



University of Pennsylvania
ScholarlyCommons

Publicly Accessible Penn Dissertations

2019

Combining Computational And Experimental Approaches To Study Disordered And Aggregation Prone Proteins

John Joseph Ferrie
University of Pennsylvania

Follow this and additional works at: <https://repository.upenn.edu/edissertations>

 Part of the [Chemistry Commons](#)

Recommended Citation

Ferrie, John Joseph, "Combining Computational And Experimental Approaches To Study Disordered And Aggregation Prone Proteins" (2019). *Publicly Accessible Penn Dissertations*. 3586.
<https://repository.upenn.edu/edissertations/3586>

This paper is posted at ScholarlyCommons. <https://repository.upenn.edu/edissertations/3586>
For more information, please contact repository@pobox.upenn.edu.

Combining Computational And Experimental Approaches To Study Disordered And Aggregation Prone Proteins

Abstract

Over the past two decades disordered proteins have become more widely recognized, challenging the canonical structure-function paradigm associated with proteins. These highly dynamic proteins have been identified across a wide range of species and play a variety of functional roles. Furthermore, the structural plasticity of these proteins gives way to their increased aggregation susceptibility, compared to canonical, well-folded proteins, placing disordered proteins at the center of many neurodegenerative diseases. Despite the increased recognition of the abundance and complexity of disordered proteins, their structural features and the mechanisms by which they transit between functional and pathological roles remains elusive. The efforts described herein focus on leveraging both experimental and computational approaches to study the structure and dynamics of these proteins. Fluorescence-based experiment have proven useful for studying these systems as the intrinsic heterogeneity of this class of proteins, which precludes the use of many traditional structural biochemistry techniques, can be accommodated. Therefore, initial efforts focused on developing new minimally perturbing fluorescence probes and coupling these probes with site-selective labeling strategies. Subsequent efforts focused on identifying methods which could predict where these probes would be tolerated to boost protein yield and avoid structural perturbation. These and other fluorescence probes were employed in Förster Resonance Energy Transfer (FRET) experiments, to study the conformational ensemble of α -synuclein, a disordered protein whose aggregation is implicated in Parkinson's Disease pathogenesis. Experimental FRET data was paired with molecular modeling in PyRosetta to simulate the conformational ensembles of α -synuclein in the presence and absence of 2 M TMAO. The accuracy of the resultant ensembles was corroborated by comparison to other experimental data. Following this initial success using experimentally constrained simulations, attention was directed towards the development of algorithms capable of generating accurate structural representations of both disordered and ordered proteins de novo. Lastly, this work showcases the utility of a high-throughput in-silico screening approach in identifying a compound that binds selectively to α -synuclein fibrils with nanomolar affinity. Overall this work highlights several computational and experimental approaches which are broadly applicable to the study of disordered and aggregation prone proteins

Degree Type

Dissertation

Degree Name

Doctor of Philosophy (PhD)

Graduate Group

Chemistry

First Advisor

Ernest J. Petersson

Keywords

amyloid, fluorescence, modeling

Subject Categories

Chemistry

COMBINING COMPUTATIONAL AND EXPERIMENTAL APPROACHES TO STUDY
DISORDERED AND AGGREGATION PRONE PROTEINS

John J. Ferrie

A DISSERTATION

in

Chemistry

Presented to the Faculties of the University of Pennsylvania

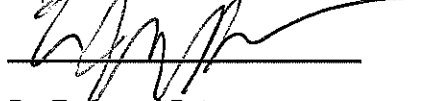
in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2019

Supervisor of Dissertation



Dr. E. James Petersson

Associate Professor of Chemistry

Graduate Group Chairperson



Dr. David W. Christianson

Roy and Diana Vagelos Professor in Chemistry and Chemical Biology

Dissertation Committee

Dr. Elizabeth Rhoades, Associate Professor of Chemistry

Dr. Jessica M. Anna, Assistant Professor of Chemistry

Dr. Zahra Fakhraai, Associate Professor of Chemistry

Dr. Jeffery G. Saven, Professor of Chemistry

Dedication

This dissertation is dedicated to all those who have helped me grow as a scientist over the course of this thesis work and all those who stood by and supported me along the way.

ACKNOWLEDGMENT

To the many people who have helped me throughout this journey a simple thanks is not enough. To friends and colleagues, new and old, this work is truly a testament not only of my own person effort, but also the effort, kindness, compassion and assistance that I have received from all of you daily. Truly this work would not exist without your tireless efforts.

First, I would like to thank Professor E. James Petersson for providing me with the opportunity to work in his lab and for all his mentorship and guidance with experiments, assistance with fellowship applications and paper writing and his career advice. I also express my deepest appreciation for the space and vote of confidence that he gave me to explore computational modeling, given no prior coding experience. No other single mission in graduate school has shaped me more than taking on this task. Also, I owe special thanks to my committee members, Professors, Rhoades, Fakhraai, Anna and Saven for all their feedback over the years and their encouragement.

The Chemistry Department at the University of Pennsylvania has provided me with so many talented colleagues and good friends. I want to specially thank my fellow classmates whom I entered graduate school with, including Michael Nicastrì, Stan Najmr and Benjamin Partridge who helped me weather many challenges throughout my graduate school career. Also, I want to thank Rebecca Wissner, who served as my mentor throughout my first year in the Petersson Lab. I was able to get a strong start in the lab because of her mentorship and the

comradery of these friends. Also, I would like to thank Professors Robert Mach, Abhinav Nath, Virginia Lee and all the other collaborators I have had the opportunity to work with over the years. I am gracious for all these opportunities and all your assistance. I would also like to thank Professor David Baker and Timothy Craven for allowing me to spend a summer at the University of Washington to better familiarize myself with the complexities of the Rosetta Modeling Suite. Additionally, I would like to thank the National Science Foundation and the Parkinson's Disease Foundation for their financial support during my graduate school career.

The Petersson lab most days feels like a family, and over my time here has grown into a tight-knit group that not only works well together but has fun and shares a lot of laughs. I have met some of my best friends in the Petersson lab such as Christopher Walters (known to many as "Old Man Walters") who was great collaborator and an even better friend, Taylor Barrett whose singing voice has no parallel and Miklos Szantai-Kis who was my desk partner from day one. More recently, I have had the pleasure of working alongside Sam Giannakoulis and Marshall Lougee, who have helped me wrap up this work. Without these and all the members of the Petersson lab my graduate school career would have been less successful and much less fun. Additionally, I would like to thank all the administration and staff of the Chemistry department

Outside of lab I have had the strong support of many friends from high school and from my undergraduate studies at The College of New Jersey.

Although too numerous to list, I want to specifically thank Ryan Janelli, Stefan Turan and Andrew Apicello whom have provided me with many fantastic pieces of advice and have shared in many fun weekends. Also, I would like to thank my family especially my parents, John and Shirley Ferrie, for all their support throughout my endeavors. Without the values they instilled in me this would not have been possible.

Lastly, I want to thank the person who stood by me day in and day out, through the ups and downs, who was there to celebrate in my victories and console me in my defeats. Elizabeth Rossiter has been the most important person to me over the course of this journey and deserves all the credit in the world for all the effort she has put in for me. I hope that this work has made her proud.

ABSTRACT

COMBINING COMPUTATIONAL AND EXPERIMENTAL APPROACHES TO STUDY DISORDERED AND AGGREGATION PRONE PROTEINS

John J. Ferrie

Professor E. James Petersson

Over the past two decades disordered proteins have become more widely recognized, challenging the canonical structure-function paradigm associated with proteins. These highly dynamic proteins have been identified across a wide range of species and play a variety of functional roles. Furthermore, the structural plasticity of these proteins gives way to their increased aggregation susceptibility, compared to canonical, well-folded proteins, placing disordered proteins at the center of many neurodegenerative diseases. Despite the increased recognition of the abundance and complexity of disordered proteins, their structural features and the mechanisms by which they transit between functional and pathological roles remains elusive. The efforts described herein focus on leveraging both experimental and computational approaches to study the structure and dynamics of these proteins. Fluorescence-based experiment have proven useful for studying these systems as the intrinsic heterogeneity of this class of proteins, which precludes the use of many traditional structural biochemistry techniques, can be accommodated. Therefore, initial efforts focused on developing new minimally perturbing fluorescence probes and coupling these probes with site-selective labeling strategies. Subsequent efforts focused on identifying methods which could predict where these probes would be tolerated to boost protein yield and avoid structural perturbation. These and other fluorescence probes were employed in Förster Resonance Energy Transfer (FRET) experiments, to study the conformational ensemble of α -synuclein, a disordered protein whose aggregation is implicated in Parkinson's Disease pathogenesis. Experimental FRET data was paired with molecular modeling in PyRosetta to simulate the conformational ensembles of α -synuclein in the presence and absence of 2 M TMAO. The accuracy of the resultant ensembles was corroborated by comparison to other experimental data. Following

this initial success using experimentally constrained simulations, attention was directed towards the development of algorithms capable of generating accurate structural representations of both disordered and ordered proteins *de novo*. Lastly, this work showcases the utility of a high-throughput *in-silico* screening approach in identifying a compound that binds selectively to α -synuclein fibrils with nanomolar affinity. Overall this work highlights several computational and experimental approaches which are broadly applicable to the study of disordered and aggregation prone proteins.

TABLE OF CONTENTS

ACKNOWLEDGMENT	iii
ABSTRACT.....	vi
LIST OF TABLES	xi
LIST OF ILLUSTRATIONS	xiii
CHAPTER 1: INTRODUCTION.....	1
§ 1.1 The Role of Amyloidogenic Proteins in Neurodegenerative Disease ...	2
§ 1.2 The Utility of Fluorescence and FRET in Studying Protein Structure.	11
§ 1.3 Computational Methods for Modeling Disordered Proteins	23
CHAPTER 2: MULTICOLOR PROTEIN FRET WITH TRYPTOPHAN, SELECTIVE COUMARIN-CYSTEINE LABELING, AND GENETIC ACRIDONYLALANINE ENCODING.....	28
§ 2.1 Introduction.....	29
§ 2.2 Characterization of Mcm-Mal and Mcm-Br	31
§ 2.2 Protein Expression, Labeling and Purification	41
§ 2.3 FRET Measurements	53
§ 2.3 Results.....	65
§ 2.6 Conclusion	82
CHAPTER 3: IMPROVING THE FLUORESCENT PROBE ACRIDONYLALANINE THROUGH A COMBINATION OF THEORY AND EXPERIMENT.	84
§ 3.1 Introduction.....	85

§ 3.2 Spectroscopic Characterization	89
§ 3.3 Results.....	101
§ 3.4 Conclusions	104
CHAPTER 4: SYSTEMATIC EVALUATION OF SOLUBLE PROTEIN EXPRESSION USING A FLUORESCENT UNNATURAL AMINO ACID REVEALS NO RELIABLE PREDICTORS OF TOLERABILITY.	
	105
§ 4.1 Introduction.....	106
§ 4.2 Experimental and Computational Methods	108
§ 4.3 Results and Discussion	111
§ 4.3 Initial Conclusions.....	119
§ 4.4 Further Algorithmic Development.....	120
CHAPTER 5: USING A FRET LIBRARY WITH MULTIPLE PROBE PAIRS TO DRIVE MONTE CARLO SIMULATIONS OF A-SYNUCLEIN.....	
	132
§ 5.1 Introduction.....	133
§ 5.2 Protein Overexpression and Purification	138
§ 5.3 FRET Data Collection and Analysis	146
§ 5.4 FCS and AFM Characterization	158
§ 5.5 Computational Modeling Procedures	161
§ 5.6 Results.....	174
§ 5.7 Discussion	206
§ 5.8 Conclusions	213
CHAPTER 6: A UNIFIED DE NOVO APPROACH FOR PREDICTING THE STRUCTURES OF ORDERED AND DISORDERED PROTEINS.	
	216

§ 6.1 Introduction.....	217
§ 6.2 Methods.....	220
§ 6.3 Calculation of Data from Ensembles	223
§ 6.3 Results.....	227
Chapter 7: HIGH-THROUGHPUT <i>IN SILICO</i> AND <i>IN VITRO</i> SCREENING FOR COMPOUNDS THAT SELECTIVELY BIND TO ALPHA-SYNUCLEIN FIBRILS.	381
§ 7.1 Introduction.....	382
§ 7.2 Methods.....	384
§ 7.3 Results.....	393
§ 7.4 Discussion	412
CHAPTER 8: FUTURE DIRECTIONS AND OUTLOOK.....	413
§ 8.1 Identification of Novel Compounds with Improved Therapeutic Potential via FRET-based Screening	414
§ 8.2 Continued Exemplar-based Screening	416
§ 8.3 Novel Approaches in Rosetta Targeting Experimental Optimization	418
BIBLIOGRAPHY	421

LIST OF TABLES

Table 2.1: Protein MALDI Masses.	47
Table 2.2: Trypsin Fragment MALDI Masses.....	47
Table 2.3: Percent of Cys Containing Protein Labeled by Mcm-Mal.....	59
Table 2.4: TCSPC Fluorescence Lifetime Values	73
Table 2.5: CaM + WpOCNC FRET	76
Table 2.6: CaM Structural Model Data.....	77
Table 2.7: Stopped-Flow Data Fitting.....	81
Table 3.1: Calculated and Observed Photophysical Parameters of Acridone Derivatives.....	93
Table 4.1: List of descriptors for Rosetta simulations.....	123
Table 5.1: MALDI Masses from Cnf-Trp Library.....	145
Table 5.2: MALDI Masses from Fam-Raz Library	146
Table 5.3a: Steady-State FRET Fitting and Distance Determination	153
Table 5.3b: Steady-State FRET Fitting and Distance Determination	154
Table 5.4a: TCSPC Data and Fitting.....	156
Table 5.4b: TCSPC Data and Fitting	157
Table 5.5: Calculated E_{FRET} values from TCSPC.....	158
Table 5.6: FCS Data	184
Table 5.7: Values from fits of Congo Red aggregation kinetics.....	205
Table 5.8: Values from fits of fluorescence polarization aggregation kinetics. .	206
Table 6.1: Values for Computing J-Couplings.....	227

Table 6.2: Comparison of Simulated α S Ensembles to Global Experimental Data.	242
Table 6.3: Comparison of Simulated α S Ensembles to Chemical Shift and RDC Data.....	244
Table 6.4: Comparison of Simulated α S Ensembles to J-Coupling Data	246
Table 6.5: Comparison of AbInitio and AbInitioVO Outputs to PDB Structures	353
Table 7.1: Comparison of K_i values in α S and $a\beta$ 42 fibrils.....	408

LIST OF ILLUSTRATIONS

Figure 1.1: Structural differences between disordered and ordered proteins.	3
Figure 1.2: Overview of the functional and pathological roles of α S.	6
Figure 1.3: Overview of Fluorescence and FRET.	14
Figure 1.4: Overview of unnatural amino acid mutagenesis.	21
Figure 1.5: Overview of PyRosetta.	26
Figure 2.1: Mcm labelling for FRET experiments.	30
Figure 2.2: Determination of Extinction Coefficient for Mcm-Mal.	32
Figure 2.3: Synthesis of Mcm-Br and Mcm-Mal.	34
Figure 2.4: Absorbance and Fluorescence Spectra of Mcm-Br.	34
Figure 2.5: LCMS Analysis of Mcm-Br Reactions with Cys.	35
Figure 2.6: Mcm-Mal Reactions with Cys and Ac-Cys and assigned products. .	36
Figure 2.7: Absorbance and Fluorescence Spectra of Mcm-Mal.	36
Figure 2.8: LCMS Analysis of Mcm-Mal Reactions with Cys and Ac-Cys.	38
Figure 2.9: Mcm-Mal Reactions with Cys and Ac-Cys after 1 and 6 h.	39
Figure 2.10: Determination of Quantum Yield for Ac-Cys Mcm Product.	40
Figure 2.11: MALDI MS Characterization of CaM Variants.	48
Figure 2.12: Trypsin Digest MALDI MS Characterization of CaM Variants.	49
Figure 2.13: MALDI MS Characterization of α S Variants.	50
Figure 2.14: Enlarged MALDI MS of labeled α S Variants.	51
Figure 2.15: Trypsin Digest MALDI MS Characterization of α S Variants.	52
Figure 2.16: Ac-CysMcm and Acd Absorption and Emission Spectra.	54

Figure 2.17: Trp and Ac-CysMcm Absorption and Emission Spectra.	54
Figure 2.18: Trp and Acd Absorption and Emission Spectra.	55
Figure 2.19: CaM Absorbance Measurements.	58
Figure 2.20: α S Absorbance Measurements.	59
Figure 2.21: Mcm-Mal turn-on experiments.	66
Figure 2.22: α S Fluorescence Spectra and FRET Fitting Results.	69
Figure 2.23: α S TMAO FRET Experiments.	70
Figure 2.24a: α S Fluorescence Spectra and FRET Fitting Results.	71
Figure 2.24b: α S Fluorescence Spectra and FRET Fitting Results.	72
Figure 2.25: Mcm/Acd FRET Data Compared to Previous Structural Models of α S Ensembles.	74
Figure 2.26: Fluorescence Assay to Determine Binding of Labeled CaM.	75
Figure 2.27: CaM Structural Models.	76
Figure 2.28: CaM Peptide binding FRET.	78
Figure 2.29: CaM Steady State FRET Measurements.	80
Figure 2.30: CaM Stopped-Flow FRET Measurements.	81
Figure 2.31: Three Color FRET and Rapid Mixing Measurements of Peptide Binding Kinetics.	82
Figure 3.1: FRET experiments and fluorescent amino acids.	89
Figure 3.2: Synthesis of nitroacridones and aminoacridones.	90
Figure 3.4: Synthesis of acridones and benzoacridones.	90
Figure 3.4: Acridone Spectra.	95
Figure 3.5: 2-Aminoacridone Spectra.	96

Figure 3.6: 4-Aminoacridone Spectra.	97
Figure 3.8: 2-Fluoroacridone, 2-Methoxyacridone, and 4-Methoxyacridone Spectra.	99
Figure 3.9: 4-Benzoacridone, 4-Fluorobenzoacridone, and 4-Methoxybenzoacridone Spectra.	100
Figure 3.10: Absorption and emission spectra of acridone derivatives.	103
Figure 4.1: Scanning a variety of positions in LexA for Acd tolerability.	112
Figure 4.2: Features associated with soluble Acd-labeled LexA proteins.	113
Figure 4.3: Features associated with soluble Acd-labeled RecA proteins.	116
Figure 4.4: Predicting protein solubility through simulation of Acd incorporation in LexA.	118
Figure 4.5: Predicting protein solubility through simulation of Acd incorporation in RecA.	119
Figure 4.6a: Comparison of Rosetta energy terms to the soluble fraction of Acd mutant LexA protein expression.	124
Figure 4.6b: Comparison of Rosetta energy terms to the soluble fraction of Acd mutant LexA protein expression.	125
Figure 4.6c: Comparison of Rosetta energy terms to the soluble fraction of Acd mutant LexA protein expression.	126
Figure 4.6d: Comparison of Rosetta energy terms to the soluble fraction of Acd mutant LexA protein expression.	127
Figure 4.7a: Comparison of Rosetta energy terms to the total expression of Acd mutant LexA proteins.	128

Figure 4.7b: Comparison of Rosetta energy terms to the total expression of Acd mutant LexA proteins.....	129
Figure 4.7c: Comparison of Rosetta energy terms to the total expression of Acd mutant LexA proteins.....	130
Figure 4.7d: Comparison of Rosetta energy terms to the total expression of Acd mutant LexA proteins.....	131
Figure 5.1: Scheme for labeled protein production.....	139
Figure 5.2: Determination of Optimal Thermodynamic Constraint Value.....	166
Figure 5.3a: Effect of Constraints on Interresidue Distance Heat Maps.....	167
Figure 5.3b: Effect of Constraints on Interresidue Distance Heat Maps.....	168
Figure 5.4: Flory Scaling Plots.....	169
Figure 5.5: 0 M TMAO EFRET Plots	172
Figure 5.6: 2 M TMAO EFRET Plots.....	172
Figure 5.7: Diffusion Coefficients.....	174
Figure 5.8: Steady-state Fluorescence Spectra of Free Fluorophores.....	176
Figure 5.9: Determination of EFRET.....	177
Figure 5.10: Comparison of experimental and simulated data.....	178
Figure 5.11a: Intermolecular FRET Lifetime Decays and Fits.....	180
Figure 5.11b: Intermolecular FRET Lifetime Decays and Fits.....	181
Figure 5.11c: Intermolecular FRET Lifetime Decays and Fits.....	182
Figure 5.12: Autocorrelation plots from FCS Data.....	183
Figure 5.13: The effect of TMAO on the aggregation of wild type α S.....	185
Figure 5.14: Distance from Polymer-Scaled Förster Equation.....	186

Figure 5.14a: Global structural summary for 0 M TMAO ensembles.	189
Figure 5.14b: Global structural summary for 0 M TMAO ensembles.	190
Figure 5.15a: Global structural summary for 2 M TMAO ensembles.	191
Figure 5.15b: Global structural summary for 2 M TMAO ensembles.	192
Figure 5.15c: Global structural summary for 2 M TMAO ensembles.	193
Figure 5.16: Analysis of α S structural ensembles.	195
Figure 5.17: Contact maps from experimentally constrained ensembles.	197
Figure 5.18: Average interresidue distance heat map.	198
Figure 5.19: Experimentally constrained ensemble Flory scaling plot.	198
Figure 5.20: Experimentally constrained ensemble DSSP analysis.	199
Figure 5.21: Intensity ratios from PRE measurements.	201
Figure 5.22: Flory scaling plot comparison to previously published ensemble.	202
Figure 5.23: CD spectra of α S in varying concentrations of TMAO.	203
Figure 5.24: Aggregation kinetics of α S monitored by Congo Red.	204
Figure 5.25: Aggregation kinetics of α S monitored by fluorescence polarization	205
Figure 6.1: Determination of optimal new Rg score term weight.	240
Figure 6.2a: Plots of Percent Helicity.	247
Figure 6.2b: Plots of Percent Helicity.	248
Figure 6.3a: Histograms of Radii of Gyration.	249
Figure 6.3b: Histograms of Radii of Gyration.	250
Figure 6.4a: Comparison of Simulated EFRET with Experimental EFRET.	251
Figure 6.4b: Comparison of Simulated EFRET with Experimental EFRET.	252

Figure 6.5a: Comparison of Simulated Distance with Experimental Distances.	253
Figure 6.5b: Comparison of Simulated Distance with Experimental Distances.	254
Figure 6.6: Comparison of Simulated Beta_PP PRE Values and Experimental PRE Values.	255
Figure 6.7: Comparison of Simulated Beta_PPFI PRE Values and Experimental PRE Values.	256
Figure 6.8: Comparison of Simulated Beta_PPFISC PRE Values and Experimental PRE Values.	257
Figure 6.9: Comparison of Simulated Beta_PPSC PRE Values and Experimental PRE Values.	258
Figure 6.10: Comparison of Simulated CenStd_PP PRE Values and Experimental PRE Values.	259
Figure 6.11: Comparison of Simulated CenStd_PPFI PRE Values and Experimental PRE Values.	260
Figure 6.12: Comparison of Simulated CenStd_Ext_PP PRE Values and Experimental PRE Values.	261
Figure 6.13: Comparison of Simulated CenStd_Ext_PPFI PRE Values and Experimental PRE Values.	262
Figure 6.14: Comparison of Simulated CenNath PRE Values and Experimental PRE Values.	263

Figure 6.15: Comparison of Simulated SimAnn_PP PRE Values and Experimental PRE Values.	264
Figure 6.16: Comparison of Simulated SimAnn_PPF1 PRE Values and Experimental PRE Values.	265
Figure 6.17: Comparison of Simulated SimAnn_PPFISC PRE Values and Experimental PRE Values.	266
Figure 6.18: Comparison of Simulated SimAnn_PPSC PRE Values and Experimental PRE Values.	267
Figure 6.19: Comparison of Simulated VDW_PP PRE Values and Experimental PRE Values.	268
Figure 6.20: Comparison of Simulated VDW_PPF1 PRE Values and Experimental PRE Values.	269
Figure 6.21: Comparison of Simulated VDW_PPFISC PRE Values and Experimental PRE Values.	270
Figure 6.22: Comparison of Simulated FloppyTail_score12 PRE Values and Experimental PRE Values.	271
Figure 6.23: Comparison of Simulated FloppyTail_ref2015 PRE Values and Experimental PRE Values.	272
Figure 6.24: Comparison of Simulated FloppyTail_Quota PRE Values and Experimental PRE Values.	273
Figure 6.25: Comparison of Simulated FloppyTail_NoFrag PRE Values and Experimental PRE Values.	274

Figure 6.26: Comparison of Simulated FloppyTail_Loops PRE Values and Experimental PRE Values.	275
Figure 6.27: Comparison of Simulated FloppyTail PRE Values and Experimental PRE Values.	276
Figure 6.28: Comparison of Simulated FloppyTail_Relax PRE Values and Experimental PRE Values.	277
Figure 6.29: Comparison of Simulated FloppyTail_Rot PRE Values and Experimental PRE Values.	278
Figure 6.30: Comparison of Simulated FloppyTail_Rot_Relax PRE Values and Experimental PRE Values.	279
Figure 6.31: Comparison of Simulated FastFloppyTail PRE Values and Experimental PRE Values.	280
Figure 6.32: Comparison of Simulated FastFloppyTail PRE Values and Experimental PRE Values.	281
Figure 6.33: Comparison of Simulated AbInitio PRE Values and Experimental PRE Values.	282
Figure 6.34: Comparison of Simulated AbInitioVO PRE Values and Experimental PRE Values.	283
Figure 6.35: Comparison of Simulated DeNovoIDP PRE Values and Experimental PRE Values.	284
Figure 6.36: Comparison of Simulated Beta_PP and Experimental NMR Chemical Shift Data.....	285

Figure 6.37: Comparison of Simulated Beta_PPFI and Experimental NMR	
Chemical Shift Data.....	286
Figure 6.38: Comparison of Simulated Beta_PPFI SC and Experimental NMR	
Chemical Shift Data.....	287
Figure 6.39: Comparison of Simulated Beta_PP SC and Experimental NMR	
Chemical Shift Data.....	288
Figure 6.40: Comparison of Simulated CenStd_PP and Experimental NMR	
Chemical Shift Data.....	289
Figure 6.41: Comparison of Simulated CenStd_PPFI and Experimental NMR	
Chemical Shift Data.....	290
Figure 6.42: Comparison of Simulated CenStd_Ext_PP and Experimental NMR	
Chemical Shift Data.....	291
Figure 6.43: Comparison of Simulated CenStd_Ext_PPFI and Experimental	
NMR Chemical Shift Data.....	292
Figure 6.44: Comparison of Simulated CenNath and Experimental NMR	
Chemical Shift Data.....	293
Figure 6.45: Comparison of Simulated SimAnn_PP and Experimental NMR	
Chemical Shift Data.....	294
Figure 6.46: Comparison of Simulated SimAnn_PPFI and Experimental NMR	
Chemical Shift Data.....	295
Figure 6.47: Comparison of Simulated SimAnn_PPFI SC and Experimental NMR	
Chemical Shift Data.....	296

Figure 6.48: Comparison of Simulated SimAnn_PPSC and Experimental NMR Chemical Shift Data.....	297
Figure 6.49: Comparison of Simulated VDW_PP and Experimental NMR Chemical Shift Data.....	298
Figure 6.50: Comparison of Simulated VDW_PPF1 and Experimental NMR Chemical Shift Data.....	299
Figure 6.51: Comparison of Simulated VDW_PPFISC and Experimental NMR Chemical Shift Data.....	300
Figure 6.52: Comparison of Simulated FloppyTail_score12 and Experimental NMR Chemical Shift Data.....	301
Figure 6.53: Comparison of Simulated FloppyTail_ref2015 and Experimental NMR Chemical Shift Data.....	302
Figure 6.54: Comparison of Simulated FloppyTail_Quota and Experimental NMR Chemical Shift Data.....	303
Figure 6.55: Comparison of Simulated FloppyTail_NoFrag and Experimental NMR Chemical Shift Data.....	304
Figure 6.56: Comparison of Simulated FloppyTail_Loops and Experimental NMR Chemical Shift Data.....	305
Figure 6.57: Comparison of Simulated FloppyTail and Experimental NMR Chemical Shift Data.....	306
Figure 6.58: Comparison of Simulated FloppyTail_Relax and Experimental NMR Chemical Shift Data.....	307

Figure 6.59: Comparison of Simulated FastFloppyTail and Experimental NMR Chemical Shift Data.....	308
Figure 6.60: Comparison of Simulated FastFloppyTail_Relax and Experimental NMR Chemical Shift Data.....	309
Figure 6.61: Comparison of Simulated FloppyTail_Rot and Experimental NMR Chemical Shift Data.....	310
Figure 6.62: Comparison of Simulated FloppyTail_Rot_Relax and Experimental NMR Chemical Shift Data.....	311
Figure 6.63: Comparison of Simulated AbInitio and Experimental NMR Chemical Shift Data.....	312
Figure 6.64: Comparison of Simulated AbInitioVO and Experimental NMR Chemical Shift Data.....	313
Figure 6.65: Comparison of Simulated DeNovoIDP and Experimental NMR Chemical Shift Data.....	314
Figure 6.66a: Comparison of Simulated and Experimental Residual Dipolar Coupling Data.....	315
Figure 6.66b: Comparison of Simulated and Experimental Residual Dipolar Coupling Data.....	316
Figure 6.66c: Comparison of Simulated and Experimental Residual Dipolar Coupling Data.....	317
Figure 6.66d: Comparison of Simulated and Experimental Residual Dipolar Coupling Data.....	318

Figure 6.66d: Comparison of Simulated and Experimental Residual Dipolar Coupling Data.....	319
Figure 6.66e: Comparison of Simulated and Experimental Residual Dipolar Coupling Data.....	320
Figure 6.66f: Comparison of Simulated and Experimental Residual Dipolar Coupling Data.....	321
Figure 6.66g: Comparison of Simulated and Experimental Residual Dipolar Coupling Data.....	322
Figure 6.67: Comparison of Simulated Beta_PP and Experimental J-Coupling Data.....	323
Figure 6.68: Comparison of Simulated Beta_PPFI and Experimental J-Coupling Data.....	324
Figure 6.69: Comparison of Simulated Beta_PPFISC and Experimental J-Coupling Data.....	325
Figure 6.70: Comparison of Simulated Beta_PPSC and Experimental J-Coupling Data.....	326
Figure 6.71: Comparison of Simulated CenStd_PP and Experimental J-Coupling Data.....	327
Figure 6.72: Comparison of Simulated CenStd_PPFI and Experimental J-Coupling Data.....	328
Figure 6.73: Comparison of Simulated CenStd_Ext_PP and Experimental J-Coupling Data.....	329

Figure 6.74: Comparison with α -Synuclein PRE Data Comparison of Simulated CenStd_Ext_PPFI and Experimental J-Coupling Data.....	330
Figure 6.75: Comparison of Simulated CenNath and Experimental J-Coupling Data.....	331
Figure 6.76: Comparison of Simulated SimAnn_PP and Experimental J-Coupling Data.....	332
Figure 6.77: Comparison of Simulated SimAnn_PPFI and Experimental J-Coupling Data.....	333
Figure 6.78: Comparison of Simulated SimAnn_PPFISC and Experimental J-Coupling Data.....	334
Figure 6.79: Comparison of Simulated SimAnn_PPSC and Experimental J-Coupling Data.....	335
Figure 6.80: Comparison of Simulated VDW_PP and Experimental J-Coupling Data.....	336
Figure 6.81: Comparison of Simulated VDW_PPFI and Experimental J-Coupling Data.....	337
Figure 6.82: Comparison of Simulated VDW_PPFISC and Experimental J-Coupling Data.....	338
Figure 6.83: Comparison of Simulated FloppyTail_score12 and Experimental J-Coupling Data.....	339
Figure 6.84: Comparison of Simulated FloppyTail_ref2015 and Experimental J-Coupling Data.....	340

Figure 6.85: Comparison of Simulated FloppyTail_Quota and Experimental J-Coupling Data.....	341
Figure 6.86: Comparison of Simulated FloppyTail_NoFrag and Experimental J-Coupling Data.....	342
Figure 6.87: Comparison of Simulated FloppyTail_Loops and Experimental J-Coupling Data.....	343
Figure 6.88: Comparison of Simulated FloppyTail and Experimental J-Coupling Data.....	344
Figure 6.89: Comparison of Simulated FloppyTail_Relax and Experimental J-Coupling Data.....	345
Figure 6.90: Comparison of Simulated FastFloppyTail and Experimental J-Coupling Data.....	346
Figure 6.91: Comparison of Simulated FastFloppyTail_Relax and Experimental J-Coupling Data.....	347
Figure 6.92: Comparison of Simulated FloppyTail_Rot and Experimental J-Coupling Data.....	348
Figure 6.93: Comparison of Simulated FloppyTail_Rot_Relax and Experimental J-Coupling Data.....	349
Figure 6.94: Comparison of Simulated Ablnitio and Experimental J-Coupling Data.....	350
Figure 6.95: Comparison of Simulated AblnitioVO and Experimental J-Coupling Data.....	351

Figure 6.96: Comparison of Simulated DeNovoIDP and Experimental J-Coupling Data.....	352
Figure 6.97: Radii of Gyration of Ordered Proteins.....	354
Figure 6.98: Radii of Gyration of Partially-Ordered Proteins.....	355
Figure 6.99: Radii of Gyration of Disordered Proteins.	356
Figure 6.100: Folding Funnel Comparison for 1bka.....	357
Figure 6.101: Folding Funnel Comparison for 1bq9.	357
Figure 6.102: Folding Funnel Comparison for 1enh.	358
Figure 6.103: Folding Funnel Comparison for 1hz6.....	358
Figure 6.104: Folding Funnel Comparison for 1pgx.....	359
Figure 6.105: Folding Funnel Comparison for 1r69.	359
Figure 6.106: Folding Funnel Comparison for 1shf.....	360
Figure 6.107: Folding Funnel Comparison for 1ubi.....	360
Figure 6.108: Folding Funnel Comparison for 5cro.	361
Figure 6.109: Folding Funnel Comparison for 1b3a.	361
Figure 6.110: Folding Funnel Comparison for 1d7q.	362
Figure 6.111: Folding Funnel Comparison for 1ejf.....	362
Figure 6.112: Folding Funnel Comparison for 1fox.....	363
Figure 6.113: Folding Funnel Comparison for 1ghc.....	363
Figure 6.114: Folding Funnel Comparison for 1lwm.	364
Figure 6.115: Folding Funnel Comparison for 1vzs.	364
Figure 6.116: Folding Funnel Comparison for 2l42.....	365
Figure 6.117: Folding Funnel Comparison for 2lsu.....	365

Figure 6.118: Folding Funnel Comparison for asyn.....	366
Figure 6.119: Folding Funnel Comparison for dsh3.....	366
Figure 6.120: Folding Funnel Comparison for ntal.....	367
Figure 6.121: Folding Funnel Comparison for paaa.	367
Figure 6.122: Folding Funnel Comparison for sic1.....	368
Figure 6.123: Folding Funnel Comparison for tauk.....	368
Figure 6.124: 1bk2 Lowest RMSD Structures.....	369
Figure 6.125: 1bq9 Lowest RMSD Structures.	369
Figure 6.126: 1enh Lowest RMSD Structures.	369
Figure 6.127: 1hz6 Lowest RMSD Structures.....	370
Figure 6.128: 1pgx Lowest RMSD Structures.....	370
Figure 6.129: 1r69 Lowest RMSD Structures.	371
Figure 6.130: 1shf Lowest RMSD Structures.....	371
Figure 6.131: 1ubi Lowest RMSD Structures.....	371
Figure 6.132: 5cro Lowest RMSD Structures.	372
Figure 6.133: 1b3a Lowest RMSD Structures.	372
Figure 6.134: 1d7q Lowest RMSD Structures.	372
Figure 6.135: 1ejf Lowest RMSD Structures.....	373
Figure 6.136: 1fox Lowest RMSD Structures.....	373
Figure 6.137: 1ghc Lowest RMSD Structures.....	373
Figure 6.138: 1lwm Lowest RMSD Structures.....	374
Figure 6.139: 1vzs Lowest RMSD Structures.....	374
Figure 6.140: 2l42 Lowest RMSD Structures.....	375

Figure 6.141: 2lsu Lowest RMSD Structures.....	375
Figure 6.142: 1b3a Lowest RMSD Full Structures.....	376
Figure 6.143: 1d7q Lowest RMSD Full Structures.....	376
Figure 6.144: 1ejf Lowest RMSD Full Structures.....	376
Figure 6.145: 1fox Lowest RMSD Full Structures.....	377
Figure 6.146: 1ghc Lowest RMSD Full Structures.....	377
Figure 6.147: 1lwm Lowest RMSD Full Structures.	377
Figure 6.148: 1vzs Lowest RMSD Full Structures. 10	378
Figure 6.149: 2l42 Lowest RMSD Full Structures.....	378
Figure 6.150: 2lsu Lowest RMSD Full Structures.....	378
Figure 6.151: Asyn Lowest RMSD Full Structures.....	379
Figure 6.152: Dsh3 Lowest RMSD Full Structures.	379
Figure 6.153: Ntal Lowest RMSD Full Structures.	379
Figure 6.154: Paaa Lowest RMSD Full Structures.	380
Figure 6.155: Sic1 Lowest RMSD Full Structures.....	380
Figure 6.156: Tauk Lowest RMSD Full Structures.....	380
Figure 7.1: Exemplar-based in-silico screening.....	395
Figure 7.2: Radioligand Structure and Initial Binding Assay.....	396
Figure 7.3. Molecular structures of 17 compounds from the Exemplar screening.	397
Figure 7.4: Site 2 Exemplar from PDB 2N0A fibril.	398
Figure 7.6: Site 9 Exemplar from PDB 2N0A fibril.	400
Figure 7.7: Top Hits from ZINC15 database for Site 9 Exemplar.....	401

Figure 7.8: Photocrosslinking of radioligand analog to α S fibrils.	403
Figure 7.9: Impact of identified molecule on α S fibril aggregation and stability.	404
Figure 7.10: SAR screen based on exemplar compound (6).	406
Figure 7.11. Molecular structure of the 39 compounds from the similarity search.	407
Figure 7.12: Synthesis route towards 61 and [125I]61.	409
Figure 7.13: Characterization of the lead compound.	410
Figure 7.14: In vitro autoradiography on mouse brain tissue sections to assess [125I]61 binding.	411

CHAPTER 1: INTRODUCTION

§ 1.1 The Role of Amyloidogenic Proteins in Neurodegenerative Disease

Introduction to Disordered Proteins

Proteins are macromolecular biopolymers that carry out a variety of tasks that give rise to cellular function. Each protein carries out a specific function or set of functions which is often correlated with its structure. The specific composition of amino acids and their arrangement, termed primary structure, allows for various secondary structures, three-dimensional structural elements which dictate the overall topology of a protein. Often, proteins can adopt multiple conformations, each required for a specific task, and can move between these states to serve several functional roles.¹ Despite the longstanding view that a protein's function is tied to its structure via a specific sequence of amino acids, the emergence of disordered proteins over the past two decades has urged this canonical model to be refined.² Intrinsically disordered proteins (IDPs) do not readily adopt a defined architecture, but instead populate a variety of folds and conformations.²⁻⁴ This ensemble of states is usually highly sensitive to stimuli, including changes in pH, ionic strength and the presence of protein partners.³ These can modulate the overall size, or radius of gyration, of a disordered protein and influence the propensity of intramolecular contact formation.³ Furthermore, in specific cases, these stimuli have been shown to elicit structural transitions by which a canonically disordered protein adopts a stable fold.²⁻⁴ Partially ordered proteins, which contain both well folded, ordered, domains in addition to intrinsically disordered regions (IDRs) also exist with many combinations of ordered and disordered domains in a single protein.²⁻⁴ Ultimately, IDPs and IDRs have been demonstrated to serve a

variety of both physiological and pathogenic roles over the past two decades, which has prompted their study and encouraged a more robust comprehension of how their structural lability gives rise their functionality.²⁻⁴

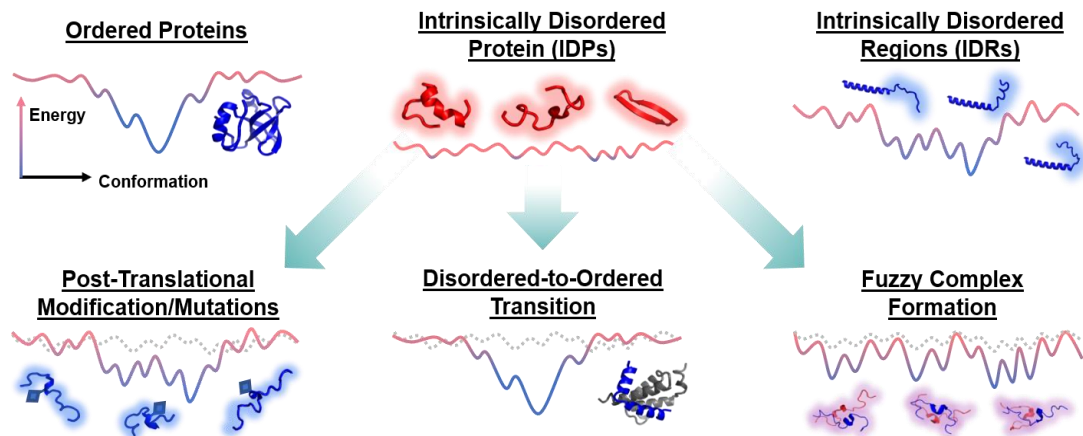


Figure 1.1: Structural differences between disordered and ordered proteins. Top: Differences between the structure and rigidity of ordered and disordered proteins as described by conformational energy landscapes. Bottom: Modulations of disordered protein ensembles.

Functional Role of Disorder

IDPs and IDRs serve a variety of physiological functions in the cell which are reliant on their ability to interact with several partners, respond to stimuli and stabilize functional conformations.² Recently, as much as 30% of the eukaryotic proteome and 50% of the human proteome have been hypothesized to be partially or fully disordered.⁵⁻⁶ As structural components, IDRs present sites on a protein surface that can be modified by post-translational modifications (PTMs) or serve as spacers between folded domains of a protein.^{2-3, 7} IDRs have been long recognized in histones for their ability to be modified by a host of PTMs which elicit

various functional and signaling consequences.⁷ Furthermore, a variety of IDPs contain molecular recognition features (MoRFs), allowing a disordered region, upon interaction with a protein or DNA partner, to transition from a disordered to an ordered state.³ For example, many transcription factors adopt helical folds when interacting with their partner DNA motif but are otherwise disordered when not complexed with DNA.⁴ Additionally, conformational ordering is also not required for interaction and many IDPs and IDRs interact through the formation of fuzzy complexes.³ These highly dynamic complexes, which have been theorized to be held together by charge complementarity and cation- π interactions, give rise to the formation of phase condensates and protein hubs.⁸⁻¹⁰ Despite these dynamics, these complexes can exist with exceptionally high affinities such as the prothymosin α – histoneH1 interaction which has a sub-nanomolar binding affinity.¹¹ Functionally, these complexes serve as the basis for many important cellular processes including transcription, DNA repair, signal transduction and assembly of the ribosome.¹²⁻¹⁴ Therefore, a more robust understanding of the underlying structure and dynamics of IDPs and IDRs will allow us to not only to better understand which architectural aspects primes them to perform these duties, but will likely also elucidate how mutations and misregulation can lead to disease.

Misfolding in Disease

Although the conformational fluidity of disordered proteins allows them to perform a wide range of tasks, this often results in the exposure of hydrophobic patches. This exposure makes this class of proteins notably aggregation prone

and amyloidogenic.⁵ Proteins fold into their canonical active structures sporadically or with the assistance of chaperones and other molecular machinery.¹ However, the folding process is tightly regulated as adoption of an incorrect topology can not only disrupt function but can serve as the basis of disease.⁵ Misfolding, or the adopt of an incorrect and sometimes pathogenic structure is often associated with protein aggregation which is implicated in several neurodegenerative diseases.^{5, 15} Prion diseases, including bovine spongiform encephalopathy, or mad cow disease, and Creutzfeldt-Jakob disease, are likely the most widely recognized protein misfolding diseases.¹⁵ Ingestion or exposure to brain tissue from a specimen containing misfolded prion protein leads to further misfolding of the prion protein endogenous to the patient, resulting in disease propagation.¹⁵ Much like prions, amyloids are proteins whose topologies allow for the template-based recruitment of other copies of the protein, generally resulting in the formation of a β -sheet rich architecture known as an amyloid fibril.¹⁶ These fibrils, along with the β -sheet rich oligomers that can form on- and off-pathway to fibril formation, are cytotoxic and have been implicated in a variety of diseases.^{5, 16} In neurodegeneration, amyloid fibrils consisting of the Huntington protein, amyloid β , α -synuclein (α S) and tau have been implicated in Huntington's Disease, Parkinson's Disease (PD), Alzheimer's Disease and Chronic Traumatic Encephalopathy (CTE), respectively.^{5, 16} Moreover, non-neurodegenerative diseases such as Type II Diabetes, associated with islet amyloid polypeptide, and cardiac amyloidosis, associated with transthyretin, are also caused by the aggregation of amyloidogenic proteins.¹⁶⁻¹⁸ Although it is well understood that the

aggregation of these proteins leads to disease, the exact mechanisms of their pathology and the processes that trigger their aggregation are not in many cases elucidated.

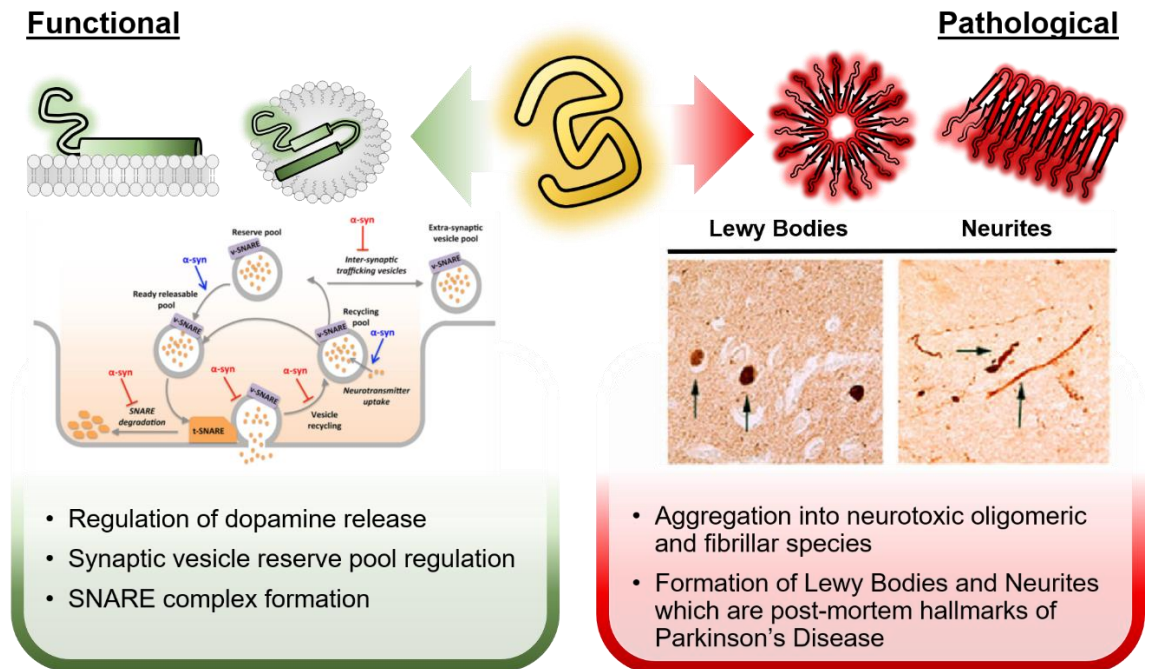


Figure 1.2: Overview of the functional and pathological roles of α S. Top: Cartoon representations of (left) helical morphologies associated with the membrane bound forms, (middle) the disordered state and (right) toxic oligomeric and fibrillar forms of α S. Middle: (left) Schematic representation of the functional role of α S and (right) histological staining of LBs and LNs. Figures adapted from Laushel et. al.¹⁹ Bottom: Bullet point summaries of the functional (left) and pathological (right) roles of α S

Physiological Role of Alpha-Synuclein

α S is member of a protein family consisting of three members and is a 140 amino acid long disordered protein.¹⁹⁻²⁰ It consists of three domains, the first being a 60 residue N-terminal domain (NTD) containing several imperfect 11 amino acid

amphipathic, KTKEGV motif, repeats.²¹ This domain is largely conserved across the other two synuclein family members, β - and γ -synuclein, and adopts a helical conformation when associated with membranes and vesicles.²¹ Following the NTD, the central domain of the protein, comprised of residues 61-94, is referred to as the non-amyloid β component of Alzheimer's disease amyloid plaque (NAC) domain.¹⁹⁻²¹ This name originates from the identification of fragments of this domain within plaques comprised primarily of amyloid β in the brains of Alzheimer's Disease patients.¹⁹⁻²⁰ The largely hydrophobic NAC domain has been shown to be essential for α S aggregation and adopts a variety of β -sheet rich architectures in the oligomers and fibrils.¹⁹⁻²¹ The final C-terminal domain (CTD) of α S boasts a significant enrichment in acidic residues and is found to be consistently disordered, regardless of the structure of the NTD or NAC domain.¹⁹⁻²¹

Despite its long-recognized role in the pathogenesis of PD, the functional role of α S has remained far more elusive.²² α S is primarily localized at the synaptic termini of neurons and is purported to serve a variety of functions and interact with several protein partners.¹⁹⁻²⁰ Although α S is primarily disordered in the cytosol, single-molecule FRET, nuclear magnetic resonance (NMR) and electron paramagnetic resonance (EPR) studies have revealed that α S can adopt a variety of helical conformations when interacting with membranes and vesicles.²³⁻²⁵ Moreover, these conformations are modulated by the curvature of the membrane.²⁴ On high curvature vesicles, α S adopts a bent helix conformation consisting of two helical stretches connected by a disordered loop, and on low curvature membranes the protein adopts a single extended helix.²⁴ Additionally,

the final C-terminal third of the protein remains disordered, regardless of the specific helical conformer.²³⁻²⁵ In the helical membrane-bound state, α S contributes to synaptic plasticity, vesicle recycling, neurotransmitter release and other phenomena.¹⁹⁻²⁰ Soluble *N*-ethylmaleimide-sensitive factor Attachment Protein Receptor (SNARE) proteins facilitate membrane fusion which leads to neurotransmitter release at synapses.²⁶ *In vitro* and cellular studies have revealed that interactions between α S and SNARE proteins, specifically VAMP-2 and synaptobrevin, promote SNARE complex formation and vesicle docking.²⁷⁻²⁹ These studies, along with knockouts of α -, β -, and γ -synuclein, suggest that α S is a nonessential component of the larger set of machinery responsible for vesicle reserve pool regulation and neurotransmitter release.¹⁹

Role of Alpha-Synuclein in Parkinson's Disease

Insoluble deposits, comprised of fibrillar α S, found in the brain have long served as a post-mortem hallmark of PD. These insoluble deposits, known as Lewy Bodies (LBs) and Lewy Neurites (LNs), are found primarily in the substantia nigra pars compacta of PD patients.^{19,22} LBs and LNs are linked with dopaminergic neuronal death which leads to the well-recognized clinical presentations of the disease.³⁰ Aggregates of α S are theorized to spread from this initial location throughout the brain over the course of disease progression.³¹⁻³² Additionally, α S aggregates have been associated with other neurodegenerative diseases, termed synucleinopathies.³³ α S aggregates take the form of LBs in dementia with Lewy Bodies, however in other synucleinopathies, such as multiple systems atrophy

(MSA), aggregates present as glia cell inclusion (GCIs).³⁴ These different aggregate morphologies are often associated with different regions of the brain, as GCIs are often found in oligodendrocytes of white matter tracts.³⁴ Although the presence of α S containing plaques is long recognized, there still exists some debate as to whether these features are causative or emerge later in disease progression, playing a potentially protective role.³⁵ However, direct administration of fibrillar α S prepared *in vitro* to cultured neurons, or intrastrially injected into the brains of model mice, has demonstrated that the presence of fibrillar α S is sufficient to elicit disease phenotypes associated with PD.³⁶⁻³⁸ Lastly, several familial mutations within the gene responsible for α S production have been identified with increased risks for PD.³⁹⁻⁴⁶ These mutations either increase the copy number of α S, thus enhancing aggregation propensity, or result in alteration of the protein sequence.³⁹⁻⁴⁰ Of the latter, five mutations to date (A30P, E446K, H50Q, G51D, and A53T) have been observed in patients and greatly increase their risk for PD.⁴¹⁻⁴⁶ These results together substantiate the need to better understand the mechanisms behind the aggregation of α S and other amyloidogenic proteins and their role in pathogenesis.

Impetus for Dissertation Studies

Despite the long-recognized role of α S aggregates in PD pathogenesis, structural and mechanistic understandings of α S aggregation and toxicity have been lacking. This gap in knowledge has not only hindered our comprehension of disease progression and onset but has also limited our ability to therapeutically

target PD or monitor the disease using molecular probes. Since the start of the work presented herein, solid-state NMR, and most impactfully cryo-electron microscopy (cryoEM), have emerged as powerful techniques for capturing the complex architectures associated with the fibrillar forms of α S.⁴⁷⁻⁵⁰

The work presented herein focuses on combining computational and experimental approaches to understand the unfolded state of α S and how the unfolded conformational ensemble can be modulated. Initially, this work highlights the development of novel minimally perturbing fluorophores, which can be employed in a variety of fluorescence-based assays.⁵¹⁻⁵² Subsequently, in-line with the long-term goal, this thesis work demonstrates how fluorescence spectroscopy can be leveraged to gain sufficient information to generate accurate experimentally constrained molecular models of the disordered state of α S.⁵³ These initial efforts led not only to improved experimental methods, but to the development of new computational approaches, capable of accurately predicting the structure of both ordered and disordered protein systems from their sequence. Lastly, this work explores the use of in silico screening methods, which use the recently available fibrillar structure of α S as inputs, to generate compounds which can be useful as radiolabelled imaging agents. Overall, the work described herein consists of a variety of experimental and computational methods designed to gain better mechanistic and structural understanding of disordered and aggregation prone proteins.

§ 1.2 The Utility of Fluorescence and FRET in Studying Protein Structure

Several different methods have been employed to study the conformations of proteins and track protein dynamics. Most of our understanding of protein structure to date has come from x-ray crystallography, NMR and, more recently, cryo-EM.^{1-2, 54-55} However, these techniques are incapable of producing structures for highly dynamic and heterogeneous systems and are generally not well-suited to cellular studies.^{2, 54, 56} Alternatively, fluorescence-based experiments boast both the ability to be performed in cells and can be used to monitor complex systems.⁵⁷ However, these advantages are juxtaposed by a steep reduction in the amount of structural detail obtained. Therefore, fluorescence-based experiments have proven as a highly useful complement to these more data rich approaches for watching protein systems in action.

Overview of Fluorescence and FRET

Fluorescence occurs when a photon interacts with a molecule resulting in the emission of a photon of a different wavelength from the molecule. As depicted in Fig. 1.3, a molecule can absorb the energy of an incident photon provided the wavelength of the light matches the energy of allowed molecular electronic/vibrational transitions. This places the molecule into an excited state. Once in an excited state, a portion of the energy is often dissipated through various non-radiative decay processes, which are often associated with molecular motion. Ultimately, fluorescence occurs when the molecule returns to the ground state, emitting the remaining energy as a photon, which is now shifted in wavelength from

the initial absorbed photon. This shift is known as the Stokes Shift, and is principally dictated by the vibrational modes of the molecule but can be affected by the molecule's surrounding environment. In addition to fluorescence, a variety of relaxation processes can result in the dissipation of energy, returning a molecule in the excited state to the ground state. By transferring energy into the surrounding environment, or through intersystem crossing, where the molecule enters the triplet state via an electronic spin flip, a molecule can relax to the ground state in the absence of fluorescence. The probability with which a photon absorbed by a molecule results in a fluorescent event is termed the quantum yield and is governed by the rates of radiative and non-radiative decay intrinsic to the molecule. Additionally, the fluorescence or excited state lifetime of a molecule is the time the molecule spends in the excited state, following an initial excitation, before emitting a photon. Much like the Stokes shift, the lifetime and quantum yield associated with a molecule's fluorescence can be modulated by the molecule's surrounding environment.⁵⁷

One photophysical phenomenon that can greatly enhance the amount of information which can be gleaned from fluorescence-based experiment is Förster Resonance Energy Transfer (FRET). Initially described by Theodor Förster in 1948, FRET involves the distance-dependent resonance energy transfer between two molecules.⁵⁸ During FRET, a donor fluorophore in the excited state is quenched, or returned to the ground state, by non-radiatively transmitting energy to an acceptor chromophore. Two major criteria dictate whether two molecules can be in resonance and therefore participate in FRET. First, the specific three-

dimensional orientation of the molecules' absorption and emission dipoles governs the dipole coupling. Second, there must exist an energetic overlap between the transition energy associated with the donor's relaxation and the excitation energy of the acceptor. This is most easily visualized as an overlap between the donor fluorophore's emission spectra and the acceptor chromophore's absorbance spectra, as is shown in Fig. 1.3. The energy transfer probability, also referred to as the FRET efficiency, has a sixth-order dependency on the intermolecular separation.⁵⁷⁻⁵⁸ Molecular pairs are often defined by their Förster radius, or distance at which the FRET efficiency is one-half, which is largely ascribed to the wavelengths associated with the pair's spectral overlap. These dependencies are mathematically defined in the latter part of this thesis.⁵⁷

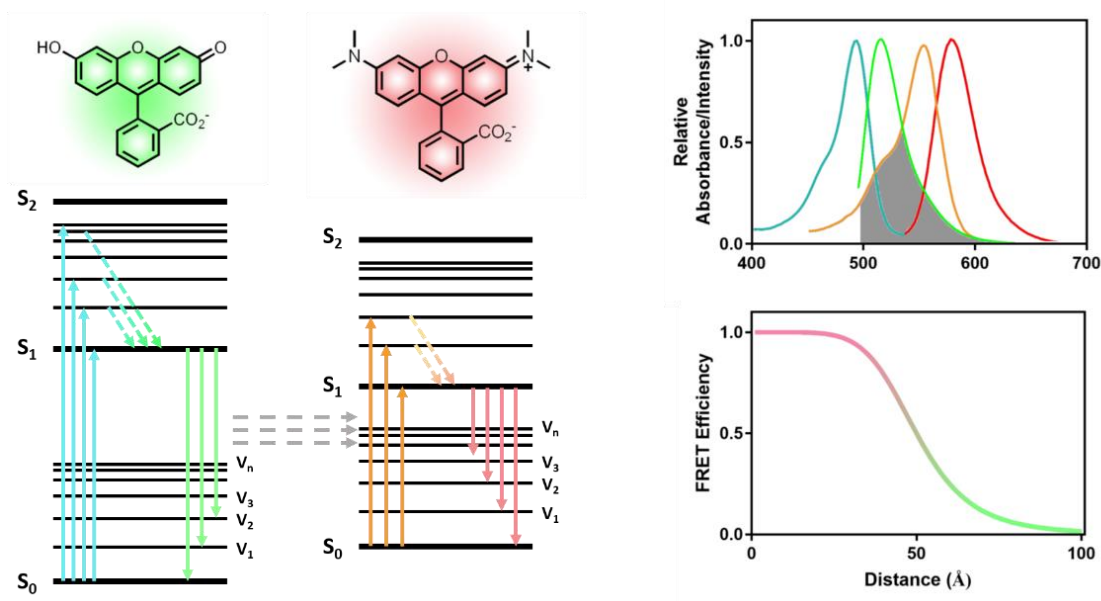


Figure 1.3: Overview of Fluorescence and FRET. Left: Jablonski diagram depicting vertical transitions associated with absorbance (upward arrows) and fluorescence (downward solid arrows) along with non-radiative relaxations (downward dashed arrows) and FRET (dashed grey arrows) between a donor (fluorescein, chemical structure, green) and acceptor (tetramethylrhodamine, chemical structure, red) fluorophore. Right Top: Absorbance (blue: fluorescein, orange: tetramethylrhodamine) and fluorescence (green: fluorescein, red: tetramethylrhodamine) spectra corresponding to Jablonski diagram transitions. Right Bottom: Plot depicting the relationship between FRET efficiency and inter-probe difference, with the coloration representing the associated fluorophore emission.

