Muyldermans, S., Wyns, L., Matagne, A., Robinson, C. V., Redfield, C. & Dobson, C. M. (2005). Reduced global cooperativity is a common feature underlying the amyloidogenicity of pathogenic lysozyme mutations. *J Mol Biol* **346**, 773-88.
25. Johnson, S. M., Wiseman, R. L., Sekijima, Y., Green, N. S., Adamski-Werner, S. L. & Kelly, J. W. (2005). Native state kinetic stabilization as a strategy to ameliorate protein misfolding diseases: a focus on the transthyretin amyloidoses. *Acc Chem Res* **38**, 911-21.
26. Hurlle, M. R., Helms, L. R., Li, L., Chan, W. & Wetzel, R. (1994). A role for destabilizing amino acid replacements in light-chain amyloidosis. *Proc Natl Acad Sci U S A* **91**, 5446-50.
27. Vassall, K. A., Stubbs, H. R., Primmer, H. A., Tong, M. S., Sullivan, S. M., Sobering, R., Srinivasan, S., Briere, L. A., Dunn, S. D., Colon, W. & Meiering, E. M. (2011). Decreased stability and increased formation of soluble aggregates by immature superoxide dismutase do not account for disease severity in ALS. *Proc Natl Acad Sci U S A* **108**, 2210-5.
28. Meiering, E. M. (2008). The threat of instability: neurodegeneration predicted by protein destabilization and aggregation propensity. *PLoS Biol* **6**, e193.
29. Chiti, F. & Dobson, C. M. (2009). Amyloid formation by globular proteins under native conditions. *Nat Chem Biol* **5**, 15-22.
30. Jahn, T. R. & Radford, S. E. (2008). Folding versus aggregation: polypeptide conformations on competing pathways. *Arch Biochem Biophys* **469**, 100-17.
31. Esteras-Chopo, A., Serrano, L. & Lopez de la Paz, M. (2005). The amyloid stretch hypothesis: recruiting proteins toward the dark side. *Proc Natl Acad Sci U S A* **102**, 16672-7.
32. Lopez de la Paz, M. & Serrano, L. (2004). Sequence determinants of amyloid fibril formation. *Proc Natl Acad Sci U S A* **101**, 87-92.
33. Maurer-Stroh, S., Debulpaep, M., Kuemmerer, N., Lopez de la Paz, M., Martins, I. C., Reumers, J., Morris, K. L., Copland, A., Serpell, L., Serrano, L., Schymkowitz, J. W. & Rousseau, F. (2010). Exploring the sequence determinants of amyloid structure using position-specific scoring matrices. *Nat Methods* **7**, 237-42.
34. Pace, C. N. (1975). The stability of globular proteins. *CRC Crit Rev Biochem* **3**, 1-43.
35. Levitsky, D. I., Pivovarova, A. V., Mikhailova, V. V. & Nikolaeva, O. P. (2008). Thermal unfolding and aggregation of actin. *Febs J* **275**, 4280-95.
36. Arnaudov, L. N. & de Vries, R. (2005). Thermally induced fibrillar aggregation of hen egg white lysozyme. *Biophys J* **88**, 515-26.
37. Vetri, V. & Militello, V. (2005). Thermal induced conformational changes involved in the aggregation pathways of beta-lactoglobulin. *Biophys Chem* **113**, 83-91.
38. Chu, H. L. & Lin, S. Y. (2001). Temperature-induced conformational changes in amyloid beta(1-40) peptide investigated by simultaneous FT-IR microspectroscopy with thermal system. *Biophys Chem* **89**, 173-80.
39. Fink, A. L., Calciano, L. J., Goto, Y., Kurotsu, T. & Palleros, D. R. (1994). Classification of acid denaturation of proteins: intermediates and unfolded states. *Biochemistry* **33**, 12504-11.

40. Narayanan, S. & Reif, B. (2005). Characterization of chemical exchange between soluble and aggregated states of beta-amyloid by solution-state NMR upon variation of salt conditions. *Biochemistry* **44**, 1444-52.
41. Sicorello, A., Torrassa, S., Soldi, G., Gianni, S., Travaglini-Allocatelli, C., Taddei, N., Relini, A. & Chiti, F. (2009). Agitation and high ionic strength induce amyloidogenesis of a folded PDZ domain in native conditions. *Biophys J* **96**, 2289-98.