Description of FRET's use in Monitoring Conformational Changes

The sensitivity of FRET, coupled with the ability of fluorescence-based experiments to be employed in a wide variety of contexts, has supported its employment to observe the structure and motions of various biopolymers (DNA, RNA, proteins). FRET had been used to track changes in protein conformation, assembly and disassembly of protein complexes, interactions between protein

partners and many other phenomena.⁵⁹ Recently, FRET has become a useful modality for studying disordered proteins, as various features associated with the heterogeneity of these systems preclude their study by several other methods.⁶⁰ X-ray crystallography, which is reliant on the formation of protein crystals, requires sampling to be highly homogenous thereby eliminating it as an option for studying disordered systems.⁵⁵⁻⁵⁶ Although cryo-EM is inherently a single-molecule technique, where data from discrete molecules can be combined, the features required for robust particle averaging prevent it from reporting on disordered structural ensembles.⁵⁴ Furthermore, in solution-phase NMR, the multiplicity of conformers, coupled with the lack of environmental distinction between residues, results in significant signal broadening and overlap.² In most cases this makes assignment of chemical shifts to individual residues difficult and, when possible, does not deliver as detailed information as would be acquired for a well-folded protein.^{2, 56} Like FRET, EPR and paramagnetic relaxation enhancement (PRE) can be used to obtain inter-residue distance information.^{3, 61} PRE experiments measure distances between a single residue and all other residues in a protein.⁶¹ Although this technique delivers significantly more data than FRET, complexities assigning chemical shifts and a working distance of $< 30 \text{ \AA}$, which would likely not capture the more extended structural conformations generally associated with disordered systems, hinder application of this technique.⁶¹ EPR provides information about the distance between two spin labeled residues on a protein along with the distribution of distances populated.⁶¹ However, much like other techniques, the significant heterogeneity and breadth of the distance distribution

between spin labels results in inaccuracies in analysis.⁶¹ Lastly, all of these techniques are either inherently *in vitro* methods or are extremely challenging to implement in a cellular context.^{55, 62} Therefore, FRET's ability to capture pairwise, inter-residue distance information, though seemingly limited, is one of the few approaches well suited to track changes in IDP conformation.

Application of FRET to IDPs

Ensemble FRET, and more often single-molecule FRET (smFRET), has been used to study a variety of disordered and unfolded protein systems.^{59-60, 63} By combining the photophysical phenomena of FRET with time-correlated single photon counting (TCSPC) and the use of microscopes, FRET measurements can be made on molecules one at a time, allowing subpopulations, which are often lost in ensemble experiments, to emerge.⁶⁰ This has been used to measure the overall size, or radius of gyration, of disordered and unfolded systems and track the effects of PTMs, mutations and environmental factors on protein size.⁵⁹⁻⁶⁰ Despite several years of debate, the accuracy of these measurements has recently been corroborated by small-angle x-ray scattering (SAXS), following technical and analytical improvements to both methods.⁶⁴⁻⁶⁶ Additionally, since smFRET-based techniques can be performed on millisecond timescales, conformational switching and other dynamic processes can be observed.⁵⁹ For example, the previously described folding phenomena associated with the interaction of α S with lipid membranes has been tracked by smFRET, as has the formation of oligomeric species en-route to fibril formation.^{24, 67} Lastly, smFRET has been utilized in a

variety of protein unfolding studies, which many researchers have used as models of disordered systems.⁶⁰

Despite the utility of FRET in studying disordered systems, the heterogeneity associated with IDPs and IDRs presents unique hurdles for analyzing FRET data.⁶⁰ Traditionally, the interpretation of FRET data for folded proteins is hindered by the assumption that the attached fluorophores are capable of diffusing isotropically.⁵⁷ This presents a challenge in accurately determining the distance between the two probes, due to inaccuracies in the computed orientation factor which impacts the transfer efficiency between two probes. However, FRET for disordered systems suffers from the opposite problem. Although these highly dynamic proteins relieve concerns associated with the orientation factor, they give rise to a new concern regarding the implicit assumption in the canonical Förster equation that the fluorophores are separated by a single distance.^{60, 65} This is not a concern for folded proteins where breadth of the distribution of any inter-residue distance is negligible but presents a clear problem for proteins which populate multiple conformations or whose populations span a significant distribution of distances.^{57, 60, 65} In part, this issue can be combatted by performing smFRET, in place of ensemble FRET, which can report on the presence of distinct subpopulations.⁵⁹⁻⁶⁰ However, this does not fully alleviate the problem as the millisecond measurement timescale is not rapid enough, nor is the signal-to-noise generally high enough, to capture information regarding the distribution breadth of sub-ensembles.⁶⁰ Therefore, most researchers utilize a modified version of the Förster equation, shown in detail in the latter parts of this thesis, adding a

probability distribution to the traditional equation.^{60, 65} Although this appears to remedy the issue, allowing a single FRET efficiency to be correlated with a distribution of states, these probability distributions often need to be supplied by the researcher.^{60, 65} Due to the topological similarities between disordered proteins and polymers, distributions from polymer physics are often employed.^{60, 65, 68} The work presented herein demonstrates how the use of multiple FRET probe pairs can facilitate distribution selection and how these distributions and FRET data can be utilized to construct molecular models of disordered proteins.

Development of Novel, Minimally Perturbing Fluorescent Probes

Depending on the experimental requisites, the effectiveness of most FRET experiments relies on the choice of appropriate fluorophores. Quantities such as quantum yield and extinction coefficient inform the overall “brightness” of the molecule, which is important in cellular and single-molecule experiments.⁶⁹ Additionally, the specific spectral range over which the pair of molecules absorb and fluoresce, not only informs their Förster Radius and ability to serve as a FRET pair, but also allows the measurement to circumvent any background signal which can complicate the measurement.⁵⁷ However, beyond the photophysical characteristics of the molecule, the overall molecular size, composition and attachment site may hinder the measurement efficacy by perturbing the system.⁷⁰⁻
⁷¹ Many cell-based experiments, as well as some *in vitro* experiments, leverage the use of fluorescent fusion proteins, such as green fluorescent protein (GFP).⁷² Although these are currently the field standard, these fluorescent proteins often

lack the photophysical criteria optimal for most experiments.⁷³ Moreover, their size is typically on the order of the size of the systems they are used to measure. This often significantly complicates the interpretation of the measurement or can perturb the systems, preventing the biomolecule's normal function. Ultimately, several groups have focused on replacing the use of GFP with Halo- and SNAP-tag fusion proteins, which can be subsequently specifically labeled with fluorophores that boast improved photophysical characteristics to enhance measurement quality.⁷²⁻⁷⁵ However, these fusion proteins still rival, and in some cases surpass, the size of GFP, making the potential for perturbation non-trivial.^{71, 74-75}

With this in mind, the Petersson Laboratory has focused on developing fluorescent probes which are minimally perturbing to the protein conformation and dynamics. This has been accomplished by focusing on small fluorescent scaffolds which can be modified by appending molecular handles, lowering the barrier for their use, and derivatizing these scaffolds to improve their photophysical characteristics.^{52, 76-78} To date, the bulk of this effort has been focused on the use of thioamides, single atom oxygen-to-sulfur substitutions, as fluorescence quenching motifs.⁷⁹⁻⁸¹ Additionally, the group has developed a derivative of acridone, acridon-2-lyalanine (Acd), which allows this small fluorophore to be directly incorporated within the primary structure of a protein via unnatural amino acid (UAA) mutagenesis.⁷⁷ These two efforts, along with the derivatization of other base scaffolds such as dimethylaminoquinoline (DMAQ), have been supplemented by identifying additional fluorophores which can serve as FRET donors or acceptors for these probes.^{51-52, 76-77} These developments, along with

the employment of UAA mutagenesis, are described in part in this thesis work and have provided several novel fluorescent tools which are minimally perturbing to protein structure.

Methods for Introducing Fluorescent Probes (UAA Mutagenesis)

One of the major techniques used in this work to generate fluorescently labeled protein constructs is unnatural amino acid (UAA) mutagenesis. This method is used herein to incorporate both natively fluorescent amino acids and amino acids that serve as biorthogonal handles for chemical labeling with synthetic dyes.^{51, 53, 78} The utility of this technique is based on the concept that by hijacking the native machinery of the cell, one can ensure precise labeling of a protein with a given probe by encoding it at the DNA level. Although more non-specific approaches are also used herein (cysteine labeling, tryptophan fluorescence, etc.) they are exclusively employed in systems where they can be leveraged site-specifically.^{53, 78}

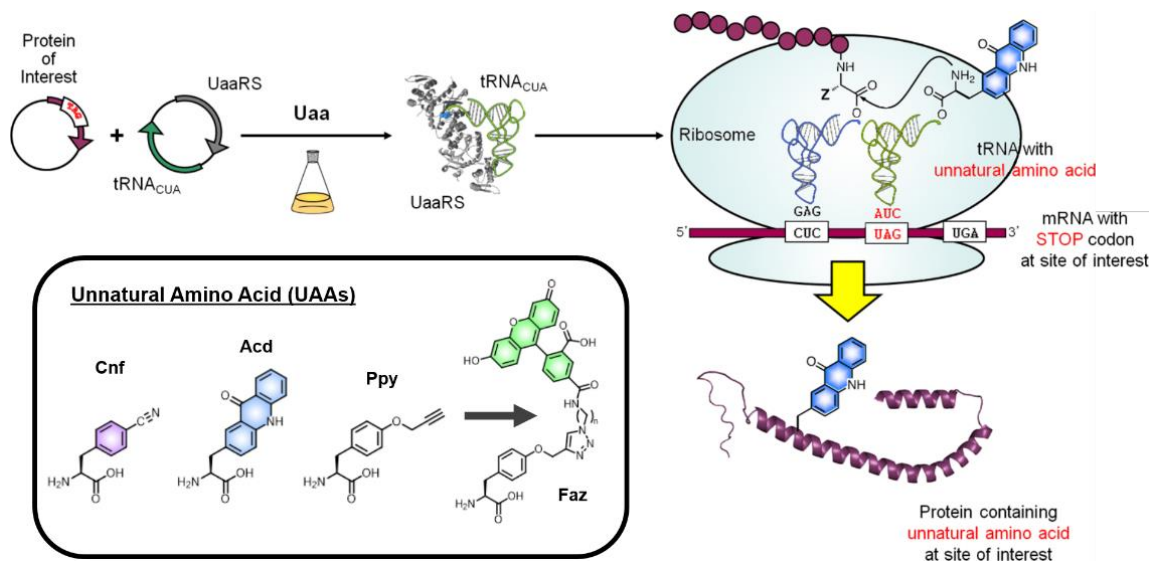


Figure 1.4: Overview of unnatural amino acid mutagenesis. Incorporation of two plasmids, one containing a protein of interested with a TAG codon and a second containing a UAA specific synthetase (UAARS) and cognate tRNA (tRNA_{CUA}) which recognizes the TAG codon, allows for site specific incorporation of an unnatural amino acid (UAA) which is supplied directly in the cellular media. Boxed: Depiction of some of the UAA used in this work which are endogenously fluorescent (Cnf, Acd) and others (Ppy) that can be labeled after being incorporated within a protein of interest.

Site-specific UAA mutagenesis was first reported by Peter Schultz and colleagues in 1989 using *in vitro* methods, and its current implementation relies on the incorporation of new machinery into the cell that has been evolved to be orthogonal to the existent components.⁸²⁻⁸⁵ During protein synthesis, each amino acid is recognized by an aminoacyl tRNA synthetase (aaRS) which is responsible for charging this amino acid to a specific tRNA.⁸²⁻⁸³ This process allows the genetic code to be converted into a protein sequence by assuring that each triplet of bases in mRNA, which is transcribed from the DNA, is read by a single tRNA that carries, and results in the translation of, a single amino acid.^{1, 83} Therefore, by encoding for incorporation of a fluorescent amino acid at the DNA level, and developing the

appropriate orthogonality of the requisite machinery, we can ensure specific incorporation of the desired probe at an exact location in a protein of interest.⁷⁸

Due to the requisite specificity, development of the aaRSs utilizes directed evolution and a new aaRS often needs to be generated for each new amino acid derivative.⁸⁶⁻⁸⁷ The directed evolution process combines positive screens, where aaRSs which can charge the appropriate tRNA are selected for and retained.⁸⁸ Subsequently, negative selections are performed where aaRSs are screened for their specificity for the desired UAA against all native or undesired amino acids.⁸⁸ Lastly, due to the multiplicity of effective orthogonal synthetases developed to date, many efforts simply screen new UAAs against libraries of existing aaRSs.⁸⁹

Although other methods exist such as the use of auxotrophic cell lines, which lack a specific amino acid, or the employment of quadruplet codons, most UAA incorporation efforts rely on suppression of an amber stop codon.⁹⁰⁻⁹¹ The amber stop codon is endogenously recognized by release factor 1 (RF1), triggering release of the growing polypeptide chain from the ribosome.⁹² Therefore, binding of the appropriate charged tRNA to mRNA on the ribosome is always in competition with RF1, resulting in a mixture of truncated protein and full-length protein containing the desired UAA.⁹²⁻⁹³ To circumvent this issue, researchers have developed RF1 knockout cell lines, and Church and colleagues have engineered an *E. Coli* cell line which both lacks RF1 and has a full genomic replacement of amber stop codons with ochre (UAA) stop codons.⁹³⁻⁹⁴ However, Petersson and colleagues developed an alternative strategy to ensure that following protein expression, only purification of the desired, full-length protein

occurs.⁹⁵ This method leverages Ni-affinity chromatography procedures via the fusion of a C-terminal intein-His6 fusion to the protein of interest. This allows for isolation of the protein by affinity chromatography exclusively in the situation where the amber stop codon is effectively suppressed and an amino acid is incorporated. Subsequently, the intein can undergo self-excision which is triggered by the addition of β -mercaptoethanol, resulting in the generation of the native C-terminal carboxylic acid at the end of the protein of interest.⁹⁵⁻⁹⁶ Overall, the methods for developing and introducing minimally perturbing fluorescence probes have provided the necessary constructs for all of the fluorescence-based experiments described herein.

§ 1.3 Computational Methods for Modeling Disordered Proteins

Since the conformational ensembles populated by disordered proteins have remained elusive and no clear single experimental approach has been identified that can report on the multitude of states, a significant effort has been put forth to supplement experimental data with computational simulations. Initially, structural ensembles were generated from self-avoiding walk and other polymer-physics based simulations. By utilizing experimental data from PRE, residual dipolar coupling and other NMR-based experiments, models containing little to no chemical information could be refined to match input data. These initial random-coil simulations generally treated proteins as self-avoiding polymers with bonding geometries serving as the only chemical consideration. Since these initial developments, a significant amount of effort has been devoted to developing both

molecular dynamics (MD) and Monte Carlo (MC) based methods to provide improved platforms that accurately capture disordered systems.

Description of previous efforts in Molecular Dynamic

MD methods, which utilize Newtonian physics to simulate molecular motions, are among the most widely employed for the study of protein dynamics. Although the force-fields and atomic parameters have been refined over many years to accurately capture protein motions in folded protein systems, initial simulations of disordered structures were far from accurate.⁹⁷⁻¹⁰⁰ These inaccuracies generally manifest as an over-population of secondary structural elements, usually helices, and over-compaction of the protein.⁹⁷⁻¹⁰⁰ Several groups, including Shaw and colleagues, focused on developing not only new force-fields but also new water models, which vastly improved the state-of-the-art.^{97, 99, 101} Despite these improvements, the required hardware and compute times for MD simulations present a high barrier to entry, limiting the use of these methods to short trajectories unless elaborate computational resources are employed.¹⁰²⁻¹⁰³ Furthermore, simulations that predict the fold of a protein in its native three-dimensional architecture have been largely intractable in MD simulations outside of a limited number of small, fast-folding proteins.⁹⁹ Overall, the requisite for sophisticated hardware, along with an inability to produce accurate folded structures from sequence, limits the utility of MD for predicting protein structure despite being the current state-of-the-art for disordered systems.

Alternatives to Molecular Dynamics

Outside of MD, there have been similar efforts to update existing tools as well as develop new packages for handling disordered proteins. In addition to the initially described, largely polymer-based, modeling approaches used to refine ensembles from experimental data, there have been several efforts to develop methods which adequately capture the complexities of various IDPs.¹⁰⁴⁻¹⁰⁸ Several packages, such as AWSEM-IDP, have developed improved coarse-grained force-fields and sampling approaches to circumvent the computational time and hardware requirements for sufficient sampling.¹⁰⁹⁻¹¹⁰ CAMPARI, developed by Pappu and colleagues, utilizes Metropolis MC sampling alongside ABSINTH, an implicit solvent model, and has proven to be the most widely accepted non-MD method to date for producing accurate disordered protein ensembles.^{101, 110-111} Although there currently exist several packages, many of which were developed during this thesis work, no package to date boasts the ability to predict the structure of both ordered and disordered proteins from primary sequence in a unified architecture.

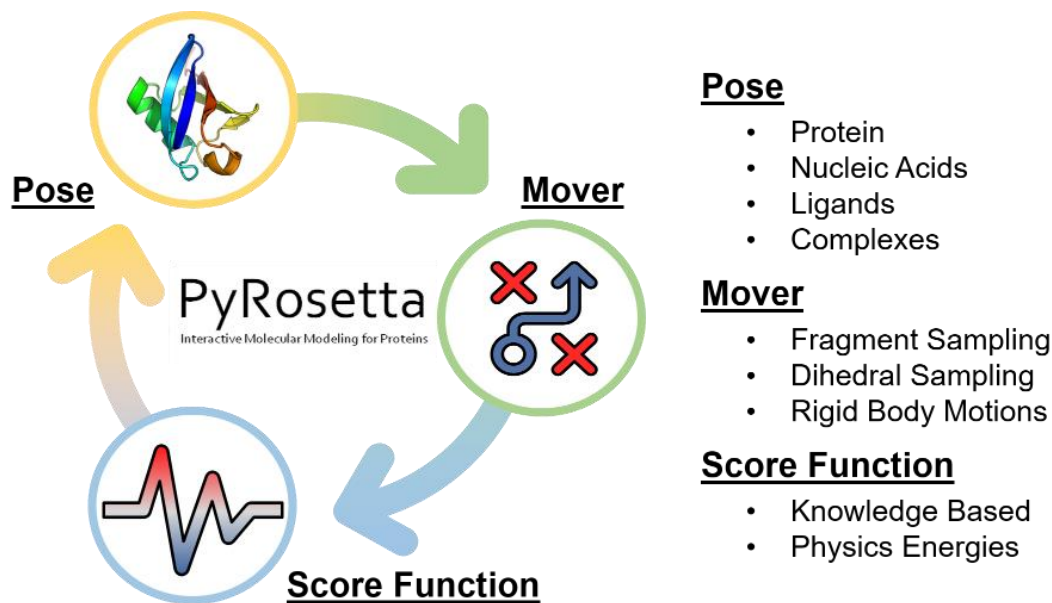


Figure 1.5: Overview of PyRosetta. Left: Depiction of the three primary components that serve as the basis for Monte-Carlo sampling in PyRosetta. Right: Examples of each of the core components of PyRosetta.¹¹²

The Rosetta Modeling Suite

One software package that has demonstrated considerable success in predicting the structure of a folded protein from its sequence is the Rosetta Modeling Suite. The Rosetta Modeling Suite was initially developed by Baker and colleagues but has continued to be improved and maintained by the over 150 developers from 60 laboratories at 23 universities that contribute to Rosetta Commons.¹¹² Although initially conceived with the goal of protein design in mind, methods in Rosetta include *de novo* protein structure prediction, protein and ligand docking, modeling with experimental data and more.¹¹²⁻¹¹³ Rosetta operates off of a relatively simplistic underlying architecture by which sampling methods, termed movers, are applied to given molecule, generally a protein, called a pose, with the

efficacy of each sampling step captured by sets of energy terms, referred to as score functions.¹¹² Although the bulk of Rosetta exists as compiled C++ packages consisting of previously devised algorithms, new algorithms can be facily devised using PyRosetta.¹¹⁴ PyRosetta, written by Gray and colleagues, provides Python wrappers for the underlying functionalities in Rosetta, allowing users to rapidly develop new methods in Python, without needing to compile.¹¹⁴ Therefore, by using existing features of Rosetta, as well as incorporating novel methods for improving sampling and scoring using functionalities in Python, the work detailed herein has focused on developing a unified architecture for the *de novo* prediction of both folded and unfolded proteins. Lastly, the latter portion of this thesis work focuses on how a sub-package of Rosetta, designed for identifying pockets on a protein surface, can be leveraged to identify potential molecular probes that bind to α S fibril structures.

**CHAPTER 2: MULTICOLOR PROTEIN FRET WITH TRYPTOPHAN,
SELECTIVE COUMARIN-CYSTEINE LABELING, AND GENETIC
ACRIDONYLALANINE ENCODING.**

The content of this chapter was originally published in *Chemical Communications*.

It is adapted here with permission from the publisher:

Multicolor protein FRET with tryptophan, selective coumarin-cysteine labeling, and genetic acridonylalanine encoding. J. J. Ferrie, N. Ieda, C. M. Haney, C. R. Walters, I. Sungwienwong, J. Yoon and E. J. Petersson, *Chem. Commun.*, 2017, **53**, 11072 DOI: 10.1039/C7CC05492K – Reproduced by permission of The Royal Society of Chemistry.

§ 2.1 Introduction

Site-specific fluorescence probes can be used to measure distances within proteins when used as part of a Förster resonance energy transfer (FRET) pair. Here we report the synthesis of a coumarin maleimide (Mcm-Mal) that is fluorogenic upon reaction with cysteine. We demonstrate that cysteine, acridonylalanine (Acd) double mutant proteins can be produced by unnatural amino acid mutagenesis and reacted with Mcm-Mal to generate Mcm/Acd labeled proteins for FRET studies. The Mcm/Acd FRET pair is minimally-perturbing, easy to install, and well-suited to studying protein distances in the 15-40 Å range. Furthermore, Mcm/Acd labeling can be combined with tryptophan fluorescence in three color FRET to monitor multiple interactions in one experiment.

Fluorescence spectroscopy methods can be valuable ways of studying protein folding and dynamics, as they allow one to observe protein motions in real time under physiological conditions.¹¹⁵⁻¹¹⁶ If two probes can be attached to the protein, one can obtain structural information using distance-dependent interactions such as Förster resonance energy transfer (FRET) and quenching by photo-induced electron transfer.⁵⁷ The development of probes that can be easily installed for these applications is an important area of bioorganic chemistry. Here, we report the synthesis of methoxycoumarin maleimide (Mcm-Mal, Fig. 2.1), a probe that can be used as part of a FRET pair with acridonylalanine (Acd). We further demonstrate that Mcm/Acd labeling can be combined with a single Trp mutant for three color FRET experiments, which can be used to simultaneously

monitor two interactions, such as a protein/protein interaction and a conformational change within one of the proteins.

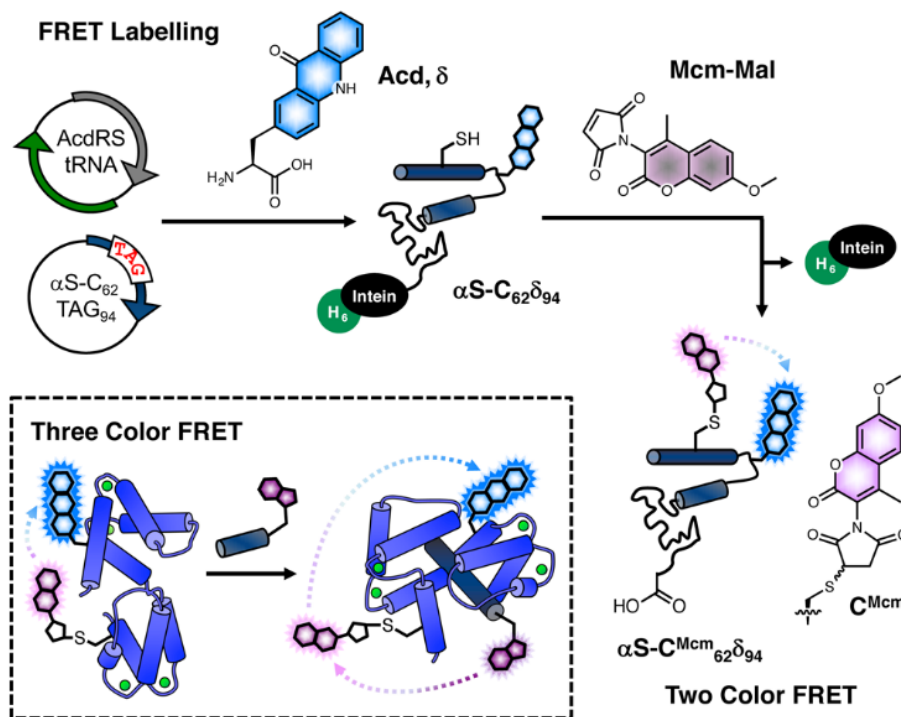


Figure 2.1: Mcm labelling for FRET experiments. Protein expressed with Cys and acridonylalanine (Acid, δ) mutants is harvested as an intein fusion and reacted with methoxycoumarin maleimide (Mcm-Mal) for two color FRET experiments. Inset: Three color FRET between Trp (purple), Mcm (pink), and Acid (cyan) can be used to detect binding (Trp excitation) and conformational change (Mcm excitation) simultaneously.

The fluorescent labeling of proteins for FRET studies has been greatly enabled by recent advances in site-specific protein modification, particularly through incorporation of unnatural amino acids (Uaas) and biorthogonal reactions.¹¹⁷ However, there are still limitations to commonly used fluorophores such as fluorescein and rhodamine that can hamper FRET experiments. Their

large size can be disruptive to protein folding, and their working range for FRET may not be suited to distance changes in the proteins. The fluorescein/ rhodamine FRET pair has a Förster radius (R_0 , the distance of half-maximal energy transfer) of ~ 50 Å, making it useful for studying distances in the 30-90 Å range. Since many important intraprotein distances are shorter than 30 Å, small short-range FRET pairs are better suited to studying these protein regions.

Our laboratory has worked to develop non-perturbing, short/medium-range probe pairs that can be selectively excited in proteins. These include fluorescent Uaas such as Acd.^{77, 118} Acd is excited at 380-400 nm with emission at 420-450 nm. We have shown that Acd can be a valuable FRET acceptor from methoxycoumarin (Mcm, excited at 325 nm) with a working range of 15-40 Å.⁷⁷ Mcm is an excellent FRET donor to Acd since it has a high extinction coefficient with a peak in the absorption spectrum that coincides with a minimum in the Acd absorption spectrum (Fig. 2.16). However, in previous studies, Mcm was introduced as 7-methoxycoumarin-4-yl-alanine, which must be incorporated by solid phase peptide synthesis.⁷⁷ We wished to introduce Mcm through selective Cys modification in order to more easily generate Mcm/Acd labelled proteins.

§ 2.2 Characterization of Mcm-Mal and Mcm-Br

Determination of Mcm-Mal Extinction Coefficient.

Mcm-Mal, Ac-Cys and Mcm-AcOH (~ 2 mg) were weighted on an analytical balance. Mcm-Mal and Mcm-AcOH were dissolved in 100 μ L DMSO while Ac-Cys was dissolved in 900 μ L 20 mM Tris 100 mM NaCl pH 7.5. Mcm-Mal was reacted

with a 100-fold excess of Ac-Cys (resulting solutions < 1 % DMSO) and allowed to react for 6 hour. Following reaction, multiple dilutions were prepared of Mcm-Mal + Ac-Cys and Mcm-AcOH and the UV-Vis absorbance of each sample was measured. The absorbance at 325 nm of each sample was plotted as function of concentration based on mass calculation and fit to a line with the extinction coefficient determined by the slope ($\epsilon_{\text{Mcm-AcOH}} = 14440 \text{ M}^{-1} \text{ cm}^{-1}$, $\epsilon_{\text{Mcm-Mal}} = 19010 \text{ M}^{-1} \text{ cm}^{-1}$).

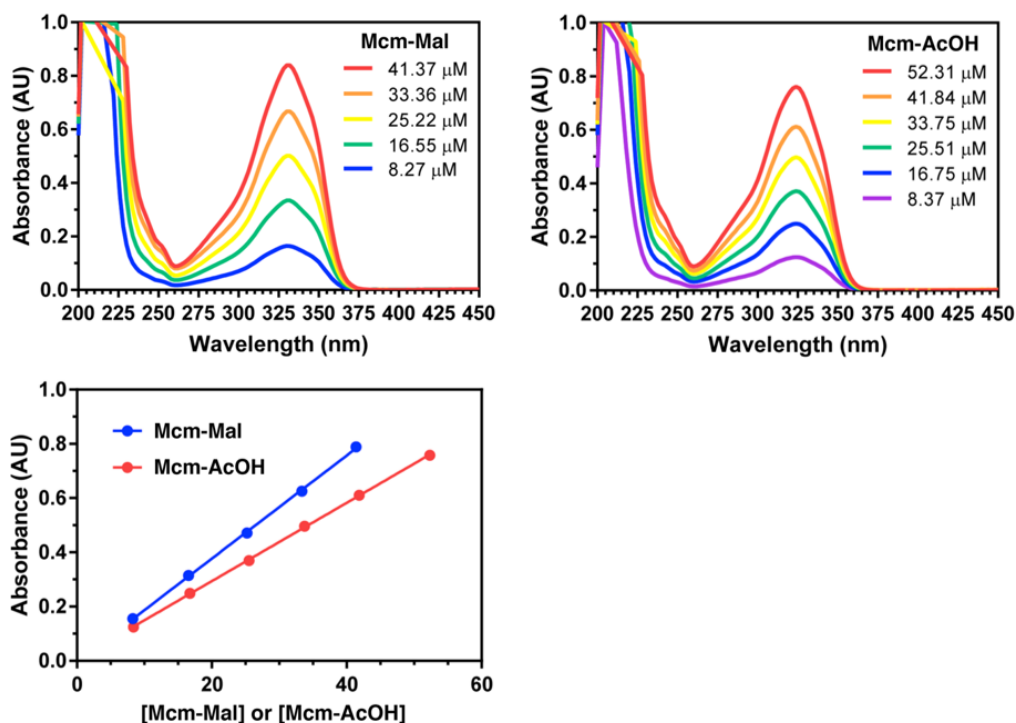


Figure 2.2: Determination of Extinction Coefficient for Mcm-Mal. Top Left: Absorbance spectra of varying concentration of Mcm-AcOH in 20 mM Tris 100 mM NaCl pH 7.5. Top Right: Absorbance spectra of varying concentration of Mcm-Mal + Ac-Cys in 20 mM Tris 100 mM NaCl pH 7.5. Bottom: Absorbance at 325 nm for varying concentrations of Mcm-Mal + Ac-Cys and Mcm-AcOH in 20 mM Tris 100 mM NaCl pH 7.5, with linear correlation for determination of extinction coefficient.

Reactions of Cys and Ac-Cys Monitored by UV/Vis, Fluorescence, and LCMS.

Stocks of Mcm-Mal and Mcm-Br were prepared in DMSO and stocks of Cys and Ac-Cys were prepared in 100 mM Tris pH 7.0 or 20 mM Tris 100 mM NaCl pH 7.5, respectively. Mcm-Mal and Mcm-Br were mixed with buffer, Cys, or Ac-Cys to final concentrations of 10 μ M of both Mcm derivative and Cys/Ac-Cys for UV-Vis absorbance, steady-state fluorescence and LRMS measurements. Reactions were allowed to take place for 6 hours following vortexing for 10 seconds prior to measurement. Mcm-Br [M+H]⁺ Calcd. 269.0/271.0, Obs. 269.0/271.0; Mcm-Mal [M+H]⁺ Calcd. 286.1, Obs. 286.2; Mcm-Mal + Cys (S4 or S5) [M+H]⁺ Calcd. 407.1, Obs. 407.2; Mcm-Mal + Ac-Cys (S6) [M+H]⁺ Calcd. 449.1, Obs. 449.4

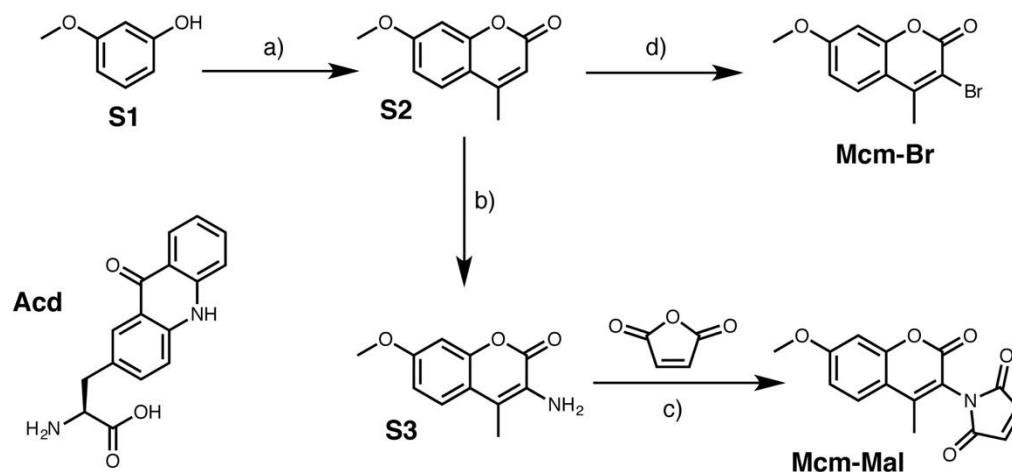


Figure 2.3: Synthesis of Mcm-Br and Mcm-Mal. a) Ethyl acetoacetate, H_2SO_4 , 65%; b) 1. Ac_2O , HNO_3 ; 2) H_2 (g), Pd/C, 32% (over two steps); c) Maleic anhydride, 84%; d) NBS, NH_4OAc , 78%

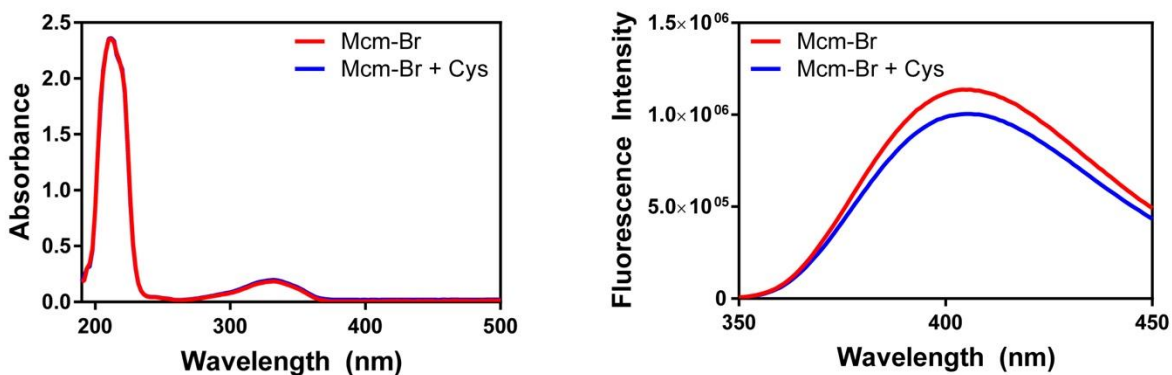


Figure 2.4: Absorbance and Fluorescence Spectra of Mcm-Br. Left: Absorbance Spectra: 10 μM Mcm-Br in Tris buffer (100 mM, pH 7.0, DMSO 1%) alone, or mixed with 10 μM Cys. Right: Fluorescence Spectra: 10 μM Mcm-Br in Tris buffer (100 mM, pH 7.0, DMSO 1%) alone, or mixed with 10 μM Cys. Fluorescence excitation at 330 nm, emission monitored at 390 nm.

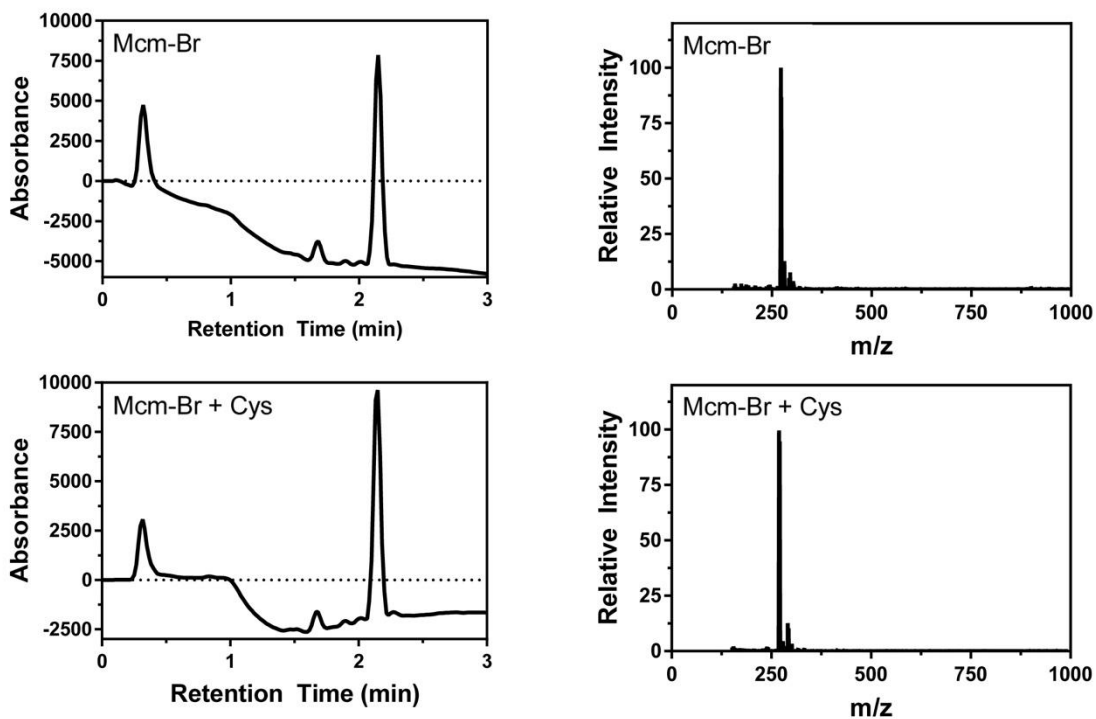


Figure 2.5: LCMS Analysis of Mcm-Br Reactions with Cys. 10 μ M Mcm-Br or Mcm-Mal in Tris buffer (100 mM, pH 7.0, DMSO 1%) alone or mixed with 10 μ M Cys. LC chromatogram monitored at 330 nm. MS spectrum obtained for largest peak in LC chromatogram.

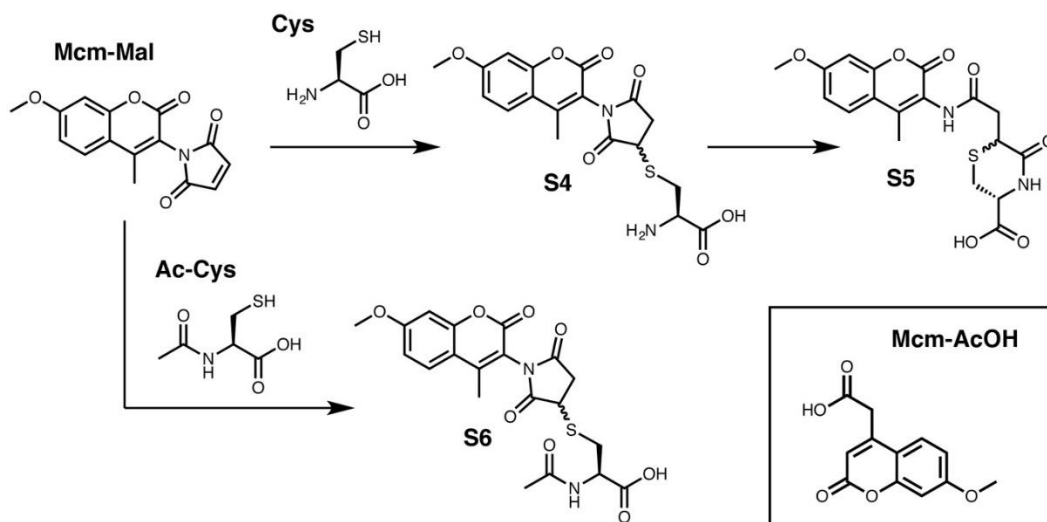


Figure 2.6: Mcm-Mal Reactions with Cys and Ac-Cys and assigned products. Inset: Mcm-AcOH.

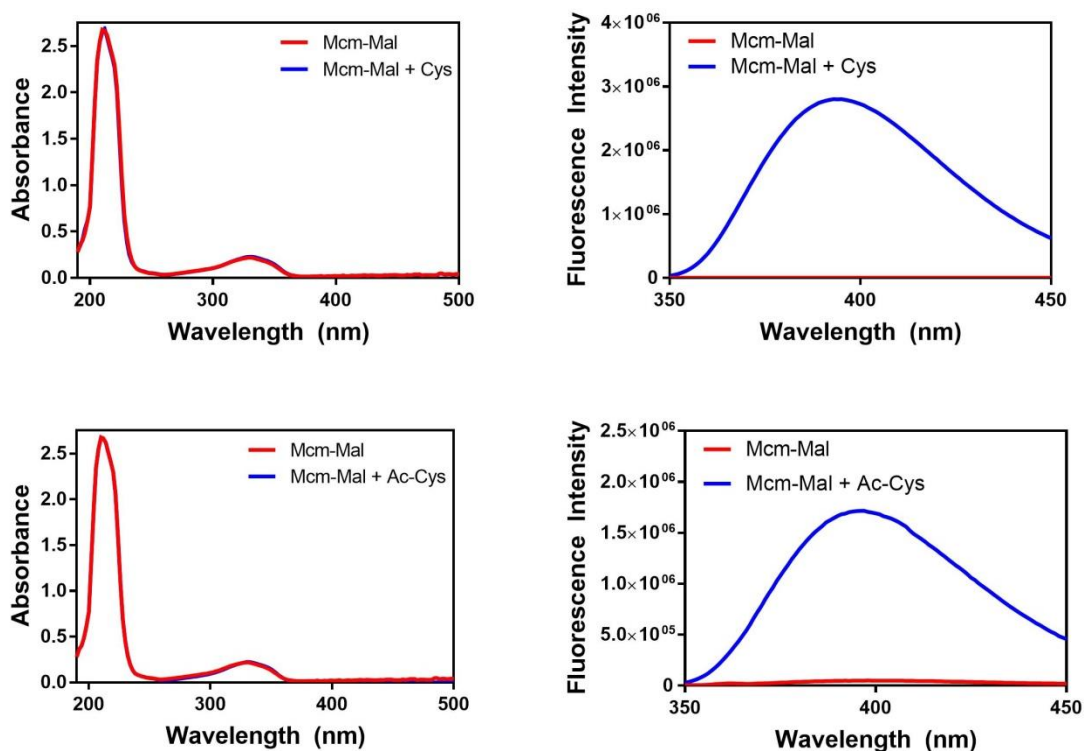


Figure 2.7: Absorbance and Fluorescence Spectra of Mcm-Mal. Left: Absorbance Spectra: 10 μ M Mcm-Mal in Tris buffer (10 mM, pH 7.0, DMSO 1%) alone, or mixed with 10 μ M Cys or Ac-Cys. Right: Fluorescence Spectra: 10 μ M Mcm-Mal in Tris buffer (100 mM, pH 7.0, DMSO 1%) alone, or mixed with 10 μ M Cys or Ac-Cys.

Fluorescence measurements were obtained using the PTI Quantmaster in 20 mM Tris, 100 mM NaCl pH 7.5 buffer with an excitation wavelength of 325 nm, measuring the emission from 350-450 nm, with 2 nm excitation and emission slit widths, a step size of 1 nm, and an integration time of 0.25 seconds per step.

Reaction Timecourses of Cys and Ac-Cys Monitored by Fluorescence.

For timecourse measurements, 2 μM dilutions were prepared for all components from DMSO stocks of Mcm-Br/Mcm-Mal and stocks of Cys and Ac-Cys in 20 mM Tris 100 mM NaCl pH 7.5. Buffer, Cys or Ac-Cys were added to aliquots of Mcm-Mal or Mcm-Br in a 96-well plate immediately prior to measurement in a Tecan M1000 plate reader producing final concentrations of 1 μM for all reaction components. Samples were excited at 325 nm and measured and the emission was monitored at 400 nm with excitation and emission slit widths of 5 nm. Measurements were taken every 20 seconds.

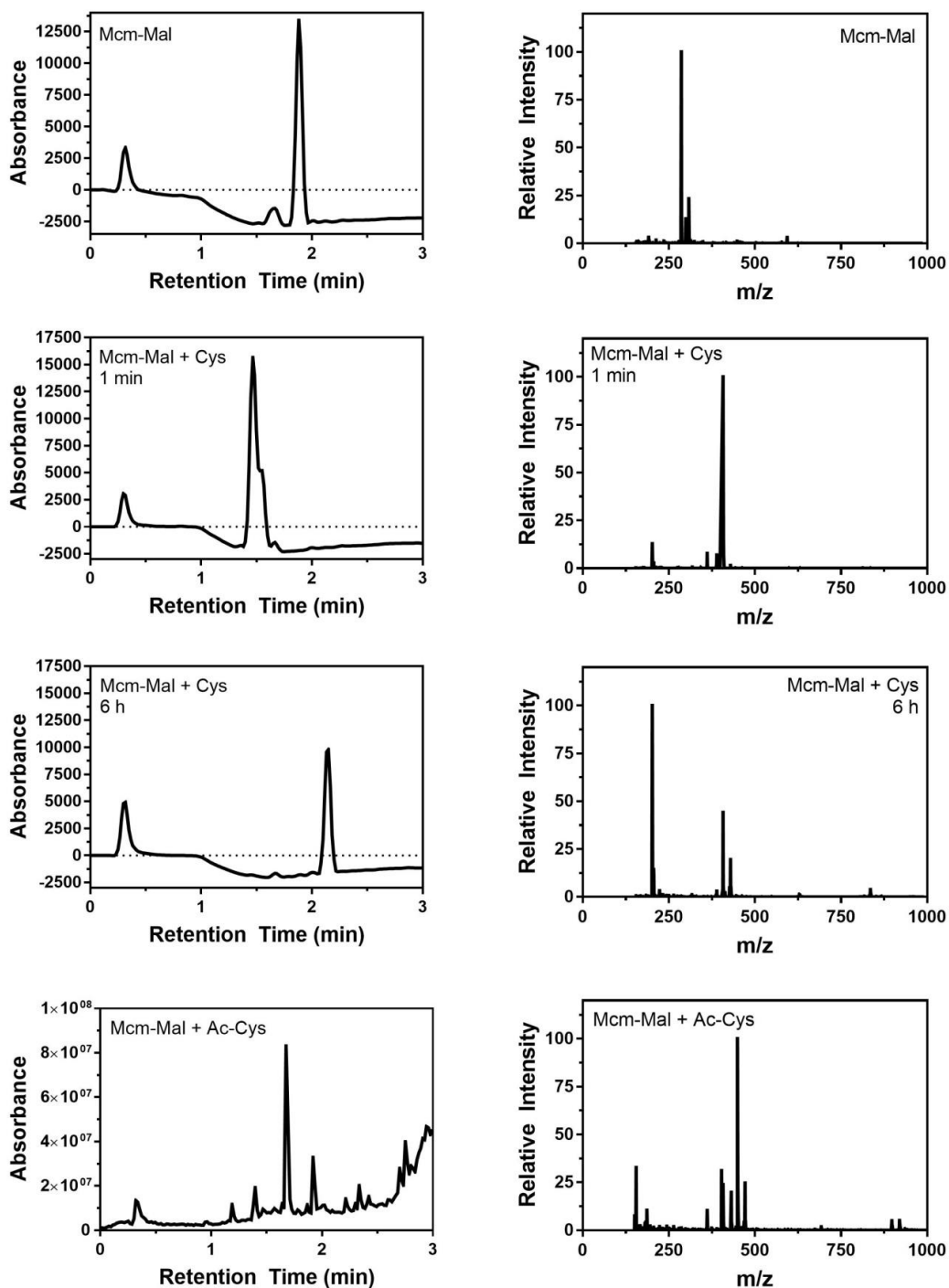


Figure 2.8: LCMS Analysis of Mcm-Mal Reactions with Cys and Ac-Cys. 10 μ M Mcm-Mal in Tris buffer (100 mM, pH 7.0, DMSO 1%) alone or mixed with 10 μ M Cys or Ac-Cys. Cys reactions were incubated for 1 minute or 6 hours. LC

chromatogram monitored at 330 nm. MS spectrum obtained for largest peak in LC chromatogram.

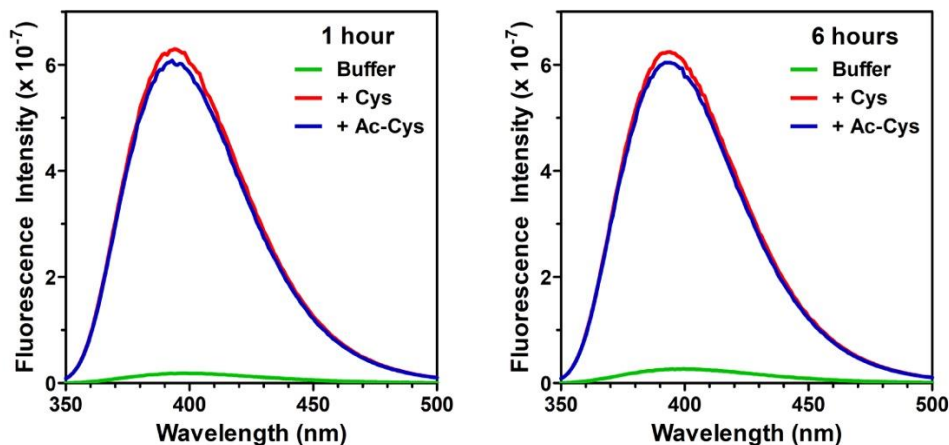


Figure 2.9: Mcm-Mal Reactions with Cys and Ac-Cys after 1 and 6 h. 10 μM Mcm-Mal in Tris buffered saline (20 mM Tris, 100 mM NaCl, pH 7.5, DMSO 1%) alone or mixed with Cys or Ac-Cys. Fluorescence excitation at 325 nm, emission 350-500 nm and acquired after 1 hour (left) or 6 hours (right).

Determination of Mcm-Mal post-reaction Quantum Yield.

Mcm-Mal was reacted with 10-fold excess Ac-Cys in 20 mM Tris 100 mM NaCl pH 7.5 (< 1% DMSO) for 6 hour. Six dilutions of reacted Mcm-Mal and 7-methoxycoumarin-4-yl-acetic acid (Mcm-AcOH, see Fig 2.6, Scheme S2 inset) ranging in absorbances from 0.05 to 0.02 were prepared. Fluorescence emission spectra were measured at an excitation of 325 nm over an emission range of 335 to 550 nm with 5 nm slit widths, a 1 nm step width and a 0.25 second integration time. The quantum yield was subsequently calculated for each Mcm-Mal/Mcm-AcOH pair by multiplying the quantum yield of Mcm-AcOH ($\Phi_{\text{Mcm-AcOH}} = 0.18$) by the ratio of the sum of the fluorescence emissions of Mcm-Mal to Mcm-AcOH, resulting in an average quantum yield of $\Phi_{\text{Mcm-Mal-Ac-Cys}} = 0.22$.

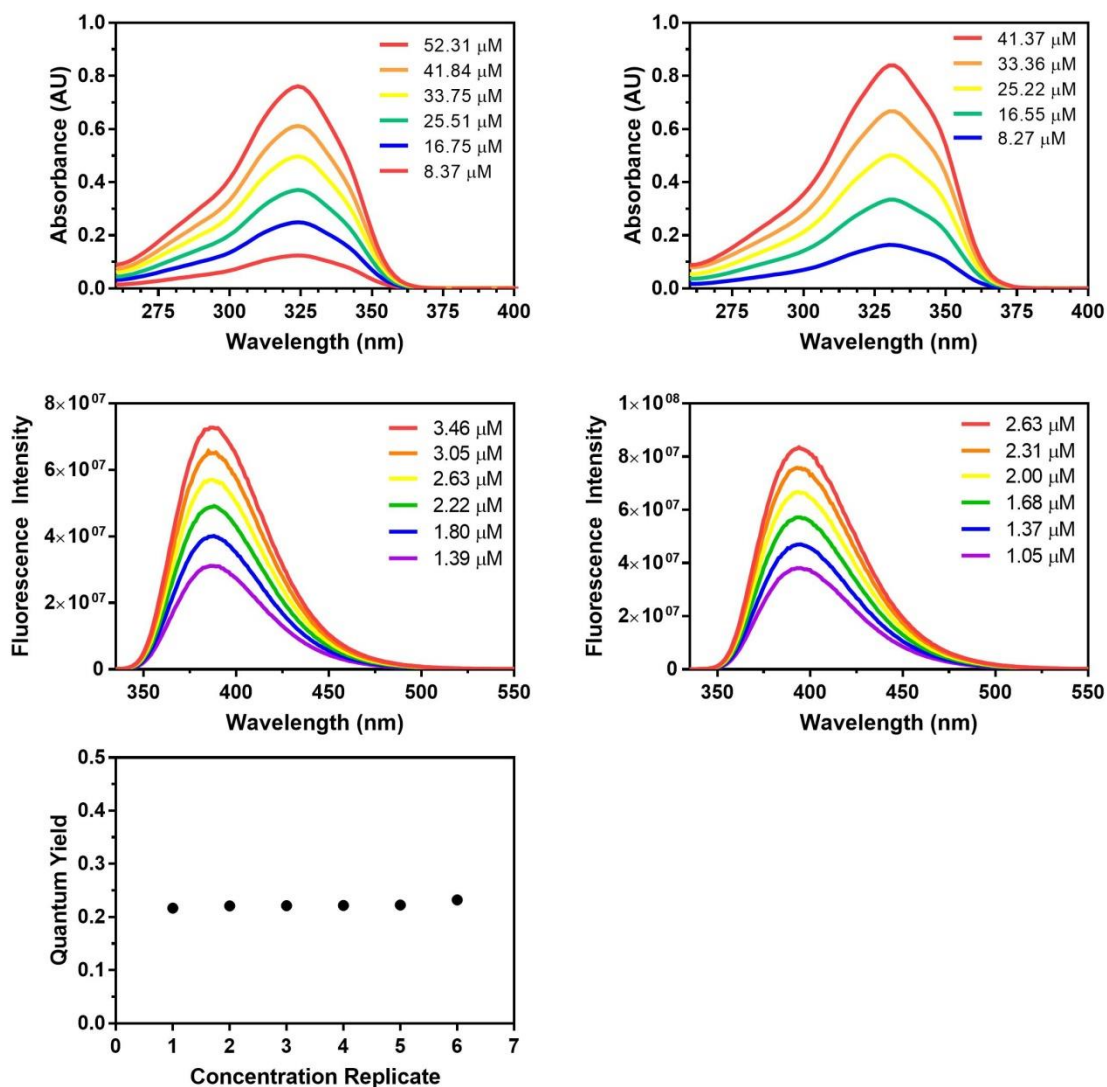


Figure 2.10: Determination of Quantum Yield for Ac-Cys Mcm Product. Top left: Absorbance spectra of aliquots of Mcm-AcOH in 20 mM Tris 100 mM NaCl pH 7.5 ranging from 0.05 to 0.02. Top right: Absorbance spectra of aliquots of Ac-Cys Mcm-Mal in 20 mM Tris 100 mM NaCl pH 7.5 ranging from 0.05 to 0.02. Middle left: Fluorescence emission spectra of Mcm-AcOH dilutions in 20 mM Tris 100 mM NaCl pH 7.5 at an excitation wavelength of 325 nm. Middle right: Fluorescence emission spectra of Ac-Cys Mcm-Mal dilutions in 20 mM Tris 100 mM NaCl pH 7.5 at an excitation wavelength of 325 nm. Bottom left: Calculated quantum yield values for each sample combination.

§ 2.2 Protein Expression, Labeling and Purification

Cloning of CaM Constructs.

The gene encoding full-length calmodulin was previously cloned into the pTXB1 vector containing a C-terminal MxeGyrA intein, followed by a His₆ purification tag.⁹⁵ A 'TAG' codon for the incorporation of Acd at Leu₁₁₂ via amber stop codon suppression and a 'TGC' codon encoding Cys at Phe₁₂ for Mcm-Mal labelling were inserted using the following sets of primers in QuikChange® PCR. In the case of the F₁₂C-L₁₁₂δ double mutant, the CaM-F₁₂C mutant was obtained first, before using the TAG₁₁₂ primers in a second round of PCR to obtain the desired plasmid.

Expression of CaM C₁₂-GyrA-H₆.

A plasmid encoding CaM F₁₂C-GyrA was transformed into BL21-Gold (DE3) *E. Coli* cells and grown against ampicillin (Amp, 100 µg/mL) on an LB-agar plate. Single colonies were picked and grown in liquid LB media (2 x 5 mL, 100 µg/mL ampicillin) with shaking (250 RPM) at 37 °C until saturation. Both primary cultures were then added to a secondary culture of autoclaved M9 media (500 mL, 100 µg/mL Amp) and grown at 37 °C, with shaking (250 RPM) until OD₆₀₀ = 0.8 (3-4 h). Isopropyl β-D-1-thiogalactopyranoside (IPTG) was then added (final concentration = 1 mM) and the temperature and shaking speed were reduced to 25 °C and 225 RPM respectively for 16 h of additional incubation.

Expression of CaM- δ_{112} and CaM-C₁₂ δ_{112} .

Plasmids encoding CaM- δ_{112} -GyrA-H₆ and CaM-C₁₂ δ_{112} -GyrA-H₆ were separately transformed into BL21 cells that also contain a plasmid encoding for an orthogonal Acd synthetase 2b (AcdRS2b, also referred to as clone A9) and tRNA_{CUA} pair described in detail elsewhere.^{95, 118} These transformations were grown against streptomycin (Strep, 100 μ g/mL) and Amp (100 μ g/mL) on an LB-agar plate. Single colonies were picked and grown in liquid LB media (2 x 5 mL each, 100 μ g/mL each of Strep and Amp) with shaking (250 RPM) until saturation. Both primary cultures were then added to a secondary culture of autoclaved M9 media (500 mL, 100 μ g/mL each of Strep and Amp) and grown at 37 °C with shaking (250 RPM) until OD₆₀₀=0.7 (3-4 hours). IPTG and Acd were then added (1 mM and 0.5 mM final concentrations, respectively) and the temperature and shaking speed were reduced to 18 °C and 225 RPM, respectively. These cultures were then incubated at these conditions for 20 hours.

Purification of CaM Constructs.

Cells were harvested by centrifugation at 5,000 RPM in a GS3 rotor and Sorvall RC-5 centrifuge for 15 minutes at 4 °C. The supernatant was discarded and the cell pellet was suspended in 20 mL lysis buffer (50 mM HEPES, pH 7.5) containing a broad spectrum protease inhibitor tablet. Resuspended cells were then lysed on ice by sonication (30 amps power, 2 second pulse, 2 second rest, 4 minutes total sonication time) and then pelleted at 13,000 RPM in an SS-34 rotor (Sorvall RC-5 centrifuge) for 15 minutes at 4 °C. The supernatant was collected

and incubated with Ni²⁺-NTA resin (2 mL column volume) for 1 h on ice with shaking. The slurry was then added to a fritted column and the liquid was allowed to flow through. The resin was then washed with 3 x 10 mL of buffer (50 mM HEPES, pH 7.5) and 3 x 10 mL of wash buffer (50 mM HEPES, 10 mM imidazole, pH 7.5). Each CaM construct was then eluted from the resin in 5 fractions each containing 3 mL of elution buffer (50 mM HEPES, 300 mM imidazole, pH 7.5). The pooled fractions were immediately subjected to intein cleavage conditions by adding β-mercaptoethanol to a final concentration of 200 mM and incubated on a rotisserie at RT for 20 hours. The resulting cleavage solution was then dialyzed against 20 mM Tris pH 8.0 (2L) overnight in preparation for anion exchange purification via FPLC. Prior to FPLC purification, 10 μL of 0.5 M TCEP Bond Breaker™ was added to the CaM-C₁₂ and CaM-C₁₂δ₁₁₂ constructs to reduce any aberrant disulfides formed between CaM proteins and/or βME. Each CaM construct was purified over a HiTrap Q column using a 120 min NaCl gradient (0.1 M to 0.8 M NaCl in 20 mM Tris, pH 8.0). Fractions containing the product peak were confirmed by MALDI and dialyzed twice against water at 4 °C (2 L, 2 hours each). These samples were then flash frozen in liquid N₂ and lyophilized to a powder.

CaM Labelling with Mcm-Mal.

A stock solution of ~25 mM Mcm-Mal was prepared by dissolving 4.6 mg of Mcm-Mal solid in 645 μL DMSO. CaM-C₁₂ and CaM-C₁₂δ₁₁₂ samples were dissolved in 1 mL of 20 mM Tris, pH 7.5. To each sample was added 10

equivalents of Mcm-Mal (based off UV-Vis quantification of protein solutions) and the samples were incubated at 37 °C with shaking (200 RPM) for 3 hours. Each sample was then diluted in 2 mL of 20 mM Tris pH 8.0 and subjected to a second round of FPLC purification using the gradient described above. Fractions were analyzed by MALDI MS and fractions containing the desired mass for Cys-Mcm conjugated product were pooled and dialyzed (2 x 2 L) against water before being flash frozen in N₂ and lyophilized to a powder.

α S-Q₆₂C and α S-E₁₁₄C Expression.

α -Synuclein (α S) mutant plasmids (pTXB1_ α S-C₆₂_Mxe-H₆ or pTXB1_ α S-C₁₁₄_Mxe-H₆) were transformed into competent *E. coli* BL21(DE3) cells and plated on LB agar plates supplemented with Amp overnight at 37 °C.¹¹⁹ Single colonies were used to inoculate 5 mL of LB media supplemented with Amp (100 μ g/mL). The primary culture was incubated at 37 °C with shaking at 250 rpm for 4 h. A single primary culture was used to inoculate 1 L of LB media containing Amp (100 mg/L) which was then grown at 37 °C with shaking at 250 rpm until it reached OD ~0.7. Expression was induced with IPTG (1 mM) and the temperature was reduced to 18 °C overnight.

α S- δ ₉₄, α S-C₆₂ δ ₉₄, and α S- δ ₉₄C₁₁₄ Expression.

α S mutant plasmids (pTXB1_ α S-TAG₉₄_Mxe-H₆, pTXB1_ α S-C₆₂TAG₉₄_Mxe-H₆, or pTXB1_ α S-TAG₉₄C₁₁₄_Mxe-H₆, generated using plasmids in Haney *et al.*) and an AcdRS2b/ tRNA plasmid were transformed into competent

E. coli BL21(DE3) cells and plated on LB agar plates supplemented with Amp overnight at 37 °C.^{78, 118} Single colonies were used to inoculate 5 mL of LB media supplemented with Amp and streptomycin (Strep, 100 µg/mL of each). The primary culture was incubated at 37 °C with shaking at 250 rpm for 4 h. A single primary culture was used to inoculate 1 L of LB media containing Amp and Strep (100 mg/L) which was grown at 37 °C with shaking at 250 rpm until it reached OD ~0.7. Expression was induced by adding IPTG and Acd (concentrations of 1 mM and 140 mg/L, respectively) which was then incubated overnight at 18 °C with shaking at 250 rpm.

Purification of α S Constructs.

Cells were harvested by centrifugation at 4,000 RPM in a GS3 rotor and Sorvall RC-5 centrifuge for 20 minutes at 4 °C. The supernatant was discarded and the cell pellet was suspended in 15 mL lysis buffer (40 mM Tris, 5 mM EDTA, pH 8.0) containing one Roche protease inhibitor cocktail pill (cOmplete mini tablets, EDTA-free, Easy Pack, Roche Cat. #04693159001). Resuspended cells were then lysed on ice by sonication (30 amps power, 1 second pulse, 1 second rest, 5 minutes total sonication time) and then pelleted at 14,000 RPM in an SS-34 rotor (Sorvall RC-5 centrifuge) for 20 minutes at 4 °C. The supernatant was collected and incubated with Ni²⁺-NTA resin (3 mL column volume) for 1 h on ice with shaking. The slurry was then added to a fritted column and the liquid was allowed to flow through. The resin was then washed with 3 x 5 mL of buffer (50 mM HEPES, pH 7.5) and 2 x 10 mL of wash buffer (50 mM HEPES, 10 mM

imidazole, pH 7.5). Constructs were then eluted from the resin in 4 fractions each containing 3 mL of elution buffer (50 mM HEPES, 300 mM imidazole, pH 7.5). The pooled fractions were immediately subjected to intein cleavage conditions by adding β -mercaptoethanol to a final concentration of 200 mM and incubated on a rotisserie at RT for 20 hours. The resulting cleavage solution was then dialyzed against 20 mM Tris pH 8.0 overnight. Prior to FPLC purification, 10 μ L of 0.5 M TCEP Bond Breaker™ was added to the Cys containing constructs to reduce any aberrant disulfides formed between α S proteins and/or β ME. Each construct was purified by ion-exchange chromatography using a HiTrap Q HP column (5 mL) on an ÄKTA FPLC using a 100 minutes NaCl gradient (0 to 500 mM NaCl in 20 mM Tris, pH 8.0). The fractions containing the product were identified by MALDI MS. α S- δ_{94} mutant was dialyzed at 4 °C against α S buffer (20 mM Tris, 100 mM NaCl, pH 7.5) overnight and stored at -80 °C in 1.5 mL aliquots and thawed once for experiments. Cys containing mutants were carried directly into the labelling step following purification.

α S Labelling with Mcm-Mal.

The protein solution after FPLC (ca. 10–20 mL) was treated with 20 μ L Bond Breaker solution. To the protein solution was added 25 mM Mcm-Mal solution in DMSO (200–300 μ L). Each reaction step was monitored by MALDI MS. After labelling with Mcm-Mal, the solution was dialyzed at 4 °C against Tris buffer (20 mM, pH 8.0) overnight. The dialyzed solution was concentrated by centrifugation with a 3 kDa cutoff filter (milliporesigma #UFC900324). The concentrated solution

was purified by HPLC with a protein C4 column (Vydac #214TP1010). After concentration by centrifugation with a 3 kDa cutoff filter, the solution was dialyzed against α S buffer (20 mM Tris, 100 mM NaCl, pH 7.5) overnight. Following dialysis, α S in buffer was stored at $-80\text{ }^{\circ}\text{C}$ in 1.5 mL aliquots and thawed once for experiments.

Table 2.1: Protein MALDI Masses.

Protein	Calc. [M+H] ⁺ or [M+Na] ⁺	Obs. [M+H] ⁺ or [M+Na] ⁺
α S-C ^{Mcm} ₆₂	14721	14731
α S-C ^{Mcm} ₆₂ δ ₉₄	14838	14844
α S-C ^{Mcm} ₁₁₄	14720	14727
α S- δ ₉₄ C ^{Mcm} ₁₁₄	14837	14844
α S- δ ₉₄	14757	14757
CaM-C ^{Mcm} ₁₂	16950	16947
CaM- δ ₁₁₂	16881	16889
CaM-C ^{Mcm} ₁₂ δ ₁₁₂	17123	17133

Table 2.2: Trypsin Fragment MALDI Masses.

Protein	Fragment	Calc. [M+H] ⁺	[M + Na] ⁺	Obs. [M+H] ⁺	[M+Na] ⁺
α S-C ^{Mcm} ₆₂	61-80	2207	2229	2206	2228
α S-C ^{Mcm} ₆₂ δ ₉₄	61-80	2207	2229	2206	2228
	81-96	1596	1618	1595	1618
α S-C ^{Mcm} ₁₁₄	103-140	4564	4586	4562	4584
α S- δ ₉₄ C ^{Mcm} ₁₁₄	81-96	1596	1618	1596	1618
	103-140	4564	4586	4561	4564
α S- δ ₉₄	81-96	1596	1618	1596	1618
CaM-C ^{Mcm} ₁₂	1-13	1763	1785	1761	1783
CaM- δ ₁₁₂	107-115	1179	1201	1180	1201
CaM-C ^{Mcm} ₁₂ δ ₁₁₂	1-13	1763	1785	1762	1784
	107-115	1179	1201	1179	1201

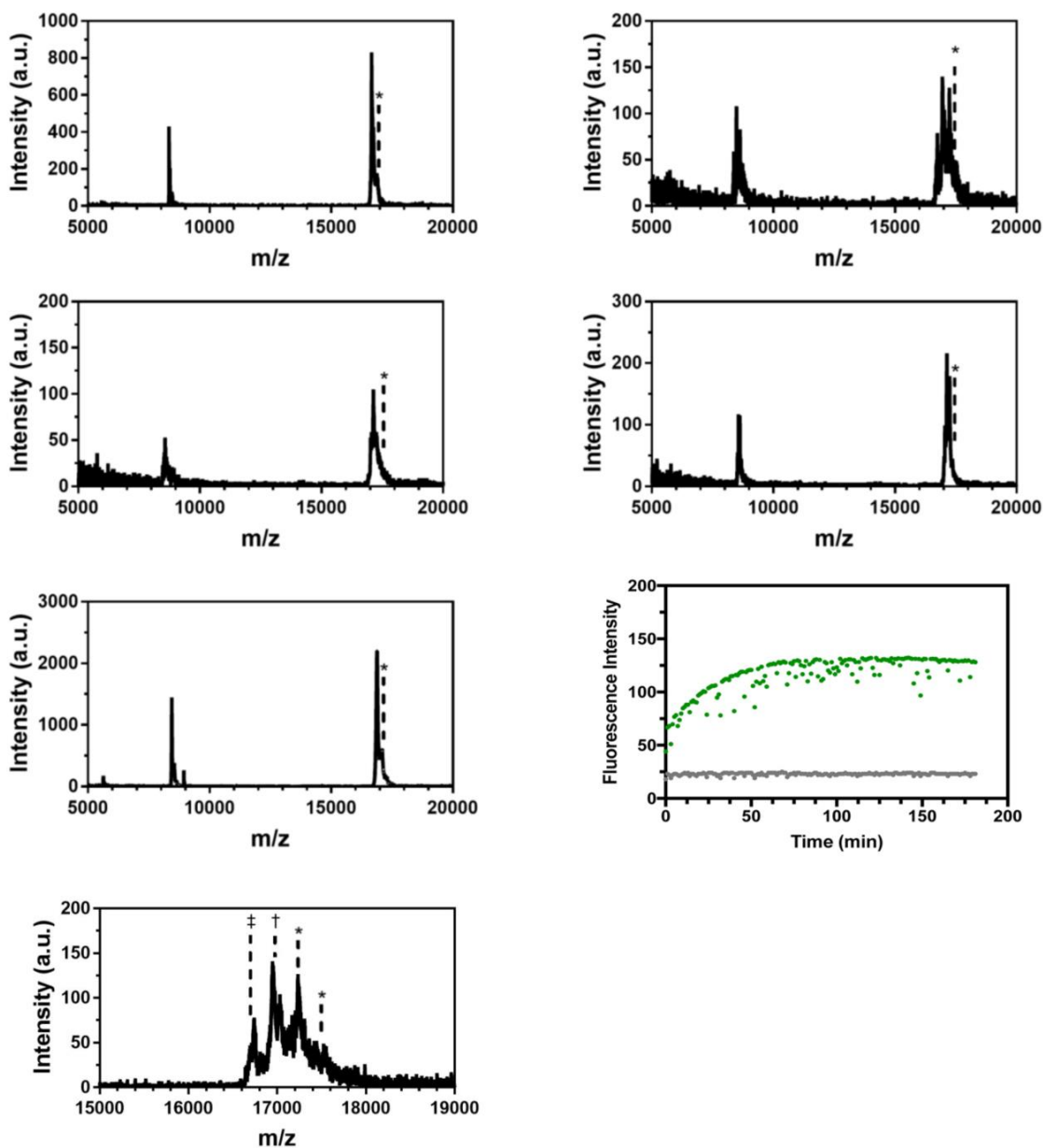


Figure 2.11: MALDI MS Characterization of CaM Variants. On each plot matrix adduct peaks are marked with *. The plots show CaM-C12 (First Row Left), CaM-CMcm12 (First Row Right), CaM-C12 δ 112 (Second Row Left), CaM-CMcm12 δ 112 (Second Row Right) and CaM- δ 112 (Third Row Left). An enlarged spectrum of the region of interest for the CaM-CMcm12 δ 112 shown where the labeled and unlabeled protein masses are indicated with † and ‡ respectively (Fourth Row Left). Fluorescence emission at 390 nm from excitation at 325 nm was monitored after mixing 35 μ M CaM-C12 with 3 equiv Mcm-Mal in Tris buffer, pH 7.5 (Third Row Right, Green: CaM-C12, Grey: Buffer control).

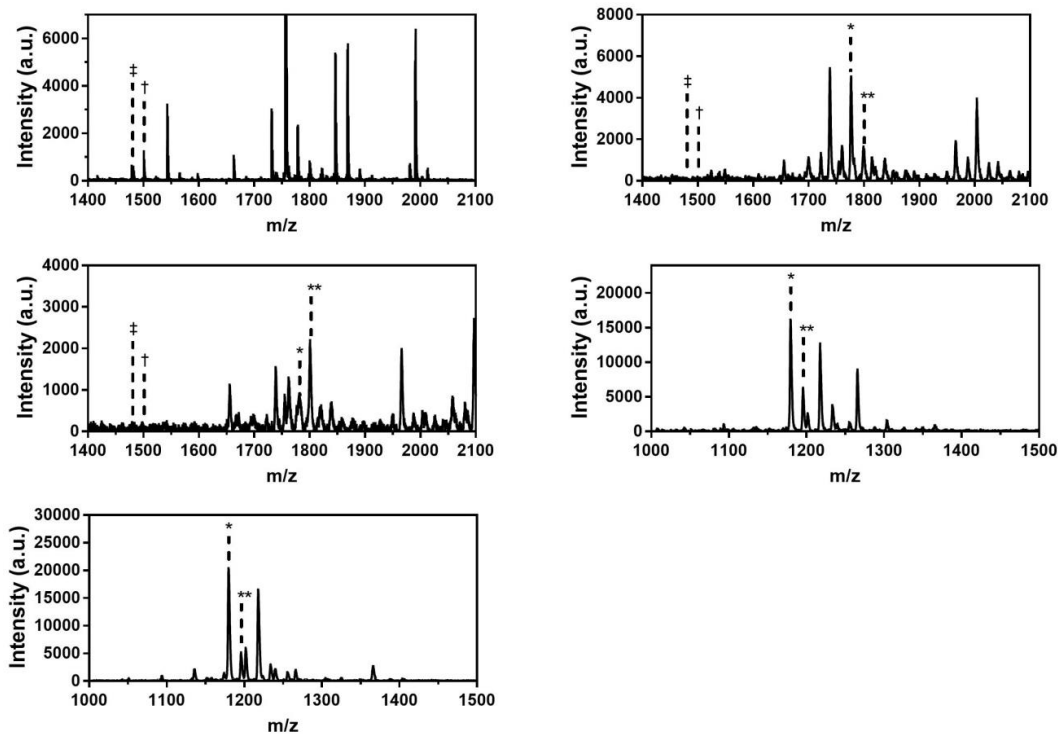


Figure 2.12: Trypsin Digest MALDI MS Characterization of CaM Variants. On each plot, the M+H peak of the fragment of interest is marked with * and the M+Na peak is marked with **, while the M+H peak corresponding to unlabeled Cys containing protein is denoted with ‡ and the M+Na peak is marked with †. The plots show the fragments of CaM1-13-C12 of CaM-C12 (Top Left), CaM1-13-CMcm12 of CaM-CMcm12 (Top Right), CaM1-13-CMcm12 of CaM-CMcm12 δ 112 (Middle Left) and CaM107-115- δ 112 of CaM-CMcm12 δ 112 (Middle Right), CaM107-115- δ 112 of CaM- δ 112 (Bottom Left).

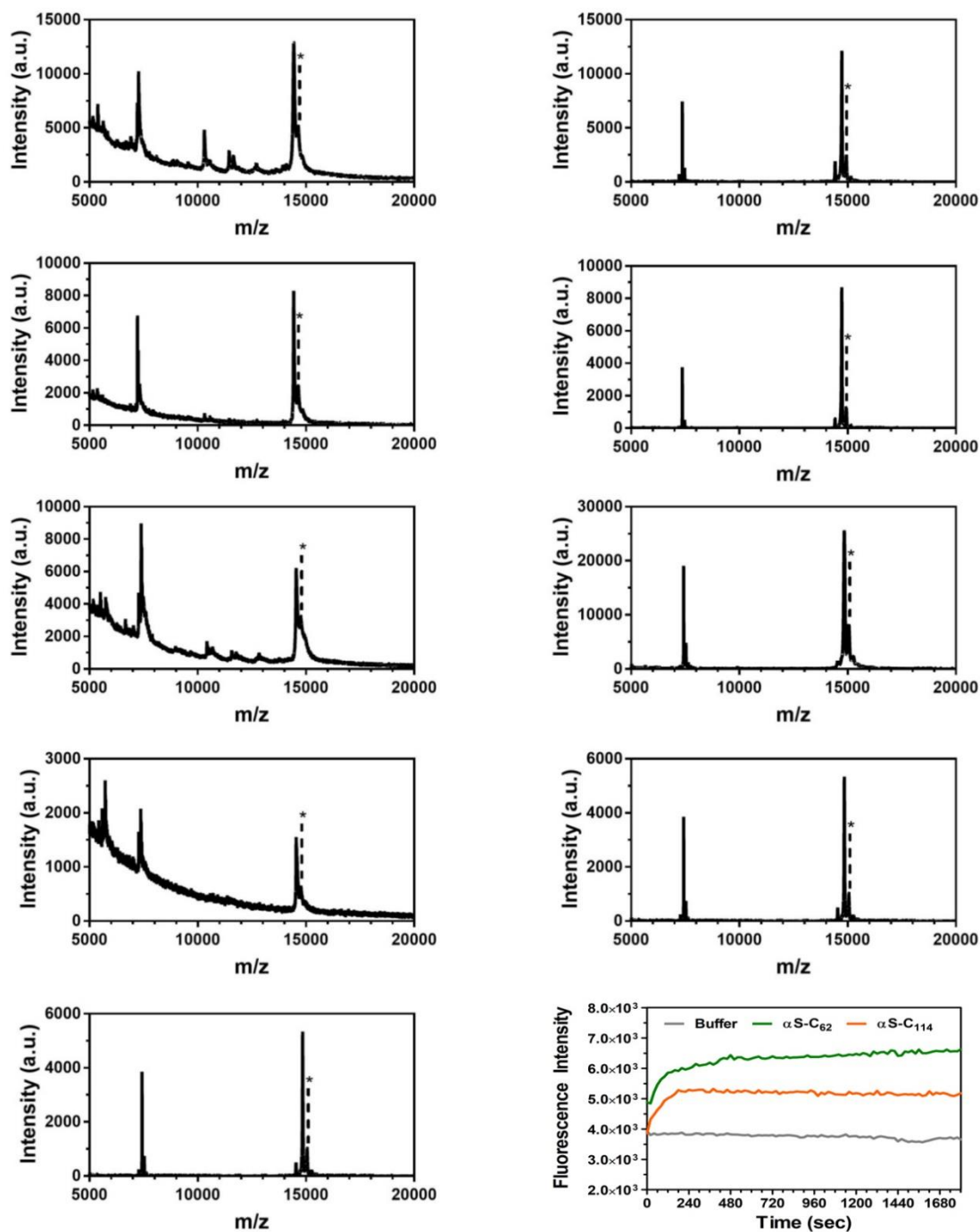


Figure 2.13: MALDI MS Characterization of α S Variants. On each plot matrix adduct peaks are marked with *. The plots show α S-C₆₂ (Row 1 Left), α S-C^{Mcm}₆₂ (Row 1 Right), α S-C₁₁₄ (Row 2 Left), α S-C^{Mcm}₁₁₄ (Row 2 Right), α S-C₆₂ δ ₉₄ (Row 3 Left), α S-C^{Mcm}₆₂ δ ₉₄ (Row 3 Right), α S- δ ₉₄C₁₁₄ (Row 4 Left), α S- δ ₉₄C^{Mcm}₁₁₄ (Row 4 Right) and α S α S- δ ₉₄ (Row 5 Left). Fluorescence emission at 400 nm from excitation at 325 nm was monitored after mixing 1 μ M α S-C₆₂ or α S-C₁₁₄ with 1

equiv Mcm-Mal in 20 mM Tris buffer, pH 7.5 (Bottom Right, Green: α S-C₆₂, Orange: α S-C₁₁₄, Grey: Buffer control).

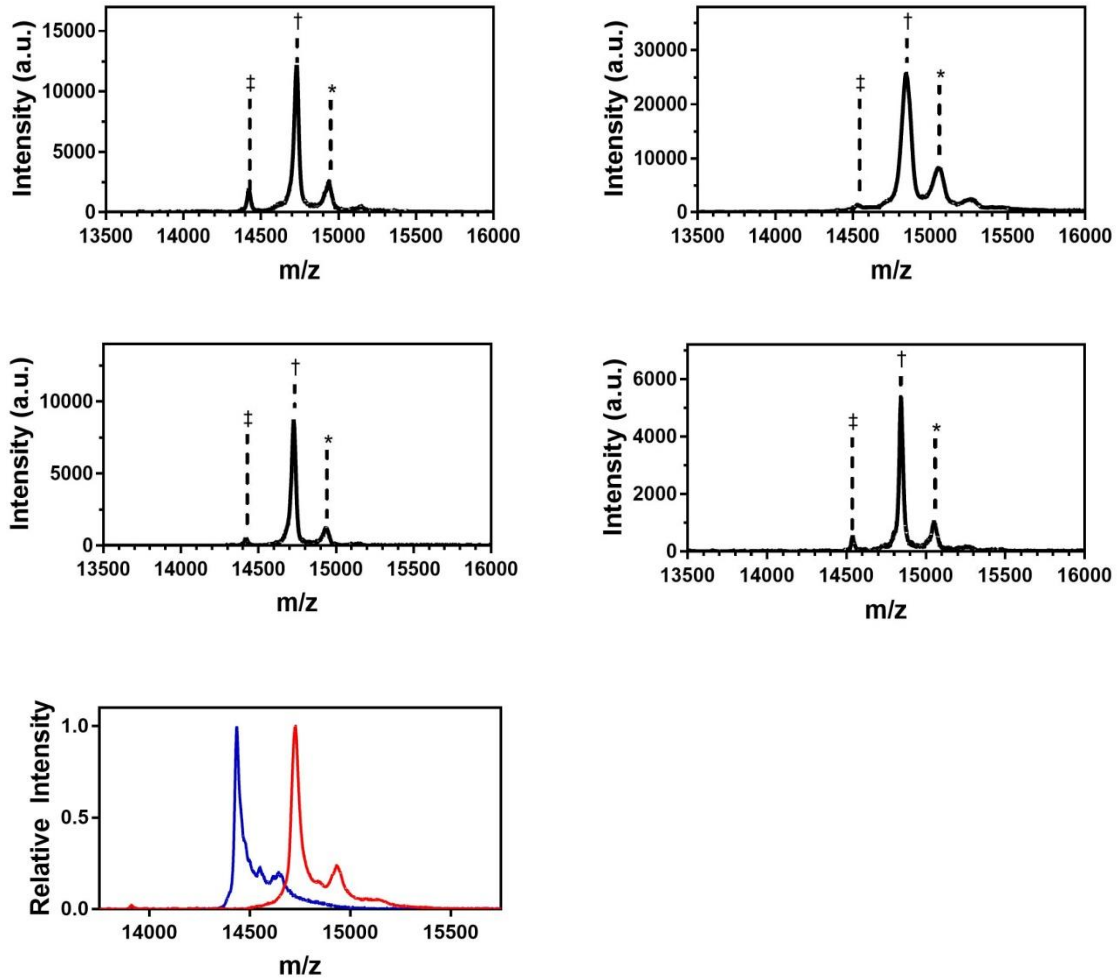


Figure 2.14: Enlarged MALDI MS of labeled α S Variants. Spectra where the region of interest is enlarged for α S-C^{Mcm}₆₂ (Top Left), α S-C^{Mcm}₆₂ δ ₉₄ (Top Right), α S-C^{Mcm}₁₁₄ (Middle Left) and α S- δ ₉₄C^{Mcm}₁₁₄ (Middle Right) where the labeled and unlabeled protein masses are indicated with † and ‡ respectively. MALDI MS of α S-C^{Mcm}₆₂ before (blue) and after 3 hr labeling reaction (red) showing extend of labeling (Bottom Left).

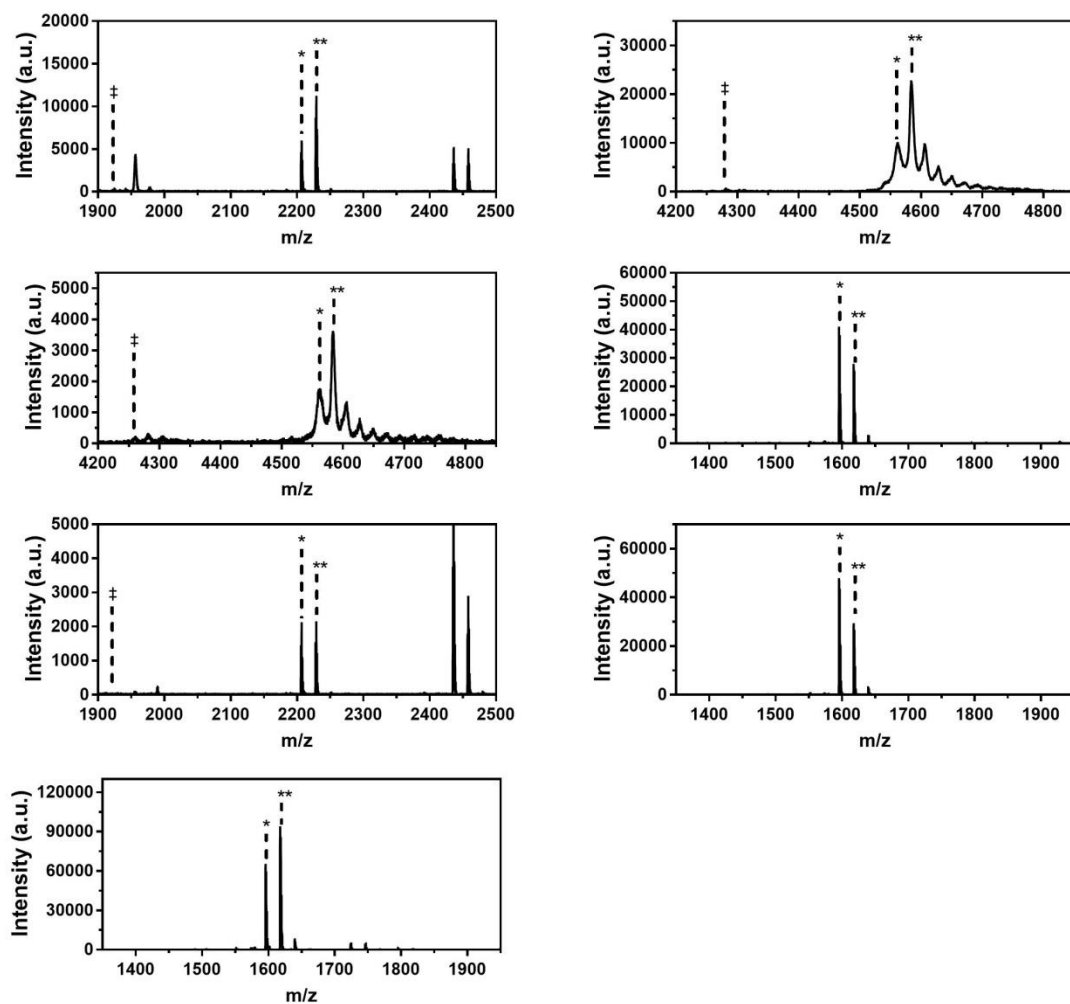


Figure 2.15: Trypsin Digest MALDI MS Characterization of α S Variants. On each plot, the M+H peak of the fragment of interest is marked with * and the M+Na peak is marked with **. The mass corresponding to unlabeled Cys containing protein is denoted with †. The plots show the fragments α S₆₁₋₈₀-C^{Mcm}₆₂ of α S-C^{Mcm}₆₂ (Top Left), α S₁₀₃₋₁₄₀-C^{Mcm}₁₁₄ of α S-C^{Mcm}₁₁₄ (Top Right), α S₁₀₃₋₁₄₀-C^{Mcm}₁₁₄ of α S- δ ₉₄C₁₁₄ (Upper Middle Left), α S₈₁₋₉₄- δ ₉₄ of α S- δ ₉₄C₁₁₄ (Upper Middle Right), α S₆₁₋₈₀-C^{Mcm}₆₂ of α S-C^{Mcm}₆₂ δ ₉₄ (Bottom Middle Left), α S₈₁₋₉₄- δ ₉₄ of α S-C^{Mcm}₆₂- δ ₉₄ (Bottom Middle Right), and α S₈₁₋₉₄- δ ₉₄ of α S- δ ₉₄ (Bottom Left).

§ 2.3 FRET Measurements

FRET Calculations

For FRET measurements, the Förster distance, R_0 , is given in Å by Equation (S1)

$$R_0^6 = \frac{9000(\ln 10)\kappa^2\Phi_D J}{128\pi^5 n^4 N_A} \quad (\text{Eq. 2.1})$$

where κ^2 is a geometrical factor that relates the orientation of the donor and acceptor transition moments, Φ_D is the quantum yield of the donor, n is the index of refraction of the solvent, N_A is Avogadro's number, and J is the spectral overlap integral defined in units of $\text{M}^{-1}\cdot\text{cm}^{-1}\cdot\text{nm}^4$. J is formally defined as

$$J = \int_0^\infty f_D(\lambda)\varepsilon_A(\lambda)\lambda^4 d\lambda \quad (\text{Eq. 2.2})$$

where $\varepsilon_A(\lambda)$ is the molar extinction coefficient of the acceptor at each wavelength λ and $f_D(\lambda)$ is the normalized donor emission spectrum given by

$$f_D(\lambda) = \frac{F_D(\lambda)}{\int_0^\infty F_D(\lambda)d\lambda} \quad (\text{Eq. 2.3})$$

where $F_{D\lambda}(\lambda)$ is the fluorescence of the donor at each wavelength λ . Substituting these results into Equation 2.1, as well as the donor (Mcm) quantum yield, 1.33 for the index of refraction of water, and 2/3 for κ^2 gives the Förster distance. These R_0 values were used to calculate FRET efficiency (E_{FRET}) as a function of distance using Equation 2.4.

$$E_{\text{FRET}} = \frac{1}{1 + \left(\frac{R}{R_0}\right)^6} \quad (\text{Eq. 2.4})$$

Here, E_{FRET} is the FRET efficiency and R is the separation between the chromophores. These values are reported in Table 2.5.

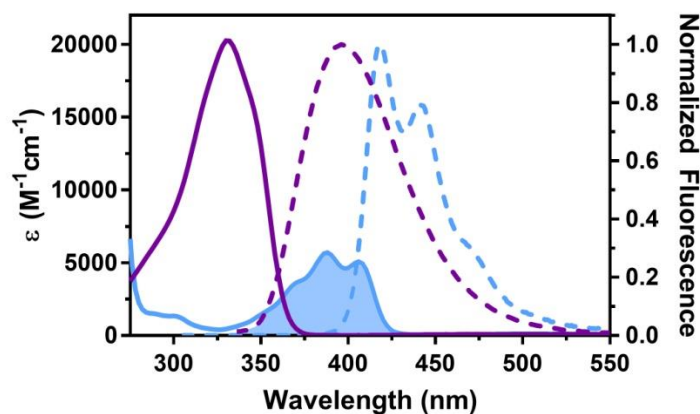


Figure 2.16: Ac-CysMcm and Acd Absorption and Emission Spectra. Ac-Cys^{Mcm} (puple) and Acd (blue) absorption spectra (solid lines) were measured in 20 mM Tris 100 mM NaCl, pH. 7.5. Ac-Cys^{Mcm} (purple) and Acd (blue) fluorescence emission spectra (dashed lines) were measured with excitation at 325 nm for Ac-Cys^{Mcm} and 386 nm for Acd. Spectral overlap between Ac-Cys^{Mcm} emission and Acd absorption is indicated by the shaded area.

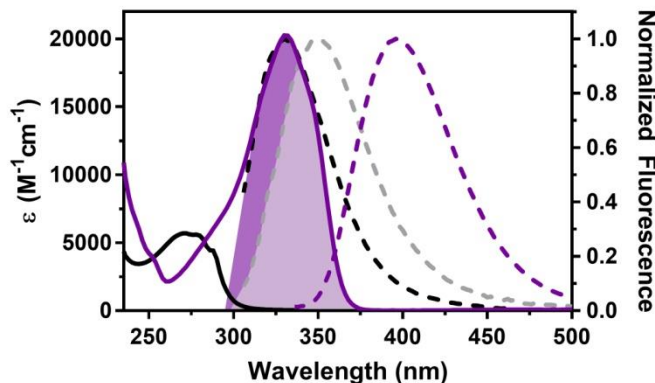


Figure 2.17: Trp and Ac-CysMcm Absorption and Emission Spectra. Ac-Cys^{Mcm} (purple) and Trp (black) absorption spectra (solid lines) were measured in 20 mM Tris 100 mM NaCl, pH. 7.5. Ac-Cys^{Mcm} (purple) and Trp (black or grey) fluorescence emission spectra (dashed lines) were measured with excitation at 325 nm for Mcm-Mal and 295 nm for Trp in the WpOCNC peptide in the absence (black) and presence (grey) of wild-type CaM to illustrate the change in overlap

due to solvation environment. Spectral overlap between Trp emission and Ac-Cys^{Mcm} absorption is indicated by the shaded areas.

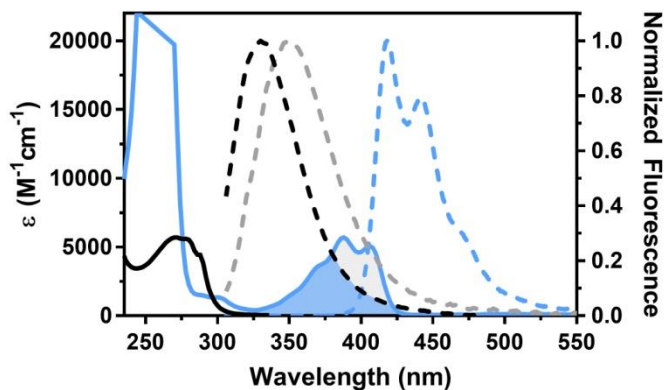


Figure 2.18: Trp and Acd Absorption and Emission Spectra. Trp (black) and Acd (blue) absorption spectra (solid lines) were measured in 20 mM Tris 100 mM NaCl, pH. 7.5. Trp (black and grey) and Acd (blue) fluorescence emission spectra (dashed lines) were measured with excitation at 386 nm for Acd and 295 nm for Trp in the WpOCNC peptide in the absence (black) and presence (grey) of wild-type CaM to illustrate the change in overlap due to solvation environment. Spectral overlap between Trp emission and Acd absorption is indicated by the shaded areas.

CaM FRET Measurements

Following preparation, dried CaM mutants along with pOCNC and WpOCNC peptides were re-dissolved in 15 mM HEPES, 140 mM KCL, and 6 mM CaCl₂, pH 6.70. Concentrations of CaM mutants were determined based on Mcm ($\epsilon_{\text{Mcm325}} = 19010 \text{ M}^{-1}\text{cm}^{-1}$) or Acd ($\epsilon_{\text{Acd386}} = 5700 \text{ M}^{-1}\text{cm}^{-1}$) absorbance while pOCNC and WpOCNC peptide concentrations were determined based on Phe ($\epsilon_{\text{Phe259}} = 189 \text{ M}^{-1}\text{cm}^{-1}$) or Trp ($\epsilon_{\text{Trp276}} = 5579 \text{ M}^{-1}\text{cm}^{-1}$) absorbance. Fluorescence measurements were taken on a PTI QuantaMaster 40 system. Trp emission spectra were collected with an excitation wavelength at 295 nm over an emission range of 305-600 nm, Mcm emission spectra were collected with an excitation wavelength at 325 nm over an emission range of 335-600 nm, and Acd emission spectra were collected with an excitation wavelength at 386 nm over an emission range of 396-600 nm. All spectra were collected with excitation and emission wavelengths set to 5 nm with a 0.25 second integration time and a 1 nm step size. CaM mutants were measured at a concentration of 0.5 μM in the absence of peptide, and the presence of pOCNC or WpOCNC peptide at a concentration of 1 μM . A single control measurement to determine the fluorescence spectrum of the bound WpOCNC peptide was taken at a 1 μM concentration of peptide and a 2 μM concentration of wild-type CaM.

To confirm that the binding affinity of CaM was not perturbed by the introduction of fluorescent probes, the binding affinity was determined for CaM-C^{Mcm}₁₂ δ ₁₁₂. This was performed by mixing labeled and wild-type CaM in varied concentrations with 1 μM WpOCNC at concentration ratios of 0.125:1, 0.25:1,

0.5:1, 1:1 and 1.5:1 of protein to peptide. Trp emission spectra were collected as previously described. For the wild-type protein, the wavelength of the maximum of the Trp fluorescence was used to quantitatively determine the amount of bound peptide, while for CaM-C^{Mcm}₁₂δ₁₁₂ the amount of bound protein was determined via the quenching in the emission at 350 nm. The normalized values were then fit with Equation 2.5 to determine K_d as in previous publications.⁷⁷

$$y = R \frac{(K_d + [P] + [L]) - \sqrt{(K_d + [P] + [L])^2 - 4[P][L]}}{2[P]} \quad (\text{Eq. 2.5})$$

In Equation 2.5 [P] and [L] are the total concentrations of the protein and peptide respectively. The K_d for the wild-type and CaM-C^{Mcm}₁₂δ₁₁₂ proteins were 7.44±6.89 and 21.8±20.5 nM respectively. Given the high affinity of the WpOCNC peptide and the 1 μM concentration in our assay, these represent only estimates of the actual K_d, with significant error. However, we take the stoichiometric binding observed in both cases to indicate that CaM labeling does not substantially disrupt peptide binding.

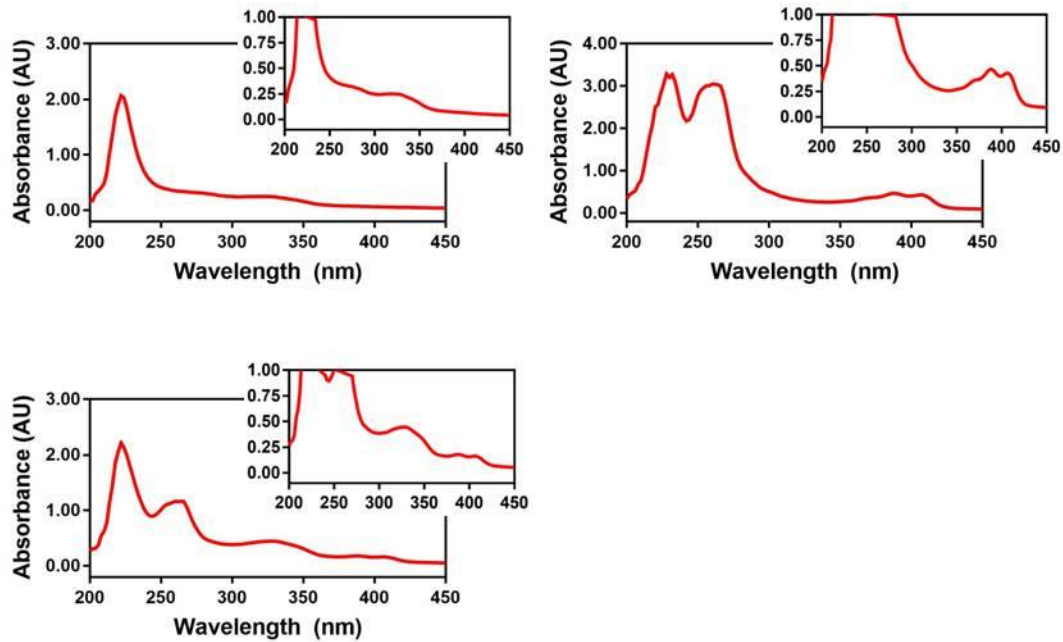


Figure 2.19: CaM Absorbance Measurements. Absorbance spectra of CaM-C^{Mcm}₁₂ (Top Left), CaM-δ₁₁₂ (Top Right) and CaM-C^{Mcm}₁₂δ₁₁₂ (Bottom Left)

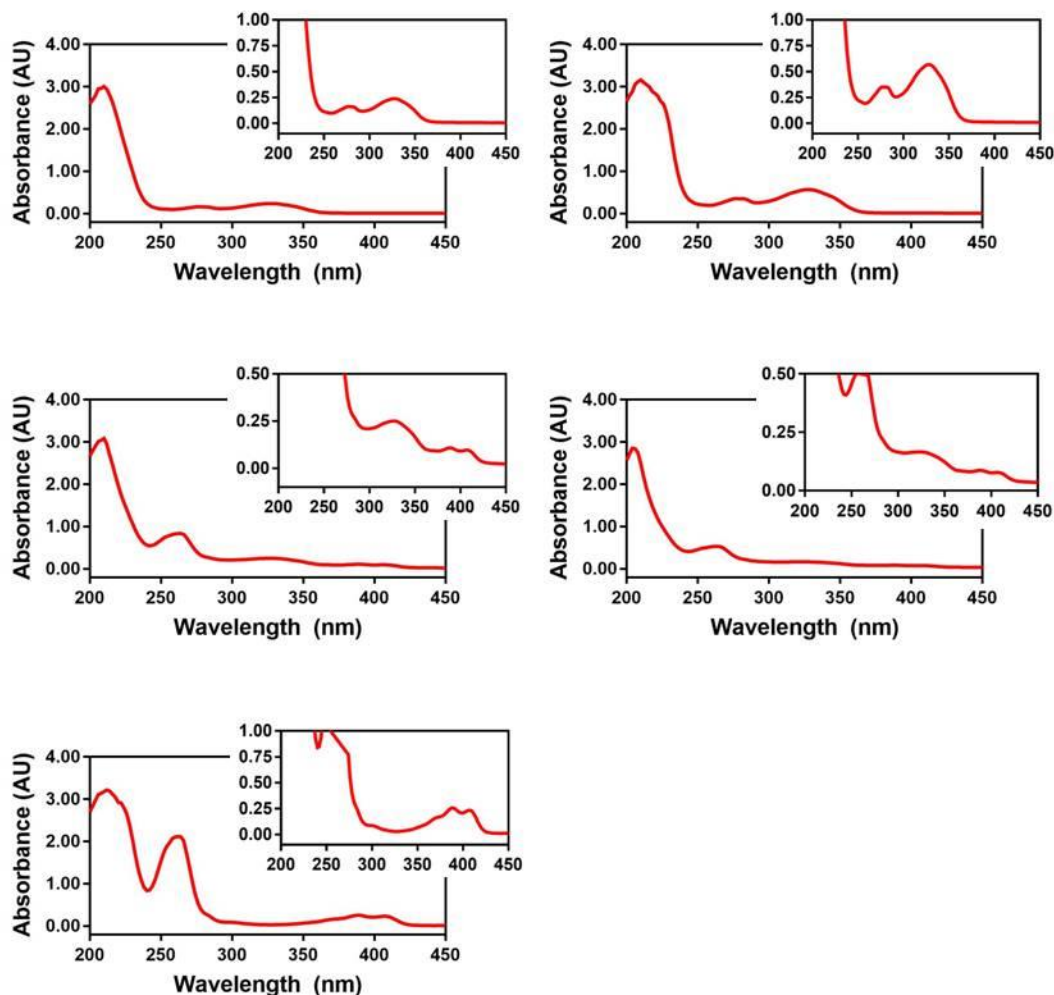


Figure 2.20: αS Absorbance Measurements. Absorbance spectra of $\alpha\text{S-C}^{\text{Mcm}_{62}}$ (Top Left), $\alpha\text{S-C}^{\text{Mcm}_{114}}$ (Top Right) $\alpha\text{S-C}^{\text{Mcm}_{62}} \delta_{94}$ (Middle Left), $\alpha\text{S-C}^{\text{Mcm}_{114}} \delta_{94}$ (Middle Right) and $\alpha\text{S-}\delta_{94}$ (Bottom Left).

Table 2.3: Percent of Cys Containing Protein Labeled by Mcm-Mal

	$\text{CaM-C}^{\text{Mcm}_{12}}\delta_{112}$	$\alpha\text{S-C}^{\text{Mcm}_{62}} \delta_{94}$	$\alpha\text{S-C}^{\text{Mcm}_{114}} \delta_{94}$
Percent Labeled	90.1 %	77.3 %	67.2 %

The percent of Cys containing protein labeled by Mcm-Mal was determined from the ratio of the concentrations of Mcm and Acd calculated from UV

absorbance spectra (Figs 2.19 and 2.20), using the previously described extinction coefficients. These values likely deviate from the true labeling percentage due to background and scattering profiles within the obtained spectra.

Stopped-Flow Measurements.

Stopped-flow experiments were performed with 1 μM Mcm/Acd labelled CaM and 2 μM pOCNC peptide. Each measurement was taken in triplicate following mixing of 20 μL of protein and 20 μL of peptide. For each mixing event the fluorescence emission of Acd (440 ± 40 nm) following excitation of Trp at 295 nm or Mcm at 325 nm. For each measurement, 15000 points were collected over a time range of 30 milliseconds. Nonlinear fits were performed in GraphPad Prism 7.00 where each measurement was fit over the entire time window to the equation

$$Y = Y_0 + Y_M \times \exp(-k \times x) \quad (\text{Eq. 2.6})$$

In Eq. 2.6, Y is the fluorescence intensity as a function of time, x, where Y_0 is the maximum intensity and Y_M is the difference between the maximum intensity and intensity at time zero. The resulting fit values for each curve and the resulting averages are seen in

Comparison of CaM FRET Data to Other Structural Data.

The calculated distances obtained from the experimental Mcm/Acd FRET measurements can be compared to published NMR structures of CaM in the peptide-bound (PDB ID 1SY9) and free (PDB ID 1X02) forms.¹²⁰⁻¹²¹ We selected three representative substructures from the ensembles of 20 low energy structures

reported for each PDB entry. Rather than simply calculate the distances between residues 12 and 112 from C α to C α , we manually placed into these PDB structures models of the C^{Mcm} chromophore (both *R* and *S* adducts) and Acd chromophore (two sidechain rotamers designated A and B in Fig. 2.26) generated from AM1 minimized structures in Gaussian 09.¹²² For Acd, the sidechain orientation was determined by making a Phe mutation at position 112 in PyMol and aligning the Acd chromophore with the Phe sidechain. For C^{Mcm}, a Cys mutant was made at position 12 in PyMol and the sulfur atom in the Mcm model was aligned with the Cys sulfur. The Mcm chromophore was then rotated about the C α -C β -S-C maleimide bond to minimize steric clashes. The Mcm and Acd models used are shown in Figure 2.26. The distances between the highlighted (black arrows) atoms were determined and averaged to identify the center-to-center distance between the two chromophores. The values from the two Acd rotamers and the *R* and *S* adducts of the Mcm chromophore are reported in Table 2.6. Ranges for these values were reported in the main text.

α S FRET Measurements

Purified C^{Mcm}-labelled, δ -labelled, and C^{Mcm}/ δ double-labelled α S proteins were dialyzed into α S buffer. Protein concentrations were determined by absorption spectroscopy (C^{Mcm}: $\epsilon_{325} = 19,010 \text{ M}^{-1}\text{cm}^{-1}$ Acd: $\epsilon_{386} = 5,700 \text{ M}^{-1}\text{cm}^{-1}$).^{77,}
¹²³ Background corrections were performed using a|e 1.2.¹²⁴ Fluorescence steady-state spectra were obtained at protein concentrations of $\sim 0.5 \text{ }\mu\text{M}$ while fluorescence lifetimes were acquired at $\sim 2 \text{ }\mu\text{M}$. Buffers containing varying

concentrations of trimethylamine *N*-oxide (TMAO) in 20 mM Tris and 100 mM NaCl were prepared on the day of the spectroscopy experiments, and the pH of each buffer was readjusted to 7.5 following the addition of TMAO. Both steady-state and lifetime measurements were performed at concentrations of 0, 2, and 4 M TMAO. Steady-state measurements were performed with direct excitation of Mcm at 325 nm, measuring emission from 335-600 nm, with all slit widths set to 5 nm, a step size of 1 nm, and an integration time of 0.25 s per step. Direct excitation of Acd was performed at 386 nm, measuring from 396-600 nm with the same slit width, step size, and integration time.

Steady State Fluorescence Data Fitting and FRET Calculation.

Due to the significant overlap of the Mcm and Acd emission spectra, deconvolution was required to determine the relative Mcm quenching due to FRET. Fitting of spectra containing Mcm and Acd double-labelled protein was performed by minimizing the square difference between the spectra from the double-labelled protein and a sum of linearly weighted single-labelled spectra using the following equation

$$\sum_{\lambda} (I(\lambda)_{DA} - (AI(\lambda)_D - BI(\lambda)_A))^2 \rightarrow \min \quad (\text{Eq. 2.7})$$

Here, $I(\lambda)_{DA}$, $I(\lambda)_D$ and $I(\lambda)_A$ represent the fluorescence intensity at a given wavelength from protein which was double-labelled, single-labelled with only the FRET donor, and single-labelled with only the FRET acceptor, respectively. A and

B are linear weights of the donor and acceptor spectra and are wavelength invariant. The Solver function in Microsoft Excel was used to vary the values of A and B to minimize the sum of this square difference across all emission wavelengths. The FRET efficiency (E_{FRET}) was then directly obtained from the linear weight of donor spectrum, A, where $A = 1 - E_{FRET}$.

For CaM experiments, distances were determined using the canonical Förster equation (Eq. 2.4). Error bars for CaM experiments were obtained via propagation of error through the determination of A and the canonical Förster equation. Interfluorophore distance values for α S experiments were calculated using a polymer scaled version of the Förster equation of the form

$$E_{FRET} = \sum_r P(r)/(1 + (r/R_0)^6) \quad (\text{Eq. 2.8})$$

were the probability distribution was set to the functional form for a Gaussian chain.

$$P(r) = 4\pi r^2 \left(\frac{3}{2\pi \langle r^2 \rangle} \right)^{3/2} \exp\left(-\frac{3}{2} \frac{r^2}{\langle r^2 \rangle}\right) \quad (\text{Eq. 2.9})$$

Error bars were obtained via propagation of experimental error through both determination of A and distance determination using the Gaussian-chain scaled Förster equation.

Fluorescence Lifetime Data Fitting and FRET Calculation.

Fluorescence lifetime measurements were acquired via time-correlated single photon counting using a 340 nm LED light source. Lifetime spectra were all acquired at 380 nm with slit widths ≤ 8 nm and collection over a 199 ns time window, which was divided into 4096 lifetime bins. Collection was terminated when a single bin reached a count of 10,000 photons. TCSPC data were fit to single or bi-exponential decays using PowerFit-10. The FRET efficiency (E_{FRET}) was calculated as one minus the ratio of the M_{cm} lifetime in the presence of Acd to the M_{cm} lifetime in the absence of Acd.

Comparison of α S FRET Data to Other Structural Data.

The calculated distances obtained from the experimental M_{cm}/Acd FRET measurements were compared to published structural ensembles. The two ensembles were previously published in Allison *et al.* and Nath *et al.*, derived from a molecular dynamics simulation restrained with paramagnetic relaxation enhancement data and a Monte Carlo simulation with constraints from FRET, respectively.^{68, 125-126} To obtain distances for comparisons, the distances between C α atoms for every residue pair in the sequence were extracted from each structure within each ensemble. Since α S is intrinsically disordered, it is unlikely that the structural ensembles represent the full distribution of states, therefore distances were averaged in a Flory-scaling like protocol similar to the analysis in Nath *et al.*⁶⁸ For each ensemble, the inter-residue distances were averaged over all residue pairs spaced by the same primary sequence separation over all

structures, rather than just averaging over a single specific residue pair over all structures. The average and standard deviation of the inter-residue distance as a function of primary sequence separation is shown in Figure 2.25, along with the probe separation distance calculated from the experimental Mcm-Acd FRET efficiency via the gaussian chain polymer-scaled version of the Förster equation. Since there are no published ensembles for the structure of α S in the presence of TMAO, this comparison could only be performed for measurements in the absence of TMAO.

§ 2.3 Results

Characterization of Mcm-Mal and Mcm-Br

We considered two ways of installing Mcm, either through conjugate addition of a 3-maleimide derivative (Mcm-Mal) or through an S_NAr reaction of a 3-bromo derivative (Mcm-Br, Fig. 2.3). Mcm-Br and Mcm-Mal were synthesized by Naoya Ieda as detailed in Figure 2.3.

We tested the reaction of Mcm-Mal and Mcm-Br with Cys in aqueous buffer and found that reactions of Mcm-Mal were much faster (complete within a few seconds at 10 μ M), with a >20-fold increase in fluorescence at 390 nm (Figs. 2.4 and 2.7). High performance liquid chromatography (HPLC) and mass spectrometry (MS) analysis indicated that the very rapid reaction initially yielded the conjugate addition product shown in Fig. 1, without maleimide ring opening (Figs. S3 and S5). This product then isomerizes on a timescale of hours to a compound of identical mass, which we attribute to the lactam form S5 (Figs. 2.6

and 2.8). This is similar to reactions recently reported for detection of Cys in its amino acid form by Tong *et al.*¹²⁷ The authors ascribed the high levels of fluorogenicity that they observed to relief of two quenching mechanisms, one by conjugate addition, and the other by opening of the maleimide ring. Since this second step involves the Cys α -amine, we also tested the reaction of Mcm-Mal with *N*-acetyl cysteine (Ac-Cys), which is more representative of Cys in proteins. The resulting >20-fold turn-on of fluorescence was comparable to that seen in reactions with Cys (Fig. 2.21) These Ac-Cys/Mcm-Mal data, together with our observations of the faster fluorescence turn-on relative to ring-opening in Cys/Mcm-Mal reactions, imply that the conjugate addition is the primary determinant of fluorescence turn-on.

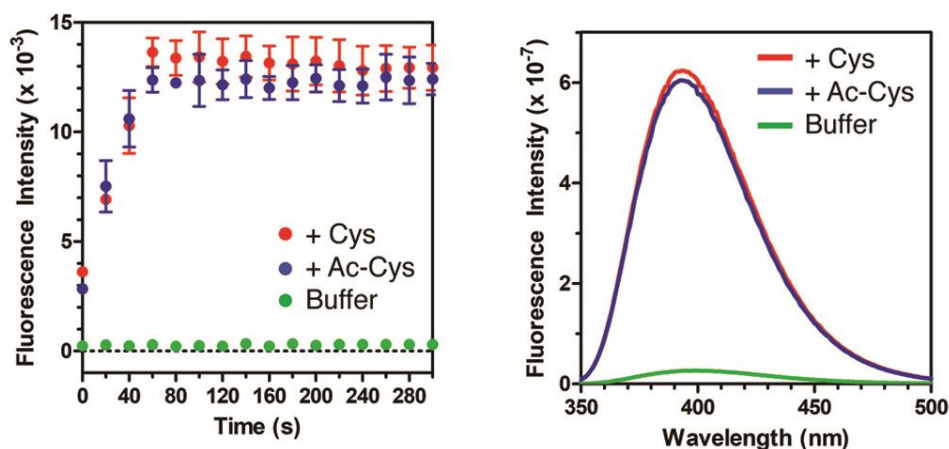


Figure 2.21: Mcm-Mal turn-on experiments. Fluorescence intensity of 10 μ M Mcm-Mal in Tris buffer (100 mM, pH 7.0, DMSO 1%) alone or mixed with 1 equiv Cys or Ac-Cys. Fluorescence excitation at 325 nm, emission monitored at 400nm; background corrected data are shown, see raw data in Fig. 2.9. Inset: Fluorescence spectra acquired after 6 h of incubation.