42. Pedersen, J. B., Fojan, P., Sorensen, J. & Petersen, S. B. (2006). Towards control of aggregational behaviour of alpha-lactalbumin at acidic pH. *J Fluoresc* **16**, 611-21.
43. Zhang, Y. & Cremer, P. S. (2006). Interactions between macromolecules and ions: The Hofmeister series. *Curr Opin Chem Biol* **10**, 658-63.
44. Yeh, V., Broering, J. M., Romanyuk, A., Chen, B., Chernoff, Y. O. & Bommarius, A. S. (2010). The Hofmeister effect on amyloid formation using yeast prion protein. *Protein Sci* **19**, 47-56.
45. Fesinmeyer, R. M., Hogan, S., Saluja, A., Brych, S. R., Kras, E., Narhi, L. O., Brems, D. N. & Gokarn, Y. R. (2009). Effect of ions on agitation- and temperature-induced aggregation reactions of antibodies. *Pharm Res* **26**, 903-13.
46. Stefani, M. (2008). Protein folding and misfolding on surfaces. *Int J Mol Sci* **9**, 2515-42.
47. Linse, S., Cabaleiro-Lago, C., Xue, W. F., Lynch, I., Lindman, S., Thulin, E., Radford, S. E. & Dawson, K. A. (2007). Nucleation of protein fibrillation by nanoparticles. *Proc Natl Acad Sci U S A* **104**, 8691-6.
48. Zhu, M., Souillac, P. O., Ionescu-Zanetti, C., Carter, S. A. & Fink, A. L. (2002). Surface-catalyzed amyloid fibril formation. *J Biol Chem* **277**, 50914-22.
49. Smith, M. I., Sharp, J. S. & Roberts, C. J. (2007). Nucleation and growth of insulin fibrils in bulk solution and at hydrophobic polystyrene surfaces. *Biophys J* **93**, 2143-51.
50. Carulla, N., Zhou, M., Arimon, M., Gairi, M., Giralte, E., Robinson, C. V. & Dobson, C. M. (2009). Experimental characterization of disordered and ordered aggregates populated during the process of amyloid fibril formation. *Proc Natl Acad Sci U S A* **106**, 7828-33.
51. Seetharaman, S. V., Prudencio, M., Karch, C., Holloway, S. P., Borchelt, D. R. & Hart, P. J. (2009). Immature copper-zinc superoxide dismutase and familial amyotrophic lateral sclerosis. *Exp Biol Med (Maywood)* **234**, 1140-54.
52. Caughey, B. & Lansbury, P. T. (2003). Protofibrils, pores, fibrils, and neurodegeneration: separating the responsible protein aggregates from the innocent bystanders. *Annu Rev Neurosci* **26**, 267-98.
53. Soto, C. & Estrada, L. D. (2008). Protein misfolding and neurodegeneration. *Arch Neurol* **65**, 184-9.
54. O'Nuallain, B. & Wetzel, R. (2002). Conformational Abs recognizing a generic amyloid fibril epitope. *Proc Natl Acad Sci U S A* **99**, 1485-90.
55. Kaye, R., Head, E., Thompson, J. L., McIntire, T. M., Milton, S. C., Cotman, C. W. & Glabe, C. G. (2003). Common structure of soluble amyloid oligomers implies common mechanism of pathogenesis. *Science* **300**, 486-9.
56. Campioni, S., Mannini, B., Zampagni, M., Pensalfini, A., Parrini, C., Evangelisti, E., Relini, A., Stefani, M., Dobson, C. M., Cecchi, C. & Chiti, F. (2010). A causative link between the structure of aberrant protein oligomers and their toxicity. *Nat Chem Biol* **6**, 140-7.
57. Winner, B., Jappelli, R., Maji, S. K., Desplats, P. A., Boyer, L., Aigner, S., Hetzer, C., Loher, T., Vilar, M., Campioni, S., Tzitzilonis, C., Soragni, A., Jessberger, S., Mira, H., Consiglio, A., Pham, E., Masliah, E., Gage, F. H. & Riek, R. (2011). In vivo demonstration that alpha-synuclein oligomers are toxic. *Proc Natl Acad Sci U S A* **108**, 4194-9.

58. Van Damme, P. & Robberecht, W. (2009). Recent advances in motor neuron disease. *Curr Opin Neurol* **22**, 486-92.
59. Tainer, J. A., Getzoff, E. D., Beem, K. M., Richardson, J. S. & Richardson, D. C. (1982). Determination and analysis of the 2 Å structure of copper, zinc superoxide dismutase. *J Mol Biol* **160**, 181-217.
60. Richardson, J., Thomas, K. A., Rubin, B. H. & Richardson, D. C. (1975). Crystal structure of bovine Cu,Zn superoxide dismutase at 3 Å resolution: chain tracing and metal ligands. *Proc Natl Acad Sci U S A* **72**, 1349-53.
61. Wang, Q., Johnson, J. L., Agar, N. Y. & Agar, J. N. (2008). Protein aggregation and protein instability govern familial amyotrophic lateral sclerosis patient survival. *PLoS Biol* **6**, e170.
62. Chattopadhyay, M. & Valentine, J. S. (2009). Aggregation of copper-zinc superoxide dismutase in familial and sporadic ALS. *Antioxid Redox Signal* **11**, 1603-14.
63. Parge, H. E., Hallewell, R. A. & Tainer, J. A. (1992). Atomic structures of wild-type and thermostable mutant recombinant human Cu,Zn superoxide dismutase. *Proc Natl Acad Sci U S A* **89**, 6109-13.
64. Furukawa, Y. & O'Halloran, T. V. (2005). Amyotrophic lateral sclerosis mutations have the greatest destabilizing effect on the apo- and reduced form of SOD1, leading to unfolding and oxidative aggregation. *J Biol Chem* **280**, 17266-74.
65. Furukawa, Y., Kaneko, K., Yamanaka, K., O'Halloran, T. V. & Nukina, N. (2008). Complete loss of post-translational modifications triggers fibrillar aggregation of SOD1 in the familial form of amyotrophic lateral sclerosis. *J Biol Chem* **283**, 24167-76.
66. Ilieva, H., Polymenidou, M. & Cleveland, D. W. (2009). Non-cell autonomous toxicity in neurodegenerative disorders: ALS and beyond. *J Cell Biol* **187**, 761-72.
67. Kerman, A., Liu, H. N., Croul, S., Bilbao, J., Rogaeva, E., Zinman, L., Robertson, J. & Chakrabarty, A. (2010). Amyotrophic lateral sclerosis is a non-amyloid disease in which extensive misfolding of SOD1 is unique to the familial form. *Acta Neuropathol* **119**, 335-344.
68. Zetterstrom, P., Stewart, H. G., Bergemalm, D., Jonsson, P. A., Graffmo, K. S., Andersen, P. M., Brannstrom, T., Oliveberg, M. & Marklund, S. L. (2007). Soluble misfolded subfractions of mutant superoxide dismutase-1s are enriched in spinal cords throughout life in murine ALS models. *Proc Natl Acad Sci U S A* **104**, 14157-62.
69. Oztug Durer, Z. A., Cohlberg, J. A., Dinh, P., Padua, S., Ehrenclou, K., Downes, S., Tan, J. K., Nakano, Y., Bowman, C. J., Hoskins, J. L., Kwon, C., Mason, A. Z., Rodriguez, J. A., Doucette, P. A., Shaw, B. F. & Selverstone Valentine, J. (2009). Loss of metal ions, disulfide reduction and mutations related to familial ALS promote formation of amyloid-like aggregates from superoxide dismutase. *PLoS One* **4**, e5004.
70. Chattopadhyay, M., Durazo, A., Sohn, S. H., Strong, C. D., Gralla, E. B., Whitelegge, J. P. & Valentine, J. S. (2008). Initiation and elongation in fibrillation of ALS-linked superoxide dismutase. *Proc Natl Acad Sci U S A* **105**, 18663-8.
71. Hwang, Y. M., Stathopoulos, P. B., Dimmick, K., Yang, H., Badiei, H. R., Tong, M. S., Rumfeldt, J. A., Chen, P., Karanassios, V. & Meiering, E. M. (2010). Nonamyloid aggregates arising from mature copper/zinc superoxide dismutases resemble those observed in amyotrophic lateral sclerosis. *J Biol Chem* **285**, 41701-11.