Following these trial reactions with small molecules, we evaluated the turn-on of Mcm-Mal fluorescence in reactions at a Cys residue in two proteins, calmodulin (CaM) and α -synuclein (α S). CaM is a calcium sensor protein that undergoes a dramatic conformational change in the presence of calcium and helical peptide binding partners.¹²⁸⁻¹²⁹ α S is a disordered protein that contributes to the pathogenesis of Parkinson's disease.^{19, 130} These provide examples of the value of FRET to monitor conformational change and protein/protein interactions. Cys mutants of CaM and α S (CaM-C₁₂, α S-C₆₂, α S-C₁₁₄) were reacted with Mcm-Mal, and fluorescence emission was monitored at 390 nm. In all cases, we observed a turn-on of fluorescence, but this was significantly less than what was observed with Cys or Ac-Cys, due in part to higher background fluorescence in protein reactions because of the need to keep the protein Cys reduced. Matrix-assisted laser desorption ionization (MALDI) MS analysis of whole proteins and trypsin digests confirmed that modification took place exclusively at the Cys, without ring-opening. (Fig. 2.11-2.15)

Use of Mcm-Acd as a FRET Pair

In order to perform FRET measurements, one must doubly label the protein with a donor/acceptor probe pair. We installed the Mcm/Acd FRET pair by using site-directed mutagenesis and amber codon suppression to generate constructs with single Cys and Acd mutations, followed by reaction of Cys with Mcm-Mal. (Fig. 2.1) The reaction progress was monitored using the turn-on of Mcm-Mal fluorescence and MALDI MS. (Figs. 2.11 and 2.13) We commonly use C-terminal

intein-His₆ tags to facilitate isolation of full-length proteins containing unnatural amino acids from truncated proteins.⁹⁵ His₆ tags can be cleaved with β -mercaptoethanol either before or after Mcm-Mal labelling. Due to the highly sensitive nature of FRET measurements, we typically perform a second purification pass by FPLC or HPLC.

We have used our doubly labeled α S and CaM constructs in proof-of-principle FRET experiments to demonstrate the value of the Mcm/Acd pair. For α S, we monitored a conformational change of the monomer induced by addition of the compacting osmolyte trimethylamine *N*-oxide (TMAO). TMAO is a naturally occurring osmolyte that is abundant in aquatic organisms, and counteracts the denaturing effects of urea required for osmotic pressure regulation.¹³¹⁻¹³² Previous work has demonstrated that α S undergoes successive compaction with exposure to increasing amounts of TMAO.¹³³⁻¹³⁵

Here, we chose to monitor TMAO-induced conformational changes in α S by introducing Cys residues at positions 62 or 114 and Acd at position 94. Single-labelled constructs containing either Mcm or Acd and the double-labelled constructs were prepared as described above. Fluorescence spectra for all of the single and double-labelled constructs were collected in 0, 2, and 4 M TMAO. Singly-labelled spectra were used to analyse the double-labelled spectra to determine FRET efficiency (E_{FRET}). This was performed by linearly weighting the singly-labelled spectra and minimizing the squared difference between the doubly-labelled spectrum and the sum of the linearly weighted single-labelled spectra at each wavelength. (Fig. 2.23) This allows one to correct for changes in quantum

yield or spectral overlap when the chromophore changes environment, and directly affords the relative M_{cm} quenching (A in Fig. 2.23) and E_{FRET} (E_{FRET} = 1-A) to determine the FRET distance through application of an appropriate form of the Förster equation.

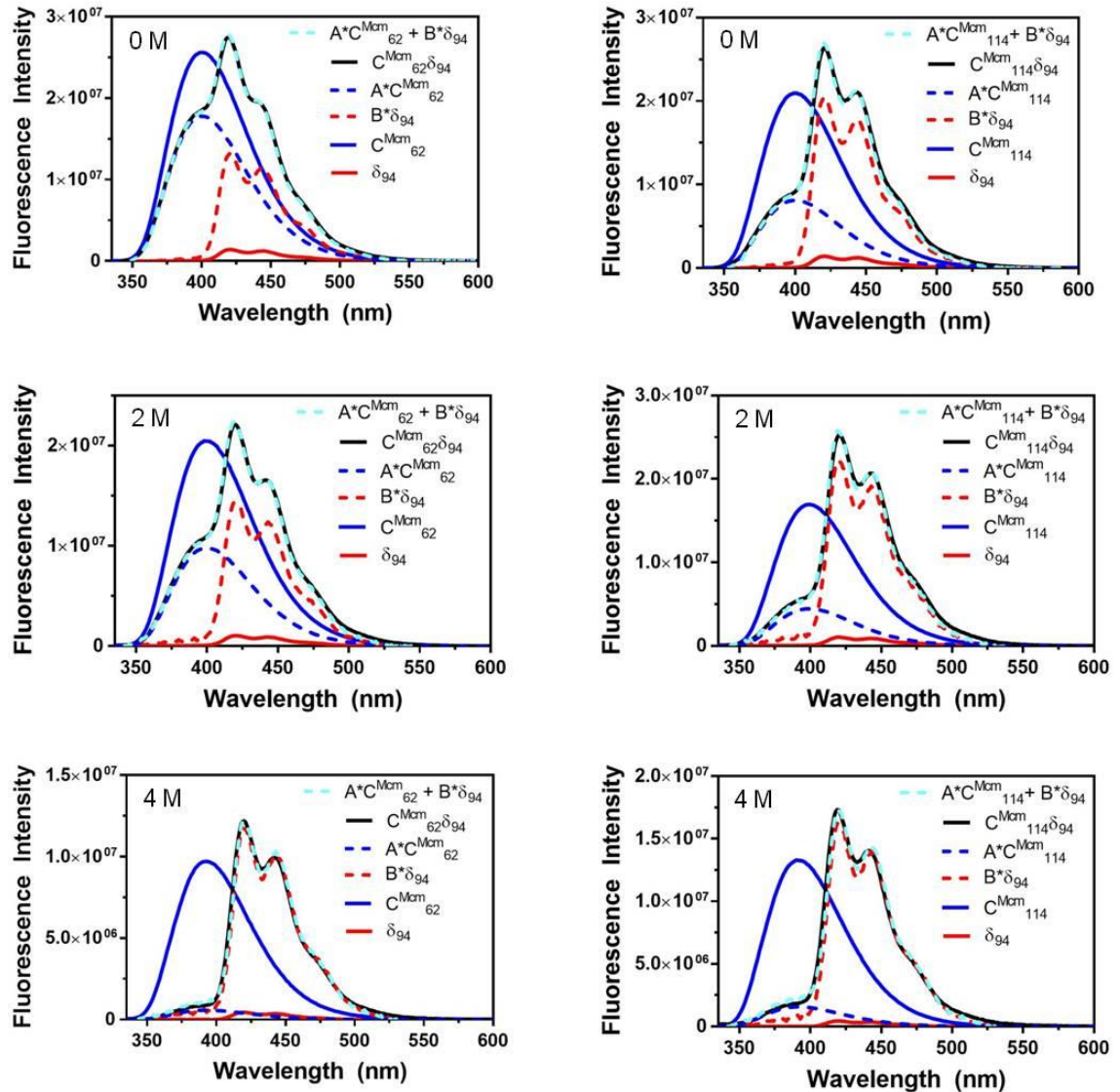


Figure 2.22: α S Fluorescence Spectra and FRET Fitting Results. Donor-only (C^{Mcm} , solid blue line), acceptor-only (δ , solid red line), or double-labelled (C^{Mcm}/δ , solid black line) spectra are shown overlaid with the results of fitting the double-labelled spectra to a weighted sum of the single-labelled spectra (light blue dashed

line). The weighted component spectra are shown in dashed blue (donor-only) and dashed red (acceptor-only) lines. Spectra were acquired in buffer containing 0, 2, or 4 M TMAO with excitation at 325 nm.

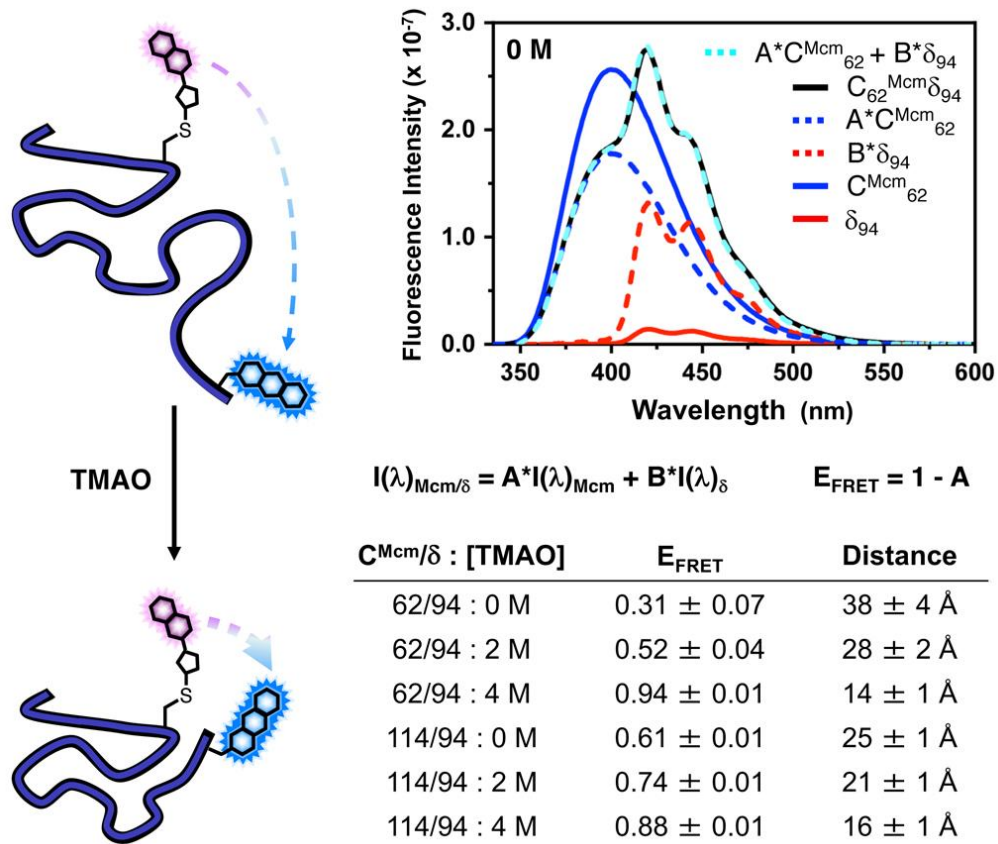


Figure 2.23: α S TMAO FRET Experiments. Left: Incubation of α S with TMAO causes protein compaction, which can be monitored by FRET. Top Right: Fluorescence emission spectra (325 nm excitation) of 1 μ M concentrations of α S- $C^{Mcm}_{62}\delta_{94}$ and the corresponding singly-labelled α S- $C^{Mcm}_{62}\delta_{94}$ and α S- $C^{Mcm}_{62}\delta_{94}$ constructs in 0 M TMAO. Deconvolution of the doubly labelled spectrum by fitting to a weighted sum of the singly-labelled spectrum are shown. Bottom Right: E_{FRET} and corresponding inter-fluorophore distances for α S constructs under varying conditions determined using spectral deconvolution.

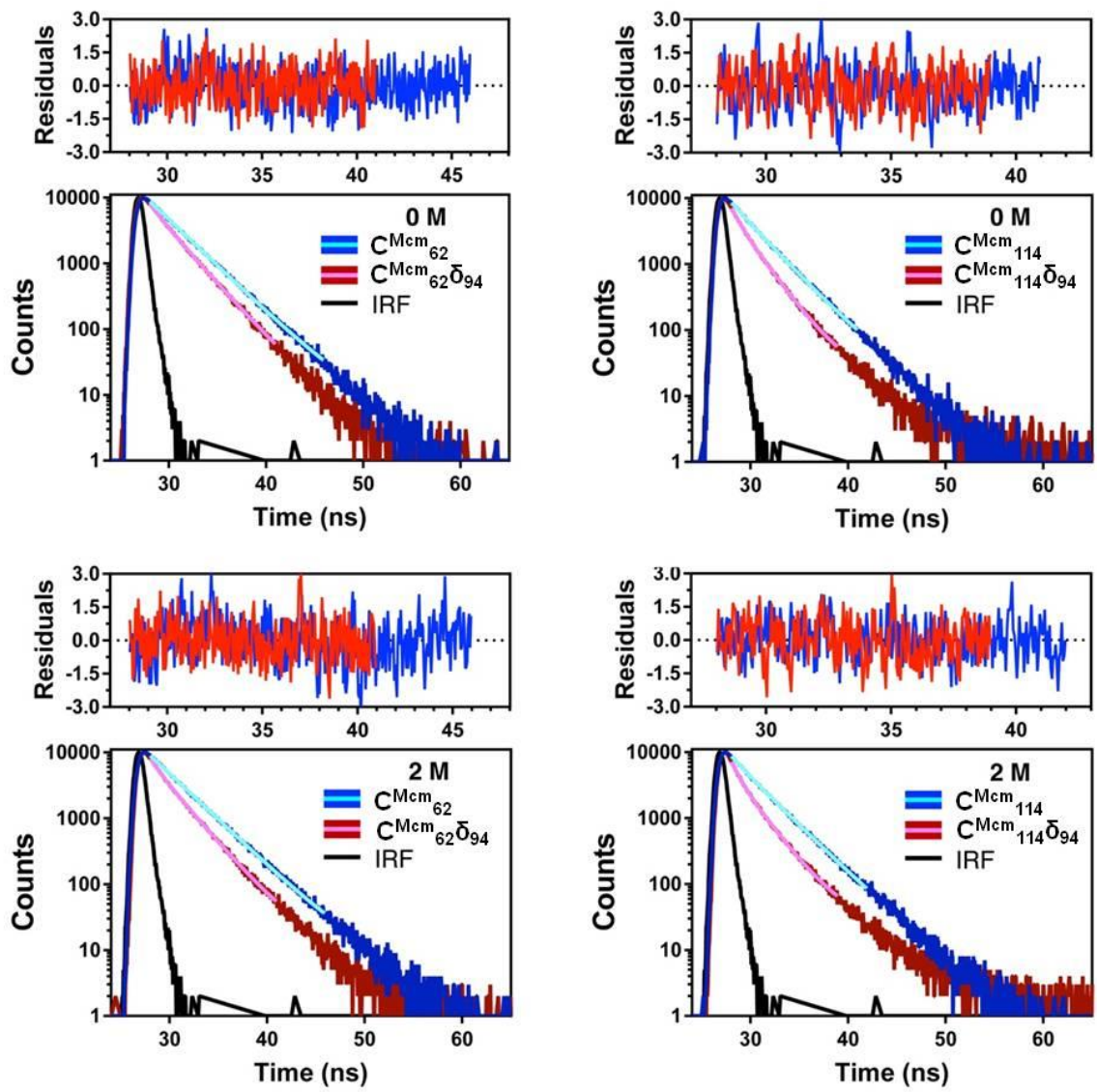


Figure 2.24a: α S Fluorescence Spectra and FRET Fitting Results. TCSPC data sets in 0 M or 2 M TMAO are shown for donor-only (C^{Mcm} , dark blue line) or double labelled (C^{Mcm}/δ , dark red line). Bi-exponential fits to the data are shown in light blue (donor only) or pink (double labelled); the instrument response function (IRF) is shown in black. Weighted residuals are shown in corresponding colors above each fluorescence lifetime plot.

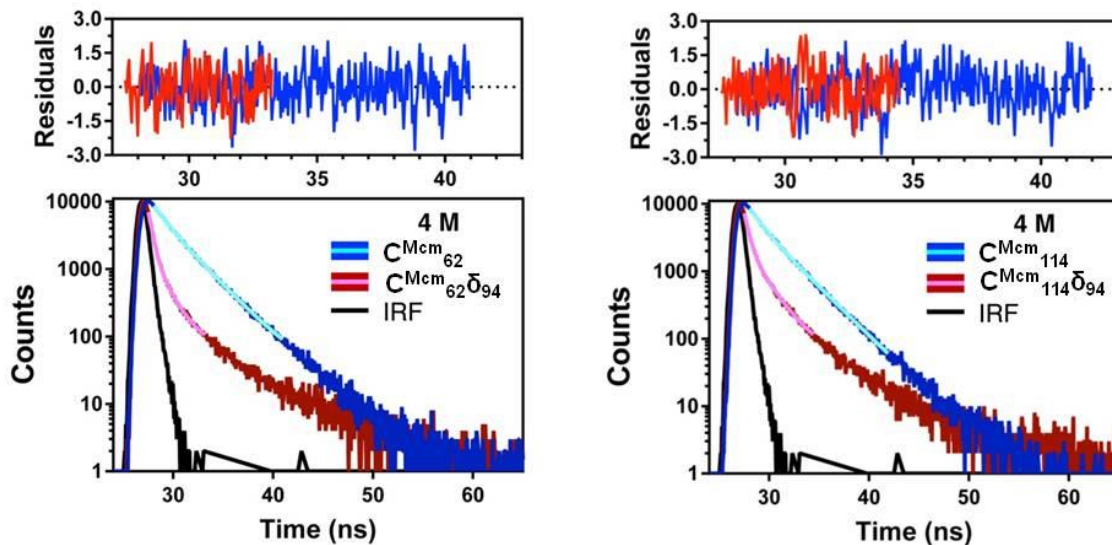


Figure 2.24b: α S Fluorescence Spectra and FRET Fitting Results. TCSPC data sets in 4 M TMAO are shown for donor-only (C^{Mcm} , dark blue line) or double labelled (C^{Mcm}/δ , dark red line). Bi-exponential fits to the data are shown in light blue (donor only) or pink (double labelled); the instrument response function (IRF) is shown in black. Weighted residuals are shown in corresponding colors above each fluorescence lifetime plot.

Table 2.4: TCSPC Fluorescence Lifetime Values

Protein	TMAO	τ_1	τ_2	Amp. 1	Amp. 2	$\tau_{Avg.}$	κ^2	E_{FRET}
$\alpha S-C^{Mcm}_{62}$	0 M	3.10	-	2.47	-	3.10	0.90	-
	2 M	3.12	-	3.03	-	3.12	1.16	-
	4 M	2.80	0.86	0.79	0.55	2.01	0.93	-
$\alpha S-C^{Mcm}_{114}$	0 M	2.85	0.95	1.01	0.25	2.47	1.20	-
	2 M	3.03	0.96	1.00	0.32	2.53	0.83	-
	4 M	3.01	1.13	0.68	0.76	2.12	0.96	-
$\alpha S-C^{Mcm}_{62}\delta_{94}$	0 M	2.74	1.29	0.93	0.50	2.23	0.84	0.28
	2 M	2.60	1.01	1.02	0.78	1.92	0.95	0.39
	4 M	1.90	0.39	0.17	2.81	0.48	0.97	0.76
$\alpha S-\delta_{94}C^{Mcm}_{114}$	0 M	2.17	0.71	0.70	0.71	1.43	1.07	0.42
	2 M	0.82	2.31	0.67	1.10	1.38	1.04	0.45
	4 M	2.07	0.42	0.28	2.68	0.57	0.99	0.73

The 0 M data are consistent with Flory scaling models of inter-residue distances in disordered proteins. (Fig. 2.25) Moreover, we observe compaction for both label positions with increasing amounts of TMAO, as previously observed in this system with other FRET pairs.^{134, 136} These results were corroborated by measuring the change in Mcm fluorescence lifetime for the two double-labelled pairs relative to the analogous Mcm single-labelled proteins.

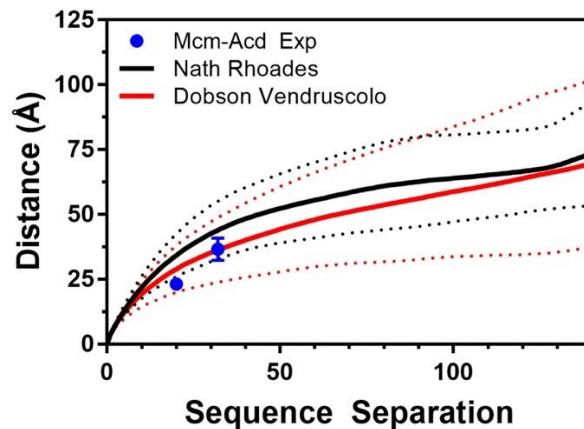


Figure 2.25: Mcm/Acd FRET Data Compared to Previous Structural Models of α S Ensembles.

For CaM, we monitored the conformational change induced by peptide binding. Acd-labelled CaM was expressed and reacted with Mcm-Mal to give CaM-C^{Mcm}₁₂ δ ₁₁₂. Based on existing crystal and NMR structures of CaM, we expected that a significant change in E_{FRET} would occur due to a decrease in distance between positions 12 and 112 upon peptide binding. (Fig. 2.28) We note that in several CaM crystal structures, an extended conformation is observed, making the 12/112 inter-residue distance >50 Å. However, NMR studies have shown that the N- and C-terminal lobes of CaM are dynamic, and that the conformations sampled in solution are more compact.¹³⁷

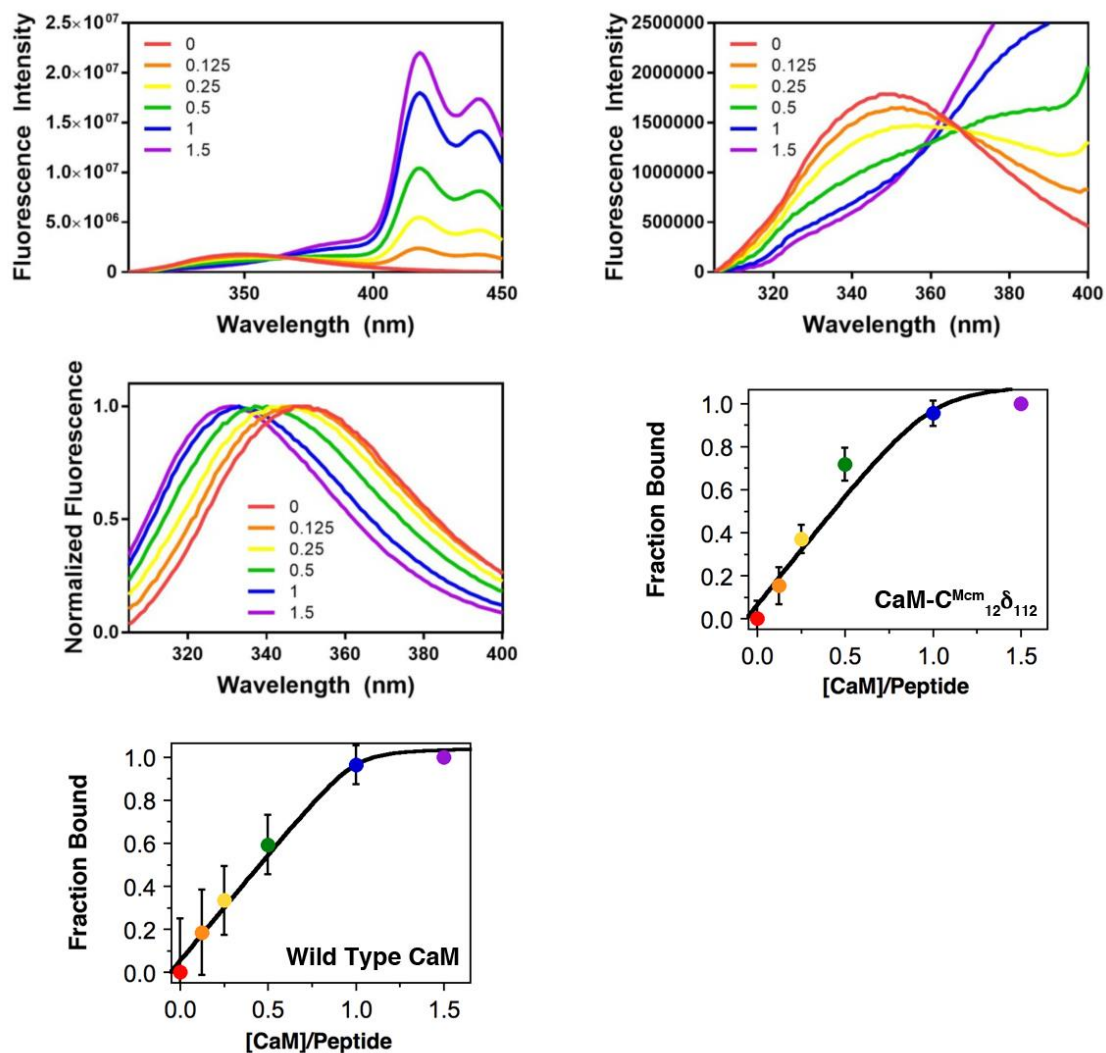


Figure 2.26: Fluorescence Assay to Determine Binding of Labeled CaM. Fluorescence spectra of WpOCNC binding to CaM-C^{Mcm}₁₂δ₁₁₂ (Top Left and Right) and CaM-WT (Middle Left) with the relative concentration of CaM:WpOCNC indicated in the legend. Plots of fraction of WpOCNC bound for both CaM-C^{Mcm}₁₂δ₁₁₂ and CaM-WT shown as a function of the concentration ratio of CaM:WpOCNC with fits from Equation (S5) (Middle Right and Bottom Left).

Table 2.5: CaM + WpOCNC FRET

	Free	Bound
Φ_D	0.16	0.17
J ($M^{-1}\cdot cm^{-1}\cdot nm^4$)	7.042×10^{13}	7.137×10^{13}
R_0 (\AA)	24.6	24.3
E_{FRET}	0.676 ± 0.002	0.405 ± 0.001
Distance (\AA)	25.9 ± 0.1	21.8 ± 0.1

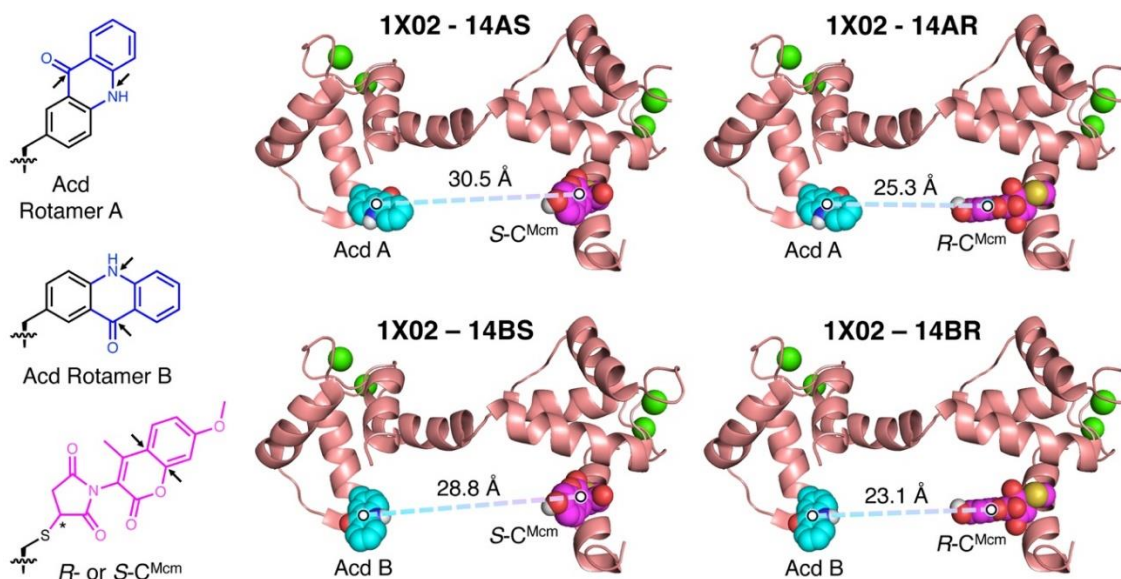


Figure 2.27: CaM Structural Models. Left: Chromophore structures used in modelling Acd or C^{Mcm} in CaM structures. Acd rotamers were superimposed on the lowest energy rotamer of a Phe₁₁₂ mutant in each CaM structure. R- or S-Mcm-Mal thiol adducts were docked onto Cys₁₂ mutants in each CaM structure. The distances between the atoms indicated by arrows were computed and averaged to determine the chromophore separation for comparison to FRET data. Right: Example of the four models of Mcm/Acd-labelled CaM generated from PDB-ID 1X02 structure 14, corresponding to Acd rotamer A or B with R- or S-C^{Mcm}. Other structures used in the analysis reported in Table 2.5 were generated in a similar fashion from structures in PDB ID 1X02 or 1SY9.

Table 2.6: CaM Structural Model Data

PDB ID	Distance (Å)	PDB ID	Distance (Å)	PDB ID	Distance (Å)
1X02 – 14AS	30.5	1X02 – 3AS	36.7	1X02 – 7AS	37.3
1X02 – 14BS	28.8	1X02 – 3BS	33.8	1X02 – 7BS	34.9
1X02 – 14AR	25.3	1X02 – 3AR	36.6	1X02 – 7AR	42.3
1X02 – 14BR	23.1	1X02 – 3BR	33.6	1X02 – 7BR	40.7
	26.9 ± 3.4		35.2 ± 1.7		38.8 ± 3.3
1SY9 – 11AS	16.5	1SY9 – 3AS	21.9	1SY9 – 17AS	20.2
1SY9 – 11BS	15.5	1SY9 – 3BS	15.2	1SY9 – 17BS	18.1
1SY9 – 11AR	16.2	1SY9 – 3AR	21.7	1SY9 – 17AR	20.0
1SY9 – 11BR	15.8	1SY9 – 3BR	19.8	1SY9 – 17BR	18.3
	16.0 ± 0.4		19.1 ± 1.1		19.6 ± 3.1

Structure names correspond to PDB ID, followed by substructure number, Acd A or B rotamer, and C^{Mcm} R or S stereochemistry at starred carbon in Fig. 2.27.

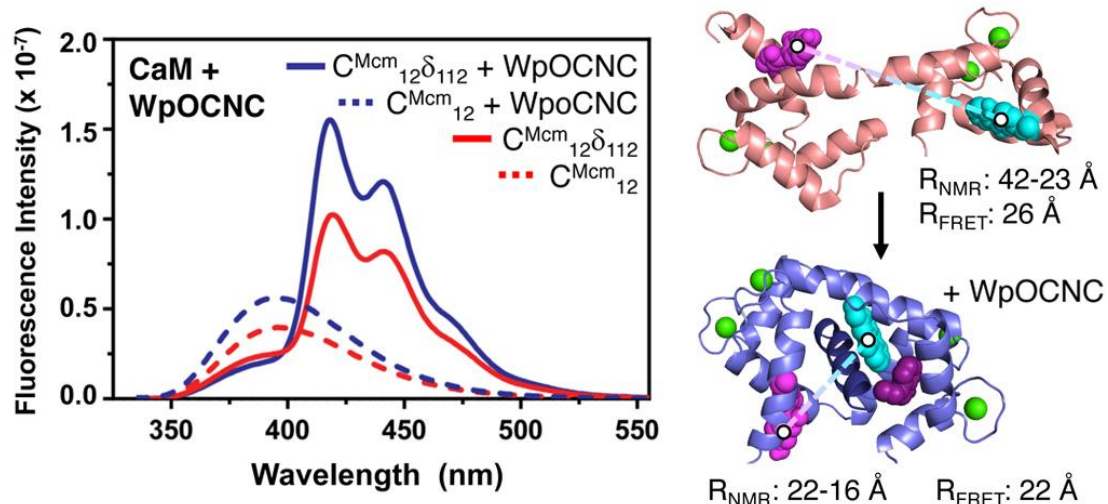


Figure 2.28: CaM Peptide binding FRET. Left: Fluorescence emission spectra of CaM-C^{Mcm}₁₂δ₁₁₂ (325 nm excitation) obtained in the presence and absence of the WpOCNC binding peptide. Note: The apparently small FRET change is the result of an opposing increase in Mcm emission upon binding. This can be accounted for by comparison to CaM-C^{Mcm}₁₂ emission with and without WpOCNC. Full spectral deconvolution to obtain E_{FRET} values is described in ESI. Right: Images of CaM in the presence and absence of the pOCNC peptide rendered from PDB entries 1X02¹²¹ and 1SY9,¹²⁰ respectively. Distances obtained from E_{FRET} measurements are consistent with the CaM conformations in these NMR studies when models of C^{Mcm} and Acd are included.

Fluorescence spectra of CaM-C^{Mcm}₁₂δ₁₁₂ were obtained in Ca²⁺-buffered solutions, and indeed we observed significant Mcm-to-Acd FRET, evidence that positions 12 and 112 were closer than 40 Å. Upon addition of WpOCNC, a Trp derivative of a known high-affinity CaM binding peptide,^{120, 138} we observed an increase in FRET. As in our αS studies, we also obtained spectra for singly-labelled CaM-C^{Mcm}₁₂ and CaM-δ₁₁₂ under identical conditions, and fit a sum of these spectra to the CaM-C^{Mcm}₁₂δ₁₁₂ spectrum to quantitatively determine E_{FRET}. The resulting inter-residue distances of 26 Å and 22 Å in the absence and presence of the WpOCNC peptide, respectively, are in reasonable agreement with Mcm-

and Acd-labelled CaM models based on existing NMR structures (Fig. 2.27 and 2.28).¹²⁰⁻¹²¹ These studies demonstrate the value of Mcm/Acd FRET in monitoring a binding induced conformational change.

Trp-Mcm-Acd Three Color FRET

The longer wavelength excitation (325 nm) and emission of Mcm and Acd allowed us to selectively observe these changes in the presence of the Trp residue in WpOCNC. However, we can also take advantage of the spectral overlap of Trp emission with both Mcm and Acd to perform three color FRET experiments, monitoring binding and conformational change events in the same experiment. Although deconvolution to extract distance information is not reliable, FRET can still be used to measure binding kinetics. Using a stopped-flow fluorometer, we observed rapid binding of WpOCNC through 295 nm excitation, monitoring Acd emission at 420 nm. In the same experimental setup, we excited at 325 nm to monitor the conformational change in CaM alone. We found that the rates of binding (E_{X295} : $338 \pm 93 \text{ s}^{-1}$) and conformational change were nearly identical (E_{X325} : $306 \pm 40 \text{ s}^{-1}$), implying that the two processes are concerted. This is consistent with previous experiments showing that CaM populates compact structures even in the absence of peptide.^{121, 137} Our data support the idea that WpOCNC stabilizes existing compact CaM conformations rather than binding and then inducing a large conformational change. We note that the rates of binding and conformational change are also consistent with previous data for CaM binding of several peptides ($\sim 500 \text{ s}^{-1}$ at $1 \mu\text{M}$ peptide and CaM).¹³⁹⁻¹⁴¹

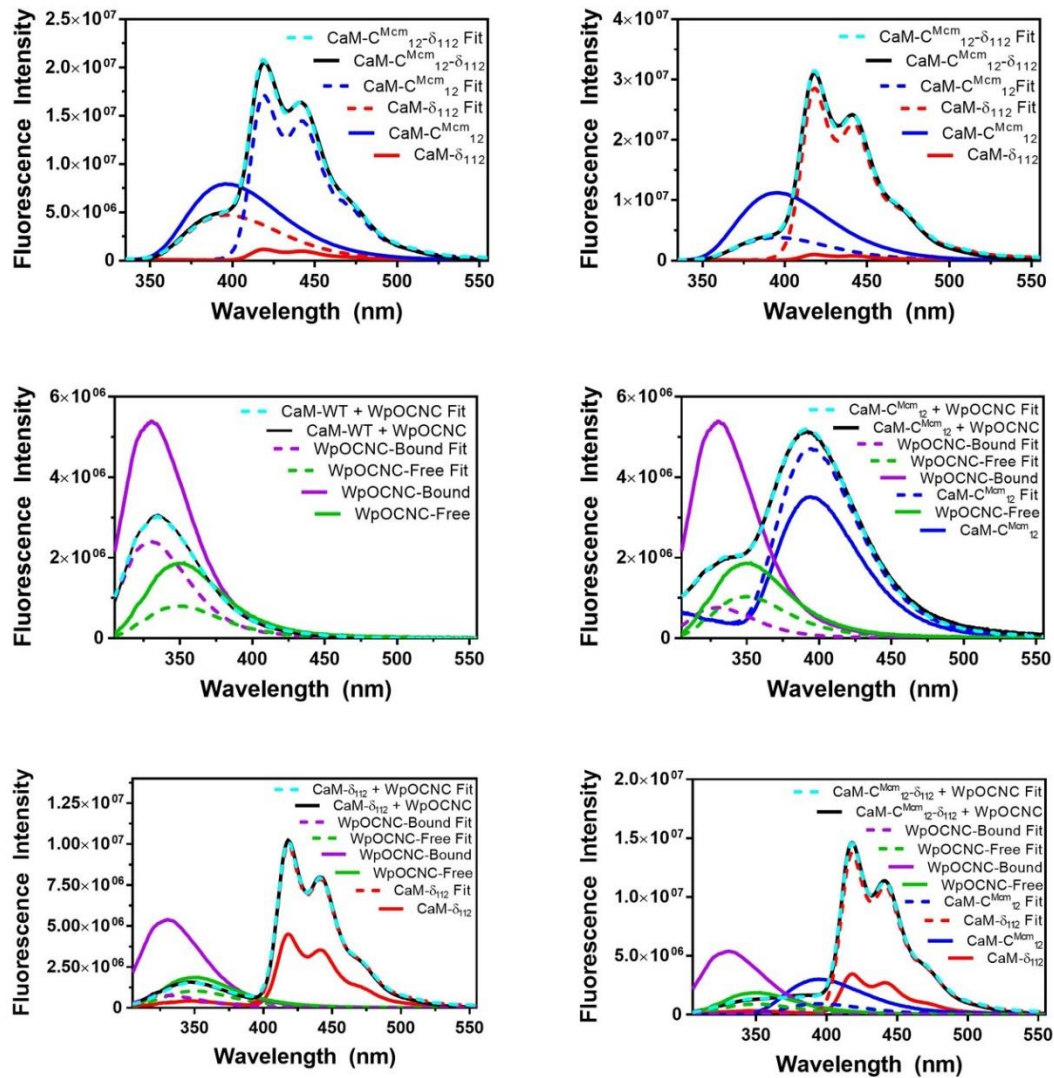


Figure 2.29: CaM Steady State FRET Measurements. Plots display the background subtracted (solid) and fit (dashed) spectra for CaM-CMcm12 δ 112 (Top Left) following excitation at 325 nm, CaM-CMcm12 δ 112 bound to WpOCNC (Top Right) following excitation at 325 nm, WpOCNC bound to wild-type CaM (Middle Left) following excitation at 295 nm, CaM-CMcm12 bound to WpOCNC (Middle Right) following excitation at 295 nm, CaM- δ 112 bound to WpOCNC (Bottom Left) following excitation at 295 nm and CaM-CMcm12 δ 112 bound to WpOCNC (Bottom Right) following excitation at 295 nm. WpOCNC-Free spectra were obtained from measurement of WpOCNC in buffer, while the WpOCNC-Bound spectrum was obtained from measuring WpOCNC bound to wild-type CaM at a 2:1 ratio of protein to peptide with final peptide concentration of 1 μ M.

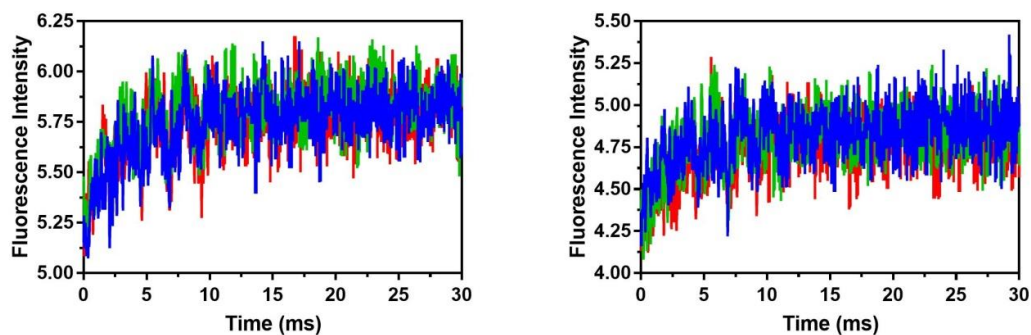


Figure 2.30: CaM Stopped-Flow FRET Measurements. Acd emission of CaM- $C^{Mcm}_{12}\delta_{112}$ during binding of WRRIAR was monitored following excitation of Mcm at 325 nm (left) or Trp at 295 nm (right). Acd emission was monitored at 440 ± 40 nm using an Edmund Optics filter. Each plot contains three single stopped-flow shots which were collected following four wasted shots. Protein and peptide were mixed in a 1:2 concentration ratio producing final concentrations of $0.5 \mu\text{M}$ protein.

Table 2.7: Stopped-Flow Data Fitting

Parameter	Ex. 295 nm	Std. Dev.	Ex. 325 nm	Std. Dev.
Y_0	4.814	0.028	5.821	0.027
Y_M	-0.517	0.072	-0.531	0.051
k	0.338	0.093	0.306	0.040

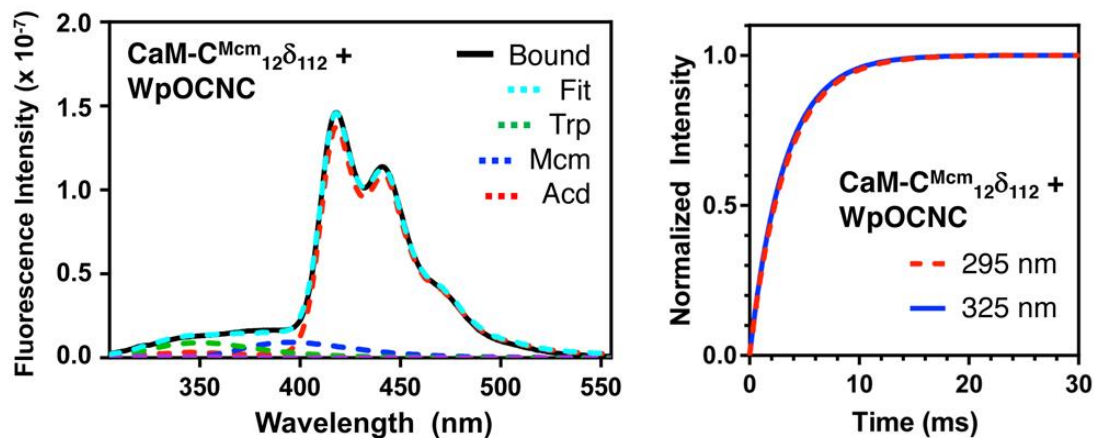


Figure 2.31: Three Color FRET and Rapid Mixing Measurements of Peptide Binding Kinetics. Left: Emission spectrum of CaM-C^{Mcm}₁₂δ₁₁₂/WpOCNC complex (295 nm excitation) shown deconvoluted with Trp, Mcm, and Acd components. Right: Overlay of fits to stopped flow fluorescence data for binding of WpOCNC to 0.5 μM CaM-C^{Mcm}₁₂δ₁₁₂. Red: 295 nm excitation, 420 nm emission; Blue: 325 nm excitation, 420 nm emission. See Figs. 2.29 and 2.30 for raw data and additional steady state spectra for FRET experiments.

§ 2.6 Conclusion

Using a combination of Cys labeling by Mcm-Mal and Acd incorporation allows one to easily introduce a FRET pair that is selectively excitable and minimally perturbing in proteins (WpOCNC K_d s are nearly identical for CaM and CaM-C^{Mcm}₁₂δ₁₁₂, Fig. 2.26). Mcm/Acd FRET pairs can be used to monitor conformational changes of 15-40 Å, a useful scale for tracking motions within protein domains or among domains of moderately sized proteins. We have shown that spectra of Mcm- or Acd-only proteins can be used to correct for changes in quantum yield or spectral shape to accurately determine E_{FRET} and interchromophore distance. While background signal from reducing agents hindered its demonstration here, the fluorescence turn-on of Mcm-Mal gives it the potential

to be used for *in situ* FRET with Acd. Additionally, Trp FRET with both Mcm and Acd can be used in combination with Mcm/Acd FRET to monitor the kinetics of two processes simultaneously. We note that previous three color (or “two step”) FRET studies have primarily been single molecule experiments using bulky probes or even fluorescent proteins.¹⁴²⁻¹⁴⁵ An example using cyanophenylalanine, Trp, and 7-azatryptophan by Gai uses smaller probes, but these are all UV wavelength probes.¹⁴⁶ The Trp/Mcm/Acd trio balances small probe size with selective excitation, and we are working to develop complementary red-shifted, minimalist fluorescence probes.

**CHAPTER 3: IMPROVING THE FLUORESCENT PROBE
ACRIDONYLALANINE THROUGH A COMBINATION OF THEORY AND
EXPERIMENT.**

The content of this chapter was originally published in the Journal of Physical Organic Chemistry. It is adapted here with permission from the publisher:

Copyright 2018 Wiley. Used with permission from Itthipol Sungwienwong, John J. Ferrie, Joomyung V. Jun, Chunxiao Liu, Taylor M. Barrett, Zachary M. Hostetler, Naoya Ieda, Amara Hendricks, Anand K. Muthusamy, Rahul M. Kohli, David M. Chenoweth, George A. Petersson and E. James Petersson, Improving the fluorescent probe acridonylalanine through a combination of theory and experiment, Journal of Physical Organic Chemistry John Wiley & Sons, Ltd.

§ 3.1 Introduction

Fluorescence spectroscopy can be a valuable tool for studying the structural dynamics of proteins and protein/protein interactions.¹¹⁶ There are several common types of protein experiments that employ fluorescence spectroscopy: folding/conformational change experiments, binding experiments, and proteolysis experiments (Fig. 3.1).¹⁴⁷⁻¹⁴⁸ Changes in fluorescence can be used to track protein structural change on the ns timescale using distance dependent chromophore interactions: either Förster resonance energy transfer (FRET) or quenching by photo-induced electron transfer (eT).¹⁴⁹⁻¹⁵¹ Such studies require fluorescent probes that enable accurate measurement on a variety of distance ranges between the two chromophores. FRET ranges are characterized by the Förster radius, R_0 , the distance of half-maximal energy transfer for any chromophore pair.¹⁵² For example, the common FRET pair fluorescein (Fam)/ tetramethylrhodamine (Tmr) has an R_0 of 47 Å and is useful to measure distances in the 30 to 90 Å range.¹⁵³ Since many inter-residue distances in proteins are shorter than this, one needs to complement the Fam/Tmr FRET pair with other probe pairs that are better suited to shorter interactions. With this in mind, the Petersson laboratory has developed a methoxycoumarinylalanine (Mcm, 1)/ acridonylalanine (Acd, 2) FRET pair for monitoring distances in the 15 to 40 Å range, and thioamide/Mcm or thioamide/Acd eT quenching pairs for short distance (<15 Å) measurements.^{51, 154}

In addition to being better suited to short distance ranges than Fam or Tmr, Acd is small enough to be directly genetically incorporated, rather than post-translationally attached, and is less likely to disrupt protein folding.¹⁵⁵ This allows

one to place a chromophore on the interior of a protein and to label proteins that cannot be reversibly unfolded and refolded (e.g., our published labeling of LexA).¹¹⁸ Having the chromophore attached by a short sidechain rather than a Cys maleimide or “click” chemistry triazole linker also reduces the positional uncertainty of the FRET probe with respect to the protein backbone. Thus, distance measurements from FRET should more reliably report on changes in protein conformation. We and others have previously shown that Acd can be a valuable probe for protein study because of its small size (222 Å³), high quantum yield in water ($\Phi = 0.95$), unusually long fluorescence lifetime ($\tau \sim 15$ ns), and high photostability.^{77, 123, 156} We have developed an engineered aminoacyl tRNA synthetase (RS)/ orthogonal tRNA pair for selective Acd incorporation by unnatural mutagenesis.⁷⁷ This has allowed us to label proteins and peptides with methoxycoumarin/Acd FRET pairs either through Mcm incorporation by solid phase peptide synthesis (SPPS) or by attachment of methoxycoumarin-maleimide to a Cys residue in a protein.⁵¹ While many aspects of the Mcm/Acd FRET pair are optimal, such as significant spectral overlap and a high Mcm extinction coefficient at 325 nm where Acd has a minimum, one disadvantage is the small Acd Stokes shift which leads to significant overlap of their emission spectra. This overlap necessitates a challenging deconvolution of the Mcm/Acd spectra in order to determine FRET efficiencies and distance measurements. Thus, an Acd derivative with a larger Stokes shift would be desirable for Mcm FRET.

In addition to improving Mcm FRET, we also wish to alter other fluorescent properties of Acd, such as red-shifting excitation and emission, increasing the

extinction coefficient, and altering solvatochromic effects to make brighter derivatives that are better suited to microscopy or single-molecule fluorescence applications. Sisido and coworkers have previously shown that some of these effects can be achieved by simply homologating Acd with a benzene ring (benzoacridonylalanine or Bad, 3) to extend the π -system.¹⁵⁷ However, we wish to make derivatives that can still be incorporated by the ribosome, and Sisido's laboratory also showed that Bad was not incorporable during in vitro translation with chemically-charged tRNA. Therefore, we will use crystal structures and computational models of our evolved AcdRS and "rules" for ribosomal permissivity established in previous in vitro translation studies^{155, 158} to restrict Acd substitutions to positions that will allow in vivo tRNA charging and incorporation into proteins.

Identifying Acd derivatives by making amino acid analogs is synthetically laborious and unnecessary given that Acd and Bad spectroscopic properties are identical to the properties of the respective chromophore cores.^{123, 156-157} Thus, to improve Acd fluorescence, we set out to make a series of acridone (5) core derivatives in order to identify derivatives with sufficiently improved properties to warrant synthesis of the amino acid form for incorporation into peptides. Previous studies of acridone derivatives have shown that many of the spectroscopic properties can easily be modulated through substituent effects, providing strong precedent for our work.^{157, 159-165} Moreover, focusing on the acridone core makes computational modeling more tractable, with the potential to further narrow the scope of synthetic work by predicting absorption and emission spectra.

Here, we prepare a series of acridone derivatives in order to validate the accuracy of our electronic structure calculations and identify a derivative, aminoacridonylalanine (Aad, 4) with substantially red-shifted emission. In addition, our calculations help to explain the origin of fine structure in acridone spectra, an explanation which is in conflict with the conclusions of previous computational and experimental Acd spectroscopy studies, but is consistent with the larger body of acridone literature. Finally, we synthesize Aad and perform initial trials toward its genetic incorporation.

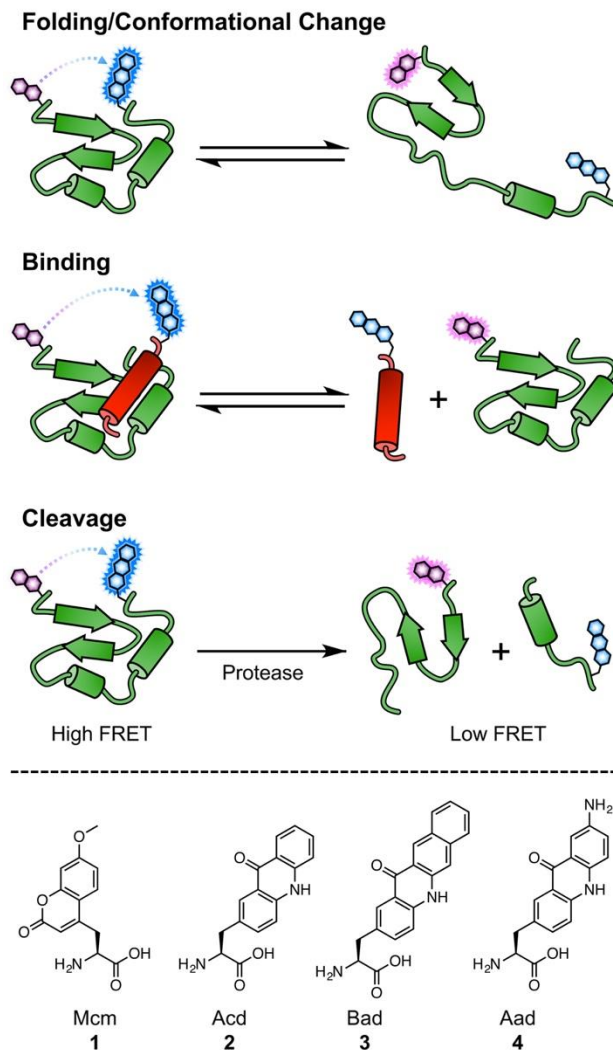


Figure 3.1: FRET experiments and fluorescent amino acids. Top: Protein conformational changes, protein–protein interactions, and proteolytic cleavage can be monitored by changes in the intra- or intermolecular distance between two FRET (Förster resonance energy transfer) probes. Bottom: Fluorescent amino acids based on 7-methoxycoumarin and acridone cores.

§ 3.2 Spectroscopic Characterization

Acridone core derivatives, except 2-methoxyacridone (**21**) which was purchased from Sigma-Aldrich, were synthesized by Itthipole Sungweinwong following the routes detailed below.

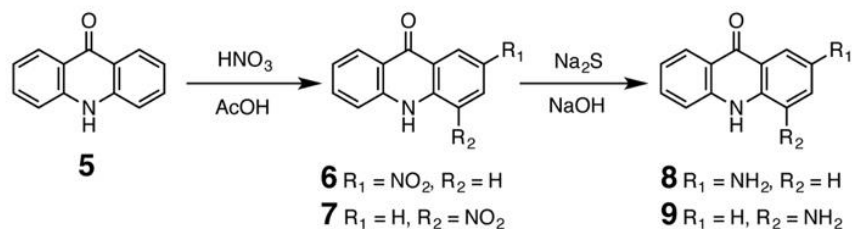


Figure 3.2: Synthesis of nitroacridones and aminoacridones.

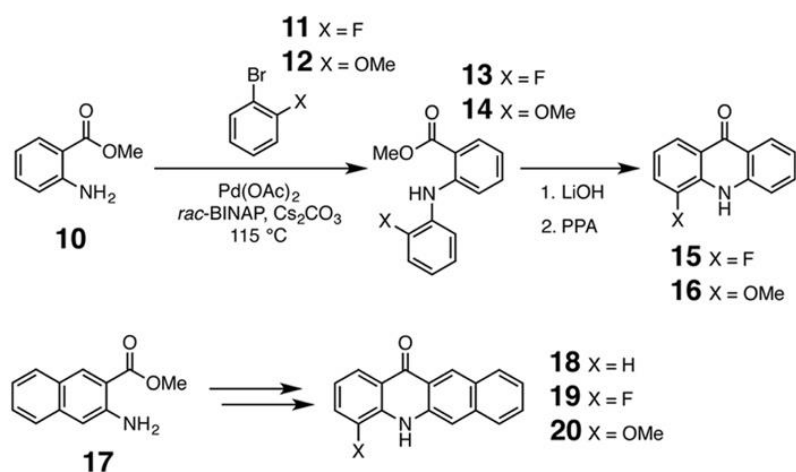


Figure 3.4: Synthesis of acridones and benzoacridones.

Prior to spectroscopic characterization, stocks of all acridone and benzoacridone compounds were prepared in acetonitrile at a concentration of 200 or 300 μM . All absorbance and fluorescence measurements were taken in 1:1 CH_3CN /buffer. Buffers include citrate buffer (89.1 mM citric acid 21.8 mM Na_2HPO_4) pH 2.6, phosphate buffered saline (PBS, 137 mM NaCl , 2.7 mM KCl , 10 mM Na_2HPO_4 , 2 mM KH_2PO_4) pH 7.4, and 3-(cyclohexylamino)-1-propanesulfonic acid (CAPS) buffer, pH 10.0. The absorbance and emission spectra of all compounds were acquired in 1:1 CH_3CN /PBS. The sensitivity of the

absorbance and emission to pH was assessed for the parent acridone core as well as the 4-NH₂ and 2-NH₂ derivatives in 1:1 CH₃CN/citrate and 1:1 CH₃CN/CAPS. Absorbance measurements were acquired at concentrations of 75 and 7.5 μM for proper visualization of the spectral profile above and below ~300 nm (4-aminoacridone absorbance measurements were taken at 7.5 and 0.75 μM).

Fluorescence measurements were acquired under identical solution conditions at concentrations of 7.5 μM for all compounds, except for the 4-NH₂, 2-NO₂, and 4-NO₂ derivatives of Acridone which were obtained at 75 μM due to a lack of brightness. Spectra were collected using excitation wavelengths matching the maximum absorbance wavelength for each analog measured. The collection window started 15 nm from the excitation wavelength and extended to 650 nm. This window was truncated to no less than a total range of 150 nm for more blue shifted chromophores. All slit widths were set to 1.5 nm and spectra were acquired with an integration time of 0.25 sec/nm. Higher signal-to-noise spectra for the 4-NH₂, 2-NO₂, and 4-NO₂ derivatives were collected keeping all other settings the same but adjusting all slit widths to 3 nm.

Extinction coefficients were calculated from absorbance measurements collected on a Tecan M1000 plate reader (Mannedorf, Switzerland). Samples containing 100, 80, 60, 40 and 20 μM chromophore were prepared in 1:1 CH₃CN/PBS solution with a total sample volume of 100 μL. Following brief vortexing, samples were loaded into a Corning CoStar black, clear bottom, 96-well plate.

Absorbance and emission spectra are reported in Figures 3.2-S9. Emission spectra are shown normalized to the acridone emission at 412 nm to approximate quantum yields. The most prominent peak positions are reported along with extinction coefficients and these relative emission intensities in Table 3.1. Calculated spectra were obtained from George Petersson and were performed in Gaussian16™ are shown for comparison. ¹⁶⁶⁻¹⁶⁸

Table 3.1: Calculated and Observed Photophysical Parameters of Acridone Derivatives.