72. Banci, L., Bertini, I., Durazo, A., Giroto, S., Gralla, E. B., Martinelli, M., Valentine, J. S., Vieru, M. & Whitelegge, J. P. (2007). Metal-free superoxide dismutase forms soluble oligomers under physiological conditions: a possible general mechanism for familial ALS. *Proc Natl Acad Sci U S A* **104**, 11263-7.

73. Karch, C. M. & Borchelt, D. R. (2008). A limited role for disulfide cross-linking in the aggregation of mutant SOD1 linked to familial amyotrophic lateral sclerosis. *J Biol Chem* **283**, 13528-37.
74. Karch, C. M., Prudencio, M., Winkler, D. D., Hart, P. J. & Borchelt, D. R. (2009). Role of mutant SOD1 disulfide oxidation and aggregation in the pathogenesis of familial ALS. *Proc Natl Acad Sci U S A* **106**, 7774-9.
75. Jonsson, P. A., Graffmo, K. S., Andersen, P. M., Brannstrom, T., Lindberg, M., Oliveberg, M. & Marklund, S. L. (2006). Disulphide-reduced superoxide dismutase-1 in CNS of transgenic amyotrophic lateral sclerosis models. *Brain* **129**, 451-64.
76. Sandelin, E., Nordlund, A., Andersen, P. M., Marklund, S. S. & Oliveberg, M. (2007). Amyotrophic lateral sclerosis-associated copper/zinc superoxide dismutase mutations preferentially reduce the repulsive charge of the proteins. *J Biol Chem* **282**, 21230-6.
77. Furukawa, Y. & O'Halloran, T. V. (2006). Posttranslational modifications in Cu,Zn-superoxide dismutase and mutations associated with amyotrophic lateral sclerosis. *Antioxid Redox Signal* **8**, 847-67.
78. Khare, S. D., Caplow, M. & Dokholyan, N. V. (2006). FALS mutations in Cu, Zn superoxide dismutase destabilize the dimer and increase dimer dissociation propensity: a large-scale thermodynamic analysis. *Amyloid* **13**, 226-35.
79. Lepock, J. R., Frey, H. E. & Hallewell, R. A. (1990). Contribution of conformational stability and reversibility of unfolding to the increased thermostability of human and bovine superoxide dismutase mutated at free cysteines. *J Biol Chem* **265**, 21612-8.
80. Pechmann, S., Levy, E. D., Tartaglia, G. G. & Vendruscolo, M. (2009). Physicochemical principles that regulate the competition between functional and dysfunctional association of proteins. *Proc Natl Acad Sci U S A* **106**, 10159-64.
81. Uversky, V. N. (2003). Protein folding revisited. A polypeptide chain at the folding-misfolding-nonfolding cross-roads: which way to go? *Cell Mol Life Sci* **60**, 1852-71.
82. Chiti, F. & Dobson, C. M. (2006). Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem* **75**, 333-66.
83. Hamley, I. W. (2007). Peptide fibrillization. *Angew Chem Int Ed Engl* **46**, 8128-47.
84. Jung, J. P., Gasiorowski, J. Z. & Collier, J. H. (2010). Fibrillar peptide gels in biotechnology and biomedicine. *Biopolymers* **94**, 49-59.
85. Tartaglia, G. G. & Vendruscolo, M. (2008). The Zyggregator method for predicting protein aggregation propensities. *Chem Soc Rev* **37**, 1395-401.
86. Trovato, A., Chiti, F., Maritan, A. & Seno, F. (2006). Insight into the structure of amyloid fibrils from the analysis of globular proteins. *PLoS Comput Biol* **2**, e170.
87. Fernandez-Escamilla, A. M., Rousseau, F., Schymkowitz, J. & Serrano, L. (2004). Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat Biotechnol* **22**, 1302-6.
88. Garbuzynskiy, S. O., Lobanov, M. Y. & Galzitskaya, O. V. (2010). FoldAmyloid: a method of prediction of amyloidogenic regions from protein sequence. *Bioinformatics* **26**, 326-32.
89. Goldschmidt, L., Teng, P. K., Riek, R. & Eisenberg, D. (2010). Identifying the amyloids, proteins capable of forming amyloid-like fibrils. *Proc Natl Acad Sci U S A* **107**, 3487-92.
90. Wang, W. (2005). Protein aggregation and its inhibition in biopharmaceutics. *Int J Pharm* **289**, 1-30.
91. Chiti, F., Taddei, N., Baroni, F., Capanni, C., Stefani, M., Ramponi, G. & Dobson, C. M. (2002). Kinetic partitioning of protein folding and aggregation. *Nat Struct Biol* **9**, 137-43.

92. Chiti, F., Calamai, M., Taddei, N., Stefani, M., Ramponi, G. & Dobson, C. M. (2002). Studies of the aggregation of mutant proteins in vitro provide insights into the genetics of amyloid diseases. *Proc Natl Acad Sci U S A* **99 Suppl 4**, 16419-26.
93. Gooch, M. D. & Stennett, D. J. (1996). Molecular basis of Alzheimer's disease. *Am J Health Syst Pharm* **53**, 1545-57; quiz 1603-4.
94. Stefani, M., Taddei, N. & Ramponi, G. (1997). Insights into acylphosphatase structure and catalytic mechanism. *Cell Mol Life Sci* **53**, 141-51.
95. Serpell, L. C. (2000). Alzheimer's amyloid fibrils: structure and assembly. *Biochim Biophys Acta* **1502**, 16-30.
96. Kaufmann, K. W., Lemmon, G. H., Deluca, S. L., Sheehan, J. H. & Meiler, J. (2010). Practically useful: what the Rosetta protein modeling suite can do for you. *Biochemistry* **49**, 2987-98.
97. Tomaselli, S., Esposito, V., Vangone, P., van Nuland, N. A., Bonvin, A. M., Guerrini, R., Tancredi, T., Temussi, P. A. & Picone, D. (2006). The alpha-to-beta conformational transition of Alzheimer's Abeta-(1-42) peptide in aqueous media is reversible: a step by step conformational analysis suggests the location of beta conformation seeding. *Chembiochem* **7**, 257-67.
98. Teng, P. K. & Eisenberg, D. (2009). Short protein segments can drive a non-fibrillizing protein into the amyloid state. *Protein Eng Des Sel* **22**, 531-6.
99. The PyMOL Molecular Graphics System Version 1.2r3pre edit. Schrödinger, LLC.
100. Routledge, K. E., Tartaglia, G. G., Platt, G. W., Vendruscolo, M. & Radford, S. E. (2009). Competition between intramolecular and intermolecular interactions in an amyloid-forming protein. *J Mol Biol* **389**, 776-86.
101. Ricagno, S., Raimondi, S., Giorgetti, S., Bellotti, V. & Bolognesi, M. (2009). Human beta-2 microglobulin W60V mutant structure: Implications for stability and amyloid aggregation. *Biochem Biophys Res Commun* **380**, 543-7.
102. Frousios, K. K., Iconomidou, V. A., Karletidi, C. M. & Hamodrakas, S. J. (2009). Amyloidogenic determinants are usually not buried. *BMC Struct Biol* **9**, 44.
103. Pastor, M. T., Esteras-Chopo, A. & Serrano, L. (2007). Hacking the code of amyloid formation: the amyloid stretch hypothesis. *Prion* **1**, 9-14.
104. Luheshi, L. M., Tartaglia, G. G., Brorsson, A. C., Pawar, A. P., Watson, I. E., Chiti, F., Vendruscolo, M., Lomas, D. A., Dobson, C. M. & Crowther, D. C. (2007). Systematic in vivo analysis of the intrinsic determinants of amyloid Beta pathogenicity. *PLoS Biol* **5**, e290.
105. Conchillo-Sole, O., de Groot, N. S., Aviles, F. X., Vendrell, J., Daura, X. & Ventura, S. (2007). AGGRESCAN: a server for the prediction and evaluation of "hot spots" of aggregation in polypeptides. *BMC Bioinformatics* **8**, 65.
106. Belli, M., Ramazzotti, M. & Chiti, F. (2011). Prediction of amyloid aggregation in vivo. *EMBO Rep* **12**, 657-63.
107. de Groot, N. S., Aviles, F. X., Vendrell, J. & Ventura, S. (2006). Mutagenesis of the central hydrophobic cluster in Abeta42 Alzheimer's peptide. Side-chain properties correlate with aggregation propensities. *Febs J* **273**, 658-68.