Compound	Calculated		Observed	
	λ_{ex}/ϵ^a	$\lambda_{em}/Int.^b$	λ_{ex}/ϵ^a	$\lambda_{em}/Int.^b$
Acd 5	253/11.4	411/1.00	254/5.22	412/1.00
	364/1.51	434/0.45	382/0.72	435/0.83
	383/2.43		398/0.70	
2-NO ₂ 6	235/5.32	-	230/3.33	-
	282/2.53		296/0.87	
	341/1.03		356/0.76	
	408/1.87		402/0.37	
4-NO ₂ 7	243/6.62	-	244/2.81	-
	292/2.56		266/0.94	
	331/1.27		328/0.58	
	464/2.06		434/0.60	
2-NH ₂ 8	252/4.93	509/0.21	256/3.38	527/0.16
	278/2.96		278/2.46	
	419/0.47		420/0.52	
4-NH ₂ 9	252/5.74	492/0.59	260/4.19	540/0.01
	299/1.20		314/0.38	
	400/0.61		400/0.59	
4-F 15	248/12.8	397/0.48	252/4.44	412/0.65
	365/1.19		378/0.51	
2-OMe 21	251/10.0	444/0.47	252/5.32	447/0.63
	271/3.61		268/5.38	
	400/0.95		396/0.82	
			414/0.82	
4-OMe 16	249/6.62	421/0.35	256/6.30	431/0.37
	375/0.62		384/0.69	
Bz 18	234/1.38	495/0.31	222/1.78	507/0.42
	269/21.7		272/12.7	
	302/1.40		296/2.07	
	446/0.81		438/0.58	
4-F-Bz 19	268/15.3	496/0.24	268/9.96	508/0.21
	446/0.61		448/0.54	
4-OMe-Bz 20	272/14.0	512/0.25	272/9.42	512/0.14
	456/0.67		454/0.61	

^aExtinction coefficients (ϵ) reported as $10^4 \text{ M}^{-1}\text{cm}^{-1}$.

^bEmission intensity (Int.) normalized the intensity of the highest emission peak of acridone for both calculated and observed spectra.

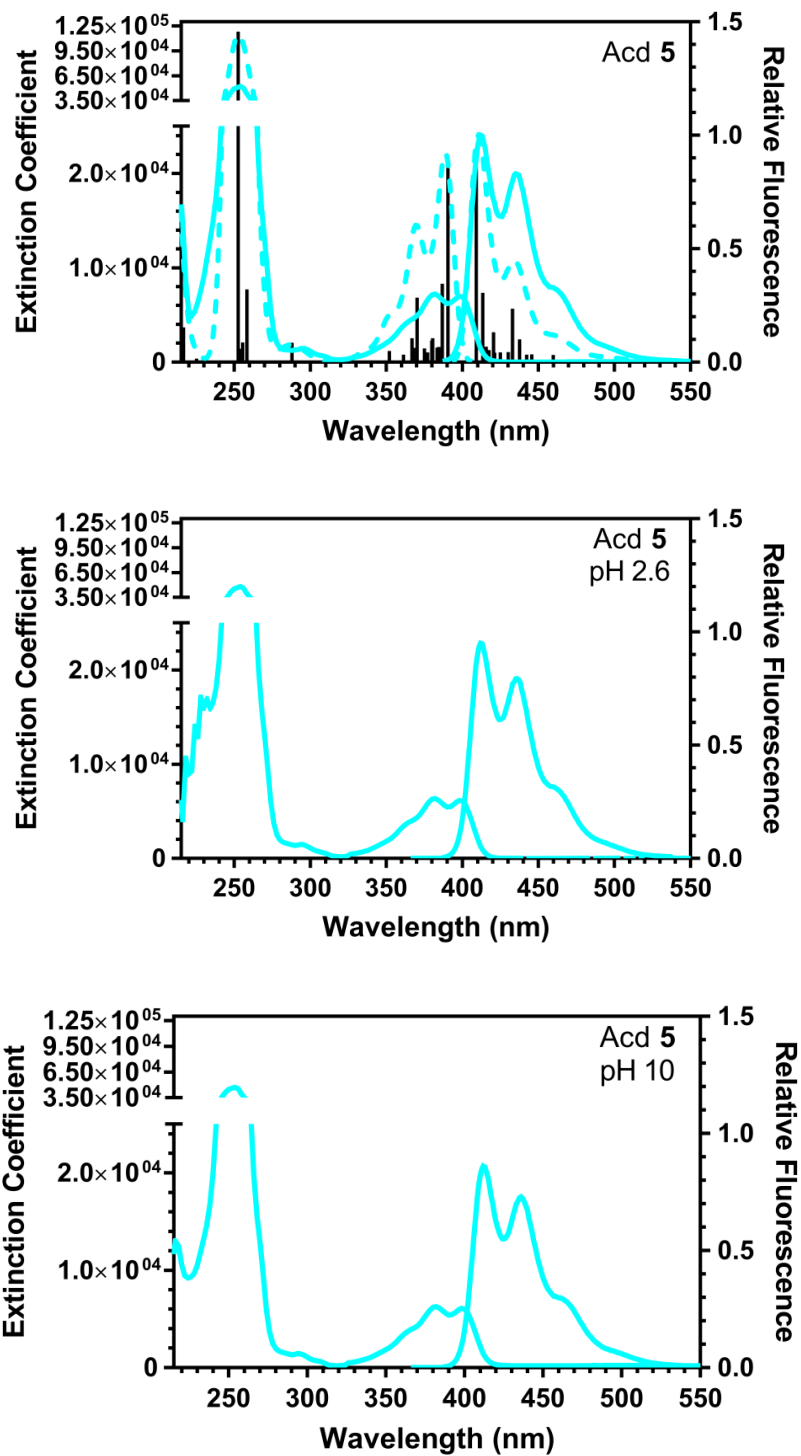


Figure 3.4: Acridone Spectra. Experimental (solid lines) and calculated (dashed lines) spectra and with vibronic transitions to/from lowest excited state (black bars) of Acridone (5) at pH 7.4, and experimental spectra at pH 2.6 and 10.0.

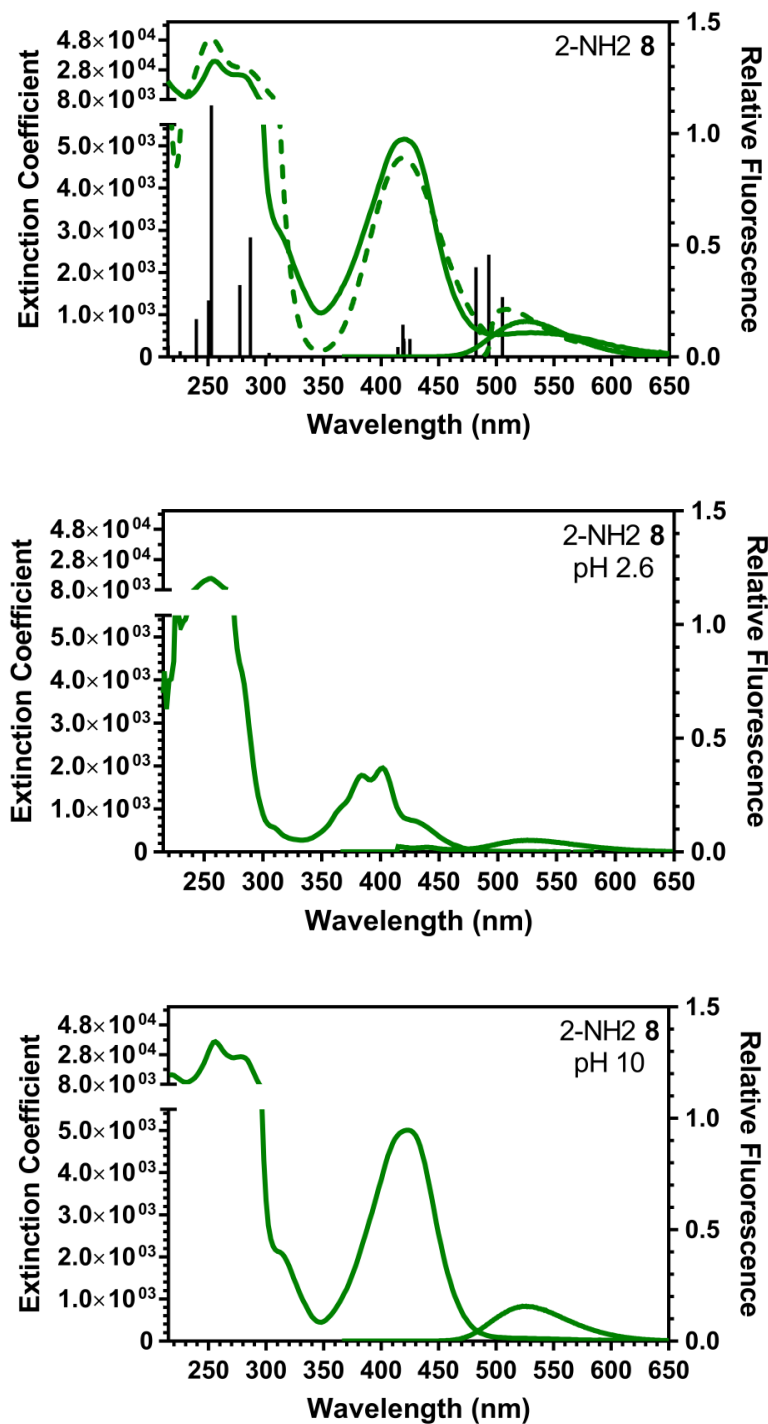


Figure 3.5: 2-Aminoacridone Spectra. Experimental (solid lines) and calculated (dashed lines) spectra and with vibronic transitions to/from lowest excited state (black bars) of 2-Aminoacridone (**8**) at pH 7.4, and experimental spectra at pH 2.6 and 10.0.

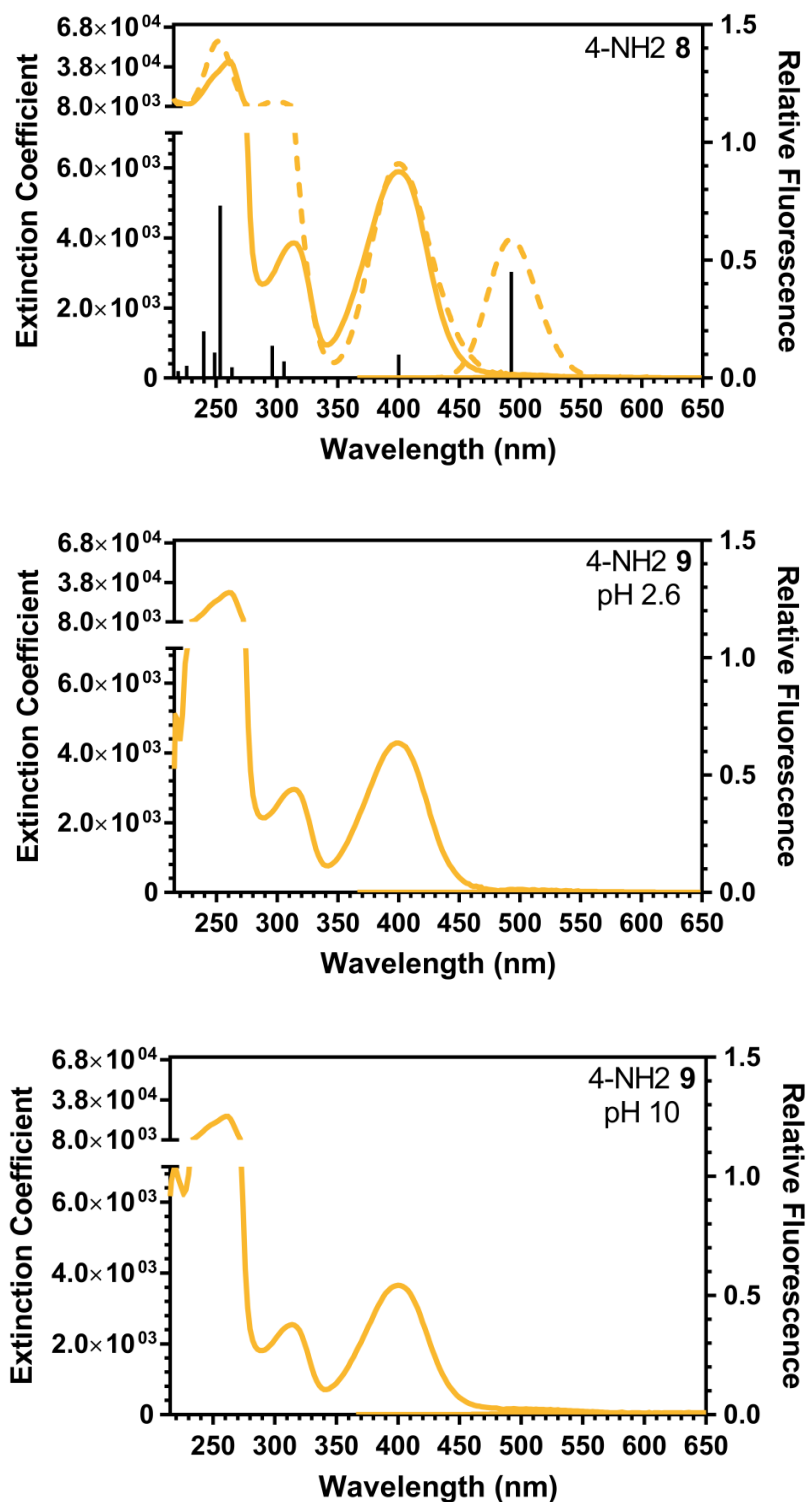


Figure 3.6: 4-Aminoacridone Spectra. Experimental (solid lines) and calculated (dashed lines) spectra and vertical transitions (black bars) of 4-Aminoacridone (9) at pH 7.4, and experimental spectra at pH 2.6 and 10.0.

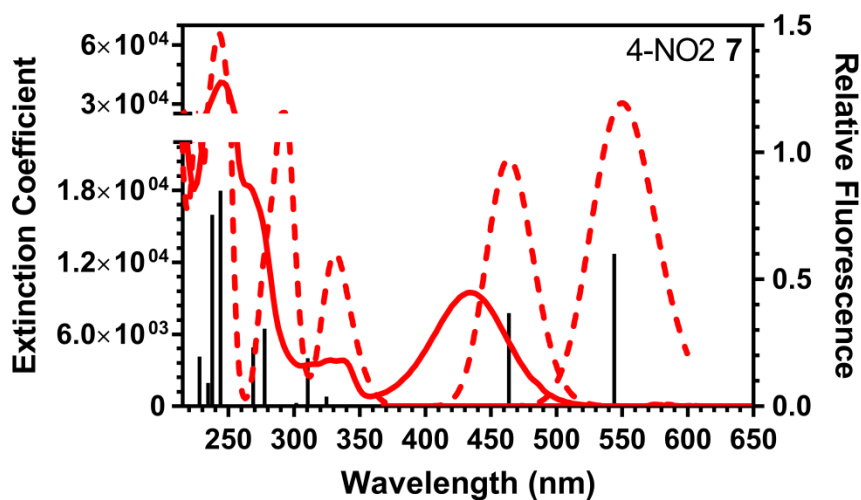
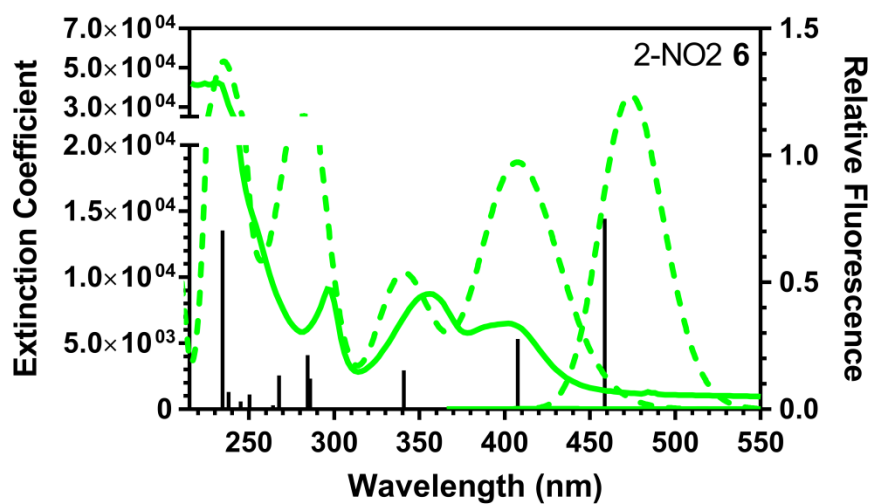


Figure 3.7: 2-Nitroacridone and 4-Nitroacridone Spectra. Experimental (solid lines) and calculated (dashed lines) spectra and vertical transitions (black bars) of 2-Nitroacridone (6) and 4-Nitroacridone (7) at pH 7.4.

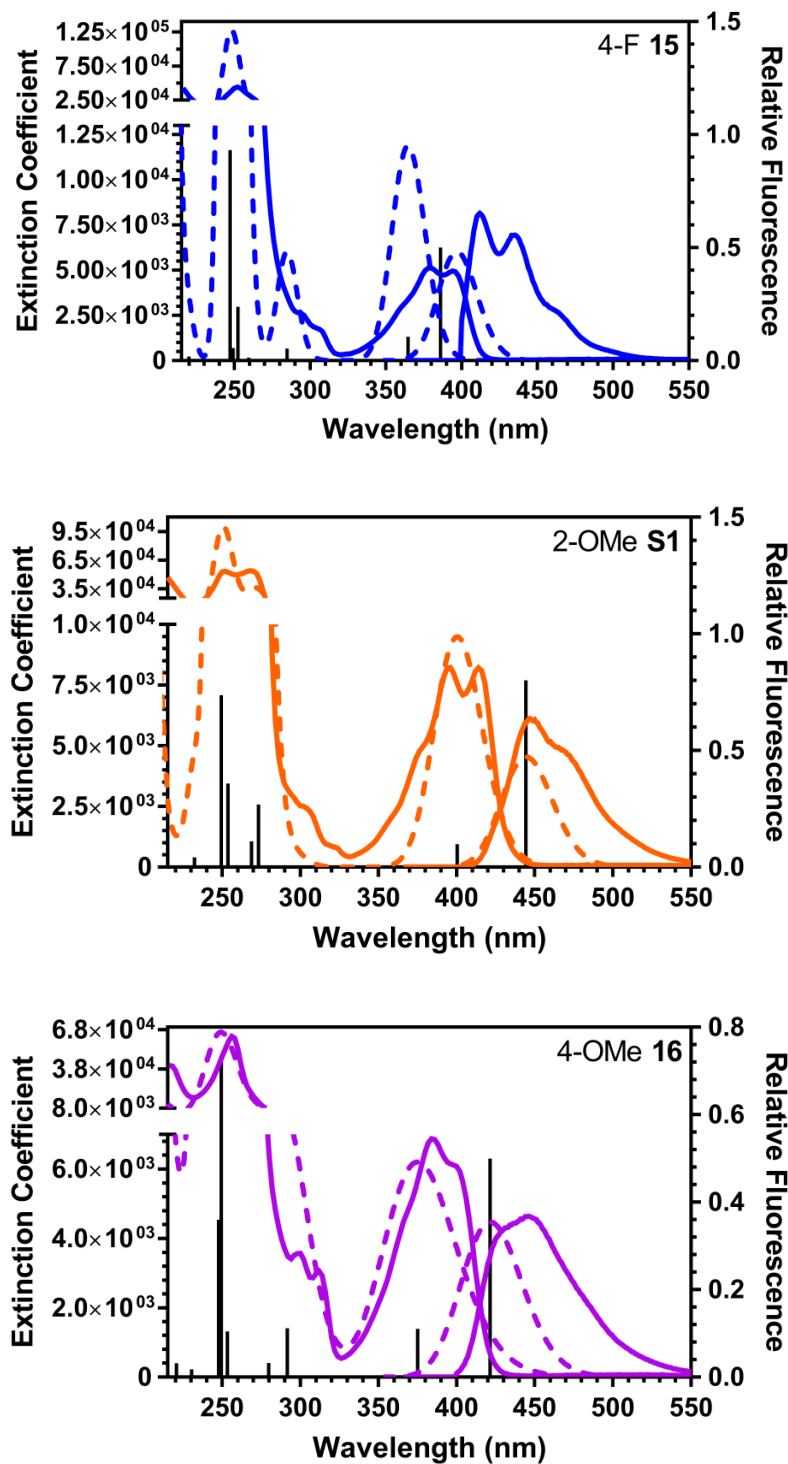


Figure 3.8: 2-Fluoroacridone, 2-Methoxyacridone, and 4-Methoxyacridone Spectra. Experimental (solid lines) and calculated (dashed lines) spectra and vertical transitions (black bars) of 2-Fluoroacridone (**15**), 2-Methoxyacridone (**21**), and 4-Methoxyacridone (**16**) at pH 7.4.

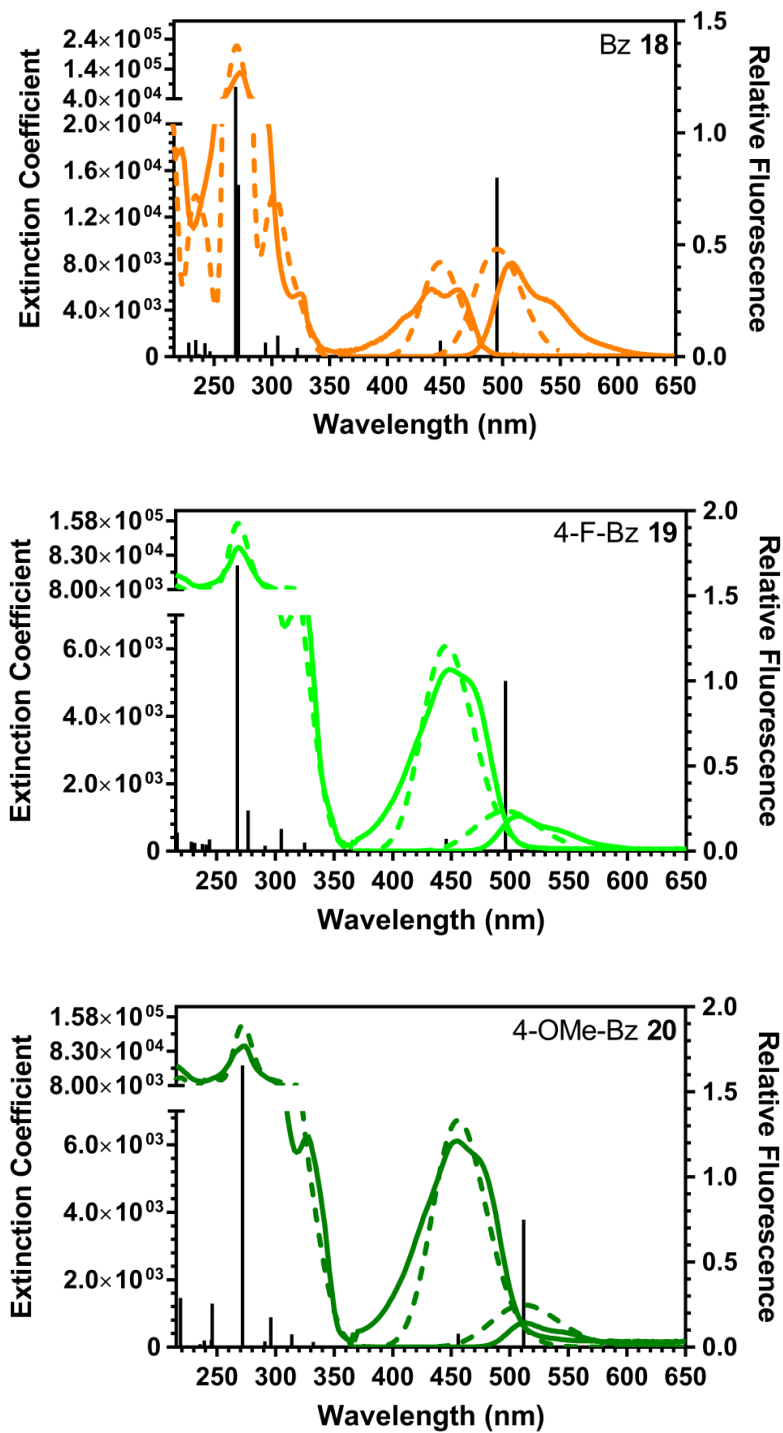


Figure 3.9: 4-Benzoacridone, 4-Fluorobenzoacridone, and 4-Methoxybenzoacridone Spectra. Experimental (solid lines) and calculated (dashed lines) spectra and vertical transitions (black bars) of 4-Benzoacridone (**18**), 4-Fluorobenzoacridone (**19**), and 4-Methoxybenzoacridone (**20**) at pH 7.4.

§ 3.3 Results

Following production of the various Acridone derivatives, we determined the absorbance and emission profiles, as well as the extinction coefficient, of each core. These data are summarized in Figure 3.10 and Table 3.1, and the spectra of each compound are reported separately in the ESI. The absorption and emission spectra are normalized in Figure 3.10 for clarity, raw spectra are shown in the ESI and extinction coefficients and emission intensities are given in Table 3.1. Since we ultimately endeavor to utilize these derivatives as fluorescent unnatural amino acids, we attempted to perform the spectroscopic characterization in phosphate-buffered saline (PBS), pH 7.4. However, due to the relatively low solubility of the acridone compounds in water, all measurements were performed in 1:1 acetonitrile/PBS. Extinction coefficients were obtained through serial dilutions of stocks of 300 μM , which is approximately the solubility limit of acridone in acetonitrile. Fluorescence emission spectra were acquired under the same solvent conditions.

The parent acridone (5) absorbance spectrum features two major peaks in the near UV region with maxima at 382 and 398 nm as well as additional features below 300 nm. Derivatization of the acridone core resulted in changes of the absorbance profile that were both functional group and position dependent. For example, introducing a methoxy group in the 2 position in 21 resulted in a ~ 15 nm red-shift in the absorbance maximum, with minimal modulation of the line shape, while introduction at the 4 position in 16 resulted in a very minimal shift, with the multi-peak profile becoming less defined. Amino modification at either the 2 or 4

position produces a singular broad feature that is significantly red-shifted. Moreover, the absorbance profile of 2-aminoacridone (8) is pH sensitive (Fig. 3.3). The spectrum takes on a single broad absorbance feature very similar to that of 4-aminoacridone (9) at high pH, which is dramatically reduced at low pH. It is important to note that both the unmodified and the 4-amino Acridone chromophores are pH insensitive (Figs. 3.4 and 3.6). Lastly, we observe that extension of the conjugated system in the case of benzoacridone (18) results in an expected shift of the absorbance to higher wavelengths, but displays an unexpected reduction in the extinction coefficient. When modified with either fluoro (19) or O-methyl (20) substituents, the absorbance profile is minimally shifted and the peaks become less well-resolved, similar to what was observed when modifying the parent acridone scaffold (15 and 16).

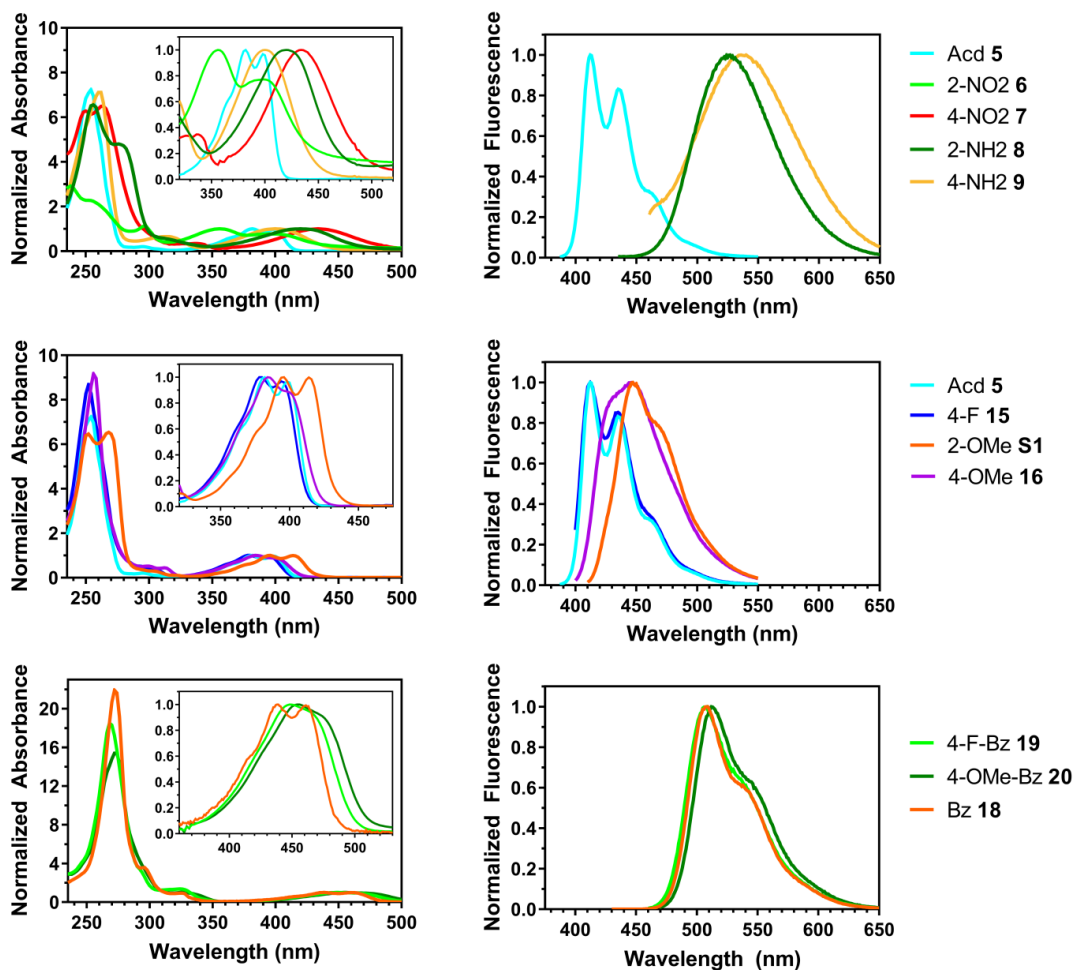


Figure 3.10: Absorption and emission spectra of acridone derivatives. Spectra determined in 1:1 acetonitrile/PBS, pH 7.4. Spectra are shown normalized to enable comparison of changes in absorption and emission maxima.

Acridone substitution also elicited changes in the emission profile and Stokes shift. The unmodified acridone emission features two major peaks at 412 and 435 nm with a minor peak around 460 nm. As in the absorbance profiles, the maxima of these peaks move to lower energy upon modification with the *O*-methyl group (**21** or **16**) with the multi-peak profile becoming less defined. Similarly, introduction of the amine functionality resulted in a significant red shift as well as

a reduction of the multi-peak line shape to a single broad emission. Both compounds **8** and **9** exhibited a dramatic increase in Stokes shift of ~100 nm compared to the acridone core. Both also displayed a sensitivity to pH, manifested as a decrease in emission intensity with decreasing pH, compared to the unmodified core whose emission is insensitive to pH changes in the 2-10 range. Finally, the benzo modified acridone analogs all feature a nearly identical emission profile ~100 nm red shifted from the emission of the parent acridone compound. The combination of benzo and fluoro or methoxy substitution further red shifts the absorption spectrum, but does not appreciably change the emission spectrum.

§ 3.4 Conclusions

We can draw several conclusions from our results. Firstly, as anticipated from previous literature reports, we are able to modulate the fluorescence of the acridone core through relatively simple substitutions. These substitutions, introduced through direct modification of acridone or through cross-coupling and cyclization, are compatible with eventual usage in generating Acd amino acid derivatives. Secondly, the close correlation (% differences λ_{ex} : 2.35, λ_{em} : 2.40) between our calculated and observed absorption and emission spectra give us confidence that we can predict the spectra for acridone derivatives, providing guidance for future synthetic efforts. This will be particularly valuable for targeting multiply-substituted derivatives, where the number of possibilities is geometrically larger and the synthesis will be more challenging. Finally, Aad appears to be superior to Acd for several fluorescence applications.

**CHAPTER 4: SYSTEMATIC EVALUATION OF SOLUBLE PROTEIN
EXPRESSION USING A FLUORESCENT UNNATURAL AMINO ACID
REVEALS NO RELIABLE PREDICTORS OF TOLERABILITY.**

The content of this chapter was originally published in *ACS Chemical Biology*. It is adapted here with permission from the publisher:

Reprinted with permission from Z. M. Hostetler, J. J. Ferrie, M. R. Bornstein, I. Sungwienwong, E. J. Petersson, and R. M. Kohli. Systematic Evaluation of Soluble Protein Expression Using a Fluorescent Unnatural Amino Acid Reveals No Reliable Predictors of Tolerability. *ACS Chem. Biol.* 13, 10, 2855-2861. Copyright 2018 American Chemical Society

§ 4.1 Introduction

Technological advances in genetic code expansion have encouraged the design of proteins with a wide range of reactive residues, post-translational modifications, photocaged groups, or intrinsic fluorophores.¹⁶⁹⁻¹⁷¹ Nonsense codon suppression using orthogonal tRNA/aminoacyl-tRNA synthetase pairs enables direct incorporation of chemically diverse unnatural amino acids (Uaas, also known as non-canonical amino acids) into proteins *in vivo*. Many efforts have sought to boost the efficiency of Uaa incorporation, including evolving more efficient aminoacyl-tRNA synthetases and recoding the *E. coli* genome to remove competing translational release factors.^{93, 172} Although these developments can improve total yields of modified proteins, factors governing the position-dependent effects of Uaa substitution on protein solubility remain understudied.

Recent reports have demonstrated that the position of a Uaa can affect the level of total protein expressed, both in cell-free and cell-based systems.^{156, 173-177} Investigations of 20 positions in IFN- α and 33 positions in VSV glycoprotein revealed varying total protein yields, from 0 to 95% of wildtype.¹⁷⁷⁻¹⁷⁸ Despite these observations, explanations for position-dependent differences in total amounts of Uaa-containing proteins have been limited, and no studies have explicitly addressed UAA incorporation versus the resulting protein solubility.

Unnatural amino acid mutagenesis could hypothetically operate under well-accepted principles that govern the effects of natural amino acid mutation. For example, substitution of a nonpolar for a polar residue within the hydrophobic core generally destabilizes proteins, whereas mutations on the solvent-exposed surface

less frequently affect solubility.¹⁷⁹⁻¹⁸⁰ Unsurprisingly, evolutionarily-conserved residues largely disfavor mutation.¹⁸¹⁻¹⁸³ Substituting bulkier and more chemically-diverse Uaas into a protein can restrict function and therefore could pose similar burdens on folding and solubility.¹⁸⁴ Nevertheless, the applicability of principles of natural amino acid mutagenesis to Uaa mutagenesis remains unknown.

Suggested guidelines or approaches for choosing Uaa-tolerant sites have been proposed. Some groups favor residues with structural similarity to the Uaa.¹⁷⁵ Others assert that candidate positions should be first assessed for mutational tolerability with natural amino acids or that proteins should be thoroughly screened by random incorporation of Uaas into protein-GFP fusions to reveal positions that label with high efficiency.^{176, 185-186} Nonetheless, the feasibility of using position-specific properties to increase soluble protein expression remains untested.

To address these open questions, we aimed to explore factors that impact Uaa incorporation and soluble protein production. By employing an intrinsically fluorescent Uaa, acridonylalanine (Acd),^{77, 118, 173} we directly detect labeled protein in cell lysate samples, overcoming the inability of past studies to measure levels of both total and soluble expressed protein. Our systematic survey of more than fifty sites across two proteins reveals that while incorporation efficiency is relatively similar, protein solubility, and by extension Uaa tolerability, varies widely across different positions. However, most position-specific physicochemical, evolutionary, and structural properties, some of which have been previously suggested to improve yield, were minimally predictive; instead, solubility more strongly associated with the identity of the protein domain. After controlling for this domain

effect, we found that only a few factors, such as a tolerance for aromatic residues, moderately trended with protein solubility. To our knowledge, this work currently represents the most systematic effort evaluating predictive factors for producing soluble Uaa-containing proteins.

§ 4.2 Experimental and Computational Methods

All experiments referenced in this chapter were performed by Zachary Hostetler, as were initial attempts to use computable characteristics (conservation, hydrophobicity, solvent accessibility, etc.) to predict Uaa mutation tolerance.

The bacterial protein LexA, a multi-domain repressor of the DNA damage response, has characteristics that made it well-suited to this broader survey. Wild-type *E. coli* LexA is well-behaved in overexpression and has previously tolerated selective unnatural amino acid (Uaa) incorporation.²² Additionally, the availability of protein crystal structures and a multiple sequence alignment for LexA enabled retrieval of position-specific properties from databases or servers that require these data as inputs. For every position in LexA, we calculated established metrics across different classes of properties: physicochemical, such as hydrophobicity; evolutionary, such as conservation; and structural, such as solvent accessibility. Using these metrics, we selected 32 positions spanning both domains of LexA, deliberately avoiding known deleterious mutants as well as the most conserved or hydrophobic positions (Figure 4.1a). Our selected positions sample the remaining metrics well (Figure 4.1b), indicating that this series is well-positioned to explore

how aromatic, accessible, or poorly-conserved residues might differentially tolerate Uaa incorporation.

Historically, measuring Uaa incorporation efficiencies *in vivo* has overlooked protein solubility issues, while labeling Uaa-containing proteins *in vitro* has suffered from incomplete sample recovery and detection. Crucially, we chose to measure both total and soluble protein levels by using the fluorescent Uaa acridonylalanine (Acd, Figure 4.1c), which already possesses an optimized tRNA/tRNA synthetase pair for *in vivo* incorporation.^{21,22} This system offers several advantages. First, Acd incorporation occurs during protein overexpression without post-translational labeling. Second, measurements of Acd fluorescence at the expected size on an SDS-PAGE gel are directly proportional to levels of protein with successfully-incorporated Acd. Finally, gel-based detection of Acd demonstrates a broad dynamic range, enabling us to detect quantitative differences in the expression of Acd-containing LexA mutants.

Expression levels for a single protein can range widely due to experimental variability, making quantitative comparison between different proteins difficult. To overcome this challenge, we overexpressed the 32 LexA mutants in the presence of both Acd and the Acd-specific tRNA/tRNA synthetase using autoinduction media for consistency in the timing and duration of protein production. Following overexpression, we measured fluorescence intensity levels of Acd-containing LexA protein in both the whole cell lysate and soluble fraction (Figure 4.1d). The use of purified Acd-containing LexA as a standard enabled quantitative and reproducible comparisons of protein amounts across independent experiments.

Simulation of Acd incorporation into LexA or RecA with Rosetta.

Prior to performing simulations, a parameter file and rotamer library were produced for Acd following a previously described method.¹⁸⁷ Starting structures for the LexA simulations were prepared from PDB 1JHE and PDB 1JHF by adding the missing residues using the remodel application in Rosetta.¹⁸⁸⁻¹⁸⁹ A blueprint file was prepared from each monomer and the primary sequence was modified to match that of the LexA expression construct. After adding the missing residues to each monomer, the dimer was reconstructed by merging the two PDB files and the resultant structure was minimized using the Relax application.¹⁹⁰ The Relax application was run by setting the jump_move, bb_move, and chi_move flags to False and using the relax:fast flag. The starting structure was selected as the lowest energy structure of 10 outputs. The same protocol was followed to produce the RecA starting structures from PDB 3CMW, omitting the remodel application step as all residues were present.¹⁹¹ For the Backrub-based method, a total 2,500 structures were produced from each starting structure. This was done by running the Backrub application in Rosetta performing 10,000 trials at 0.6 kT to generate each output structure.¹⁹² The total energy was computed for each member of the ensemble following the single-site mutation to Acd and global repacking in PyRosetta.¹¹⁴ For RecA, all mutations were performed and assessed within a single monomeric unit (residues 967-1299) within the multimer. The total energy was averaged across all members of the single ensemble for RecA and across all members of both ensembles for LexA. LexA simulations based on the relax-based algorithm were performed in PyRosetta using the same initial structures as starting

points. The method consisted exclusively of the FastRelax mover constrained to the starting coordinates using the 'lbfgs_armiho_nonmonotone' min_type and a maximum of 200 iterations. A total of five output were produced for each mutation and the energy was averaged across all outputs for both starting structures for a given site. All methods were run using the 'beta_nov15' score function weights.

§ 4.3 Results and Discussion

Observing that the position of Acd can substantially impact protein solubility, we next asked which of the properties that ostensibly affect Uaa tolerability might correlate with solubility (Fig. 4.1). We fitted the soluble fraction as a response variable to each property in individual linear regression models. For almost all of the properties we evaluated, the explained variability (adj. R^2) was about 5% or less, indicating that if any property-specific effect exists, it is insubstantial and likely below our ability to detect with a sample size of 32. We note that particular properties—such as accessibility, conservation, and hydrophobicity—did not explain any substantial variation in our data, despite past suggestions that choosing accessible, less-conserved, and chemically-similar residues may yield more soluble Uaa-containing protein (

Figure 4.1c).

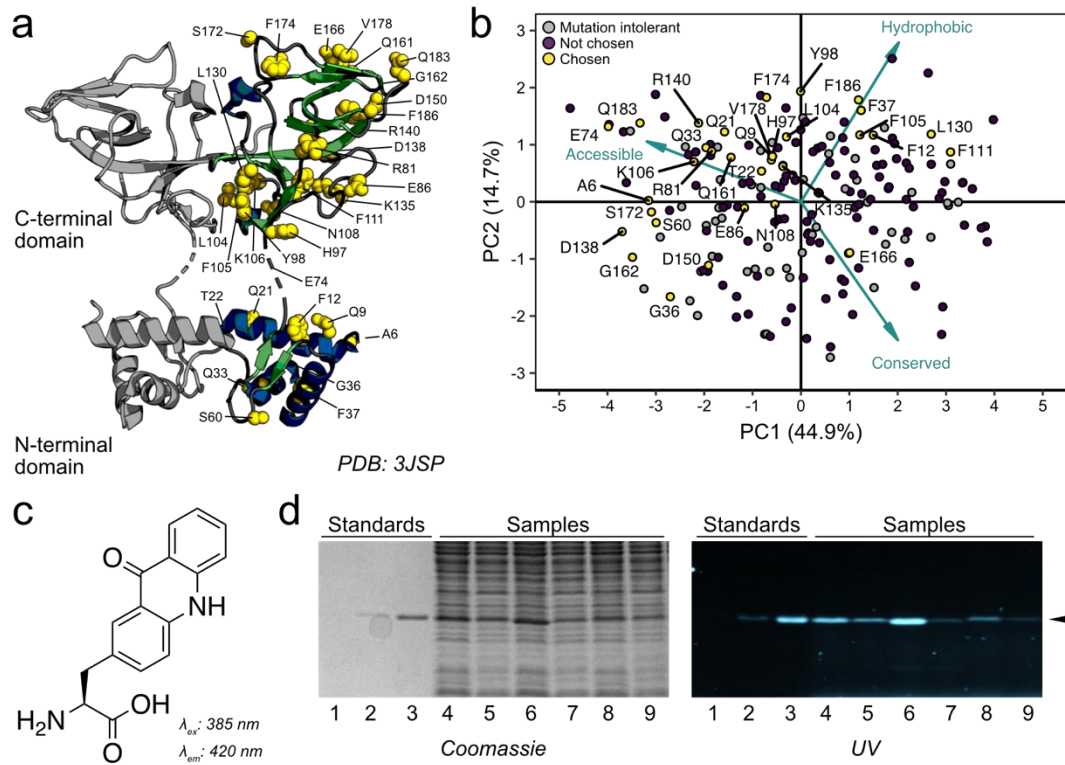


Figure 4.1: Scanning a variety of positions in LexA for Acd tolerability. (a) Positions chosen for Uaa incorporation in the LexA dimer. Chosen positions are depicted in yellow, α -helices in blue, and β -sheets in green. (b) Principal component analysis (PCA) of LexA positions determined by multiple structural, evolutionary, and physicochemical properties (see methods). All residues in LexA were scored and plotted against the first two principal components, with positions chosen for Uaa incorporation highlighted in yellow. Arrow segments represent a few notable variables among those used in PCA loaded onto the plotted data. (c) Chemical structure of Acd with indicated excitation and emission peaks. (d) Acd-labeled LexA samples visualized in 15% SDS-PAGE gels by Coomassie staining (left) or UV excitation (right). Lanes 1–3 show purified LexA standards. Lanes 4–11 show paired total and soluble fractions from four individual mutants as representative examples.

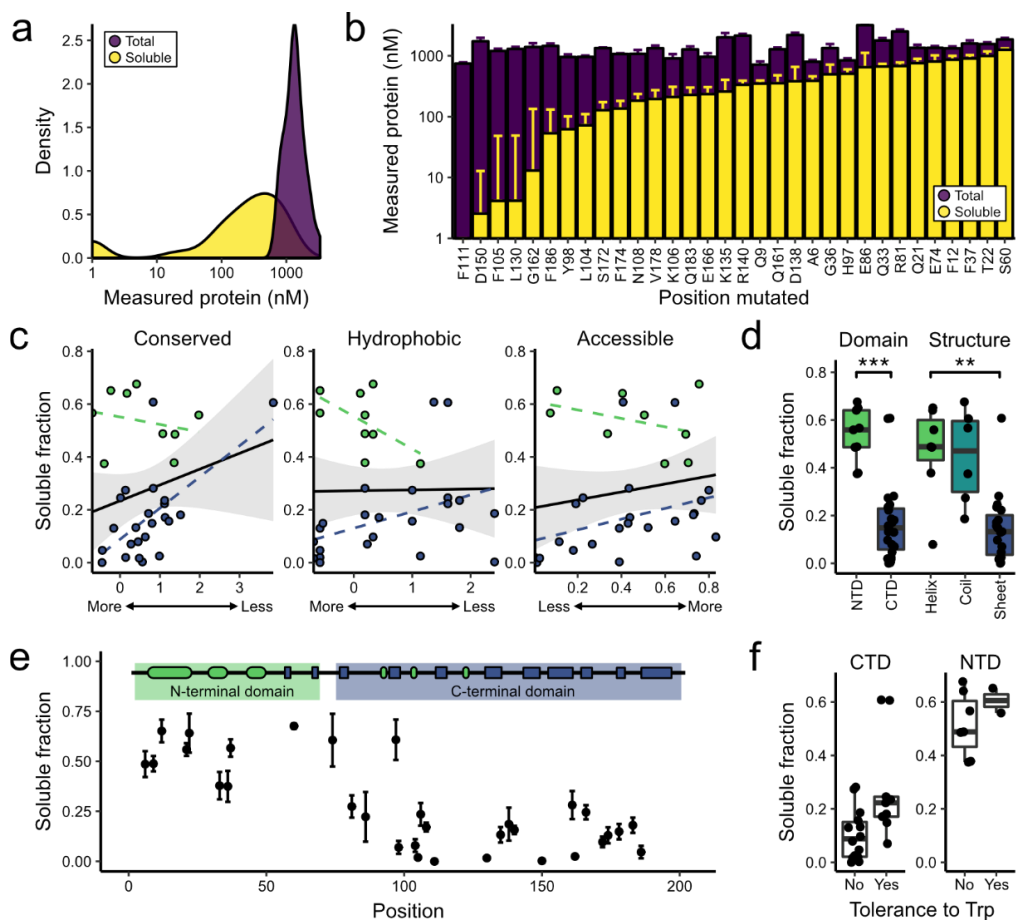


Figure 4.2: Features associated with soluble Acd-labeled LexA proteins. (Previous Page) (a) Smoothed density plots of \log_{10} -transformed amounts of total protein or soluble protein. (b) Average \log_{10} -transformed soluble protein amounts overlaid on average \log_{10} -transformed total protein amounts for each mutant. Error bars indicate the standard deviation from three individual replicates each derived from separate clones. (c) Plots of the average fraction of soluble protein as a function of three selected parameters: conservation, hydrophobicity, and accessibility. Fits for the entire LexA dataset to individual linear regression models yield best fit lines (solid black) and 95% confidence intervals (shaded gray). Fits of data from each separate LexA domain yield best fit lines for the NTD (dashed green) or CTD (dashed blue). (d) Boxplots comparing the average fraction of soluble protein against either domain or secondary structure, with individual averages overlaid. Differences between groups were evaluated using Tukey's HSD test for multiple pairwise comparisons (** = p -value < 0.01; *** = p -value < 0.001). (e) Plot of the average fraction of soluble protein as a function of position in the LexA sequence, with error bars indicating the standard deviation from three replicates. Above, the secondary and tertiary structure of LexA is indicated; α -helices are depicted as green ovals and β -sheets as blue rectangles. (f) Separate boxplots for each LexA domain indicating the relationship between average fraction of soluble protein and

evolutionary tolerance at each position to tryptophan, as one example of an aromatic residue.

Conspicuously, several highly-correlated properties each explained around 50% of the variability in our data, including individual residue position (adj. $R^2 = 0.53$), secondary structure (adj. $R^2 = 0.45$), and overall protein domain (adj. $R^2 = 0.53$) (

Figure 4.2d and

Figure 4.2e). Specifically, we obtained more soluble protein when Acd was incorporated within the first 74 residues of LexA, which includes all three of the α -helices that comprise the N-terminal domain. By contrast, Acd incorporation within the β -sheets of the C-terminal domain resulted in much lower proportions of soluble protein. The nearly uniform secondary structure composition of each domain limited our ability to interpret whether Acd tolerability is due to local secondary structure effects or global protein domain stabilities.

Studying Acd incorporation in a distinct protein scaffold with mixed α/β character could help dissect the similar effects we observed from the highly-correlated domain and secondary structure factors with LexA. Thus, we extended our survey to RecA, a bacterial ATPase that binds LexA to suppress its repressor function.²⁶ We selected positions in *E. coli* RecA that satisfied one or more criteria: high accessibility, low conservation, few inter-residue contacts, or prior functional tolerance to mutation (Figure 4.3a).²⁷ After expressing these mutants with Acd and measuring protein amounts, we again observed greater variability in

logarithmically-transformed soluble protein levels (mean = 3.42, SD = 0.40) compared to total protein levels (mean = 3.72, SD = 0.17) (Figure 4.3b and 4.3c). Similar to LexA, most properties examined did not explain much variation in the fractions of soluble protein (Figure 4.3d), with the exception that solubility modestly trended with domain type and tolerance to aromatics. However, unlike in LexA, no clear relationship existed between protein solubility and type of secondary structure (Figure 4.3e), a result consistent with a more limited prior survey of GFP.⁸ This survey in RecA bolsters a model in which the intrinsic Uaa tolerability of a protein domain remains the key obstacle for the production of soluble protein.

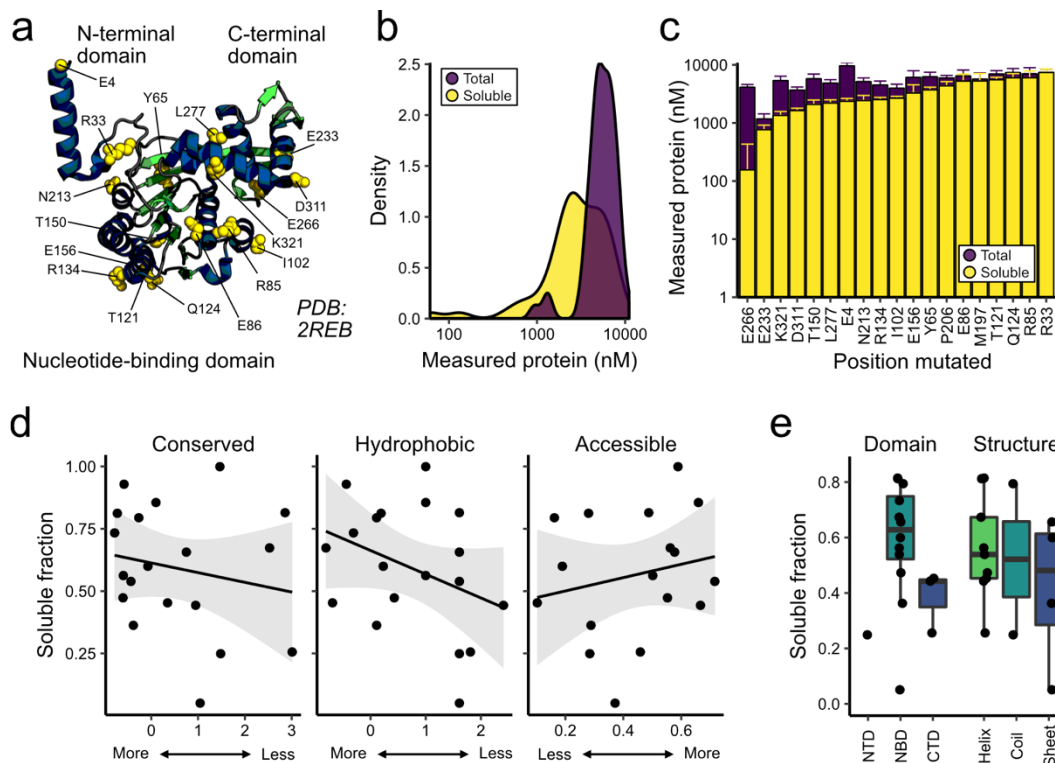


Figure 4.3: Features associated with soluble Acd-labeled RecA proteins. (a) Positions chosen for Acd incorporation in RecA. Chosen positions are depicted in yellow, α -helices in blue, and β -sheets in green. (b) Smoothed density plots of \log_{10} -transformed amounts of total protein or soluble protein. (c) Average \log_{10} -transformed soluble protein amounts overlaid on average \log_{10} -transformed total protein amounts for each mutant. Error bars indicate the standard deviation from three individual replicates each derived from separate clones. (d) Plots of the average fraction of soluble protein as a function of three selected parameters: conservation, hydrophobicity, and accessibility. Fits to individual linear regression models yield best fit lines (solid black) and 95% confidence intervals (shaded gray). (e) Boxplots comparing the average fraction of soluble protein against domain or secondary structure, with individual averages overlaid.

Searching for easily-determined properties that correlate with Acid tolerability may have eliminated from consideration more complicated properties with higher predictive ability. Additionally, linear regression modeling may have over-simplified the inter-dependence of certain properties and protein solubility.

Previously, Rosetta modeling has predicted the $\Delta\Delta G$ associated with a particular mutation and identified tolerated mutations within a protein.¹⁹²⁻¹⁹⁴ Speculating that Rosetta modeling could recapitulate our experimental results, we used the Rosetta Modeling Suite to simulate the resulting energy associated with Acd incorporation in LexA or RecA. However, we observed no significant correlations between simulated energies and soluble fractions of LexA or RecA (Figure 4.4 and Figure 4.5). Incidentally, we noted that nearly all high-energy positions in LexA experimentally yielded insoluble protein and may therefore have been useful in filtering out those positions; however, we did not observe a similar energy threshold effect for RecA. Accordingly, further refinement towards predicting Uaa incorporation using Rosetta is required in order to recapitulate experimental data and exclude higher-energy and lower-solubility mutants.

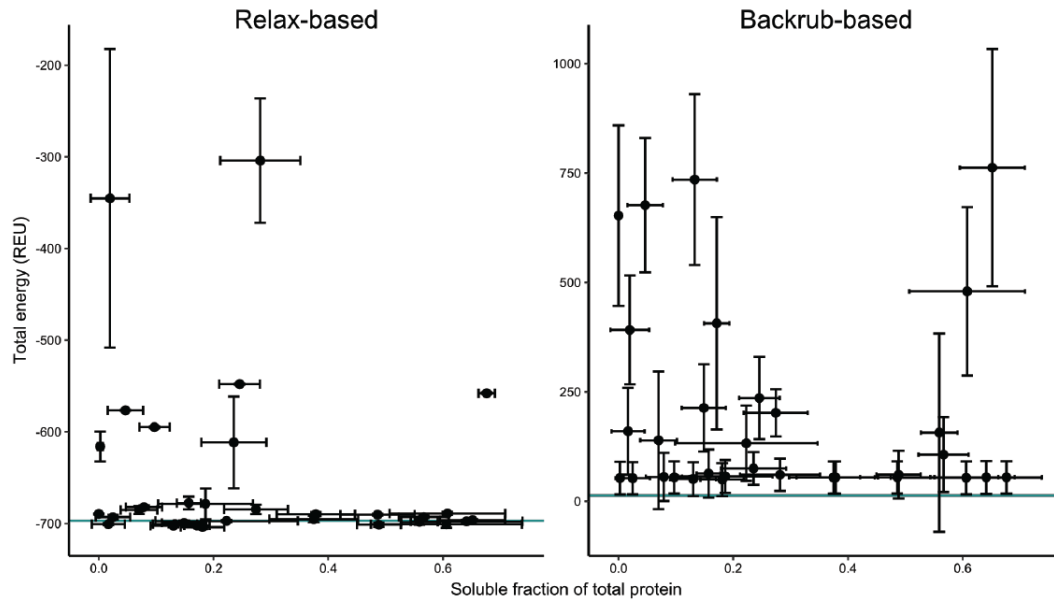


Figure 4.4: Predicting protein solubility through simulation of Acd incorporation in LexA. Scatterplots of the total energies in Rosetta Energy Units (REU) from simulating Acd incorporation in LexA as a function of the soluble fraction of total protein. Rosetta energies were obtained by performing each single mutation on a relaxed structure of LexA derived from one of two previously published structures (PDB: 1JHE or 1JHF), using either a Relax-based (left) or Backrub-based (right) method. The total energy of each LexA mutant was computed following mutation of the residue of interest to Acd either by minimizing of the energy using a relax-based protocol or following repacking of all residues for each member of an ensemble of LexA structures. Each point represents the average of the two different simulations, with vertical error bars representing standard deviations. The solid turquoise line represents the average energy of energy-minimized LexA without any Acd mutation.

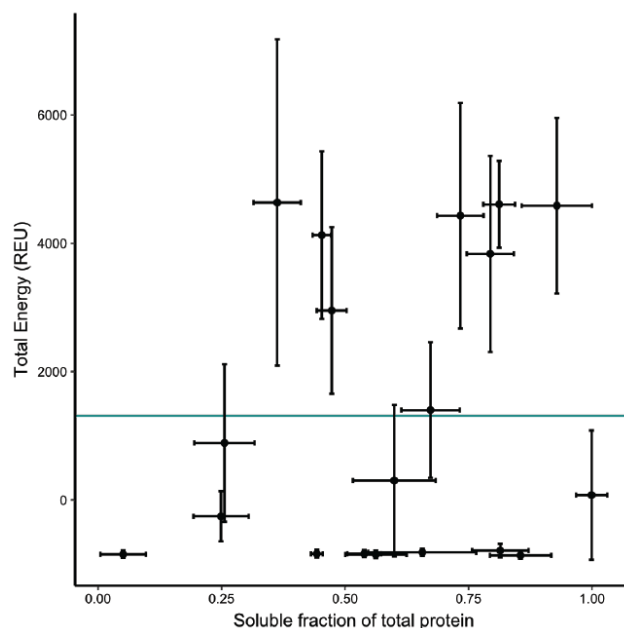


Figure 4.5: Predicting protein solubility through simulation of Acd incorporation in RecA. Scatterplot of the total energies in Rosetta Energy Units (REU) from simulating Acd incorporation in RecA as a function of the soluble fraction of total protein. Rosetta energies were obtained by performing each single mutation on each member of a 2,500 structure RecA ensemble generated using the Backrub application. Separate ensembles were generated from the previously published structure (PDB: 3CMW). The total energy of each RecA mutant was computed after mutating the residue of interest to Acd and repacking all residues in RecA. Each point represents the average energy computed across all members of the different simulations, with vertical error bars representing standard deviations. The solid turquoise line represents the average energy of energy-minimized RecA without any Acd mutation.