108. Van Nostrand, W. E., Melchor, J. P., Cho, H. S., Greenberg, S. M. & Rebeck, G. W. (2001). Pathogenic effects of D23N Iowa mutant amyloid beta -protein. *J Biol Chem* **276**, 32860-6.
109. Miravalle, L., Tokuda, T., Chiarle, R., Giaccone, G., Bugiani, O., Tagliavini, F., Frangione, B. & Ghiso, J. (2000). Substitutions at codon 22 of Alzheimer's abeta peptide induce diverse conformational changes and apoptotic effects in human cerebral endothelial cells. *J Biol Chem* **275**, 27110-6.

110. Nilsberth, C., Westlind-Danielsson, A., Eckman, C. B., Condron, M. M., Axelman, K., Forsell, C., Sten, C., Luthman, J., Teplow, D. B., Younkin, S. G., Naslund, J. & Lannfelt, L. (2001). The 'Arctic' APP mutation (E693G) causes Alzheimer's disease by enhanced Abeta protofibril formation. *Nat Neurosci* **4**, 887-93.
111. Biancalana, M. & Koide, S. (2010). Molecular mechanism of Thioflavin-T binding to amyloid fibrils. *Biochim Biophys Acta* **1804**, 1405-12.
112. Vassar, P. S. & Culling, C. F. (1959). Fluorescent stains, with special reference to amyloid and connective tissues. *Arch Pathol* **68**, 487-98.
113. Naiki, H., Higuchi, K., Hosokawa, M. & Takeda, T. (1989). Fluorometric determination of amyloid fibrils in vitro using the fluorescent dye, thioflavin T1. *Anal Biochem* **177**, 244-9.
114. Santos, N. C. & Castanho, M. A. (1996). Teaching light scattering spectroscopy: the dimension and shape of tobacco mosaic virus. *Biophys J* **71**, 1641-50.
115. Lomakin, A., Teplow, D. B. & Benedek, G. B. (2005). Quasielastic light scattering for protein assembly studies. *Methods Mol Biol* **299**, 153-74.
116. Lomakin, A., Benedek, G. B. & Teplow, D. B. (1999). Monitoring protein assembly using quasielastic light scattering spectroscopy. *Methods Enzymol* **309**, 429-59.
117. Stathopoulos, P. B. (2005). Stability and Aggregation of Amyotrophic Lateral Sclerosis-Associated Mutant Copper, Zinc Superoxide Dismutases, University of Waterloo.
118. Kayatekin, C., Zitzewitz, J. A. & Matthews, C. R. (2010). Disulfide-reduced ALS variants of Cu, Zn superoxide dismutase exhibit increased populations of unfolded species. *J Mol Biol* **398**, 320-31.
119. Naeem, A. & Fazili, N. A. (2011). Defective Protein Folding and Aggregation as the Basis of Neurodegenerative Diseases: The Darker Aspect of Proteins. *Cell Biochem Biophys*.
120. Khurana, R., Gillespie, J. R., Talapatra, A., Minert, L. J., Ionescu-Zanetti, C., Millett, I. & Fink, A. L. (2001). Partially folded intermediates as critical precursors of light chain amyloid fibrils and amorphous aggregates. *Biochemistry* **40**, 3525-35.
121. Smith, D. P., Jones, S., Serpell, L. C., Sunde, M. & Radford, S. E. (2003). A systematic investigation into the effect of protein destabilisation on beta 2-microglobulin amyloid formation. *J Mol Biol* **330**, 943-54.
122. Ramshini, H., Parrini, C., Relini, A., Zampagni, M., Mannini, B., Pesce, A., Saboury, A. A., Nemat-Gorgani, M. & Chiti, F. (2011). Large proteins have a great tendency to aggregate but a low propensity to form amyloid fibrils. *PLoS One* **6**, e16075.
123. Otzen, D. E., Kristensen, O. & Oliveberg, M. (2000). Designed protein tetramer zipped together with a hydrophobic Alzheimer homology: a structural clue to amyloid assembly. *Proc Natl Acad Sci U S A* **97**, 9907-12.
124. Rousseau, F., Serrano, L. & Schymkowitz, J. W. (2006). How evolutionary pressure against protein aggregation shaped chaperone specificity. *J Mol Biol* **355**, 1037-47.
125. Chi, E. Y., Krishnan, S., Kendrick, B. S., Chang, B. S., Carpenter, J. F. & Randolph, T. W. (2003). Roles of conformational stability and colloidal stability in the aggregation of recombinant human granulocyte colony-stimulating factor. *Protein Sci* **12**, 903-13.
126. Stryer, L. (1965). The interaction of a naphthalene dye with apomyoglobin and apohemoglobin. A fluorescent probe of non-polar binding sites. *J Mol Biol* **13**, 482-95.
127. Hawe, A., Sutter, M. & Jiskoot, W. (2008). Extrinsic fluorescent dyes as tools for protein characterization. *Pharm Res* **25**, 1487-99.

128. Gasymov, O. K., Abduragimov, A. R. & Glasgow, B. J. (2007). Characterization of fluorescence of ANS-tear lipocalin complex: evidence for multiple-binding modes. *Photochem Photobiol* **83**, 1405-14.
129. Bertoncini, C. W. & Celej, M. S. (2011). Small molecule fluorescent probes for the detection of amyloid self-assembly in vitro and in vivo. *Curr Protein Pept Sci* **12**, 205-20.
130. Bolognesi, B., Kumita, J. R., Barros, T. P., Esbjorner, E. K., Luheshi, L. M., Crowther, D. C., Wilson, M. R., Dobson, C. M., Favrin, G. & Yerbury, J. J. (2010). ANS binding reveals common features of cytotoxic amyloid species. *ACS Chem Biol* **5**, 735-40.
131. Bajaj, H., Sharma, V. K., Badkar, A., Zeng, D., Nema, S. & Kalonia, D. S. (2006). Protein structural conformation and not second virial coefficient relates to long-term irreversible aggregation of a monoclonal antibody and ovalbumin in solution. *Pharm Res* **23**, 1382-94.
132. Le Brun, V., Friess, W., Bassarab, S. & Garidel, P. (2010). Correlation of protein-protein interactions as assessed by affinity chromatography with colloidal protein stability: a case study with lysozyme. *Pharm Dev Technol* **15**, 421-30.
133. Saluja, A., Fesinmeyer, R. M., Hogan, S., Brems, D. N. & Gokarn, Y. R. (2010). Diffusion and sedimentation interaction parameters for measuring the second virial coefficient and their utility as predictors of protein aggregation. *Biophys J* **99**, 2657-65.
134. Lenhoff, A. M., Dumetz, A. C., Snellinger-O'Brien, A. M. & Kaler, E. W. (2007). Patterns of protein - protein interactions in salt solutions and implications for protein crystallization. *Protein Science* **16**, 1867-1877.
135. Stathopoulos, P. B., Rumfeldt, J. A., Scholz, G. A., Irani, R. A., Frey, H. E., Hallewell, R. A., Lepock, J. R. & Meiering, E. M. (2003). Cu/Zn superoxide dismutase mutants associated with amyotrophic lateral sclerosis show enhanced formation of aggregates in vitro. *Proc Natl Acad Sci U S A* **100**, 7021-6.
136. Rumfeldt, J. A., Stathopoulos, P. B., Chakraborty, A., Lepock, J. R. & Meiering, E. M. (2006). Mechanism and thermodynamics of guanidinium chloride-induced denaturation of ALS-associated mutant Cu,Zn superoxide dismutases. *J Mol Biol* **355**, 106-23.
137. Rumfeldt, J. A. (2006). Thermodynamics, Kinetics and Structural Dynamics of Amyotrophic Lateral Sclerosis-Associated Mutant Copper-Zinc Superoxide Dismutases, University of Waterloo.
138. Lyons, T. J., Liu, H., Goto, J. J., Nersissian, A., Roe, J. A., Graden, J. A., Cafe, C., Ellerby, L. M., Bredesen, D. E., Gralla, E. B. & Valentine, J. S. (1996). Mutations in copper-zinc superoxide dismutase that cause amyotrophic lateral sclerosis alter the zinc binding site and the redox behavior of the protein. *Proc Natl Acad Sci U S A* **93**, 12240-4.
139. Vassall, K. A. (2009). Folding and Stability Studies on Amyotrophic Lateral Sclerosis-Associated Cu, Zn apo Superoxide dismutases, University of Waterloo.