§ 4.3 Initial Conclusions

The expression of soluble protein is a major bottleneck for the study of protein function. Here, we leveraged the fluorescence of Acd to study how protein solubility is impacted by Uaa mutagenesis. In two bacterial proteins, we demonstrated the dramatic impact that Uaa position has on protein solubility. Surprisingly, a number of amino acid properties that purportedly contribute to Uaa

tolerability—including low evolutionary conservation, similar hydrophobic character, or high surface accessibility—were unreliable predictors of protein solubility. Instead, these inconsistent relationships suggest that consideration of specific amino acid features for successful Uaa mutagenesis is less critical than previously thought. Rather, we speculate that the Uaa tolerability of a protein domain may matter more. Our results also emphasize a continued need to explore, through theory and experiment, the steric and chemical burdens different Uaas pose to the expression of soluble protein. In the absence of reliable predictors or refined simulation algorithms for Uaa tolerability, a chemical biologist pursuing Uaa incorporation in a new protein, as of now, should broaden rather than narrow the types of residues screened for Uaa tolerability when possible.

§ 4.4 Further Algorithmic Development

Following the previously described efforts to identify a metric or algorithm which can adequately predict the soluble fraction of Uaa containing protein, further attempts were made to develop an effective Rosetta-based method. The previous simulations leveraged different sampling approaches but utilized a single common score function. Since all Rosetta score functions are a composite of several different energy terms which describe various physical characteristics of the system, I sought to identify which aspects of scoring along with which isolable energy terms best captured the variance in the data. This was accomplished by re-performing the previously described Backrub-based and Relax-based simulations on the LexA 1JHE PDB structure and computing the score contribution

from each energy term across the entire protein, each monomer of the LexA dimer, each protein domain and on a per-residue basis. Table 4.1 details the nomenclature employed in Figures 4.6 – 4.7 comparing the results of these simulations to the previously acquired total expression and soluble fraction experimental data.

Although the previous methods using the fully compiled “beta_nov15” Rosetta energy function did not have a sufficient correlation with the protein soluble fraction data, the investigation of individual energy terms (Fig 4.6a-d) revealed that several terms may serve as useful predictors in a threshold-based prediction scheme. Decomposition of the “beta_nov15” score function in its constituent energy terms revealed that the top 20 highly correlated terms, shown in Figs. 4.ba-d, all display a stepwise relationship with the fraction soluble data. Curiously, the bulk of these correlates emerge on domain-level scoring, which supports the findings of the previous work which demonstrated that domain identification was a strong predictor in LexA of position tolerance to amino acid substitution. Furthermore, it is important to note that almost all sampling schemes are represented in the top 20 correlates indicating that even though different sampling approaches lead to different energy terms which are significant, all sampling schemes are capable of producing some form of prediction.

These findings are further reinforced when investigating the potential correlations with the total expression values for Acd mutant LexA variants. Compared with the soluble fraction data, the total expression data has a significantly reduced variance and, unsurprisingly, the single-term correlates from

Rosetta appear to perform less effectively for the total expression dataset. However, the individual energy terms no longer show stepwise correlations (Fig 4.7a-d) demonstrating that linear regression of combined terms may deliver a highly effective metric. Lastly, several of the top 20 correlates originate from the energy values associated with the native residue in LexA which is selected for mutation (Fig 4.7a-b). If a significant correlation can be gleaned from simply scoring the input structure, a very rapid algorithm that does not require *in silico* mutation and sampling may be capable of predicting components of the LexA dataset.

Ultimately these findings encourage further investigation into the feasibility of employing sampling and scoring methods endogenous to the Rosetta Modeling Suite to predict which sites are more favorable for mutations to Uaas. Overall, through the use of PyRosetta, development of new score functions through linear regressions of existing energy terms and the ability to make slight modifications to current sampling schemes should lead to an effective predictive algorithm.

Table 4.1: List of descriptors for Rosetta simulations

Sampling Schemes	
Backrub_Pre	Mutation of residue to Acd is performed prior to running 100 independent backrub trajectories
Backrub_Post	Mutation of residue to Acd is performed prior after running 100 independent backrub trajectories on the unmutated proteins
Relax_Cst	Following mutation of the residue to Acd, the whole protein is relaxed in the presence of constraints based on the starting structure
Relax_No_Cst	Following mutation of the residue to Acd, the whole protein is relaxed in the absence of constraints
Scoring Granularity	
Full_Length	All residues in the protein are used to compute the score
Monomer_One/Monomer_Two	Residues from a single chain (A/B) are used to compute the score
Domain_One/Domain_Two	All residues in the N-terminal or C-terminal domain from a single chain (A/B), whichever contains the amino acid of interest, is used to compute the score
Residue	The single residue site post mutation is used to compute the score
WT	The single residue site pre mutation is used to compute the score
Energy Terms	
Total_score	"beta_nov15" containing weighted contributions from all energy terms assayed

* All energy terms are described in detail in Alford *et. al.*¹⁹⁴

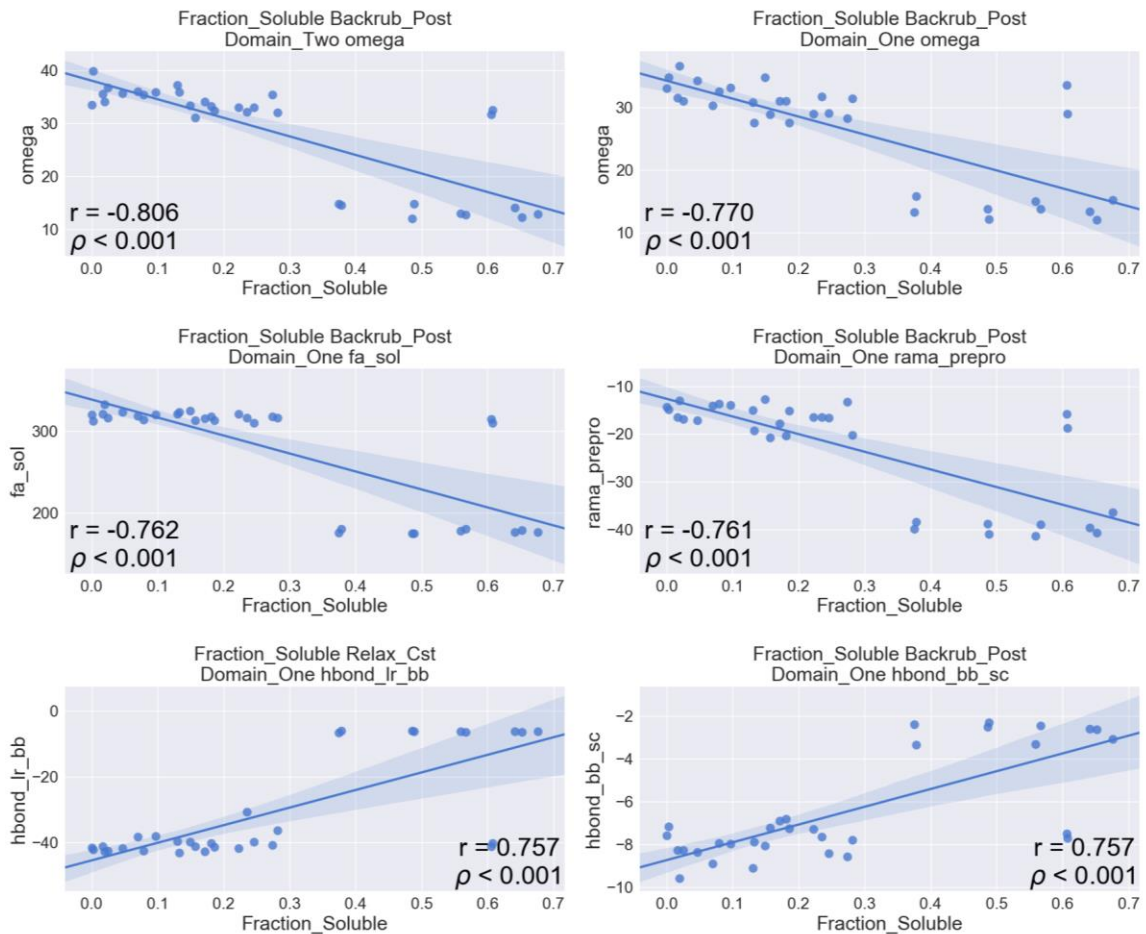


Figure 4.6a: Comparison of Rosetta energy terms to the soluble fraction of Acd mutant LexA protein expression. Depicted are members of the top-20 correlates of single energy terms from Rosetta simulations to the fraction of soluble protein from expression of single-point Acd mutants of LexA. Figure titles (experimental data, sampling scheme, scoring granularity, energy term) shown above each plot follow the nomenclature detailed in Table 4.1.

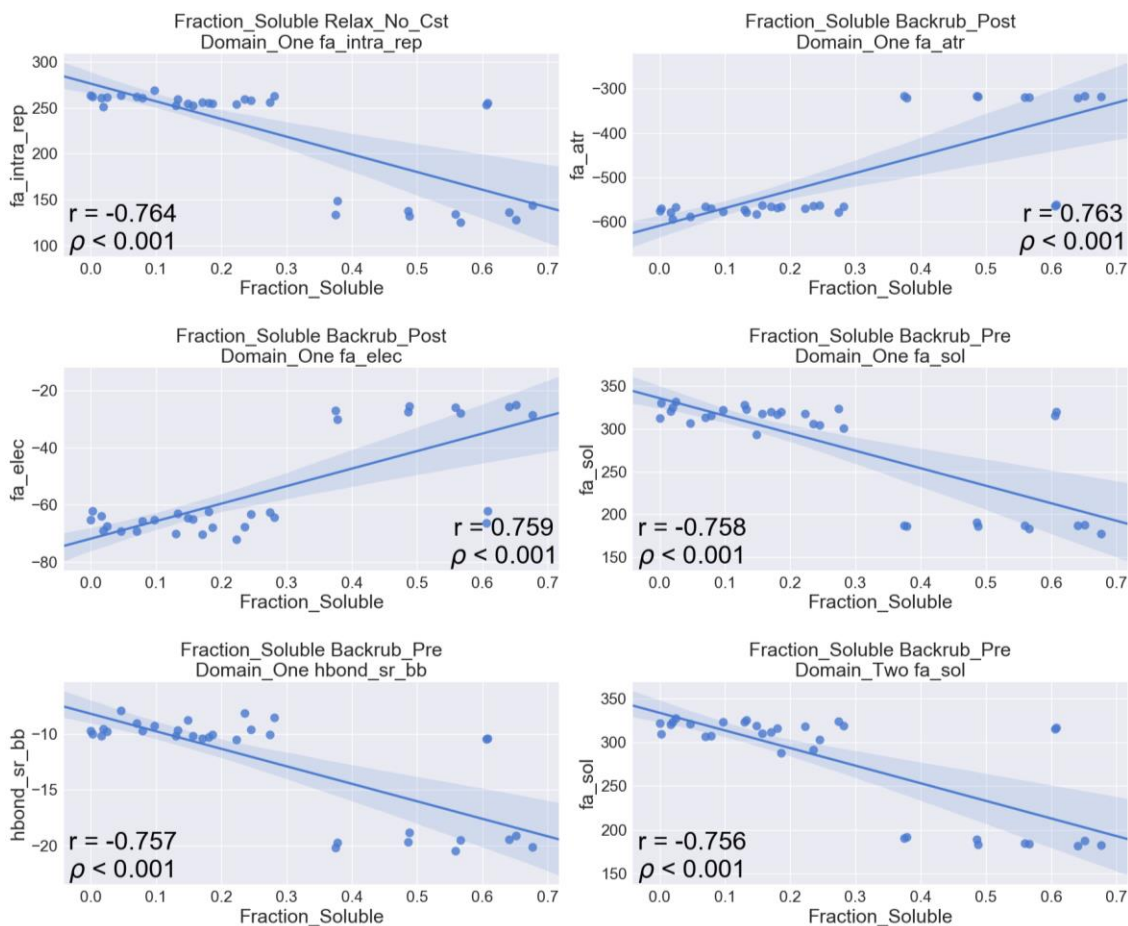


Figure 4.6b: Comparison of Rosetta energy terms to the soluble fraction of Acd mutant LexA protein expression. Depicted are members of the top-20 correlates of single energy terms from Rosetta simulations to the fraction of soluble protein from expression of single-point Acd mutants of LexA. Figure titles (experimental data, sampling scheme, scoring granularity, energy term) shown above each plot follow the nomenclature detailed in Table 4.1.

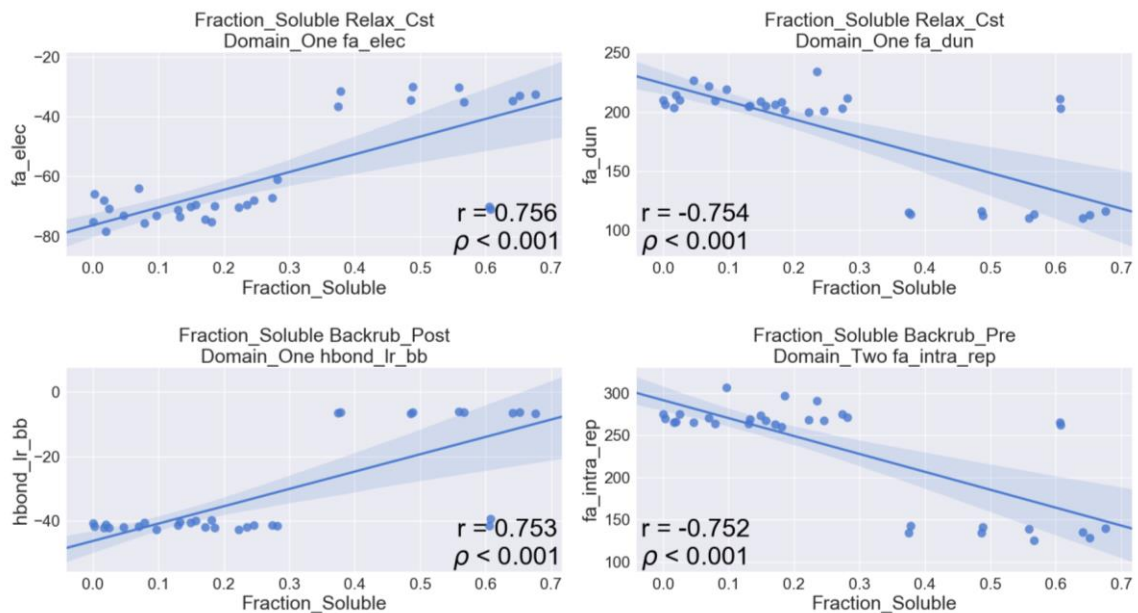


Figure 4.6c: Comparison of Rosetta energy terms to the soluble fraction of Acd mutant LexA protein expression. Depicted are members of the top-20 correlates of single energy terms from Rosetta simulations to the fraction of soluble protein from expression of single-point Acd mutants of LexA. Figure titles (experimental data, sampling scheme, scoring granularity, energy term) shown above each plot follow the nomenclature detailed in Table 4.1.

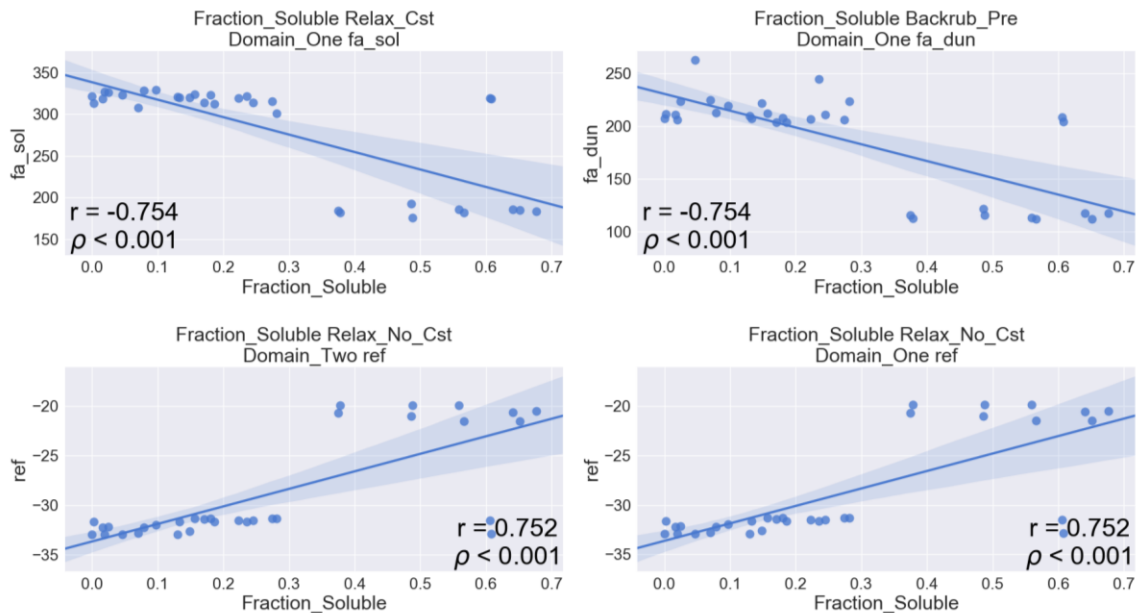


Figure 4.6d: Comparison of Rosetta energy terms to the soluble fraction of Acd mutant LexA protein expression. Depicted are members of the top-20 correlates of single energy terms from Rosetta simulations to the fraction of soluble protein from expression of single-point Acd mutants of LexA. Figure titles (experimental data, sampling scheme, scoring granularity, energy term) shown above each plot follow the nomenclature detailed in Table 4.1.

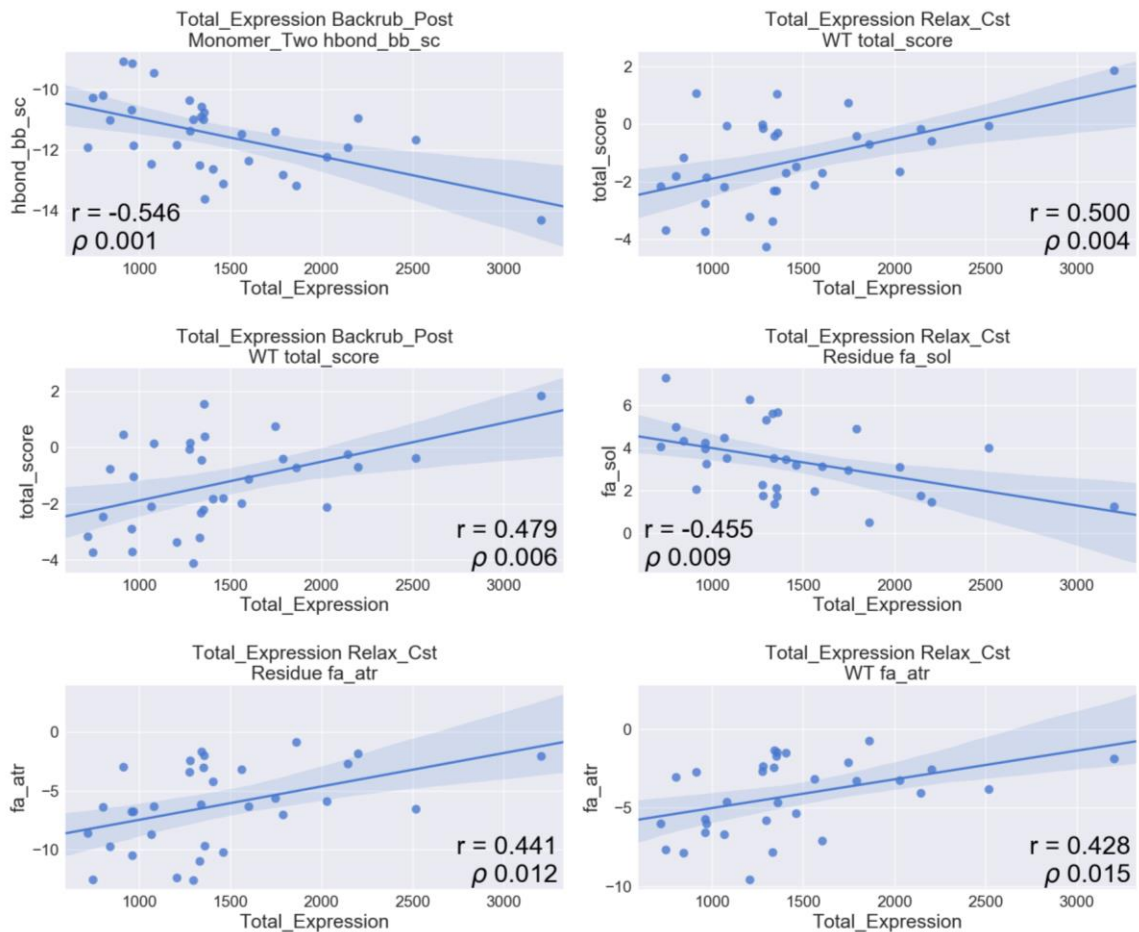


Figure 4.7a: Comparison of Rosetta energy terms to the total expression of Acd mutant LexA proteins. Depicted are members of the top-20 correlates of single energy terms from Rosetta simulations to the total protein yield from expression of single-point Acd mutants of LexA. Figure titles (experimental data, sampling scheme, scoring granularity, energy term) shown above each plot follow the nomenclature detailed in Table 4.1.

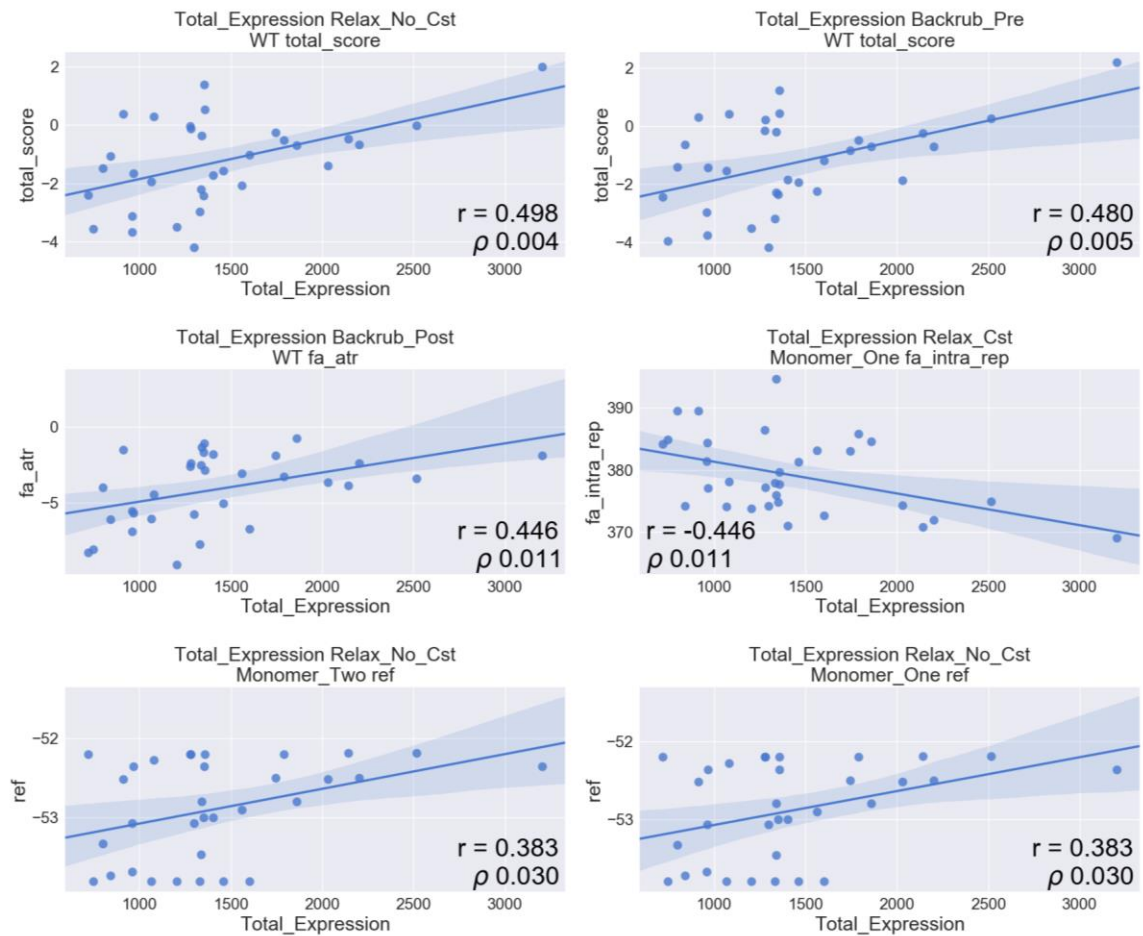


Figure 4.7b: Comparison of Rosetta energy terms to the total expression of Acd mutant LexA proteins. Depicted are members of the top-20 correlates of single energy terms from Rosetta simulations to the total protein yield from expression of single-point Acd mutants of LexA. Figure titles (experimental data, sampling scheme, scoring granularity, energy term) shown above each plot follow the nomenclature detailed in Table 4.1.

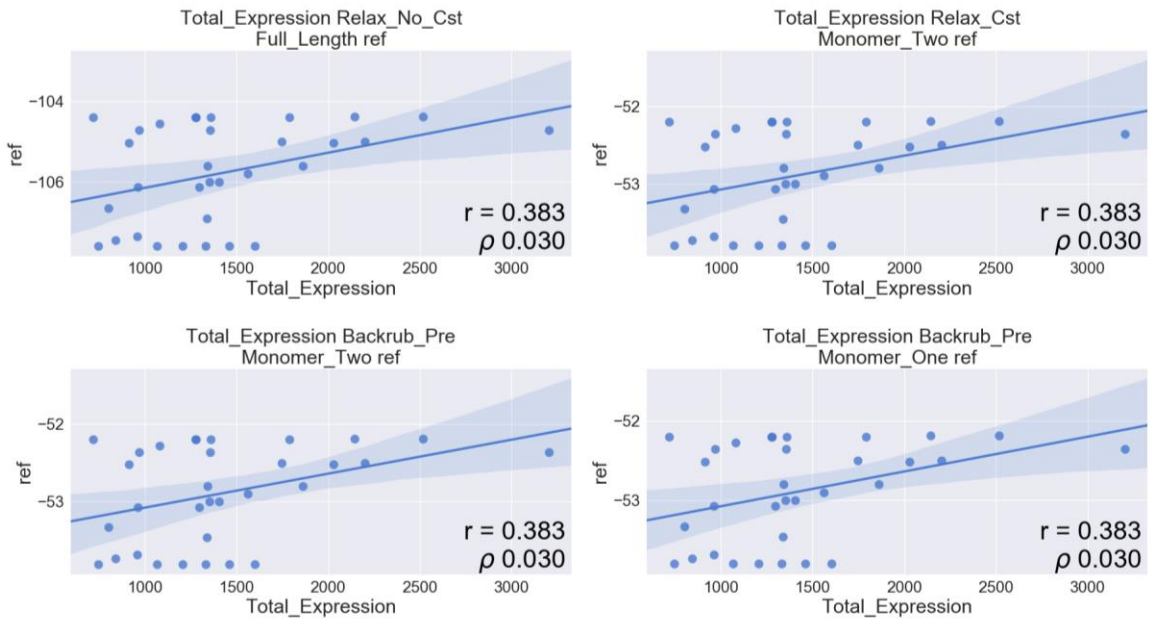


Figure 4.7c: Comparison of Rosetta energy terms to the total expression of Acd mutant LexA proteins. Depicted are members of the top-20 correlates of single energy terms from Rosetta simulations to the total protein yield from expression of single-point Acd mutants of LexA. Figure titles (experimental data, sampling scheme, scoring granularity, energy term) shown above each plot follow the nomenclature detailed in Table 4.1.

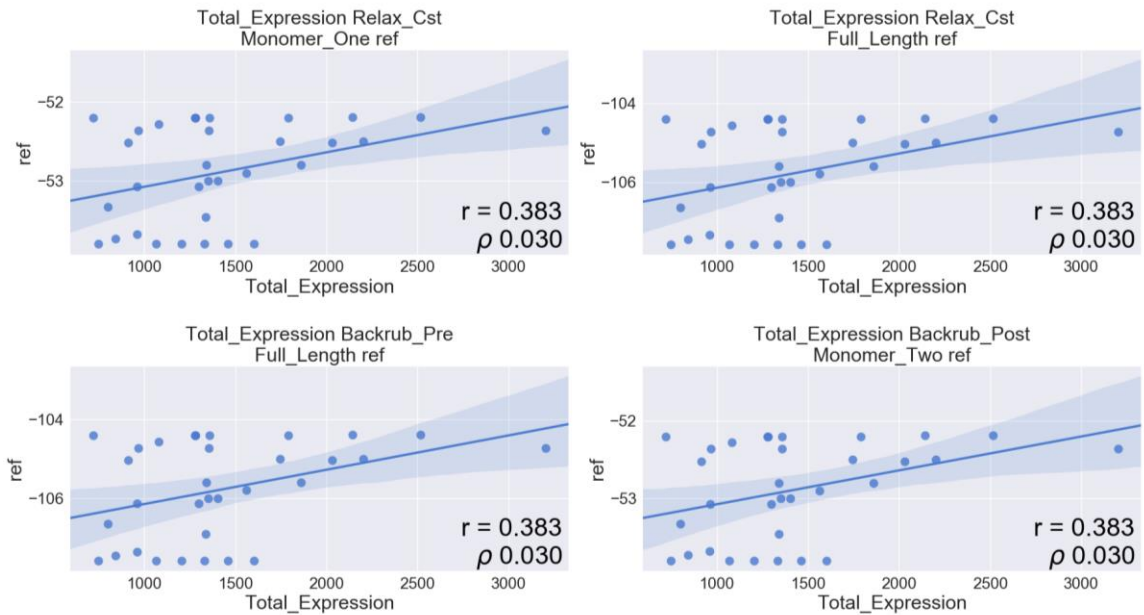


Figure 4.7d: Comparison of Rosetta energy terms to the total expression of Acd mutant LexA proteins. Depicted are members of the top-20 correlates of single energy terms from Rosetta simulations to the total protein yield from expression of single-point Acd mutants of LexA. Figure titles (experimental data, sampling scheme, scoring granularity, energy term) shown above each plot follow the nomenclature detailed in Table 4.1.

CHAPTER 5: USING A FRET LIBRARY WITH MULTIPLE PROBE PAIRS TO DRIVE MONTE CARLO SIMULATIONS OF ALPHA-SYNUCLEIN.

The content of this chapter was originally published in *Biophysical Journal*. It is adapted here with permission from the publisher:

This article was published in *Biophysical Journal*, 114, John J. Ferrie, Conor M. Haney, Jimin Yoon, Buyan Pan, Yi-Chih Lin, Zahra Fakhraai, Elizabeth Rhoades, Abhinav Nath, and E. James Petersson, Using a FRET Library with Multiple Probe Pairs to Drive Monte Carlo Simulations of α -Synulcein, 53-64, Copyright Biophysical Society 2017.

§ 5.1 Introduction

Intrinsically disordered proteins (IDPs) and proteins containing disordered regions are exceptionally responsive to changes in solution conditions, making them prone to misfolding and aggregation. One such IDP is α -synuclein (α S), a 140-amino acid neuronal protein the aggregation of which is implicated in Parkinson's Disease pathogenesis¹⁹⁻²⁰. α S is primarily expressed at presynaptic termini and is suspected to play roles in regulating neurotransmitter release and maintaining synaptic function and plasticity¹⁹⁵. Likely the most recognized aspect of α S is its ability to misfold and self-associate, resulting in the production of toxic amyloid fibrils. These fibrils are the primary components of Lewy Bodies, which have been long been recognized as a post mortem hallmark of Parkinson's Disease¹⁹. The protein is comprised of three domains: the N-terminal domain (residues 1-60), comprised of four imperfect 11-amino acid repeats featuring a KTKEGV motif featured in amphipathic helices; the non-amyloid β component, or NAC domain (residues 61-95), which contains two additional KTKEGV repeats and forms the β -sheet rich core of amyloid aggregates; and the acidic C-terminal domain, which is highly charged and considered largely disordered. The first two domains adopt an α -helical structure in the presence of lipid membranes or detergent micelles, while in solution the α S monomer is largely disordered^{20, 23-25, 196-197}. The structural plasticity of α S, and most IDPs, is attributed to a lack of hydrophobic residues and an excess of charged residues, producing systems that are self-repulsive and unable to form a collapsed hydrophobic core¹⁹⁷. Despite these characteristics, α -synuclein has been shown to be partially collapsed and to

deviate from a true random coil structure^{104, 108, 126, 198-200}. Moreover, the dynamic structure of α S has high environmental sensitivity that has largely hindered development of a cohesive characterization of the structural ensemble of monomeric α S in solution^{64, 68, 104, 108, 126, 197, 199-202}. Solution conditions including temperature, salt concentration, and pH all have substantial effects on the partial folding and collapse of α S. Furthermore, cosolvents and osmolytes have been shown to drive conformational readjustment of α S^{197, 203-206}. In particular, seminal work by Uversky, Fink and coworkers demonstrated the ability of trimethylamine-*N*-oxide (TMAO) to promote compaction and partial helical folding of α S monomers, and in high concentrations drive the formation of helical oligomers¹³³. Intriguingly, these helical conformations are morphologically distinct from the helical structure formed on membranes, and the presence of osmolytes has been shown to affect fibril formation¹³⁴.

TMAO is a naturally occurring amphiphilic osmolyte that is found in several aquatic organisms, where it counteracts the destabilizing effects of high concentrations of urea required for regulation of osmotic pressure^{131-132, 207}. Experimental and theoretical efforts have afforded an effective model by which preferential exclusion of TMAO from the protein backbone and sequestration of water by TMAO promotes the formation of intramolecular hydrogen bonds and a reduction in exposed surface area for the protein^{30, 208-213}. Solution studies of TMAO have revealed that the large 4.67 D dipole moment prompts significant water ordering around each molecule²¹⁴. Comprised of a total of ~13 water molecules, direct coordination of water to the oxygen along with formation of a

clathrate like structure about the methyl groups produces a first solvation shell with a 6 Å radius and elicits an excluded volume effect that entropically drives protein compaction^{208, 214-216}. In addition to the effects of excluded volume and water arrangement, TMAO has been proposed to act as a nanocrowder and also serves as a poor solvent of the peptide backbone²¹⁷⁻²¹⁸. Single-molecule Förster resonance energy transfer (FRET) has been previously used by Deniz and coworkers to investigate the compact structure of α S in TMAO¹³⁴⁻¹³⁵. These studies were performed with a limited set of measurements and demonstrated that successive compaction of α S occurs with increasing concentrations of TMAO while maintaining a single, broadly distributed conformational state. Moreover, the work by Deniz and coworkers showed that despite the emergence of a partially helical secondary structure, the folding pathway and resulting structure were distinct from helices formed on membranes. Although a putative structural ensemble was not proposed in this work, our laboratory and others have demonstrated the ability to apply distances and distributions obtained from single molecule FRET data as constraints for Monte Carlo (MC) simulations⁶⁸. In the context of the present problem of TMAO compaction of α S, we wish to further develop a method for employing FRET constraints to generate experimentally-constrained models of intrinsically disordered protein ensembles.

Although the use of FRET data has not been extensively explored as a basis for generating structural ensembles of IDPs, the application of other long-range measurements has demonstrated success. Work by Forman-Kay, Zweckstetter, Blackledge and others has shown the efficacy of utilizing data from

paramagnetic relaxation enhancement (PRE) NMR experiments as restraints for simulating ensembles of IDPs^{61, 107-108, 199-200}. Dobson, Vendruscolo and Eliezer have also applied NMR-based methods to study the disordered ensemble of α S^{104, 126}. Moreover, efforts by Langen have demonstrated the similar usefulness of electron paramagnetic resonance (EPR)^{196, 219}. To date, most studies have focused on the application of PRE data, which lacks description of the underlying distribution of states⁶¹. Single molecule FRET has demonstrated an exceptional ability to visualize subpopulations of disordered ensembles⁶⁰. Furthermore, Best and Schuler have addressed some of the major concerns surrounding distance extraction from FRET data^{64, 202}. Studies of chemically denatured ubiquitin have shown that FRET and SAXS data afford comparable molecular sizes and that the inclusion of large hydrophobic probe molecules does not significantly impact the structural ensemble²⁰². Moreover, recent work by Schuler has verified that distributions extracted from single molecule FRET data display a high degree of agreement with distributions compiled from structural ensembles generated from molecular dynamics simulations restrained with NMR and SAXS data⁶⁴.

In this report, we focus on elucidating the structural changes associated with TMAO induced compaction of α S by combining MC simulations in PyRosetta with experimental constraints from ensemble FRET measurements. Although single molecule FRET measurements provide additional information regarding the number of distributions as well as the distribution breadth associated with a particular average value, the photophysical requisites for these measurements (i.e., bright, visible wavelength fluorophore pairs, which tend to have 30-70 Å

working ranges) generally preclude accurate measurement of distances below ~30 Å. This limitation is important, as we wish to generate atomically-detailed computational models of α S and therefore need short distance constraints for our simulations. Based on the aforementioned results from Deniz and coworkers, we assume that the observed ensemble FRET efficiencies are resultant from single distributions, well described by polymer physics models, allowing us to rely on distances obtained from single distribution analysis^{60, 134-135}. Furthermore, we propose that the most effective set of constraints would encompass not only long distances, such as those traditionally afforded from single molecule FRET, but also short distances, closer to those obtained via PRE measurements, which should more effectively limit the conformational variability within the ensemble.

Previously, we have studied α S compaction using the *p*-cyanophenylalanine (Cnf) and thioamide probe pair, which has a short Förster radius (R_0 , the distance of half-maximal energy transfer) of 18 Å^{136, 220}. We showed that increases in FRET efficiency (E_{FRET}) indicating compaction are observable as α S is incubated with increasing concentrations of TMAO. However, these Cnf-thioamide labeled proteins required semi-synthesis through native chemical ligation, a method with insufficient throughput for FRET library construction. Herein, we employ Cnf-tryptophan (Trp) as a genetically incorporable, short range probe pair allowing for more facile production of a library of labeled constructs for measuring distinct intramolecular distances in the presence of TMAO²²¹⁻²²². Additionally, we have incorporated a second, longer range probe pair, fluorescein-5-maleimide (Fam) conjugated to cysteine (Cys) and tetramethylrhodamine azide

(Raz) conjugated to an *O*-propargyl tyrosine (Ppy) unnatural amino acid ⁷⁸. The Fam-Raz pair has a working distance around 50 Å and provides accuracy in ranges equivalent to those accessible through single molecule FRET.

By applying measurements from these two complementary FRET libraries as weighted constraints for atomically-detailed MC simulations in PyRosetta, we construct model conformational ensembles that agree well with experimental data such as fluorescence correlation spectroscopy (FCS), reporting on the overall average size of the protein. This strategy of using FRET data to direct all atom models in MC simulations provides a means to generate structure-guided hypotheses for allosteric transitions in α S in order to understand the effects of changes in environment or interactions with ligands.

§ 5.2 Protein Overexpression and Purification

Protein expression was performed in *E. Coli*. where unnatural amino acids were incorporated via amber stop codon suppression and traceless purification was facilitated via attachment of a C-terminal intein containing a C-terminal His-tag (Fig. 5.1). For the Cnf-Trp library, all native tyrosine residues were mutated to phenylalanine to assure that all energy transfer occurred exclusively between Cnf and Trp.

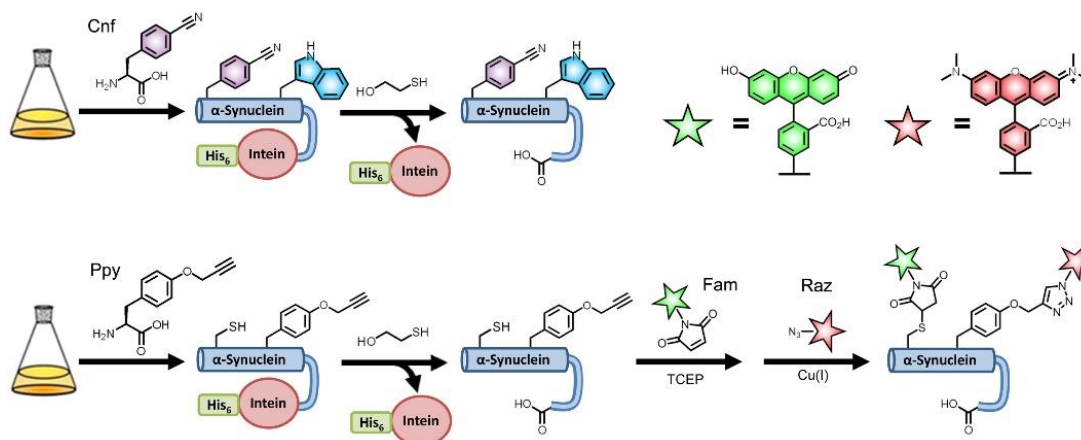


Figure 5.1: Scheme for labeled protein production. Top: Direct incorporation of p-cyanophenylalanine (Cnf) and tryptophan (Trp) via unnatural amino acid mutagenesis and conventional mutagenesis, respectively. Bottom: Incorporation of Cysteine (Cys) and O-propargyl tyrosine (Ppy) with subsequent labeling by fluorescein-5-maleimide (Fam) and tetramethylrhodamine azide (Raz), respectively.

General Expression Protocol for α S-intein-H₆ Fusion Proteins:

A pTXB1 plasmid encoding for the α S-intein-H₆ fusion protein was transformed into *E. coli* BL21-Gold (DE3) cells. The DNA GyrA intein from Mxe fusion construct has been previously reported⁷⁸. Transformed cells were selected based on ampicillin (Amp) (100 mg/L) resistance, encoded on the pTXB1 plasmid containing the protein of interest. Single colonies were selected and used to inoculate 5 mL primary cultures of LB media, and were grown at 37 °C, shaking at 250 rpm in the presence of 100 μ g/mL Amp. A secondary culture of LB media containing 100 mg/L Amp was inoculated with a single saturated primary culture, and was grown at 37 °C while shaking at 250 rpm. After reaching an OD₆₀₀ of 0.7-1.0, 0.5 mM isopropyl β -D-1-thiogalactopyranoside (IPTG) was added to the

culture to induce expression. Following induction the culture continued to grow overnight at 18 °C overnight.

Expression of α S-intein-H₆ Proteins Containing Unnatural Amino Acids:

The procedure for the expression of mutants containing Cnf or Ppy was identical to the general protocol above with the following noted exceptions. The pTXB1 plasmid coding for the protein of interest contained an amber stop codon (TAG) at the intended site for unnatural amino acid introduction was co-transformed with a plasmid encoding a pDULE2-pFX plasmid containing an orthogonal aminoacyl synthetase/tRNA pair ⁷⁸. pDULE2-pFX encodes streptomycin (Strep) resistance, and cells containing both plasmids were selected for based on Strep (100mg/L) resistance in addition to Amp resistance. Primary cultures were also grown in the presence of both Strep (100 µg/mL) and Amp (100 µg/mL). In place of LB, secondary cultures were grown in M9 minimal media containing 6 g Na₂HPO₄, 3 g KH₂PO₄, 0.5 g NaCl and 1 mL of 2 M MgSO₄, 1 mL of 1mg/mL FeCl₂ (in 1.0 M HCl), 1 mL of 15 mg/mL, 2 mL of 10% Yeast Extract, 12.5 mL 40% glucose (w/v) in 1L of autoclaved water, along with 100 mg/L Amp and 100 mg/L Strep. Lastly, once the secondary culture reached an OD₆₀₀ of 0.7-1.0, 0.8 mM of the unnatural amino acid was added to culture prior to induction with IPTG. Expression was performed overnight at 18 °C.

General Purification Protocol for Cnf-Trp containing α S-intein-H6 Fusion Proteins:

Cells were harvested via centrifugation at 5000 rpm for 20 minutes at 4 °C with a GS3 rotor on a Sorvall RC-5 centrifuge. The supernatant was decanted away from the cell pellet, and the pellet was resuspended in 20 mL resuspension

buffer containing 40 mM Tris, 5 mM EDTA, pH 8.3 and a protease inhibitor tablet (Roche cOmplete mini tabs, EDTA free). Cells were lysed by sonication at an amplitude of 30 for 5 minutes with 1 second on 1 second off. Lysate was subsequently centrifuged at 14000 rpm for 20 minutes using an SS-34 rotor on a Sorvall RC-5 centrifuge. His tagged protein was isolated from the supernatant via nickel affinity. Ni-NTA resin (3 mL CV) was incubated with the supernatant on ice for 1 hour. Following incubation the mixture was loaded into a fritted column, and the lysate was drained. The resin was initially washed with 15 mL of 50 mM HEPES, pH 7.5 followed by 20 mL of 50 mM HEPES, 5 mM imidazole, pH 7.5, prior to elution of the protein of interest with 12 mL of 50 mM HEPES, 300 mM imidazole, pH 7.5. Cleavage of the intein was performed via addition of β -mercaptoethanol (β ME) to a final concentration of 200 mM and incubation at room temperature for 18 hours on a rotisserie. Following cleavage of the intein, removal of imidazole and β ME was facilitated by dialysis into 20 mM tris base pH 8.0 at 4 °C overnight. The undesired intein was removed by a second Ni column (3 mL CV). The flow through containing the protein of interest was collected after 1 hour incubation with the Ni resin on ice.

General Labeling Protocol for Fam-Raz containing α S-intein-H₆ Fusion Proteins:

α S variants containing Cys and/or Ppy were expressed as described above and purified by Ni-NTA affinity. Following the second purification subsequent to intein cleavage, the desired protein was dialyzed into 20 mM Tris, pH 8.0 overnight at 4 °C. The protein was then labeled in this semi-crude state in one (donor-only

or acceptor-only) or two (doubly-labeled) labeling steps as described below. Proteins lacking an unnatural amino acid (α S-Cys) were quantified by UV-Vis spectroscopy using $\epsilon_{280} = 5120 \text{ M}^{-1} \text{ cm}^{-1}$ with the exception of α S-Cys₁₃₆, which was quantified using $\epsilon_{280} = 3840 \text{ M}^{-1} \text{ cm}^{-1}$ ²²³. Protein variants containing Ppy were quantified using the BCA assay with bovine serum albumin (BSA) standards generated by two-fold serial dilution in water from 2 mg/mL to 0.125 mg/mL. Labeling reactions of proteins containing Cys were carried out by addition of BondBreaker® TCEP solution to a final concentration of 1 mM and incubation at room temperature for 10-15 minutes. Following this time, fluorescein-5-maleimide (Fam) was added from a 25 mM stock in DMSO to a final concentration equal to five-fold excess relative to protein (typical dye concentration was ~500-800 μ M). The labeling reaction was shielded from light by wrapping in aluminum foil and incubated at 37 °C. The labeling reaction was monitored by MALDI-TOF MS following 5-fold dilution of the reaction into water. Following complete disappearance of unlabeled protein as determined by MALDI, the labeled product was dialyzed overnight at 4 °C in 20 mM Tris pH 8.0. Proteins containing Ppy were labeled via copper(I) catalyzed azide-alkyne cycloaddition as described below; doubly-labeled proteins were first labeled with Fam as described above and subsequently by Raz. Azide-alkyne labeling reactions were performed by preparation of copper(I)-tris(3-hydroxypropyltriazolylmethyl)amine (THPTA) catalyst mixture by using 80 mM CuSO₄ in water (1.25 μ L/mL protein solution) to which was added 50 mM THPTA in water (30 μ L/mL protein solution) and then 100 mM sodium ascorbate in water (30 μ L/mL protein solution). This solution was

incubated at room temperature for 10-15 minutes prior to addition to the protein solution. Protein was labeled by addition of five equivalents of Raz from a 50 mM stock in DMSO (typical dye concentrations were ~300-500 μ M), followed by addition of the Cu(I) catalyst mixture. Labeling reactions were then shielded from light and incubated at 37 °C and reaction progress monitored by MALDI-TOF MS following 5-fold dilution of the reaction in water. Upon completion of the reaction, the labeled product was dialyzed into 20 mM Tris pH 8.0 overnight at 4 °C to remove excess dye.

Purification Protocol for All α S Proteins:

All proteins were then purified by anion-exchange chromatography on an ÄKTA FPLC system using HiTrap Q HP columns and elution between 20 mM Tris pH 8.0 (buffer A) and 20 mM Tris, 1 M NaCl pH 8.0 (buffer B). Fractions containing the desired protein were identified by MALDI-TOF MS and were pooled. All proteins from the Fam-Raz library as well as poorly expressing member of the Cnf-Trp library were concentrated using Amicon Ultra 3 kDa filters to a total volume of \leq 1 mL. Further purification was then performed using a Varian HPLC system and a Vydac C4 TP reverse-phase semi-preparative column with a flow rate of 4 mL/min and gradient between 0.1% TFA in water (solvent A) and 0.1% TFA in acetonitrile (solvent B). Fractions containing the desired product were identified by MALDI-TOF MS and judged to be of high purity were combined and diluted a minimum of 5-fold by addition of 20 mM Tris, 100 mM NaCl pH 7.5 and subsequently concentrated via Amicon μ Ltra 3 kDa centrifugal filters to a total

volume of ≤ 5 mL, diluted 3-5 fold and re-concentrated; the final dilution and concentration was repeated twice. All proteins were then assessed for purity and integrity of the fluorescent label by MALDI-TOF analysis of whole protein and tryptic fragments.

Table 5.1: MALDI Masses from Cnf-Trp Library

Construct	Full Length		Tyrpsin Digest		Obs. [M+H] ⁺	Exp. [M+Na] ⁺	Obs. [M+Na] ⁺
	Exp. [M+H] ⁺	Obs. [M+H] ⁺	Fragment	Exp. [M+H] ⁺			
α S-Cnf ³⁹	14421	14422.4	35-43	960.5	960.7	982.5	982.7
α S-Cnf ⁶⁴	14421	14421.4	81-96	1504.7	1504.0	1526.7	1526.0
α S-Cnf ¹²⁵	14421	14421.7	103-140	4266.4	4265.9	4288.4	4287.9
α S-Cnf ¹³⁶	14421	14421	103-140	4266.4	4265.9	4288.4	4287.9
α S-Trp ²⁴	14454	14455.6	24-32	888.5	888.6	910.5	910.6
α S-Trp ³⁹	14435	14434.7	35-43	974.5	974.8	996.5	996.8
α S-Trp ⁶²	14454	14454	61-80	1986.1	1986.5	2008.1	2008.5
α S-Trp ⁷⁹	14454	14454.3	61-80	1986.1	1986.4	2008.1	2008.4
α S-Trp ⁹⁴	14435	14435.4	81-96	1517.8	1518.1	1539.8	1540.1
α S-Trp ¹²⁵	14435	14434.8	103-140	4280.5	4280.0	4302.5	4302.0
α S-Cnf ³⁹ -Trp ²⁴	14479	14479.5	35-43	960.5	960.7	982.5	982.7
			24-32	888.5	888.7	910.5	910.7
α S-Cnf ³⁹ -Trp ⁶²	14479	14480.1	35-43	960.5	960.7	982.5	982.7
			61-80	1986.1	1986.4	2008.1	2008.4
α S-Cnf ³⁹ -Trp ⁷⁹	14479	14479.3	35-43	960.5	960.7	982.5	982.7
			61-80	1986.1	1986.4	2008.1	2008.4
α S-Cnf ³⁹ -Trp ⁹⁴	14460	14461.6	35-43	960.5	960.7	982.5	982.7
			81-96	1517.8	1518.1	1539.8	1540.1
α S-Cnf ⁶⁴ -Trp ²⁴	14479	14480.2	81-96	1504.7	1504.1	1526.7	1526.0
			24-32	888.5	888.6	910.5	910.6
α S-Cnf ⁶⁴ -Trp ³⁹	14460	14461.4	81-97	1504.7	1504.1	1526.7	1526.0
			35-43	974.5	974.7	996.5	996.7
α S-Cnf ⁶⁴ -Trp ⁶²	14479	14480.1	81-98	1504.7	1504.1	1526.7	1526.1
			61-80	1986.1	1986.4	2008.1	2008.4
α S-Cnf ⁶⁴ -Trp ⁷⁹	14479	14480.1	81-99	1504.7	1504.1	1526.7	1526.1
			61-80	1986.1	1986.4	2008.1	2008.4
α S-Cnf ¹²⁵ -Trp ²⁴	14479	14480.1	103-140	4266.4	4268.7	4288.4	4290.7
			24-32	888.5	888.6	910.5	910.6
α S-Cnf ¹²⁵ -Trp ³⁹	14460	14461.5	103-140	4266.4	4265.8	4288.4	4287.8
			35-43	974.5	974.7	996.5	996.7
α S-Cnf ¹²⁵ -Trp ⁶²	14479	14479.6	103-140	4266.4	4265.3	4288.4	4287.3
			61-80	1986.1	1986.4	2008.1	2008.4
α S-Cnf ¹²⁵ -Trp ⁷⁹	14479	14480	103-140	4266.4	4265.5	4288.4	4287.5
			61-80	1986.1	1986.4	2008.1	2008.4
α S-Cnf ¹²⁵ -Trp ⁹⁴	14460	14460	103-140	4266.4	4265.7	4288.4	4287.7
			81-96	1517.8	1518.1	1539.8	1540.0
α S-Cnf ¹³⁶ -Trp ²⁴	14479	14478.5	103-140	4266.4	4265.9	4288.4	4287.9
			24-32	888.5	888.6	910.5	910.6
α S-Cnf ¹³⁶ -Trp ³⁹	14460	14460.9	103-140	4266.4	4266.2	4288.4	4288.2
			35-43	974.5	974.7	996.5	996.7
α S-Cnf ¹³⁶ -Trp ⁹⁴	14460	14461.6	103-140	4266.4	4265.5	4288.4	4287.5
			81-96	1517.8	1518.1	1539.8	1540.1
α S-Cnf ¹³⁶ -Trp ¹²⁵	14460	14460.9	103-140	4305.5	4304.8	4327.5	4326.8

Table 5.2: MALDI Masses from Fam-Raz Library

Construct	Full Length		Trypsin Digest				
	Exp. [M+H] ⁺	Obs. [M+H] ⁺	Fragment	Exp. [M+H] ⁺	Obs. [M+H] ⁺	Exp. [M+Na] ⁺	Obs. [M+Na] ⁺
α S-Raz ⁹⁴	15027	15027.8	81-96	2044.0	2046.6	2066.0	2068.6
α S-Raz ¹³⁶	15011	15013.3	103-140	4839.0	4837.4	4861.0	4859.4
α S-Fam ⁹ -Raz ¹³⁶	15474	15475.1	7-10	865.3	865.9	887.3	887.9
			103-140	4839.0	4837.1	4861.0	4859.1
α S-Fam ⁹ -Raz ⁹⁴	15488	15488.1	7-10	865.3	865.6	887.3	887.6
			81-96	2044.0	2044.6	2066.0	2066.6
α S-Fam ²⁴ -Raz ⁹⁴	15447	15450.7	24-32	1250.5	1250.3	1272.4	1272.3
			81-96	2044.0	2044.8	2066.0	2066.8
α S-Fam ⁴² -Raz ⁹⁴	15488	15490.1	35-43	1412.6	1413.7	1434.6	1435.7
			81-96	2044.0	2045.3	2066.0	2067.3
α S-Fam ⁶² -Raz ⁹⁴	15447	15448.5	61-80	2348.1	-	2370.1	2371.7
			81-96	2044.0	2046.4	2066.0	2068.4
α S-Fam ⁸⁷ -Raz ⁹⁴	15488	15488.4	81-96	2505.1	2506.6	2527.0	2528.6
α S-Fam ¹¹⁴ -Raz ⁹⁴	15446	15446.5	103-140	4704.8	-	4726.8	4728.8
			81-96	2044.0	2046.4	2066.0	2068.4
α S-Fam ¹²³ -Raz ⁹⁴	15446	15446.1	103-140	4704.8	-	4726.8	4728.9
			81-96	2044.0	2046.4	2066.0	2068.4
α S-Fam ¹³⁶ -Raz ⁹⁴	15410	15413.8	103-140	4668.8	-	4690.7	4694.8
			81-96	2044.0	2046.4	2066.0	2068.4

§ 5.3 FRET Data Collection and Analysis

TMAO Assay:

All labeled α S variants were dialyzed into 20 mM tris, 100 mM NaCl, pH 7.5. Buffers containing TMAO were also prepared with 20 mM tris, 100 mM NaCl, and the pH was adjusted to 7.5 following the addition of TMAO. Concentrations for the Cnf-Trp library was determined using the Sigma-Aldrich FluoroProfile Quantification Kit, while concentrations for Fam-Raz constructs were determined via UV-Vis absorbance. Steady-state measurements for the Cnf-Trp library and time correlated single photon counting (TCSPC) measurements of the Fam-Raz

library were performed at 1 μ M, while steady-state measurements for the Fam-Raz library were performed at 100 nM. Prior to each measurement, labeled protein was mixed with TMAO containing buffer and briefly vortexed. Measurements were taken in triplicate at final concentrations of 0, 2 and 4 M TMAO at 20 °C. Steady-state measurements for the Cnf-Trp library were collected with an excitation of 240 nm over an emission range of 275 - 410 nm with excitation and emission slits set to 5 nm and 1 nm step size collecting for 0.75 seconds per step, exciting primarily the Cnf fluorophore. Spectra were collected for direct excitation of Trp with an excitation of 280 nm over an emission range of 310 - 410 nm with excitation and emission slits set to 5 nm and 1 nm step size collecting for 0.75 seconds per step. Measurements of the Fam-Raz library were collected with an excitation of 486 nm over an emission range of 495 - 700 nm with excitation and emission slits set to 5 nm and 1 nm step size collecting for 0.25 seconds per step. Direct excitation of Raz was performed by exciting at 555 and measuring over an emission range of 565 - 700 nm with a 1 nm step size and a collection time of 0.25 seconds per step.

All TCSPC measurements of fluorescence lifetime decays were collected using a pulsed LED with a maximum emission at 486 nm. Fluorescence was collected at 515 nm with the slit widths adjusted for each measurement to keep the ACD value between 1 - 3 % of the SYNC value. The instrument response function (IRF) was collected for each slit width used for collection. For these experiments labeled α S was mixed with TMAO containing buffer to final protein concentrations of 2 μ M and 0, 2 or 4 M TMAO. Additionally, formation of aggregates in buffer or TMAO were assayed by mixing single-labeled α S containing FAM and single-

labeled α S containing Raz in equimolar concentrations with a final protein concentration of 2 μ M in 0, 2 and 4 M TMAO.