140. Han, J. C. & Han, G. Y. (1994). A procedure for quantitative determination of tris(2-carboxyethyl)phosphine, an odorless reducing agent more stable and effective than dithiothreitol. *Anal Biochem* **220**, 5-10.
141. Getz, E. B., Xiao, M., Chakraborty, T., Cooke, R. & Selvin, P. R. (1999). A comparison between the sulfhydryl reductants tris(2-carboxyethyl)phosphine and dithiothreitol for use in protein biochemistry. *Anal Biochem* **273**, 73-80.
142. Rhee, S. S. & Burke, D. H. (2004). Tris(2-carboxyethyl)phosphine stabilization of RNA: comparison with dithiothreitol for use with nucleic acid and thiophosphoryl chemistry. *Anal Biochem* **325**, 137-43.

143. Yancey, P. H., Blake, W. R. & Conley, J. (2002). Unusual organic osmolytes in deep-sea animals: adaptations to hydrostatic pressure and other perturbants. *Comp Biochem Physiol A Mol Integr Physiol* **133**, 667-76.
144. Welch, W. J. & Brown, C. R. (1996). Influence of molecular and chemical chaperones on protein folding. *Cell Stress Chaperones* **1**, 109-15.
145. Loria, J. P. & Doan-Nguyen, V. (2007). The effects of cosolutes on protein dynamics: The reversal of denaturant-induced protein fluctuations by trimethylamine N-oxide. *Protein Science* **16**, 20-29.
146. Ahmad, F., Singh, L. R., Poddar, N. K., Dar, T. A., Rahman, S. & Kumar, R. (2011). Forty Years of Research on Osmolyte-Induced Protein Folding and Stability. *Journal of the Iranian Chemical Society* **8**, 1-23.
147. Banci, L., Bertini, I., Boca, M., Girotto, S., Martinelli, M., Valentine, J. S. & Vieru, M. (2008). SOD1 and Amyotrophic Lateral Sclerosis: Mutations and Oligomerization. *Plos One* **3**.
148. Yamamoto, K., Yagi, H., Ozawa, D., Sasahara, K., Naiki, H. & Goto, Y. (2008). Thiol compounds inhibit the formation of amyloid fibrils by beta 2-microglobulin at neutral pH. *J Mol Biol* **376**, 258-68.
149. Mello, C. C. & Barrick, D. (2003). Measuring the stability of partly folded proteins using TMAO. *Protein Sci* **12**, 1522-9.
150. Hornberg, A., Logan, D. T., Marklund, S. L. & Oliveberg, M. (2007). The coupling between disulphide status, metallation and dimer interface strength in Cu/Zn superoxide dismutase. *J Mol Biol* **365**, 333-42.
151. Vassall, K. A., Stathopoulos, P. B., Rumfeldt, J. A., Lepock, J. R. & Meiering, E. M. (2006). Equilibrium thermodynamic analysis of amyotrophic lateral sclerosis-associated mutant apo Cu,Zn superoxide dismutases. *Biochemistry* **45**, 7366-79.
152. Cobos, E. S. & Radford, S. E. (2006). Sulfate-induced effects in the on-pathway intermediate of the bacterial immunity protein Im7. *Biochemistry* **45**, 2274-82.
153. Behrens, S. H. & Grier, D. G. (2001). The charge of glass and silica surfaces. *Journal of Chemical Physics* **115**, 6716-6721.
154. Banci, L., Bertini, I., Boca, M., Girotto, S., Martinelli, M., Valentine, J. S. & Vieru, M. (2008). SOD1 and amyotrophic lateral sclerosis: mutations and oligomerization. *PLoS One* **3**, e1677.
155. Olsen, S. N., Andersen, K. B., Randolph, T. W., Carpenter, J. F. & Westh, P. (2009). Role of electrostatic repulsion on colloidal stability of Bacillus halmopalus alpha-amylase. *Biochim Biophys Acta* **1794**, 1058-65.
156. Giurleo, J. T., He, X. & Talaga, D. S. (2008). Beta-lactoglobulin assembles into amyloid through sequential aggregated intermediates. *J Mol Biol* **381**, 1332-48.
157. Tartaglia, G. G., Pawar, A. P., Campioni, S., Dobson, C. M., Chiti, F. & Vendruscolo, M. (2008). Prediction of aggregation-prone regions in structured proteins. *J Mol Biol* **380**, 425-36.