Fitting Steady-State Data:

Following data collection, the single labeled spectra were used to quantify the degree of energy transfer. First, the spectral overlap of the donor and acceptor were deconvoluted by fitting the double labeled spectrum with the linear sum of the individual donor and acceptor-labeled spectra. Fitting was performed by minimizing the total least squared difference using the Excel Solver feature to adjust the constants A and B :

$$\sum_{\lambda} (I(\lambda)_{DA} - AI(\lambda)_D - BI(\lambda)_A)^2 \rightarrow \min \quad (\text{Eq. 5.1})$$

$$I(\lambda)_{DA} = AI(\lambda)_D + BI(\lambda)_A \quad (\text{Eq. 5.2})$$

Here, $I(\lambda)_{DA}$, $I(\lambda)_D$ and $I(\lambda)_A$ are the wavelength dependent fluorescence intensities of the double-labeled, single labeled protein containing the donor, and single-labeled protein containing the acceptor fluorophore, respectively. Solutions to Eq. 5.2 were obtained by utilizing the Excel Solver functionality. This procedure was performed at each TMAO concentration, thus accounting for any changes in quantum yield or spectral shifting. The linear contributions of the single-labeled construct containing the donor only, A , and the contribution from the single-labeled construct containing the acceptor only, B , were both used to independently

calculate the E_{FRET} through Eq. 5.3 and 5.4 and combined in a weighted average via Eq. 5.5:

$$E_D = (1 - A) \quad (\text{Eq. 5.3})$$

$$E_A = (B - 1) \frac{\varepsilon_A}{\varepsilon_D} \quad (\text{Eq. 5.4})$$

$$E_{FRET} = \left(\frac{1}{S_D} + \frac{1}{S_A} \right)^{-1} \left(\frac{E_D}{S_D} + \frac{E_A}{S_A} \right) \quad (\text{Eq. 5.5})$$

In Eqs. 5.3-5.5, E_D and E_A are E_{FRET} values calculated from the donor and acceptor weights respectively. Additionally, Eq. 5.4 requires ratio of the extinction coefficient for the acceptor, ε_A , to the donor, ε_D , which are detailed in Tables 5.3a and 5.3b and have been determined from the absorption spectrum of each fluorophores scaled using published extinction coefficients (Cnf $\varepsilon_{240}=13,921 \text{ M}^{-1}\text{cm}^{-1}$ ²²¹, Trp $\varepsilon_{278}=5700 \text{ M}^{-1}\text{cm}^{-1}$ ²²¹, FAM $\varepsilon_{494}=68,000 \text{ M}^{-1}\text{cm}^{-1}$ ²²⁴ and Raz $\varepsilon_{555}=87,000 \text{ M}^{-1}\text{cm}^{-1}$ as indicated by the manufacturer) . The E_{FRET} values from the donor and acceptor were then used to compute a weighted average E_{FRET} value by using the inverse of the experimental error to weight each E_{FRET} value, where S_D and S_A represent the donor and acceptor propagated error respectively.

$$S_A = A \times \sqrt{(S_{I_A} \times B / I_D)^2 + (S_{I_{DL}} / I_D)^2 + (S_{I_D} \times (I_{DL} - B \times I_A) / I_D^2)^2} \quad (\text{Eq. 5.6})$$

$$S_B = B \times \sqrt{(S_{I_D} \times S / I_A)^2 + (S_{I_{DL}} / I_A)^2 + (S_{I_A} \times (I_{DL} - A \times I_D) / I_A^2)^2} \quad (\text{Eq. 5.7})$$

$$S_{EFRET} = \sqrt{2} / \left(\frac{1}{S_A} + \frac{1}{S_B} \right) \quad (\text{Eq. 5.8})$$

Lastly, S_{EFRET} represents the propagated error of the calculated weighted average E_{FRET} value.

Accurate calculation of R_0 was required for calculation of interresidue distances from the determined E_{FRET} . R_0 was calculated using the equation:

$$R_0^6 = (9 \ln(10) \kappa^2 \Phi_D J) / (128 \pi^2 N_A n^4) \quad (\text{Eq. 5.9})$$

Here, N_A is Avogadro's number, κ^2 is the dipole orientation factor, approximated at 2/3, Φ_D is the quantum yield of the donor, J is the spectral overlap integral between the emission of the donor and the absorbance of the acceptor and n is the refractive index of the medium. The overlap integral of the donor fluorescence and acceptor absorbance spectra for each fluorophore pair was determined empirically from the absorbance and emission spectra of the free fluorophores in buffer through application of the integral:

$$J = \int_0^\infty f_D(\lambda) \varepsilon_A(\lambda) \lambda^4 d\lambda \quad (\text{Eq. 5.10})$$

where $f_D(\lambda)$ is the normalized donor emission, $\varepsilon_A(\lambda)$ is the molar extinction coefficient of the acceptor, at each wavelength (λ). The normalized donor emission is given by:

$$f_D(\lambda) = \frac{F_D(\lambda)}{\int_0^\infty F_D(\lambda) d\lambda} \quad (\text{Eq. 5.11})$$

where $F_D(\lambda)$ is the fluorescence emission spectrum of the donor dye. Following determination of R_0 average distance values from the polymer scaled Förster equation (Eq. 5.12), using either Eq 5.13 or Eq 5.14 to determine $P_n(r,x)$, along with the associated error and constraint weights for each constraining function, were determined in Wolfram Mathematica.

$$E_{FRET} = \sum_r P_n(r, x) / (1 + (r/R_0)^6) \quad (\text{Eq. 5.12})$$

$$P_1(r, x) = 4\pi r \left(\frac{3}{2\pi x^2} \right)^{3/2} \exp\left(-\frac{3}{2} \frac{r^2}{x^2}\right) \quad (\text{Eq. 5.13})$$

$$P_2(r, x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2} \frac{(r-x)^2}{\sigma^2}\right) \quad (\text{Eq. 5.14})$$

Fluorescence measurements of labeled proteins have shown that the quantum yield is sensitive to the local environment, as well as TMAO concentration. Therefore, changes in quantum yield must also be taken into account to effectively determine R_0 via Eq. 5.9. The quantum yield was calculated by fitting the emission spectrum of the free fluorophore in buffer, without TMAO, to the emission spectrum of each labeled analog at each concentration of TMAO, again using a linear least squared difference approach.

$$\sum_{\lambda} (I_D - C \times I_{Dye})^2 \quad (\text{Eq. 5.15})$$

The sum in Eq. 5.15 was minimized using the Excel solver function by adjusting the constant, C. Here, I_D and I_{Dye} represent the sum of the fluorescence

intensity over all wavelengths of the labeled protein and the free fluorophore, respectively. The donor quantum yield is then defined as:

$$\Phi_D = C \times \Phi_0 \quad (\text{Eq. 5.16})$$

where Φ_0 is the quantum yield of the free fluorophore. By using this empirically adjusted quantum yield in the calculation of R_0 , we effectively reduce inaccuracies in the determined interchromophore distance arising from changes in quantum yield. Lastly, the distance of interest, R , is determined from the above variables as described in the main text.

Finally, error was propagated through the calculation of the interprobe distance. This was performed by determining the inverse function of $P_n(r,x)$, here represented as $F(E_{FRET})$, using Wolfram Mathematica as shown in Eq. 5.17.

$$F(E_{FRET}) = P_n^{-1}(r, x) \quad (\text{Eq. 5.17})$$

The error is then simply propagated following the determination of the partial derivative of $P_n^{-1}(E_{FRET})$ with respect to the E_{FRET} variable as shown in Eq. 5.18.

$$S_{Distance} = S_{EFRET} \times \frac{\delta F(E_{FRET})}{\delta E_{FRET}} \quad (\text{Eq. 5.18})$$

Table 5.3a: Steady-State FRET Fitting and Distance Determination

Pair	PSS	TMAO	Q ₀	R ₀	ϵ_A/ϵ_D	A	E _D	S _A	B	E _A	S _B	E _{FRET}	S _{FRET}	GaussianChain		Semi-Empirical	
														r (Å)	S _{Dist}	r (Å)	S _{Dist}
Cnf ⁶⁹ -Trp ²⁴	15	0M	0.10	16.7	0.13	0.587	0.413	0.094	1.619	0.080	0.005	0.096	0.006	39.1	0.8	26.4	0.4
	15	2M	0.04	14.6	0.13	0.613	0.387	0.140	1.782	0.102	0.008	0.116	0.010	31.3	0.9	21.4	0.4
	15	4M	0.04	14.2	0.13	0.233	0.767	0.146	1.390	0.051	0.010	0.094	0.013	30.0	1.0	20.7	0.5
Cnf ⁶⁹ -Trp ⁶²	23	0M	0.10	16.7	0.13	0.599	0.401	0.138	1.841	0.109	0.006	0.121	0.008	36.5	0.7	25.2	0.3
	23	2M	0.04	14.6	0.13	0.535	0.465	0.218	2.173	0.153	0.007	0.163	0.010	27.8	0.7	19.9	0.3
	23	4M	0.04	14.2	0.13	0.155	0.845	0.209	1.386	0.050	0.010	0.085	0.013	31.7	1.3	21.4	0.6
Cnf ⁶⁹ -Trp ⁷⁹	42	0M	0.10	16.7	0.13	0.675	0.325	0.078	1.201	0.026	0.005	0.044	0.007	52.3	2.2	32.6	1.0
	42	2M	0.04	14.6	0.13	0.589	0.411	0.198	1.564	0.073	0.011	0.090	0.014	34.6	1.5	22.9	0.7
	42	4M	0.04	14.2	0.13	0.111	0.889	0.177	1.382	0.050	0.009	0.089	0.012	31.0	0.5	21.1	0.2
Cnf ⁶⁹ -Trp ⁹⁴	55	0M	0.10	16.7	0.13	0.920	0.080	0.097	1.697	0.091	0.010	0.090	0.012	44.1	2.3	28.8	1.1
	55	2M	0.04	14.6	0.13	0.950	0.050	0.206	2.194	0.155	0.025	0.144	0.031	33.2	3.1	22.3	1.4
	55	4M	0.04	14.2	0.13	0.250	0.750	0.168	1.938	0.122	0.006	0.145	0.009	25.5	0.5	18.6	0.2
Cnf ⁶⁴ -Trp ²⁴	70	0M	0.08	16.2	0.13	1.104	-0.104	0.033	1.156	0.020	0.007	0.020	0.007	74.1	9.5	42.5	4.7
	70	2M	0.04	14.4	0.13	0.916	0.084	0.059	1.122	0.016	0.010	0.026	0.012	58.4	8.9	33.2	4.0
	70	4M	0.03	13.8	0.13	0.350	0.650	0.080	1.496	0.064	0.009	0.123	0.011	27.4	1.0	19.2	0.4
Cnf ⁶⁴ -Trp ³⁹	55	0M	0.08	16.2	0.13	0.945	0.055	0.097	1.655	0.085	0.013	0.082	0.016	44.8	3.3	28.7	1.5
	55	2M	0.04	14.4	0.13	0.741	0.259	0.088	1.496	0.064	0.008	0.081	0.010	37.0	1.7	23.9	0.7
	55	4M	0.03	13.8	0.13	0.265	0.735	0.094	2.002	0.130	0.006	0.166	0.008	23.3	0.5	17.4	0.2
Cnf ⁶⁴ -Trp ⁶²	32	0M	0.08	16.2	0.13	0.950	0.050	0.080	2.401	0.182	0.009	0.169	0.011	34.4	1.1	23.9	0.5
	32	2M	0.04	14.4	0.13	0.853	0.147	0.135	3.059	0.288	0.012	0.258	0.015	24.8	0.7	18.4	0.3
	32	4M	0.03	13.8	0.13	0.305	0.695	0.092	1.920	0.120	0.006	0.153	0.008	24.2	0.5	17.8	0.2
Cnf ⁶⁴ -Trp ⁷⁹	15	0M	0.08	16.2	0.13	0.543	0.457	0.134	2.045	0.136	0.006	0.149	0.008	32.2	0.5	22.9	0.2
	15	2M	0.04	14.4	0.13	0.453	0.547	0.182	2.114	0.145	0.008	0.162	0.011	27.2	0.6	19.5	0.3
	15	4M	0.03	13.8	0.13	0.226	0.774	0.103	1.543	0.071	0.008	0.119	0.010	26.4	0.8	18.8	0.4
Cnf ¹²⁵ -Trp ²⁴	101	0M	0.05	15.0	0.13	0.965	0.035	0.057	1.423	0.055	0.009	0.052	0.011	48.9	4.0	29.6	1.7
	101	2M	0.02	13.3	0.13	0.842	0.158	0.105	2.116	0.145	0.024	0.148	0.028	28.3	2.0	19.3	0.8
	101	4M	0.03	14.1	0.13	0.440	0.560	0.121	1.785	0.102	0.013	0.146	0.017	27.3	0.8	19.4	0.4
Cnf ¹²⁵ -Trp ³⁹	86	0M	0.05	15.0	0.13	0.853	0.147	0.035	1.361	0.047	0.003	0.055	0.004	45.7	1.4	28.2	0.6
	86	2M	0.02	13.3	0.13	0.884	0.116	0.032	1.598	0.078	0.005	0.083	0.006	36.2	0.9	22.7	0.4
	86	4M	0.03	14.1	0.13	0.254	0.746	0.037	1.729	0.095	0.003	0.141	0.004	27.8	0.5	19.6	0.2
Cnf ¹²⁵ -Trp ⁶²	63	0M	0.05	15.0	0.13	0.777	0.223	0.080	1.358	0.047	0.009	0.064	0.011	42.3	2.5	26.7	1.1
	63	2M	0.02	13.3	0.13	0.860	0.140	0.163	1.691	0.090	0.029	0.097	0.034	32.8	3.8	21.2	1.6
	63	4M	0.03	14.1	0.13	0.328	0.672	0.146	1.598	0.078	0.008	0.110	0.011	28.4	0.8	19.9	0.4
Cnf ¹²⁵ -Trp ⁷⁹	46	0M	0.05	15.0	0.13	0.619	0.381	0.009	1.116	0.015	0.001	0.045	0.001	46.0	0.3	28.3	0.1
	46	2M	0.02	13.3	0.13	0.633	0.367	0.015	1.527	0.068	0.002	0.096	0.002	31.4	0.2	20.7	0.1
	46	4M	0.03	14.1	0.13	0.177	0.823	0.015	1.336	0.044	0.001	0.088	0.001	30.5	0.1	20.8	0.0
Cnf ¹²⁵ -Trp ⁹⁴	31	0M	0.05	15.0	0.13	0.782	0.218	0.098	1.523	0.068	0.010	0.082	0.013	39.3	2.3	25.3	1.0
	31	2M	0.02	13.3	0.13	0.737	0.263	0.166	1.728	0.095	0.021	0.113	0.026	30.1	2.4	20.1	1.0
	31	4M	0.03	14.1	0.13	0.258	0.742	0.140	1.729	0.095	0.008	0.128	0.010	26.3	0.7	18.9	0.3

Table 5.3b: Steady-State FRET Fitting and Distance Determination

Pair	PSS	TMAO	Q ₀	R ₀	ϵ_A/ϵ_D	A	E _D	S _A	B	E _A	S _B	E _{FRET}	S _{EFFRET}	Gaussian Chain		Semi-Empirical	
														r (Å)	S _{Dist}	r (Å)	S _{Dist}
Cnf ¹³⁶ -Trp ²⁴	112	0M	0.10	17.0	0.13	1.065	-0.065	0.050	1.243	0.032	0.009	0.032	0.009	66.3	7.3	39.5	3.5
	112	2M	0.04	14.6	0.13	0.947	0.053	0.114	1.362	0.047	0.020	0.048	0.024	48.7	8.5	29.2	3.7
	112	4M	0.03	13.7	0.13	0.607	0.393	0.160	1.663	0.086	0.014	0.111	0.018	27.7	1.6	19.3	0.7
Cnf ¹³⁶ -Trp ³⁹	97	0M	0.10	17.0	0.13	0.909	0.091	0.072	1.571	0.074	0.007	0.076	0.009	47.5	2.0	30.5	0.9
	97	2M	0.04	14.6	0.13	0.761	0.239	0.129	1.391	0.051	0.011	0.065	0.014	40.4	2.8	25.5	1.3
	97	4M	0.03	13.7	0.13	0.458	0.542	0.201	1.730	0.095	0.011	0.118	0.015	26.2	1.1	18.6	0.5
Cnf ¹³⁶ -Trp ⁹⁴	42	0M	0.10	17.0	0.13	0.973	0.027	0.077	1.752	0.098	0.008	0.091	0.010	45.2	2.0	29.5	0.9
	42	2M	0.04	14.6	0.13	0.909	0.091	0.163	1.786	0.102	0.017	0.101	0.022	36.8	3.1	23.9	1.4
	42	4M	0.03	13.7	0.13	0.511	0.489	0.199	1.954	0.124	0.009	0.140	0.012	24.9	1.0	18.1	0.5
Cnf ¹³⁶ -Trp ¹²⁵	11	0M	0.10	17.0	0.13	0.946	0.054	0.085	1.693	0.090	0.014	0.085	0.017	46.6	3.6	30.1	1.7
	11	2M	0.04	14.6	0.13	0.973	0.027	0.089	1.607	0.079	0.018	0.070	0.022	43.5	4.8	26.9	2.1
	11	4M	0.03	13.7	0.13	0.577	0.423	0.147	1.887	0.115	0.013	0.140	0.017	25.2	1.0	18.2	0.4
Fam ⁹ -Raz ⁹⁴	85	0M	0.49	54.3	0.24	0.590	0.410	0.009	3.932	0.717	0.075	0.444	0.012	69.5	1.2	59.8	0.8
	85	2M	0.33	50.9	0.24	0.384	0.616	0.005	4.439	0.841	0.033	0.643	0.006	49.3	0.4	45.6	0.3
	85	4M	0.03	33.9	0.24	0.037	0.963	0.010	4.961	0.968	0.123	0.964	0.013	16.4	1.2	18.5	1.0
Fam ²⁴ -Raz ⁹⁴	70	0M	0.57	55.6	0.24	0.477	0.523	0.023	4.325	0.813	0.219	0.551	0.030	60.5	2.5	54.3	1.7
	70	2M	0.38	52.1	0.24	0.305	0.695	0.005	4.454	0.844	0.043	0.711	0.007	45.8	0.5	43.5	0.3
	70	4M	0.05	36.8	0.24	0.075	0.925	0.007	4.938	0.963	0.138	0.927	0.009	20.9	0.6	22.8	0.5
Fam ⁴² -Raz ⁹⁴	52	0M	0.52	54.9	0.24	0.357	0.643	0.008	3.864	0.700	0.068	0.649	0.010	51.8	0.7	48.2	0.5
	52	2M	0.40	52.6	0.24	0.229	0.771	0.004	4.051	0.746	0.038	0.769	0.005	41.4	0.4	40.7	0.3
	52	4M	0.05	37.1	0.24	0.093	0.907	0.008	6.634	1.378	0.174	0.907	0.011	22.3	0.7	23.9	0.5
Fam ⁶² -Raz ⁹⁴	32	0M	0.54	55.2	0.24	0.301	0.699	0.003	3.955	0.722	0.019	0.702	0.004	48.2	0.2	45.9	0.2
	32	2M	0.45	53.5	0.24	0.238	0.762	0.003	4.116	0.762	0.021	0.762	0.004	42.7	0.3	43.0	0.2
	32	4M	0.16	45.3	0.24	0.027	0.973	0.003	6.541	1.355	0.220	0.973	0.003	20.6	0.5	23.3	0.4
Fam ⁸⁷ -Raz ⁹⁴	7	0M	0.51	54.7	0.24	0.102	0.898	0.001	1.154	0.038	0.010	0.807	0.002	34.8	0.1	36.1	0.1
	7	2M	0.35	51.5	0.24	0.117	0.883	0.002	2.244	0.304	0.015	0.817	0.002	33.4	0.2	34.6	0.2
	7	4M	0.04	36.3	0.24	0.034	0.966	0.009	7.093	1.490	0.176	0.966	0.009	17.3	1.0	19.6	0.8
Fam ¹¹⁴ -Raz ⁹⁴	20	0M	0.59	56.0	0.24	0.023	0.977	0.001	2.486	0.363	0.012	0.920	0.002	25.4	0.3	28.4	0.3
	20	2M	0.38	52.1	0.24	0.022	0.978	0.001	3.635	0.644	0.013	0.942	0.002	23.1	0.4	26.0	0.4
	20	4M	0.08	39.7	0.24	0.043	0.957	0.006	8.040	1.721	0.199	0.957	0.006	20.0	0.6	22.2	0.6
Fam ¹²³ -Raz ⁹⁴	29	0M	0.42	53.0	0.24	0.276	0.724	0.006	2.821	0.445	0.040	0.689	0.007	45.2	0.5	43.4	0.4
	29	2M	0.36	51.5	0.24	0.210	0.790	0.003	3.205	0.539	0.023	0.762	0.004	39.6	0.3	39.2	0.2
	29	4M	0.09	41.2	0.24	0.108	0.892	0.005	8.941	1.941	0.220	0.892	0.005	25.8	0.3	27.2	0.3
Fam ¹³⁶ -Raz ⁹⁴	42	0M	0.56	55.5	0.24	0.164	0.836	0.002	3.393	0.585	0.015	0.811	0.002	39.4	0.2	39.7	0.1
	42	2M	0.42	52.9	0.24	0.116	0.884	0.004	4.536	0.864	0.032	0.882	0.004	33.7	0.4	35.1	0.3
	42	4M	0.19	46.2	0.24	0.116	0.884	0.002	6.291	1.293	0.195	0.884	0.002	29.5	0.2	30.8	0.1
Fam ⁹ -Raz ¹³⁶	127	0M	0.55	55.5	0.24	0.925	0.075	0.006	3.558	0.625	0.048	0.134	0.007	142.7	4.8	108.1	5.3
	127	2M	0.40	52.6	0.24	0.729	0.271	0.013	4.500	0.856	0.090	0.348	0.017	82.7	2.7	67.9	1.7
	127	4M	0.06	37.9	0.24	0.059	0.941	0.005	4.517	0.860	0.036	0.931	0.006	20.4	0.6	22.5	0.5

Fitting Lifetime Data:

Lifetime data were fit using PowerFit10 distributed by PTI. Each decay was fit to a single or double exponential decay where the time regime was selected to minimize the chi-squared values and the residuals. E_{FRET} was determined from $1 - (T_{DA}/T_D)$ where T_{DA} and T_D are the lifetimes for double-labeled and donor-only constructs respectively. For biexponential decays both the amplitude average and the intensity average lifetimes were used to calculate E_{FRET} values for comparison to the E_{FRET} values extracted from steady-state measurement.

Table 5.4a: TCSPC Data and Fitting

Construct	TMAO	T ₁	T ₂	Amplitude ₁	Amplitude ₂	Intensity ₁	Intensity ₂	T _{Amp. Avg.}	T _{Int. Avg.}	χ ²
αS-Fam ⁹	0 M	3.97E+00 ± 5.02E-03	-	1.15E+00 ± 8.16E-03	-	1.00E+00	-	3.97E+00	3.97E+00	9.57E-01
	2 M	3.98E+00 ± 7.48E-03	-	7.66E-01 ± 3.70E-02	-	1.00E+00	-	3.98E+00	3.98E+00	9.40E-01
	4 M	3.03E+00 ± 6.06E-02	1.27E+00 ± 5.36E-02	2.79E-01 ± 2.29E-02	1.10E+00 ± 1.05E-01	2.01E-01	7.99E-01	1.63E+00	1.94E+00	8.64E-01
	0 M	3.99202 ± 5.72E-03	-	7.93E-01 ± 1.34E-02	-	1.00E+00	-	3.99202	3.99202	9.86E-01
αS-Fam ²⁴	2 M	4.02857 ± 7.13E-03	-	7.77E-01 ± 2.35E-02	-	1.00E+00	-	4.02857	4.02857	1.02E+00
	4 M	2.68377 ± 8.74E-02	1.16144 ± 4.39E-02	1.77E-01 ± 1.88E-02	9.12E-01 ± 2.45E-02	3.10E-01	6.90E-01	1.40886	1.63276	9.64E-01
	0 M	4.03269 ± 7.48E-03	-	7.83E-01 ± 2.63E-02	-	1.00E+00	-	4.03269	4.03269	9.49E-01
	2 M	4.02E+00 ± 6.07E-03	-	8.06E-01 ± 2.58E-02	-	1.00E+00	-	4.02E+00	4.02E+00	1.1847
αS-Fam ⁴²	4 M	2.88E+00 ± 5.03E-02	1.12E+00 ± 2.36E-02	1.88E-01 ± 1.01E-02	8.98E-01 ± 5.93E-03	3.50E-01	6.50E-01	1.42E+00	1.73E+00	1.02E+00
	0 M	4.14E+00 ± 6.45E-03	-	9.65E-01 ± 6.35E-02	-	1.00E+00	-	4.14E+00	4.14E+00	9.82E-01
	2 M	4.13E+00 ± 5.64E-03	-	9.77E-01 ± 7.47E-03	-	1.00E+00	-	4.13E+00	4.13E+00	1.10E+00
	4 M	3.54E+00 ± 5.82E-03	1.46E+00 ± 5.23E-03	3.33E-01 ± 1.50E-03	1.02E+00 ± 6.77E-03	4.43E-01	5.57E-01	1.97E+00	2.38E+00	9.18E-01
αS-Fam ⁸⁷	0 M	4.16E+00 ± 8.86E-03	-	7.72E-01 ± 5.40E-02	-	1.00E+00	-	4.16E+00	4.16E+00	1.08E+00
	2 M	4.12E+00 ± 8.53E-03	-	7.21E-01 ± 3.60E-02	-	1.00E+00	-	4.12E+00	4.12E+00	8.96E-01
	4 M	3.07E+00 ± 7.69E-02	1.15E+00 ± 4.94E-02	2.36E-01 ± 1.47E-02	7.52E-01 ± 9.02E-03	4.57E-01	5.43E-01	1.61E+00	2.03E+00	1.30E+00
	0 M	4.04E+00 ± 6.09E-03	-	1.35E+00 ± 1.14E-02	-	1.00E+00	-	4.04E+00	4.04E+00	1.01E+00
αS-Fam ¹¹⁴	2 M	4.07E+00 ± 5.90E-03	-	1.52E+00 ± 3.87E-02	-	1.00E+00	-	4.07E+00	4.07E+00	9.03E-01
	4 M	3.95E+00 ± 5.31E-03	1.92E+00 ± 1.19E-03	1.21E-01 ± 4.47E-04	1.61E+00 ± 2.65E-03	1.34E-01	8.66E-01	2.06E+00	2.19E+00	1.05E+00
	0 M	3.64E+00 ± 4.64E-03	-	1.21E+00 ± 8.97E-03	-	1.00E+00	-	3.64E+00	3.64E+00	9.56E-01
	2 M	3.81E+00 ± 1.52E-03	-	8.70E-01 ± 6.82E-04	-	1.00E+00	-	3.81E+00	3.81E+00	1.02E+00
αS-Fam ¹²³	4 M	3.15E+00 ± 1.26E-01	1.56E+00 ± 9.23E-02	2.39E-01 ± 3.34E-02	7.03E-01 ± 5.11E-02	4.06E-01	5.94E-01	1.96E+00	2.20E+00	9.78E-01
	0 M	3.77E+00 ± 7.07E-03	-	9.01E-01 ± 4.82E-02	-	1.00E+00	-	3.77E+00	3.77E+00	9.87E-01
	2 M	3.82E+00 ± 6.62E-03	-	8.28E-01 ± 3.28E-02	-	1.00E+00	-	3.82E+00	3.82E+00	1.08E+00
	4 M	3.51E+00 ± 4.72E-02	1.47E+00 ± 8.81E-02	4.82E-01 ± 4.91E-02	9.22E-01 ± 1.48E-01	5.55E-01	4.45E-01	2.17E+00	2.60E+00	9.44E-01
αS-Fam ⁹ -Raz ⁴	0 M	3.25E+00 ± 5.53E-03	-	2.62E+00 ± 2.91E-01	-	1.00E+00	-	3.25E+00	3.25E+00	9.41E-01
	2 M	3.67E+00 ± 3.10E-02	2.03E+00 ± 4.24E-02	4.60E-01 ± 1.80E-02	7.19E-01 ± 2.91E-02	5.37E-01	4.63E-01	2.67E+00	2.91E+00	1.10E+00
	4 M	3.88E-01 ± 1.79E-02	-	5.97E+00 ± 4.08E-01	-	1.00E+00	-	3.88E-01	3.88E-01	1.19E+00
	0 M	4.57E+00 ± 1.44E-02	3.01E+00 ± 7.35E-03	1.58E-01 ± 1.55E-03	6.29E-01 ± 3.18E-03	2.75E-01	7.25E-01	3.32E+00	3.44E+00	9.50E-01
αS-Fam ⁹ -Raz ¹³⁶	2 M	3.41E+00 ± 2.47E-03	-	3.40E+00 ± 9.08E-03	-	1.00E+00	-	3.41E+00	3.41E+00	1.04E+00
	4 M	3.87E+00 ± 1.23E-02	5.41E-01 ± 4.71E-02	2.93E-01 ± 5.70E-03	3.06E+00 ± 5.15E-01	4.07E-01	5.93E-01	8.32E-01	1.90E+00	1.11E+00
	0 M	3.94E+00 ± 1.63E-02	2.51E+00 ± 1.82E-02	4.34E-01 ± 7.96E-03	9.94E-01 ± 1.65E-02	4.07E-01	5.93E-01	2.94E+00	3.09E+00	9.46E-01
	2 M	3.47E+00 ± 3.68E-02	1.34E+00 ± 8.49E-02	5.65E-01 ± 5.40E-02	9.89E-01 ± 1.58E-01	5.96E-01	4.04E-01	2.12E+00	2.61E+00	8.80E-01
αS-Fam ²⁴ -Raz ⁴	4 M	3.48E+00 ± 2.52E-01	4.27E-01 ± 3.18E-02	1.89E-02 ± 2.02E-03	3.46E+00 ± 4.48E-01	4.28E-02	9.57E-01	4.43E-01	5.57E-01	1.13E+00
	0 M	3.77E+00 ± 1.48E-02	2.09E+00 ± 1.84E-02	3.89E-01 ± 5.91E-03	7.53E-01 ± 1.31E-02	4.81E-01	5.19E-01	2.66E+00	2.90E+00	1.04E+00
	2 M	3.53E+00 ± 2.01E-02	9.99E-01 ± 4.10E-02	3.49E-01 ± 2.10E-02	9.01E-01 ± 1.07E-01	5.78E-01	4.22E-01	1.70E+00	2.46E+00	1.03E+00
	4 M	3.82E+00 ± 2.90E-01	5.50E-01 ± 1.54E-02	1.51E-02 ± 1.58E-03	2.17E+00 ± 1.39E-01	4.62E-02	9.54E-01	5.73E-01	7.02E-01	9.94E-01

Table 5.4b: TCSPC Data and Fitting

Construct	TMAO	T ₁	T ₂	Amplitude ₁	Amplitude ₂	Intensity ₁	Intensity ₂	I _{Amp. Avg.}	I _{Int. Avg.}	χ ²
αS-Fam ⁶² -Raz ⁹⁴	0 M	4.02E+00 ± 3.27E-02	1.71E+00 ± 6.02E-02	3.90E-01 ± 1.25E-02	6.57E-01 ± 3.40E-02	5.82E-01	4.18E-01	2.57E+00	3.06E+00	8.68E-01
	2 M	3.98E+00 ± 2.25E-02	1.57E+00 ± 8.34E-02	5.09E-01 ± 1.33E-02	4.21E-01 ± 1.95E-02	7.54E-01	2.46E-01	2.89E+00	3.39E+00	9.24E-01
	4 M	3.77E+00 ± 2.18E-02	2.35E-01 ± 3.29E-02	1.12E-01 ± 1.57E-03	3.47E+00 ± 5.95E-01	3.40E-01	6.60E-01	3.46E-01	1.44E+00	9.16E-01
αS-Fam ⁸⁷ -Raz ⁹⁴	0 M	4.04E+00 ± 1.42E-02	7.74E-01 ± 2.87E-02	3.98E-01 ± 1.64E-02	1.40E+00 ± 1.90E-01	5.98E-01	4.02E-01	1.50E+00	2.73E+00	1.09E+00
	2 M	3.97E+00 ± 1.57E-02	4.83E-01 ± 3.72E-02	2.83E-01 ± 6.14E-03	2.51E+00 ± 3.68E-01	4.81E-01	5.19E-01	8.37E-01	2.16E+00	1.01E+00
	4 M	4.86E-01 ± 6.49E-03	-	3.04E+00 ± 5.30E-02	-	1.00E+00	-	4.86E-01	4.86E-01	1.06E+00
αS-Fam ¹¹⁴ -Raz ⁹⁴	0 M	3.25E+00 ± 1.88E-02	7.56E-01 ± 3.34E-03	6.22E-02 ± 7.98E-04	1.99E+00 ± 1.81E-02	1.19E-01	8.81E-01	8.31E-01	1.05E+00	1.01E+00
	2 M	3.14E+00 ± 7.13E-02	6.01E-01 ± 1.70E-02	4.98E-02 ± 2.27E-03	1.57E+00 ± 9.54E-02	1.43E-01	8.57E-01	6.79E-01	9.63E-01	9.38E-01
	4 M	4.95E-01 ± 1.90E-02	-	2.63E+00 ± 2.12E-01	-	1.00E+00	-	4.95E-01	4.95E-01	8.61E-01
αS-Fam ¹²³ -Raz ⁹⁴	0 M	3.86E+00 ± 5.96E-02	1.81E+00 ± 1.04E-01	5.11E-01 ± 3.72E-02	7.49E-01 ± 7.82E-02	5.93E-01	4.07E-01	2.65E+00	3.03E+00	1.00E+00
	2 M	3.87E+00 ± 1.08E-02	1.32E+00 ± 3.04E-02	5.84E-01 ± 4.95E-03	8.56E-01 ± 3.90E-02	6.67E-01	3.33E-01	2.35E+00	3.02E+00	1.09E+00
	4 M	5.25E-01 ± 1.06E-02	-	2.18E+00 ± 1.13E-01	-	1.00E+00	-	5.25E-01	5.25E-01	9.64E-01
αS-Fam ¹³⁶ -Raz ⁹⁴	0 M	2.78E+00 ± 1.00E-02	1.39E+00 ± 1.99E-02	5.25E-01 ± 6.82E-03	7.57E-01 ± 1.93E-02	5.81E-01	4.19E-01	1.96E+00	2.20E+00	1.02E+00
	2 M	2.75E+00 ± 4.57E-02	9.53E-01 ± 4.36E-02	2.63E-01 ± 1.92E-02	1.03E+00 ± 1.04E-01	4.24E-01	5.76E-01	1.32E+00	1.71E+00	1.05E+00
	4 M	3.67E+00 ± 9.63E-01	5.80E-01 ± 2.70E-02	1.52E-02 ± 4.99E-03	5.65E+00 ± 5.38E-01	1.68E-02	9.83E-01	5.89E-01	6.32E-01	8.18E-01
αS-Fam ⁹ +	0 M	3.84E+00 ± 1.12E-02	-	7.97E-01 ± 3.15E-02	-	1.00E+00	-	3.84E+00	3.84E+00	1.06E+00
	2 M	3.89E+00 ± 1.08E-02	-	7.06E-01 ± 1.43E-02	-	1.00E+00	-	3.89E+00	3.89E+00	9.95E-01
	4 M	1.87E+00 ± 2.43E-01	6.12E-01 ± 5.06E-02	1.46E-01 ± 4.68E-02	1.75E+00 ± 1.05E-01	2.02E-01	7.98E-01	7.08E-01	8.66E-01	8.82E-01
αS-Fam ²⁴ +	0 M	3.96E+00 ± 1.12E-02	-	7.40E-01 ± 1.73E-02	-	1.00E+00	-	3.96E+00	3.96E+00	1.03E+00
	2 M	3.98E+00 ± 1.19E-02	-	6.94E-01 ± 1.78E-02	-	1.00E+00	-	3.98E+00	3.98E+00	1.05E+00
	4 M	1.09E+01 ± 3.81E+00	1.09E+00 ± 1.75E-02	2.78E-02 ± 2.99E-03	2.21E+00 ± 1.39E-01	1.12E-01	8.88E-01	1.21E+00	2.20E+00	1.06E+00
αS-Fam ¹² +	0 M	3.99E+00 ± 1.30E-02	-	7.16E-01 ± 2.35E-02	-	1.00E+00	-	3.99E+00	3.99E+00	8.63E-01
	2 M	3.96E+00 ± 8.80E-03	-	6.91E-01 ± 1.12E-02	-	1.00E+00	-	3.96E+00	3.96E+00	1.15E+00
	4 M	3.01E+00 ± 7.68E-03	7.91E-01 ± 9.01E-04	7.68E-02 ± 3.12E-04	1.37E+00 ± 2.59E-03	1.76E-01	8.24E-01	9.08E-01	1.18E+00	9.02E-01
αS-Fam ⁶² +	0 M	4.15E+00 ± 1.32E-02	-	6.91E-01 ± 1.50E-02	-	1.00E+00	-	4.15E+00	4.15E+00	9.06E-01
	2 M	4.09E+00 ± 1.01E-02	-	6.70E-01 ± 9.15E-03	-	1.00E+00	-	4.09E+00	4.09E+00	9.62E-01
	4 M	2.84E+00 ± 4.19E-03	7.47E-01 ± 1.18E-03	1.55E-01 ± 3.86E-04	1.22E+00 ± 3.23E-03	3.26E-01	6.74E-01	9.84E-01	1.43E+00	1.06E+00
αS-Fam ¹¹⁴ +	0 M	4.00E+00 ± 1.26E-02	-	7.22E-01 ± 1.72E-02	-	1.00E+00	-	4.00E+00	4.00E+00	1.11E+00
	2 M	4.03E+00 ± 1.43E-02	-	7.52E-01 ± 2.54E-02	-	1.00E+00	-	4.03E+00	4.03E+00	9.12E-01
	4 M	2.70E+00 ± 2.64E-02	8.19E-01 ± 9.15E-03	1.46E-01 ± 2.92E-03	1.12E+00 ± 1.48E-02	3.01E-01	6.99E-01	1.04E+00	1.39E+00	8.68E-01
αS-Fam ¹²³ +	0 M	3.58E+00 ± 1.07E-02	-	8.99E-01 ± 3.99E-02	-	1.00E+00	-	3.58E+00	3.58E+00	1.02E+00
	2 M	3.78E+00 ± 1.13E-02	-	8.35E-01 ± 3.10E-02	-	1.00E+00	-	3.78E+00	3.78E+00	1.07E+00
	4 M	2.60E+00 ± 1.18E-01	7.56E-01 ± 6.47E-02	2.23E-01 ± 1.88E-02	9.39E-01 ± 2.30E-02	4.50E-01	5.50E-01	1.11E+00	1.59E+00	9.34E-01
αS-Fam ¹³⁶ +	0 M	3.72E+00 ± 9.63E-03	-	8.29E-01 ± 2.93E-02	-	1.00E+00	-	3.72E+00	3.72E+00	9.76E-01
	2 M	3.79E+00 ± 1.18E-02	-	9.81E-01 ± 6.79E-02	-	1.00E+00	-	3.79E+00	3.79E+00	1.05E+00
	4 M	3.79E+00 ± 1.00E-03	7.55E-01 ± 5.69E-04	2.70E-01 ± 1.01E-04	8.85E-01 ± 1.25E-03	6.05E-01	3.95E-01	1.46E+00	2.59E+00	9.35E-01

Table 5.5: Calculated E_{FRET} values from TCSPC

Construct	Intramolecular			Construct	Intermolecular		
	TMAO	$E_{\text{FRET}}^{\text{Amp}}$	$E_{\text{FRET}}^{\text{Int}}$		TMAO	$E_{\text{FRET}}^{\text{Amp}}$	$E_{\text{FRET}}^{\text{Int}}$
$\alpha\text{S-Fam}^9\text{-Raz}^{94}$	0 M	0.18	0.18	$\alpha\text{S-Fam}^9 +$ $\alpha\text{S-Raz}^{94}$	0 M	0.03	0.03
	2 M	0.33	0.27		2 M	0.02	0.02
	4 M	0.45	0.55		4 M	0.57	0.55
$\alpha\text{S-Fam}^9\text{-Raz}^{136}$	0 M	0.16	0.13	$\alpha\text{S-Fam}^{24} +$ $\alpha\text{S-Raz}^{94}$	0 M	0.01	0.01
	2 M	0.14	0.14		2 M	0.01	0.01
	4 M	0.49	0.02		4 M	0.14	-0.34
$\alpha\text{S-Fam}^{24}\text{-Raz}^{94}$	0 M	0.26	0.23	$\alpha\text{S-Fam}_{42} +$ $\alpha\text{S-Raz}^{94}$	0 M	0.01	0.01
	2 M	0.47	0.35		2 M	0.01	0.01
	4 M	0.69	0.66		4 M	0.36	0.32
$\alpha\text{S-Fam}^{42}\text{-Raz}^{94}$	0 M	0.34	0.28	$\alpha\text{S-Fam}_{62} +$ $\alpha\text{S-Raz}^{94}$	0 M	0.00	0.00
	2 M	0.58	0.39		2 M	0.01	0.01
	4 M	0.60	0.60		4 M	0.39	0.29
$\alpha\text{S-Fam}^{62}\text{-Raz}^{94}$	0 M	0.38	0.26	$\alpha\text{S-Fam}_{114} +$ $\alpha\text{S-Raz}^{94}$	0 M	0.01	0.01
	2 M	0.30	0.18		2 M	0.01	0.01
	4 M	0.82	0.40		4 M	0.50	0.37
$\alpha\text{S-Fam}^{87}\text{-Raz}^{94}$	0 M	0.64	0.34	$\alpha\text{S-Fam}_{123} +$ $\alpha\text{S-Raz}^{94}$	0 M	0.02	0.02
	2 M	0.80	0.48		2 M	0.01	0.01
	4 M	0.70	0.76		4 M	0.43	0.28
$\alpha\text{S-Fam}^{114}\text{-Raz}^{94}$	0 M	0.79	0.74	$\alpha\text{S-Fam}_{136} +$ $\alpha\text{S-Raz}^{94}$	0 M	0.01	0.01
	2 M	0.83	0.76		2 M	0.01	0.01
	4 M	0.76	0.77		4 M	0.32	0.00
$\alpha\text{S-Fam}^{123}\text{-Raz}^{94}$	0 M	0.27	0.17				
	2 M	0.38	0.21				
	4 M	0.73	0.76				
$\alpha\text{S-Fam}^{136}\text{-Raz}^{94}$	0 M	0.48	0.42				
	2 M	0.65	0.55				
	4 M	0.73	0.76				

§ 5.4 FCS and AFM Characterization

FCS Measurements:

FCS measurements were done at 20°C on a lab-built instrument based on an Olympus IX71 microscope. A continuous emission 488 nm DPSS 50 mW laser (Spectra-Physics, Santa Clara, CA) was adjusted to 4.5 μW power just prior to entering the microscope. Fluorescence was collected through the objective and

separated from the excitation laser using a Z488rdc long pass dichroic and an HQ600/200m bandpass filter (Chroma, Bellows Falls, VT). Fluorescence was focused onto the aperture of a 50 μm optical fiber coupled to an avalanche photodiode (Perkin Elmer, Waltham, MA). 10 autocorrelation curves of 10 seconds each were taken using a digital correlator (Flex03Q-12, correlator.com, Bridgewater, NJ). Fitting was done using MATLAB (The MathWorks, Natick, MA).

Eight-well chambered coverglasses (Nunc, Rochester, NY) were cleaned by plasma treatment followed by incubation with polylysine-conjugated polyethylene glycol (PEG-PLL). PEG-PLL was prepared from a modified Pierce PEGylation protocol (Pierce, Rockford, IL). After overnight incubation with PEG-PLL, chambers were rinsed with Millipore water and stored until use for measurements. Measurements were done by first replacing the water with solutions of $\sim 400 \mu\text{M}$ wild type α -synuclein in each concentration of TMAO to prevent adsorption of labeled protein. 20 nM of Alexa Fluor 488 or labeled α -synuclein were added into a chamber, and each sample was mixed by pipetting immediately before measurement. The free dye measurements were used to obtain reference diffusion times.

The observed fluorescence fluctuations were autocorrelated and the raw autocorrelation data was fit to the equation:

$$G(\tau) = \frac{1}{N} \left(\frac{1}{1 + \frac{\tau}{\tau_{\alpha S}}} \times \left(\frac{1}{1 + \frac{s^2 \tau}{\tau_{\alpha S}}} \right)^{1/2} \right) \quad (\text{Eq. 5.19, S15})$$

N is the number of molecules in the focal volume, τ is the delay time, $\tau_{\alpha S}$ is the time spent by the sample in the focal volume, and s represents the eccentricity of the confocal volume, and is fixed to 0.2. The diffusion time of the protein, $\tau_{\alpha S}$, is extracted from fit.

AFM Measurements:

Three droplets of αS solution were incubated on the mica substrate for 5 minutes. To remove the excess solution, the samples were blotted by Kimwipe, gently rinsed with 2 mL Milli-Q water, and then dried using a weak nitrogen stream. During the sample preparation, the substrate was tilted 30-45° to prevent the solution from flowing backwards. The samples were imaged in tapping mode using a Keysight 5500 AFM instrument (Keysight Technologies) equipped with a closed-loop scanner. Rotated silicon probes with aluminum reflex coating (BudgetSensors, Tap-300G, resonance frequency ~300 kHz, tip radius <10 nm, force constant 40 N/m) were used to record topographic, amplitude, and phase images with 512 x 512 pixel resolution. The AFM images were analyzed by Gwyddion package. A third-order polynomial was used to flatten the background for topographic images. The volumetric analyses of globular structures were performed using Gwyddion software.

§ 5.5 Computational Modeling Procedures

Semi-Empirical Probability Distribution:

This simulation was run utilizing the same format as the 0 M TMAO unconstrained simulation where the score function at each step was replaced simply with a single repulsive van der Waals term. For the combined set of outputs, all segments which are spaced by the same number of residues in the primary sequence were fit to a normal distribution. From this an empirical relationship was constructed between the distance spanning any two parts of the protein and the average breadth, σ .

PyRosetta Simulations:

All simulations were performed in PyRosetta on the University of Pennsylvania School of Arts and Science General Purpose Cluster. The simulation format followed a basic simulated annealing procedure detailed in the general script attached. In short, the initial protein structure was randomized by sampling random backbone ϕ/ψ angles against a score function consisting solely of repulsive van der Waals energy terms while the protein was represented in the course-grained centroid model. Following randomization, constraints were applied with continued sampling of ϕ/ψ angles in centroid model while increasing the score function complexity from score0-score3 ending with the score3 function as well as long and short range hydrogen bonding terms (hbond_srbb and hbond_lrbb). Lastly, the protein representation was switched into full-atom and backbone torsion angles were sampled along with side-chain χ angles while the 'beta' score function

was applied with the addition of constraints. Acceptance for a set of moves was determined as usual in a Monte Carlo simulation by comparing the sum of the score energy and constraint energy of the previous structure to the newly generated structure. The lowest energy structure generated from each of these steps was retained and used as the starting structure for the next simulation step. The final output structure as determined from the sum of the total 'beta' score function energy and the constraint energy.

Several different methods were employed in an attempt to capture the influence of TMAO on the protein structure in addition to applying constraints. Simulations were performed where the solvation term (fa_{sol}) was removed from the score function since solvation effects from TMAO introduction could not be accounted for directly. Moreover, fragment insertions was employed alongside ϕ/ψ angle sampling in an effort to incorporate the possible formation of significant secondary structure. Fragment libraries were prepared from the primary sequence of wild-type αS using Robetta, a protein prediction server which uses Rosetta software and was developed and supported by the Rosetta Commons and the Baker Lab ²²⁵. Modifications to the base method for each simulation for the 2 M TMAO data set are detailed with the reported results.

In order to produce a simulation which accurately represented the FRET data, each distance constraint was implemented with a "knowledge" weighting factor, γ . This knowledge weighting factor, γ , is inversely proportional to the standard deviation in the calculated distance, based on a given version of the polymer-scaled Förster equation.

$$\gamma(E_{FRET}) = \frac{N}{S_{DA} \times \delta x(E_{FRET}) / \delta E_{FRET}} \quad (\text{Eq. 5.20, S16})$$

$$N = \left(\int_{E_{FRET}} \frac{1}{S_{DA} \times \delta x(E_{FRET}) / \delta E_{FRET}} d(E_{FRET}) \right)^{-1} \quad (\text{Eq. 5.21, S17})$$

Constraints were applied using one of two functions based on the probability distribution functions used to interpret the FRET data. The Gaussian chain (Eq. 5.13) and normal distribution (Eq. 5.14) functions were transformed into Eq. 5.22 and 5.23 respectively to generate potential energy functions termed the Gaussian chain constraint and the harmonic constraint functions respectively.

$$U_1(r) = a \times \gamma \times \log \left(4\pi r \left(\frac{3}{2\pi \langle r^2 \rangle} \right)^{3/2} \exp \left(-\frac{3}{2} \frac{r^2}{\langle r^2 \rangle} \right) \right) \quad (\text{Eq. 5.22, S18})$$

$$U_2(r) = a \times \gamma \times \log \left(\frac{1}{\sqrt{2\pi\sigma^2}} \exp \left(-\frac{1}{2} \frac{(r - \langle r^2 \rangle)^2}{\sigma^2} \right) \right) \quad (\text{Eq. 5.23, S19})$$

The additional term, a , in both equations represents the “thermodynamic” weighting factor of the constraints with respect to the Rosetta score function. The optimal thermodynamic weighting factor was determined empirically by running trial simulations with a set to values from 0.25-5. The breadth, σ for the harmonic restraints (Eq. 5.23) was determined from a relationship between the average distance and the normal distribution breadth derived from the simulated structures used to generate the semi-empirical constraints. The optimal weight (a) was determined to be 1, since at that value constraints were satisfied but the resulting

ensembles were not overconstrained so as to yield physically unreasonable conformations. For each interpretation of the data, both constraint functions were applied to determine the efficacy of the constraining method and the data interpretation. Simulations assessing the constraint weight and the method of constraining generated 120 outputs, where the lowest 100 structures were used in analysis due to the lack of differences in energy in the outputs. Final simulations of the Gaussian chain interpreted and Gaussian chain constrained data produced 1020 structures where the lowest 1000 were used for interpretation.

Simulation Output Analysis:

HydroPro was used to calculate the diffusion coefficients and radii of gyration.²²⁶ For analysis of 2 M simulated structures in HydroPro, the input viscosity was multiplied by the ratio of the 0 M:2 M diffusion times of AlexaFluor488. This effectively accounted for the increase in viscosity in 2 M versus 0 M TMAO. Following determination in HydroPro, the diffusion coefficient was converted for comparison with the diffusion coefficient determined from FCS using a previously published conversion equation detailed in Eq. 5.24, while the diffusion coefficient was determined from the measured diffusion time via Eq. 5.25²²⁷.

$$D_{Calibrated} = \frac{D_{HydroPro}^{+0.582}}{1.08} \quad (\text{Eq. 5.24})$$

$$D_P = D_{AF488} \frac{\tau_{AF488}}{\tau_P} \quad (\text{Eq. 5.25})$$

In these equations, D_P , D_{AF488} , $D_{HydroPro}$ and $D_{Calibrated}$ are the diffusion coefficients of the protein of interest, AlexaFluor488, as determined from literature, from HydroPro and the experimentally calibrated diffusion coefficient for comparison to D_P respectively. Additionally, T_{AF488} and T_P are the experimentally derived diffusion times of AlexaFluor488 and the protein of interest respectively.

All distances required for Flory scaling plots, heat maps and comparisons to experimental FRET data were extracted from C-alpha to C-alpha distances for each residue using Python or BioPython and all calculations were performed using NumPy. For E_{FRET} determination, distance values from each member of a given ensemble were transformed into E_{FRET} values using the classical Förster equation prior to averaging to capture the explicit distance probability distribution of the ensemble. The same method was used to calculate comparisons to paramagnetic relaxation enhancement data, where distances were extracted from C-alpha to amide proton distance. Distances converted into intensity ratios as described by Piana *et. al.* and the intensity ratios were averaged to calculate the final values⁹⁷.

Impact of the Thermodynamic Constraint Weight:

Assessment of the thermodynamic constraint weight effects were assessed prior to application of the knowledge constraint weight. This was done in order to assess the maximum impact of the constraint set. Knowledge constraints applied to the full simulation were scaled to a maximum value determined by the thermodynamic constraint as shown in Eq. 5.22 and Eq. 5.23, which is why

determination of the maximum impact of the constraint set was crucial. Experimentally constrained simulations were run as described below with the omission of side chain rotamer packing using constraints where the distances were determined by using the Gaussian chain version of the polymer scaled Förster equation and were applied using the Gaussian chain derived constraining function.

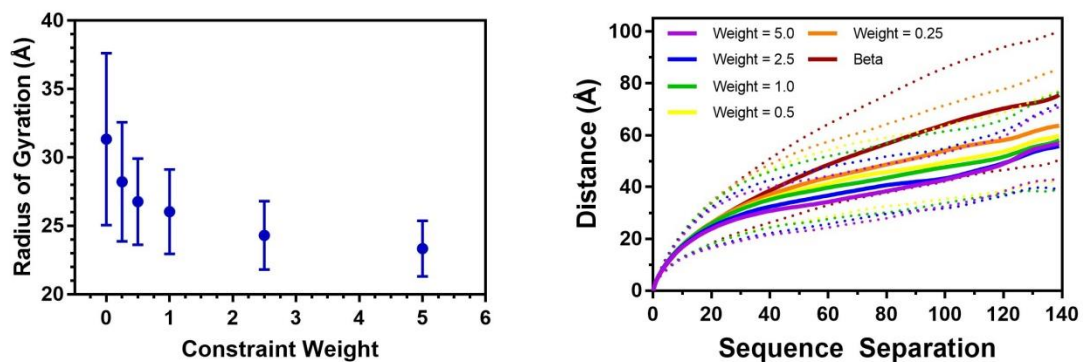


Figure 5.2: Determination of Optimal Thermodynamic Constraint Value. Radius of gyration (left) and average Flory scaling behavior (right) of simulated structures with varying values of the thermodynamic constraint weight, α .

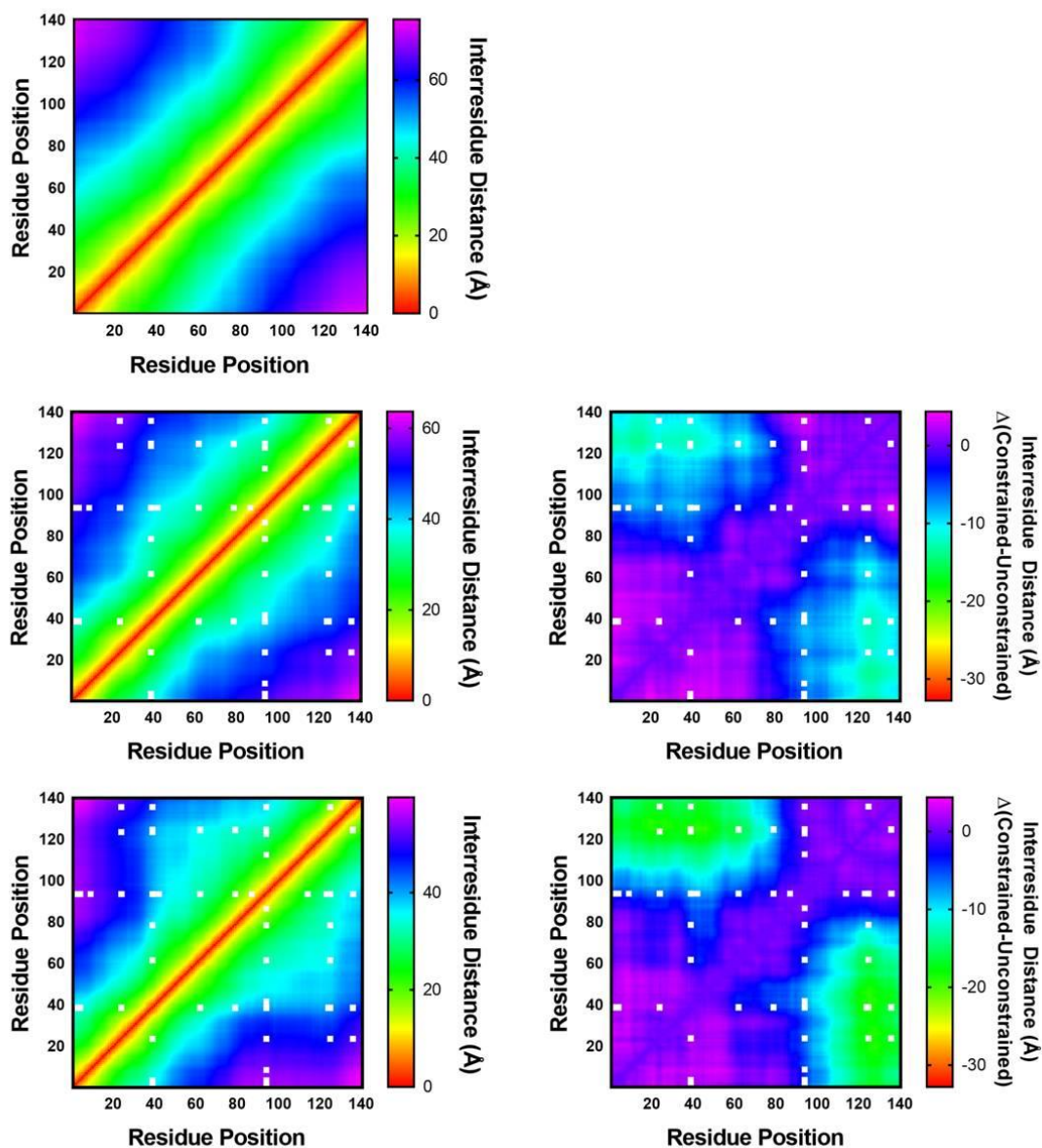


Figure 5.3a: Effect of Constraints on Interresidue Distance Heat Maps. Heat maps depicting the average interresidue distance for each thermodynamic constraint weight, α , (left) and average difference in interresidue distance between constrained and unconstrained simulations (right). Heat maps depict averages from an unconstrained simulation (top row) and simulations with thermodynamic constraint weights set to 0.25 (middle row) and 0.5 (bottom row).

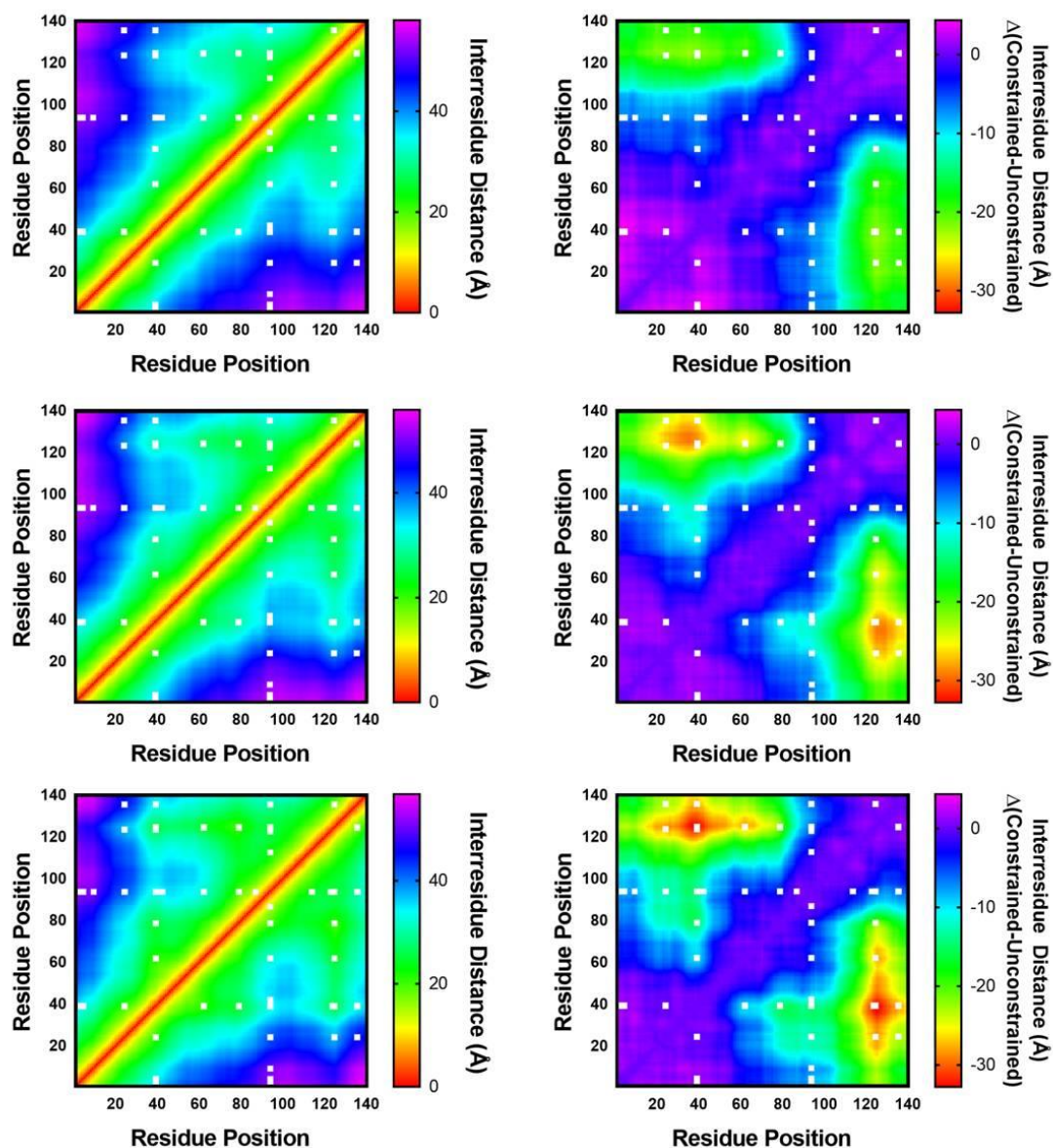


Figure 5.3b: Effect of Constraints on Interresidue Distance Heat Maps. Heat maps depicting the average interresidue distance for each thermodynamic constraint weight, α , (left) and average difference in interresidue distance between constrained and unconstrained simulations (right). Heat maps depict averages from simulations with thermodynamic constraint weights set to 1.0 (top row), 2.5 (middle row) and 5.0 (bottom row).

Assessment of the Constraint Functional Form and Sampling Method:

Simulations were performed utilizing distances from the Gaussian chain (Eq. 5.13) and semi-empirical (Eq. 5.14) FRET distribution analyses in conjunction with either the Gaussian chain (Eq. 5.22) or harmonic (Eq. 5.23) constraints. This was done to assure that no bias was introduced by selecting a particular constraining method. No significant differences in the resulting ensembles were observed between simulations run with different constraint data or constraining functions when comparing the results in Flory-scaling plots (Fig. 5.4), or plots of computed E_{FRET} values (Fig. 5.5-5.6), diffusion coefficients (Fig. 5.7), or “heat maps” of global structural analyses (Fig. 5.14a-5.14b).

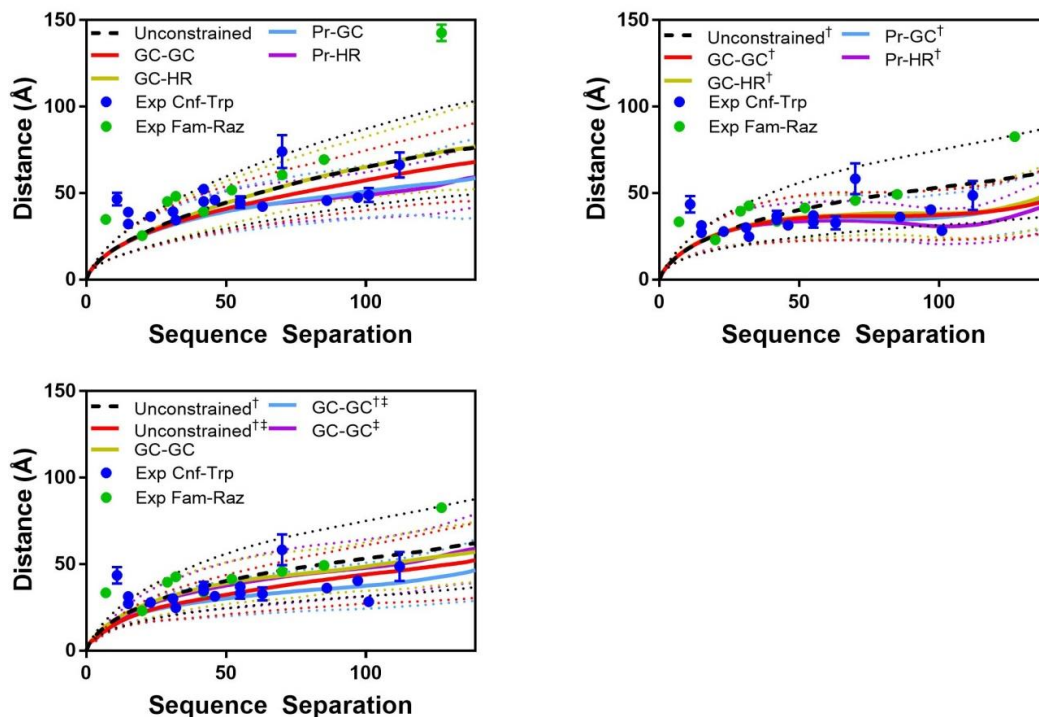


Figure 5.4: Flory Scaling Plots. Plots of Flory scaling determined from simulated ensembles constrained with distance constraints from measurements in 0 M (top

left) and 2 M (top right/bottom left) TMAO. Names in the legends represent the probability distribution used in the polymer-scaled Förster equation for obtaining distances (Gaussian-chain = GC or semi-empirical = Pr) followed by the constraining function employed (Gaussian-chain = GC or harmonic potential = HR). Simulations marked with † indicate that the weight of the solvation term (fa_{sol}) was set to 0 while simulations marked with ‡ indicated that fragment insertion was incorporated within the search.

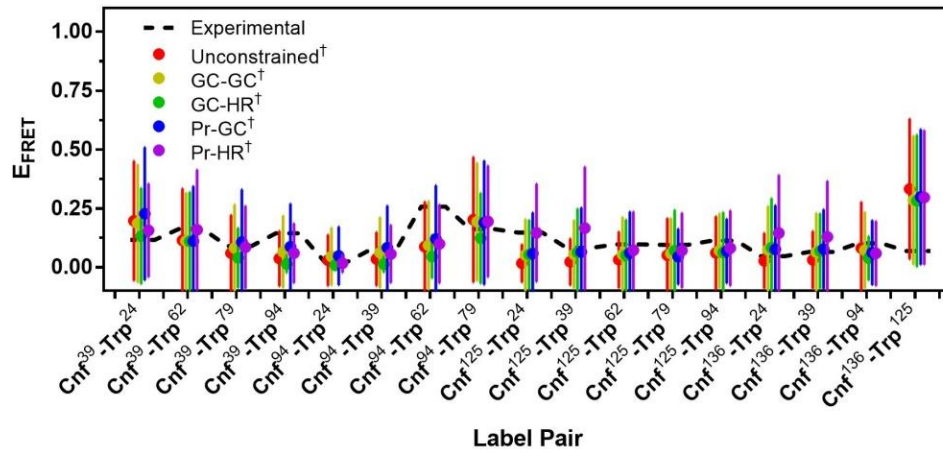
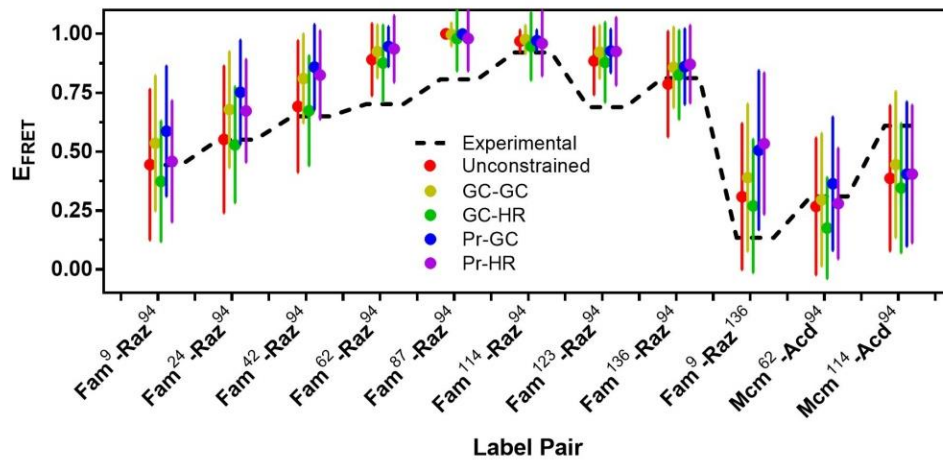
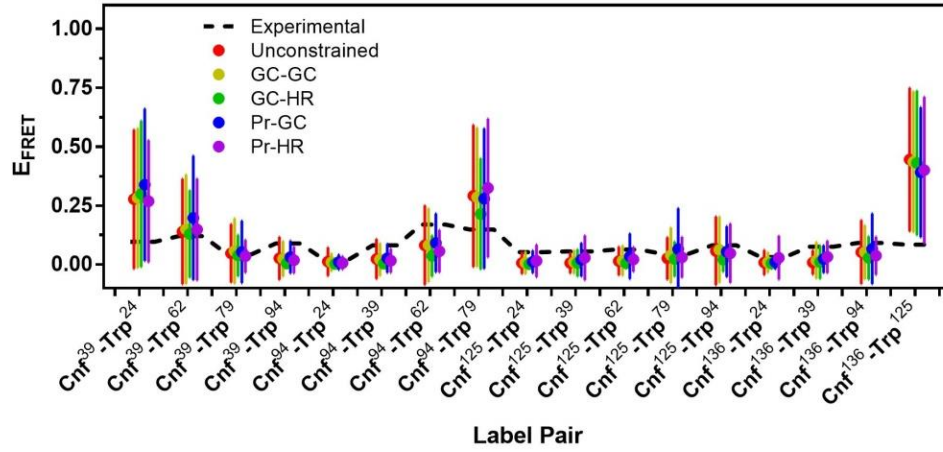
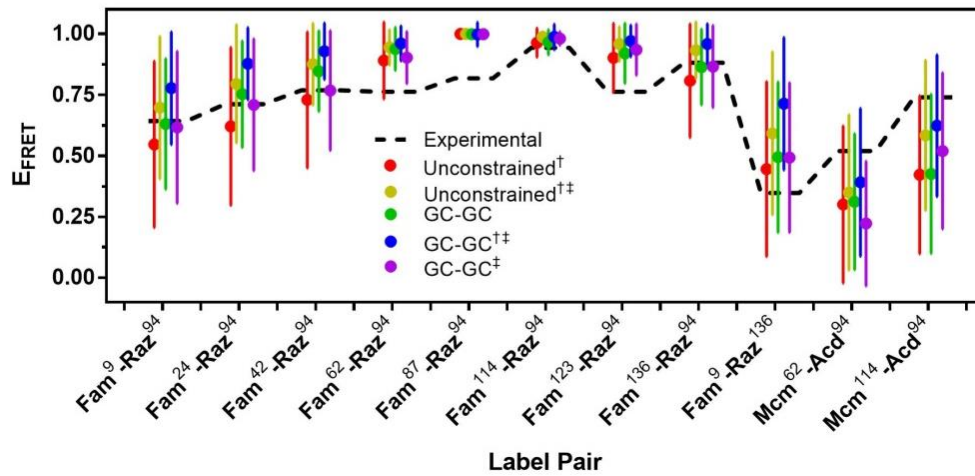
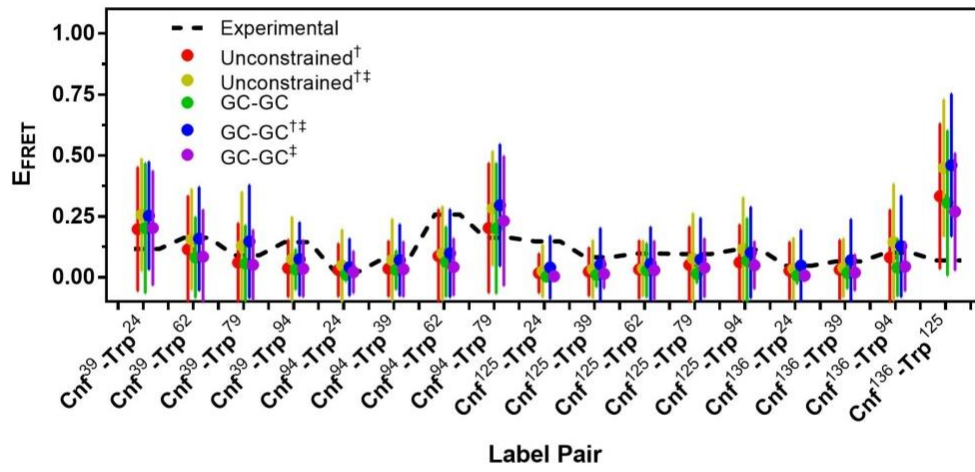
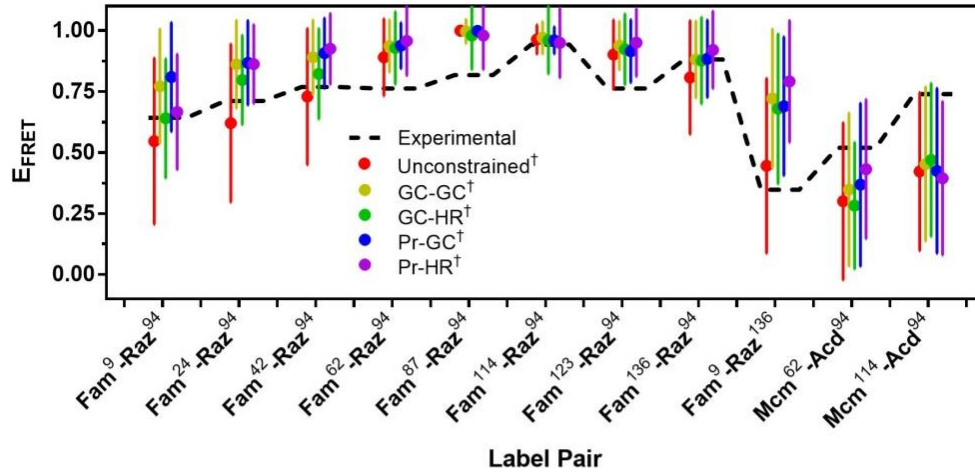


Figure 5.5: 0 M TMAO EFRET Plots. (Previous Page) Plots show the average and standard deviation of the EFRET values determined from simulated ensembles constrained with distance constraints from measurements in 0 M (top/middle) and 2 M (bottom) TMAO. Names in the legends represent the probability distribution used in the polymer-scaled Förster equation for obtaining distances (Gaussian-chain = GC or semi-empirical = Pr) followed by the constraining function employed (Gaussian-chain = GC or harmonic potential = HR). Simulations marked with † indicate that the weight of the solvation term (f_{a_sol}) was set to 0 while simulations marked with ‡ indicated that fragment insertion was incorporated within the search.

Figure 5.6: 2 M TMAO EFRET Plots. (Next Page) Plots show the average and standard deviation of the EFRET values determined from simulated ensembles constrained with distance constraints from measurements in 2 M TMAO. Names in the legends represent the probability distribution used in the polymer-scaled Förster equation for obtaining distances (Gaussian-chain = GC or semi-empirical = Pr) followed by the constraining function employed (Gaussian-chain = GC or harmonic potential = HR). Simulations marked with † indicate that the weight of the solvation term (f_{a_sol}) was set to 0 while simulations marked with ‡ indicated that fragment insertion was incorporated within the search.



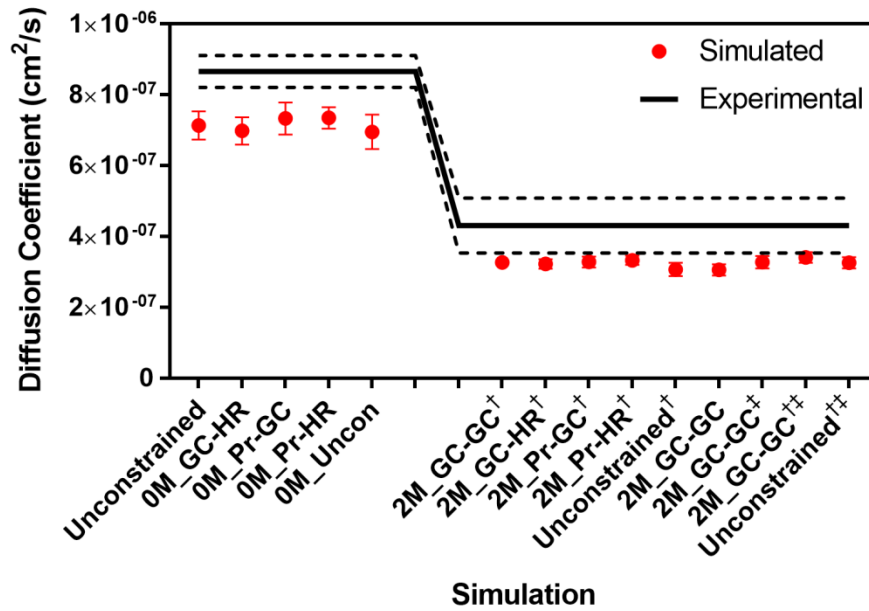


Figure 5.7: Diffusion Coefficients. Plots show the average diffusion coefficient determined from FCS data (black line) and from various simulations (red points). Simulation names represent the concentration of TMAO in which measurements were taken, the probability distribution used in the polymer-scaled Förster equation for obtaining distances (Gaussian-chain = GC or semi-empirical = Pr) and the constraining function employed (Gaussian-chain = GC or harmonic potential = HR), respectively. Simulations marked with † indicate that the weight of the solvation term (fa_{sol}) was set to 0 while simulations marked with ‡ indicated that fragment insertion was incorporated within the search.

§ 5.6 Results

We began by generating two libraries of proteins, labelled with either Cnf-Trp or Fam-Raz pairs, and making FRET measurements in varying concentrations of TMAO. For both FRET pairs, we observed changes in photophysical parameters impacting the extraction of distance information from FRET data that necessitated performing control fluorescence measurements using singly-labelled proteins. When calculating R_0 (Eq. 5.9) for these experiments, there were four parameters of interest, Φ_D , J , κ^2 and n , which represent the quantum yield of the

FRET donor, the overlap integral between donor emission and acceptor absorbance, the orientation factor between the two fluorophores, and the refractive index of the solution, respectively. Of these parameters, we determined that changes in Φ_D as a function of environment contributed most significantly to changes in R_0 for both Cnf and Fam. Additionally, Zheng *et al.* previously demonstrated that for disordered proteins the orientations of the fluorophores are sufficiently isotropic to warrant the approximation of $\kappa^2 = 2/3$, which is the value for an isotropic distribution of orientations⁶⁴. Fluorescence measurements made on donor-only and acceptor-only proteins enabled us to not only account for changes in Φ_D to enhance our accuracy in calculating R_0 , but also allowed us to easily overcome the difficulties of interpreting highly overlapped spectra and extract distances from E_{FRET} measurements for use in modeling.

Cnf-Trp library

The Cnf-Trp construct library consisted of a total of 27 proteins, 17 double-labeled and 10 single-labeled α S mutants, spanning 16 unique intramolecular distances. Acquiring concentration-matched emission spectra for each single-labeled construct in varying TMAO conditions allowed for tracking of changes in the quantum yield and spectral shape of Cnf and Trp emission. The sensitivity of these photophysical properties to increasing concentrations of TMAO was initially assessed for the free amino acids (Fig. 5.8).

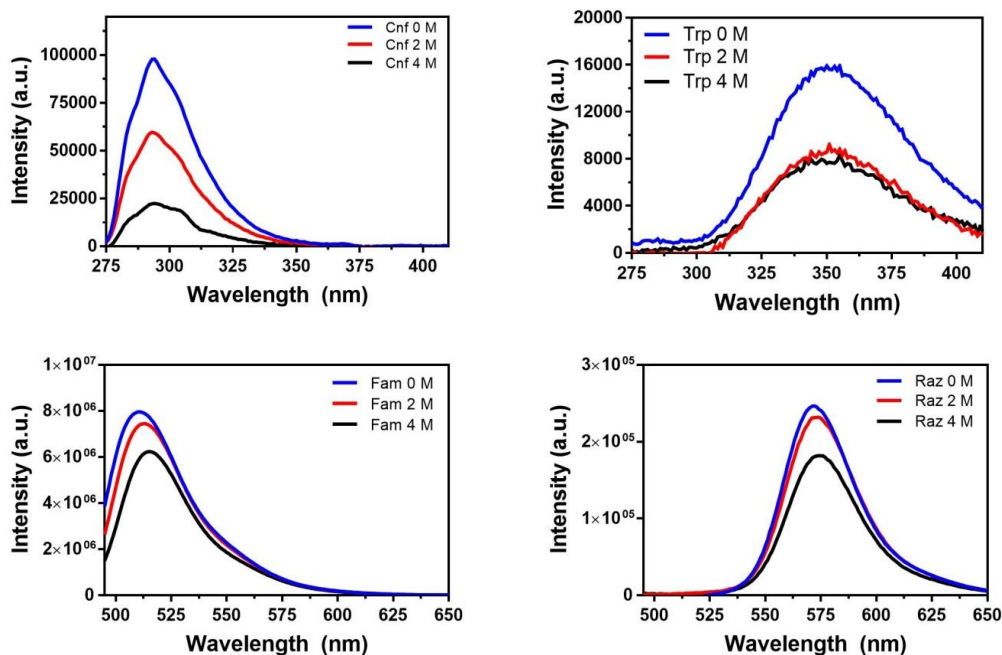


Figure 5.8: Steady-state Fluorescence Spectra of Free Fluorophores. Plots show background subtracted spectra of Cnf (top left), Trp (top right), Fam (bottom left) and Raz (bottom right) for 0 (blue), 2 (red) and 4 M TMAO (black).

Although a decrease in the quantum yield was observed for both Cnf and Trp, no major change in emission maximum was observed in the Trp spectrum. In contrast to the free amino acid measurements, Trp-containing α S mutants successively blue-shifted in increasing concentrations of TMAO, as a result of changes in local environment upon compaction. Thus, measuring single-labeled spectra was not only crucial for tracking spectral changes due to environmental effects, but was also essential for deconvoluting the highly overlapped Cnf and Trp emission spectra in constructs containing both fluorophores. E_{FRET} values were obtained for each probe pair in buffer containing 0, 2, and 4 M TMAO. Figure 5.9 shows the deconvolution of a 0 M measurement as well as a 2 M measurement for the Cnf₁₂₅-Trp₉₄ pair. As expected, E_{FRET} increases, correlating with the

expected decrease in inter-residue distance resulting from compaction in TMAO. Without TMAO present, in most cases this probe pair is not able to accurately capture the apparent intraresidue distance as seen in low E_{FRET} values outside of the most reliable FRET range ($E_{FRET} = 0.3-0.7$). This is most clearly observed in Figure 5.10, where the distances extracted from these data are largely invariant above sequence separations of 50 residues. However, in cases where the probes are relatively close in primary sequence, the utility of this short-range pair is clearly demonstrated. In the presence of 2 M, and especially 4 M TMAO, a significant number of the probed regions display E_{FRET} values within the optimal range for accurate distance determination.

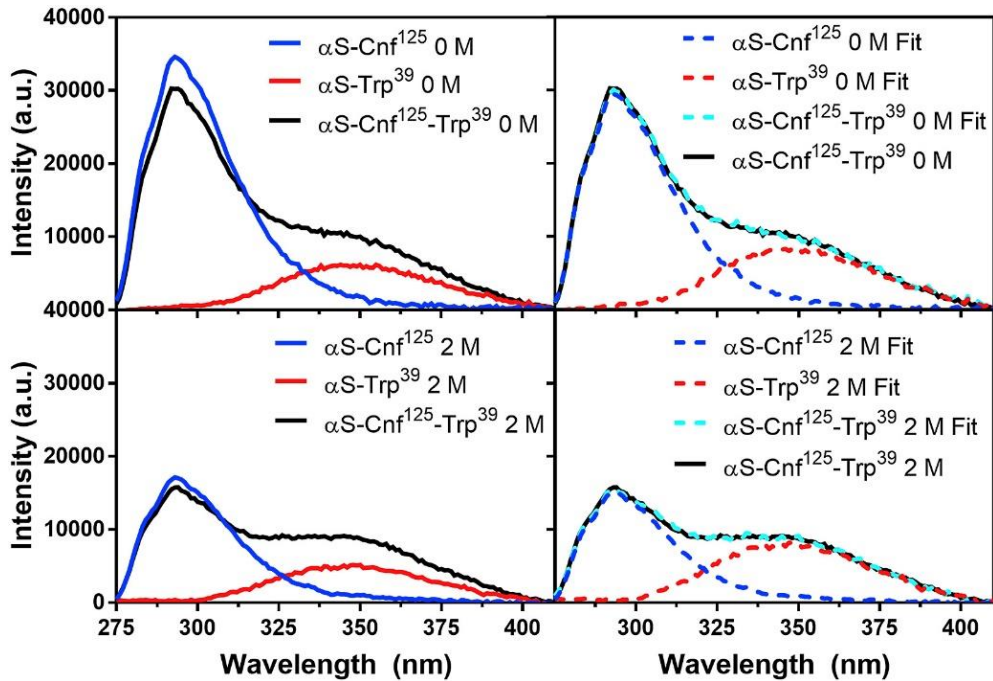


Figure 5.9: Determination of EFRET. Left: Background subtracted fluorescence emission spectra of constructs labeled with Cnf, Trp, or both Cnf and Trp in 0 M (top) and 2 M TMAO (bottom). Right: Double-labeled Cnf-Trp spectrum compared to weighted sum of Cnf-only and Trp-only spectra, along with the contributions from

each single-labeled spectrum shown for 0 M (top) and 2 M TMAO (bottom) spectra.

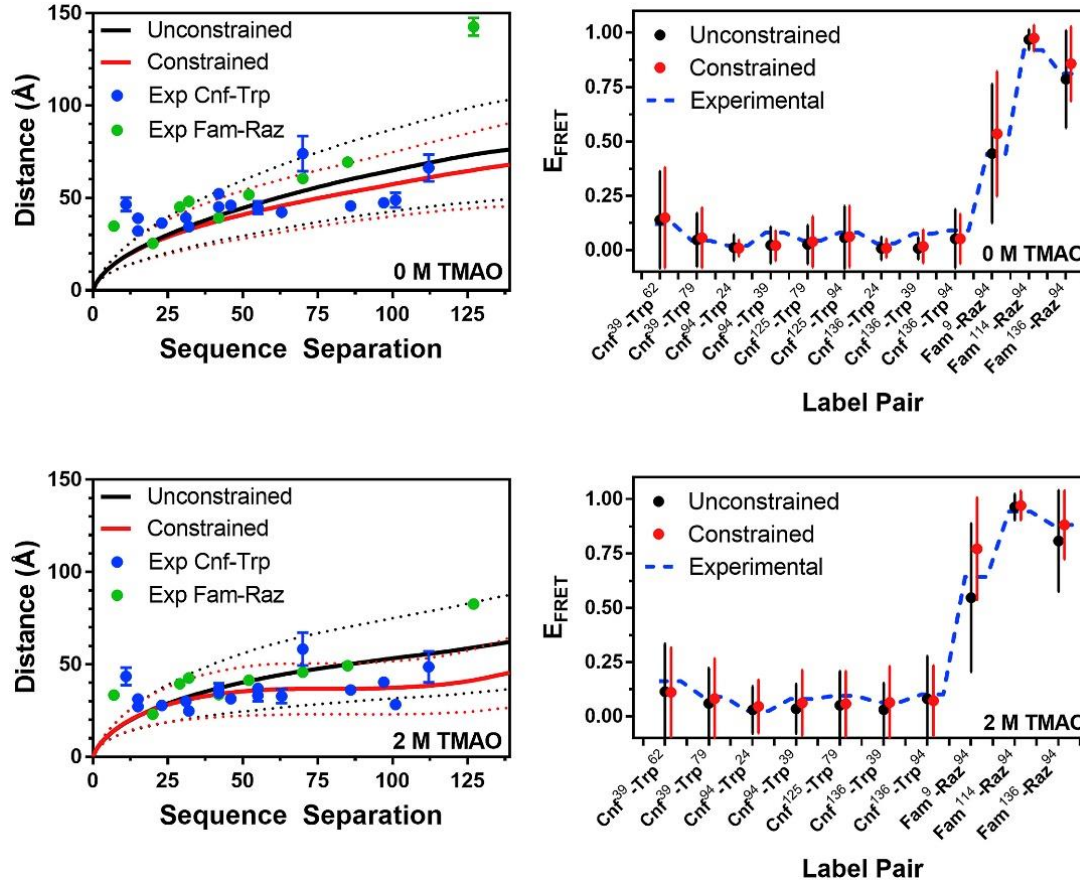


Figure 5.10: Comparison of experimental and simulated data. Left: Distances extracted from EFRET measurements of the Cnf-Trp and Fam-Raz libraries shown with interresidue distances for a given primary sequence separation in unconstrained or constrained MC simulations. The average (solid line) and standard deviation (dashed line) of interresidue distances are shown for the simulated ensembles. Right: Experimental and simulated EFRET values. Dashes represent average EFRET values for each pair of labeled residues, with interconnecting lines to guide the eye. The average (point) and standard deviation (line) of EFRET values obtained based on interresidue separations in the unconstrained or constrained simulated ensembles.

Fam-Raz library

To complement our Cnf-Trp data, we also performed FRET measurements with a probe pair with a longer working range. A set of 21 constructs, consisting of

10 dual-labeled and 11 single-labeled analogs containing Fam and/or Raz, was produced to accurately monitor long range distance changes during the compaction of α S by TMAO. In cases where exceedingly low E_{FRET} was observed with the short-range probe pair, we observed efficiencies much closer to the optimal efficiency range for the Fam-Raz pair. This was most significant in the absence of TMAO. However, our measurements approached the short end of the working range for the Fam-Raz pair in some instances in 2 M TMAO and exhausted its utility in 4 M TMAO.

In 4 M TMAO, the long working range of this probe pair was effective for assessing the formation of oligomers, which were reported by Uversky *et al.* When mixing Fam and Raz single-labeled species in 4 M TMAO, there was a significant reduction in the lifetime of the Fam labeled construct, consistent with intermolecular FRET (Figs. 5.11a-5.11c).

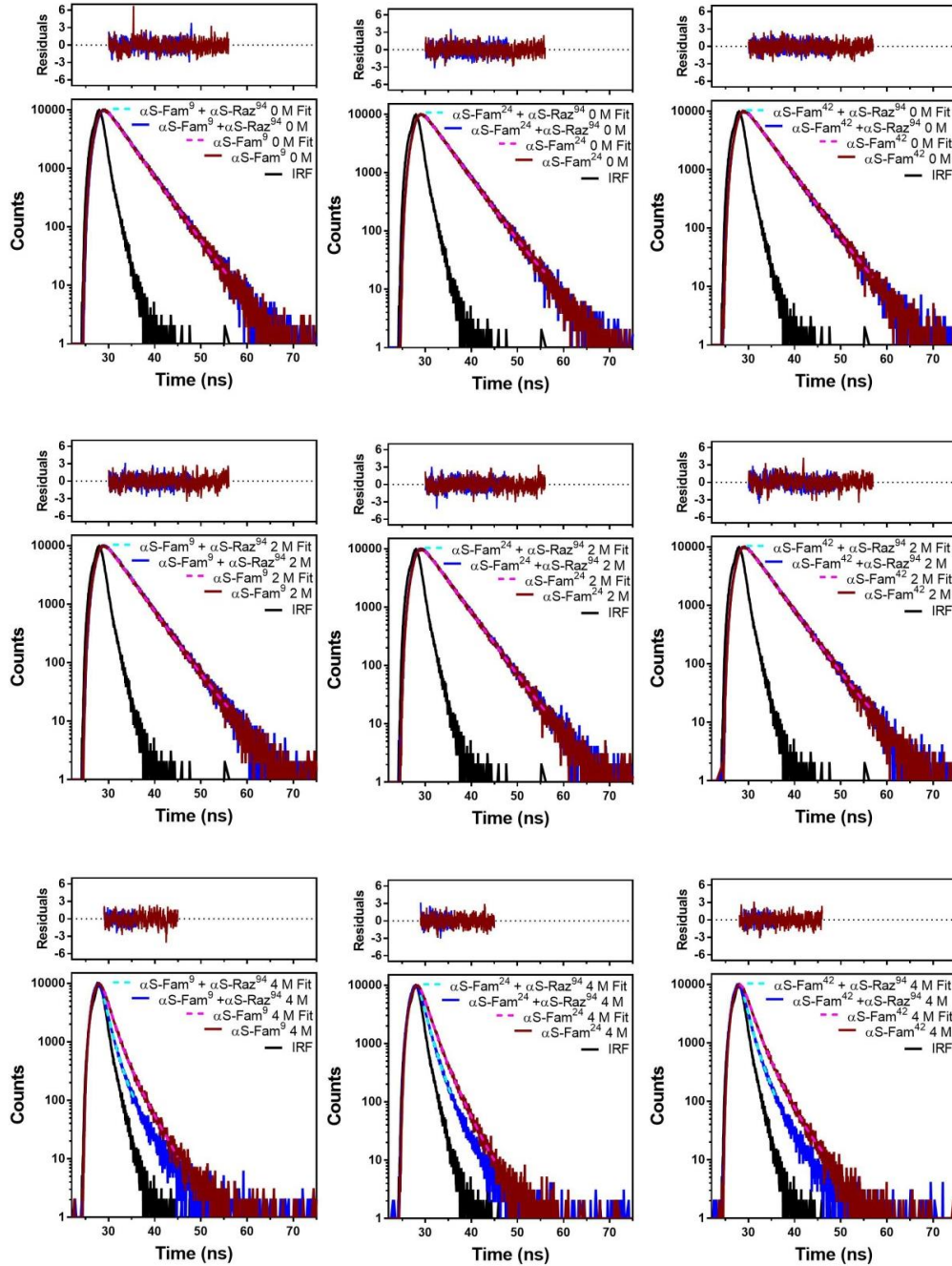


Figure 5.11a: Intermolecular FRET Lifetime Decays and Fits. Plots show lifetime decays and fits of (Left) α S-Fam9 + α S-Raz94, (Middle) α S-Fam9 + α S-Raz136 and (Right) α S-Fam24 + α S-Raz94 in (Top-Bottom) 0, 2 and 4 M TMAO. Decay of single-labeled Fam⁹ constuct decays are shown before (red) and after (blue) the addition of the Raz containing construct with fits shown as dashed lines.

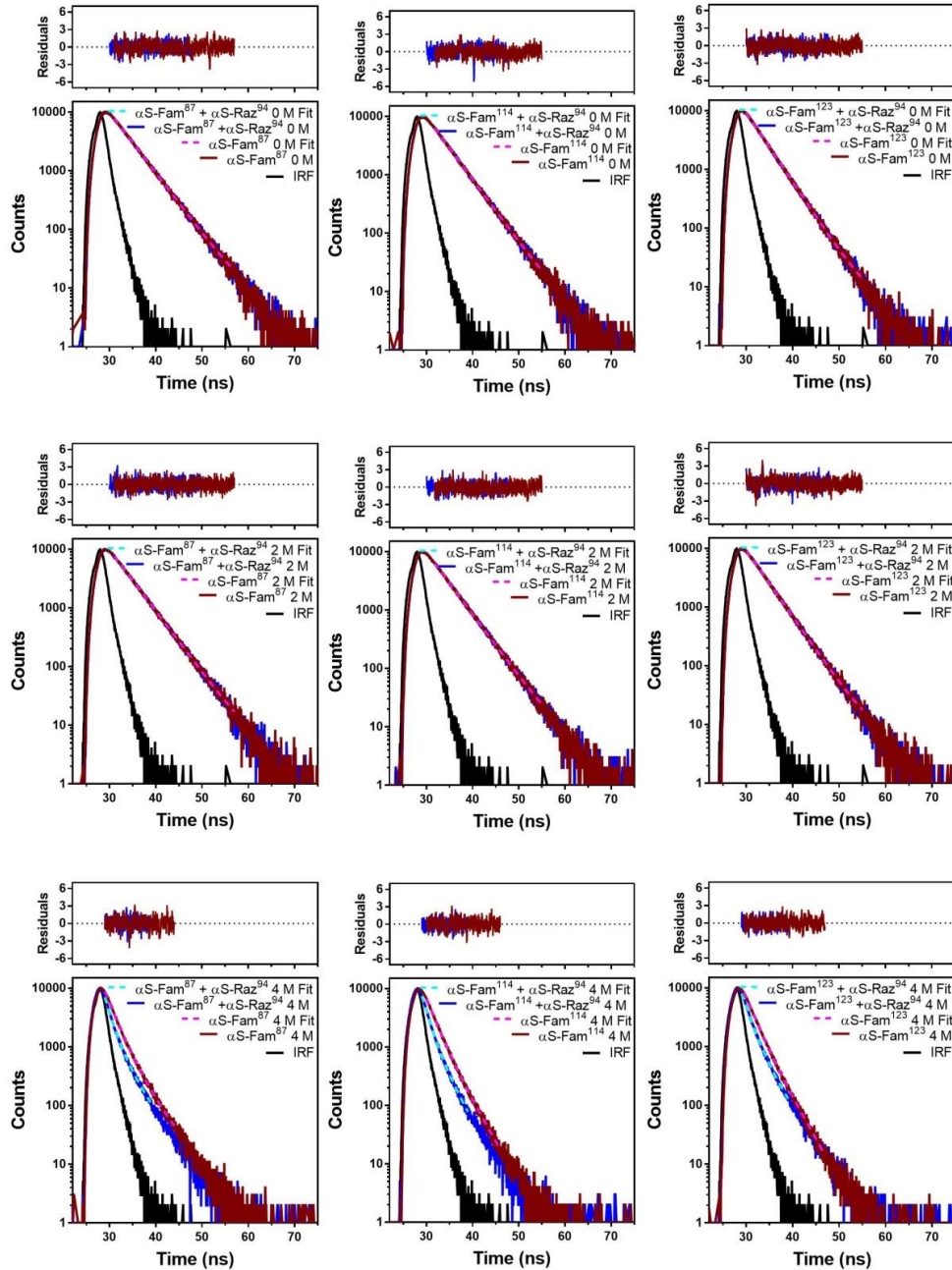


Figure 5.11b: Intermolecular FRET Lifetime Decays and Fits. Plots show lifetime decays and fits of (Left) α S-Fam87 + α S-Raz94, (Middle) α S-Fam114 + α S-Raz136 and (Right) α S-Fam123 + α S-Raz94 in (Top-Bottom) 0, 2 and 4 M TMAO. Decay of single-labeled Fam construct decays are shown before (red) and after (blue) the addition of the Raz containing construct with fits shown as dashed lines.

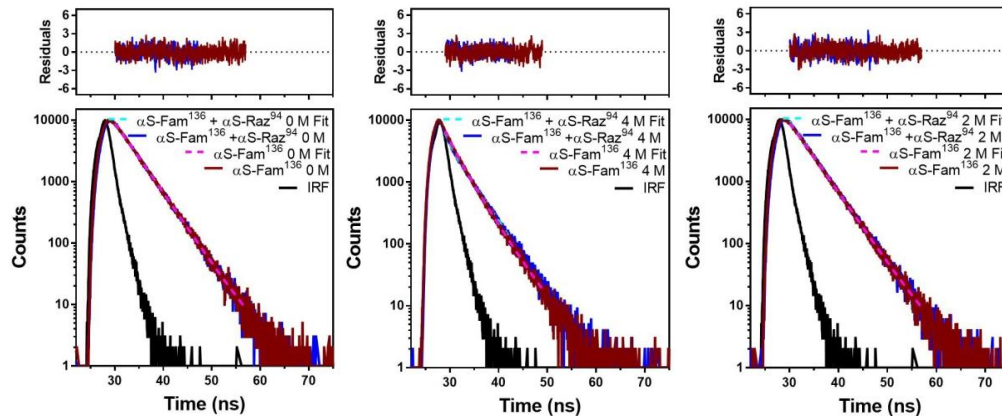


Figure 5.11c: Intermolecular FRET Lifetime Decays and Fits. Plots show lifetime decays and fits of α S-Fam136 + α S-Raz94 in (Left-Right) 0, 2 and 4 M TMAO. Decay of single-labeled Fam construct decays are shown before (red) and after (blue) the addition of the Raz containing construct with fits shown as dashed lines.

In buffer or 2 M TMAO, there was no observed change in the lifetime of the Fam labeled protein due to the presence of the protein containing Raz. FCS and AFM measurements also demonstrate the formation of oligomers in 4 M TMAO (Figs. 5.12–5.13). As a result of these observations, we decided to forego modeling the 4 M structure due to the complexity of deconvoluting intra- and intermolecular FRET in oligomers.

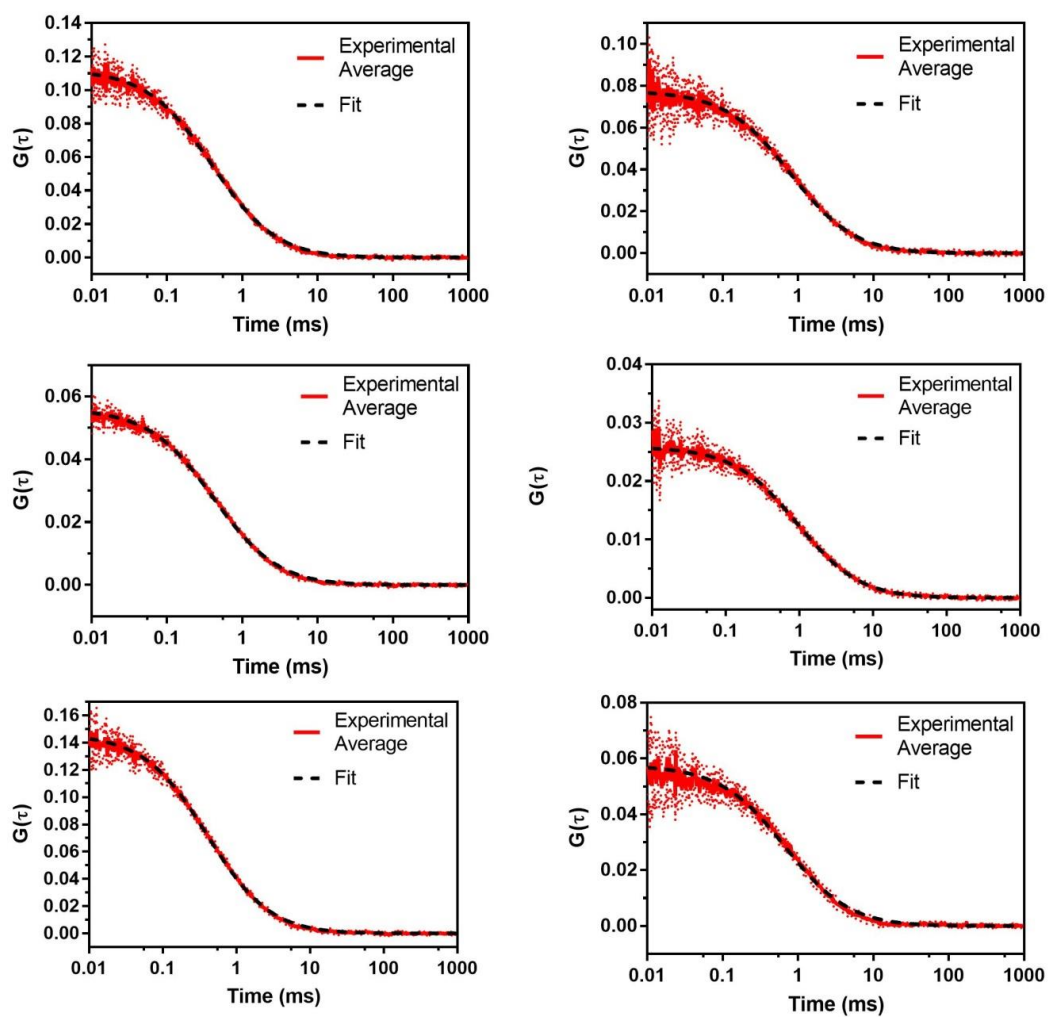


Figure 5.12: Autocorrelation plots from FCS Data. Plots show the average autocorrelation decays from 10 measurements (red) with fits (black-dashed) for α S-AF4889 (Top), α S-AF488114 (Middle) and α S-AF488130 (Left) in 0 M (Left) and 2 M (Right) TMAO.

Table 5.6: FCS Data

Position (TMAO)	9 (0 M)	114 (0 M)	130 (0 M)
N	8.9006	17.809	6.8298
Delay Time	0.401	0.42034	0.40662
Intensity	59007	193000	51376
αS	0.43 ± 0.027	0.418 ± 0.010	0.411 ± 0.010
Position (TMAO)	9 (2 M)	114 (2 M)	130 (2 M)
N	12.889	38.693	17.384
Delay Time	0.79827	0.94794	0.68422
Intensity	44824	163510	54824
αS	0.822 ± 0.007	0.946 ± 0.035	0.779 ± 0.037

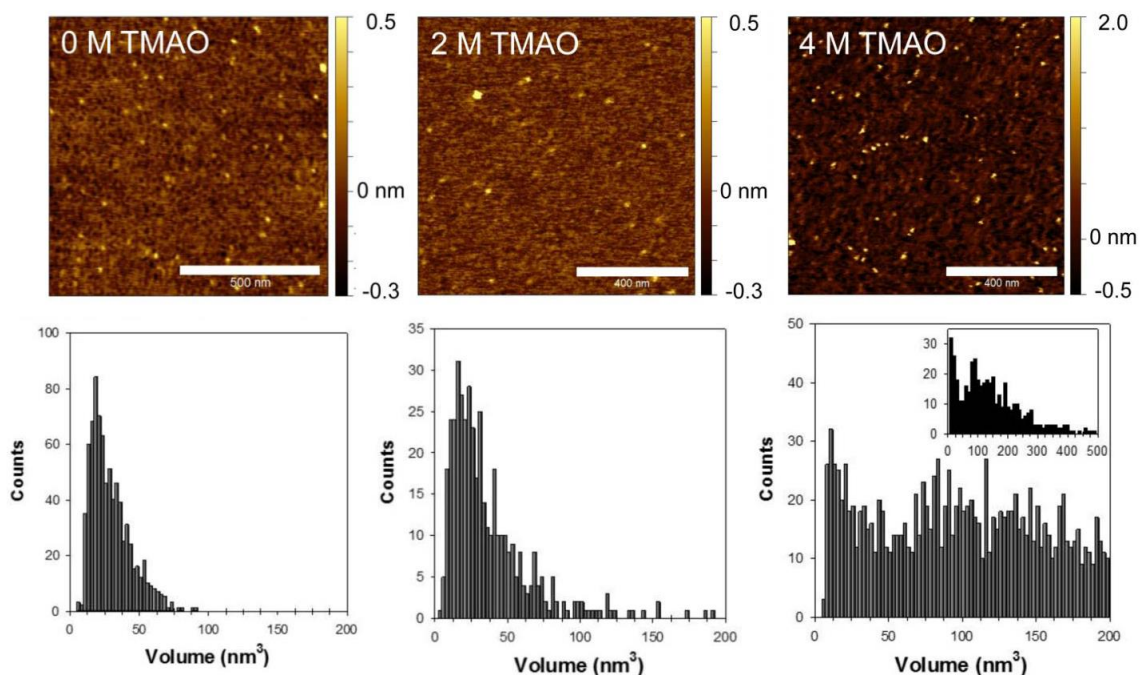


Figure 5.13: The effect of TMAO on the aggregation of wild type α S. Top: AFM images of the dried samples. Bottom: corresponding volume histograms of the α S aggregates. The α S concentration used for these samples are (left) 10 μ M, (middle) 0.1 μ M, and (right) 0.5 μ M. Each volume histogram was obtained with accumulated data from more than five AFM images screened at various area on the sample. No α S fibrillization is observed, which is consistent with the fact that α S fibrillization in solution can't occur in low concentrations and under static conditions, and no agitation or shaking of the solution were performed here.

Interpretation of FRET data

Because α S is disordered, the interresidue distance separating any given label pair is widely distributed. To accurately obtain an average distance value, this distribution needs to be taken into account. Therefore, the corresponding interresidue distances were calculated using a polymer-scaled Förster equation, Eq. 5.12, taking into account changes in R_0 resulting from changes in quantum yield for each labeling position. Distances were calculated using both the Gaussian

chain (Eq. 5.13) and semiempirical (Eq. 5.14) models, which produced significantly different sets of results. Conceptually, two positions labeled with two FRET pairs with different R_0 values should have distinct FRET efficiencies, but comparable extracted distances. Since the Cnf-Trp and Fam-Raz libraries contained a limited number of labeled positions in common, for comparison the distances were plotted in Fig. 5.11 as a function of probe pair primary sequence separation.

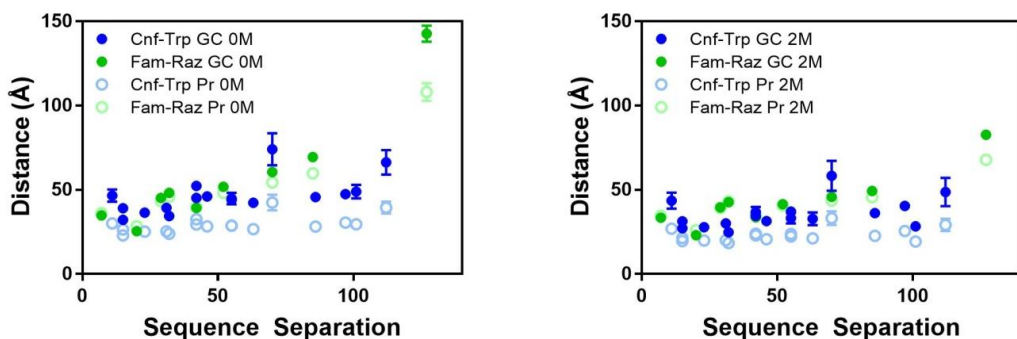


Figure 5.14: Distance from Polymer-Scaled Förster Equation. Plot show distances obtained from steady-state derived EFRET values through the Gaussian chain (GC) or semi-empirical (Pr) forms of the polymer-scaled Förster equation. Plots are shown for data obtained in 0 M (left) and 2 M (right) TMAO.

We reasoned that the consistency between data from the Cnf-Trp and Fam-Raz libraries on a Flory scaling plot should help us to identify the more accurate probability distribution. In Fig. 5.14, it is clear that there is substantially better agreement between the Cnf-Trp and Fam-Raz data sets when applying the Gaussian chain distribution to extract distance data from the E_{FRET} values.

Simulations

After data collection, simulations were performed in PyRosetta to generate structural ensembles that represent the changes observed by FRET. Before introducing FRET-based constraints, we optimized a PyRosetta script to effectively produce ensembles of structures in general agreement with previously published radius of gyration (R_g) and radius of hydration (R_h) values^{205, 228}. Simulations were performed where the weighting of the FRET constraints relative to other Rosetta energy terms (a , Eqs. 5.22 and 5.23) was varied, and different shapes for the FRET constraint function were assessed. These constraint function shapes were based on the Gaussian chain (Eq. 5.13) or semiempirical (Eq. 5.14) distance distributions used in the FRET analyses. Given the greater consistency between Cnf-Trp and Fam-Raz data when using the Gaussian chain probability distribution (above), we favored using this function for implementing constraints as well. However, we tested all combinations to ensure that no bias was introduced based on the pairing of distribution functions used in FRET data analysis and in constraint implementation. The results of these tests are detailed in Figs. 5.2-5.7, 5.14a-5.15c and are discussed further below. We found that setting the weight of the constraints relative to the other score function components to unity allowed the FRET constraints to influence the structure without overconstraining, and that reasonable variations of the shape of the constraint function did not dramatically influence the simulation results. Thus, in the main text, we exclusively report simulations performed using the Gaussian chain distribution, with all other simulations reported in the Supporting Material. The 2 M TMAO simulations were

performed with the solvation term removed in an effort to account for the significant change in solvation. Unconstrained simulations performed with this altered score function provided a more compact starting point for introducing FRET constraints, whereas constrained ensembles were noticeably more compact (Figs. 5.4-5.7).

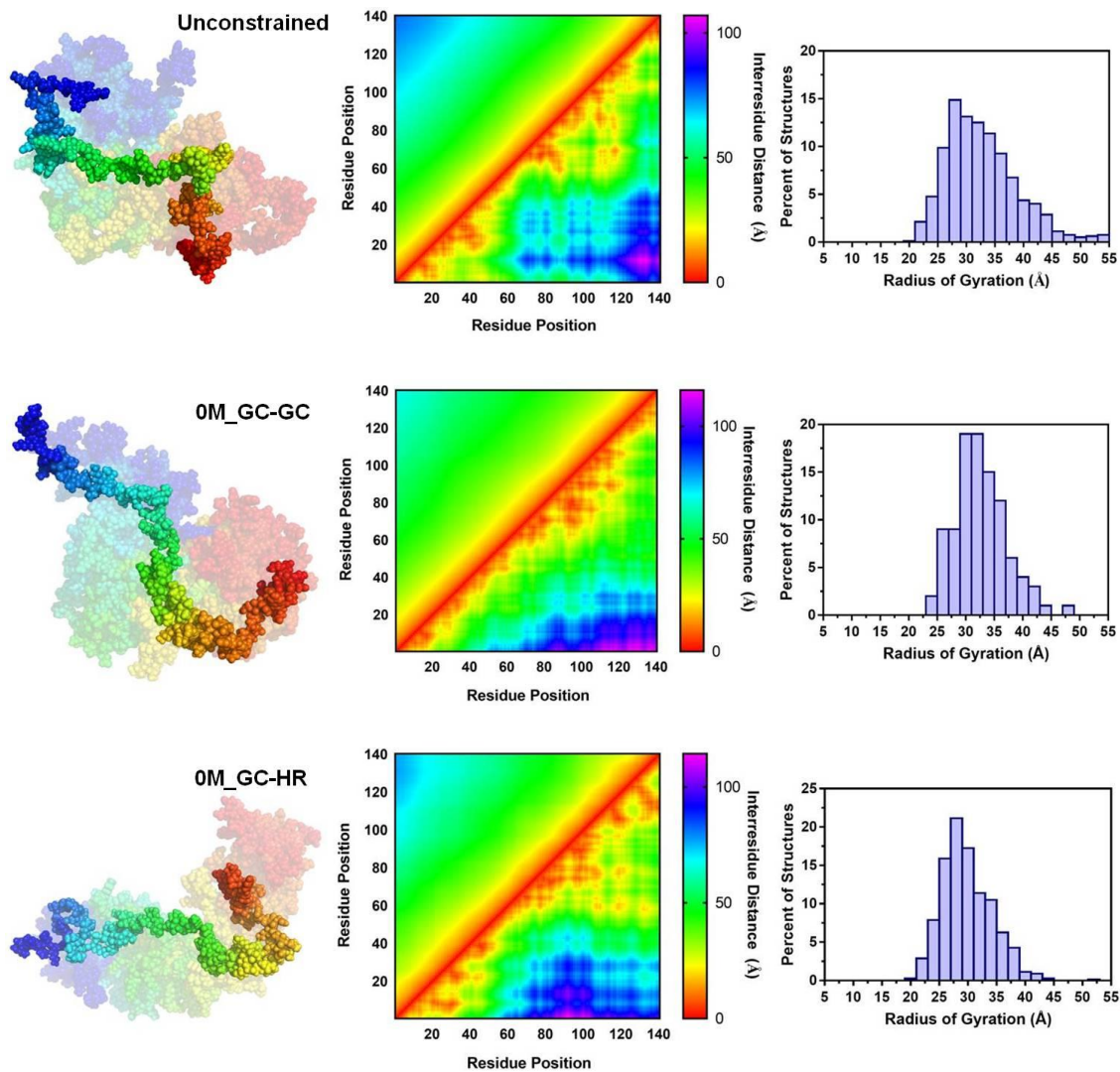


Figure 5.14a: Global structural summary for 0 M TMAO ensembles. Plots are derived from unconstrained (top set), 0M_GC-GC (middle set) and 0M_GC-HR (bottom set) simulations. Simulation names represent the concentration of TMAO in which measurements were taken, the probability distribution used in the polymer-scaled Förster equation for obtaining distances (Gaussian-chain = GC or semi-empirical = Pr) and the constraining function employed (Gaussian-chain = GC or harmonic potential = HR), respectively. Structures (left) show the 10 lowest energy structures faded with the single lowest energy structure darkened (blue - red, N-term - C-term). Heat map (middle) shows the average inter-residue distances from all output structures (above diagonal) and the inter-residue distances from the single lowest energy structure (below diagonal) as a function of residue pair. Histogram (right) of the percent of output structures for a given radius of gyration with bin widths of 2 Å.

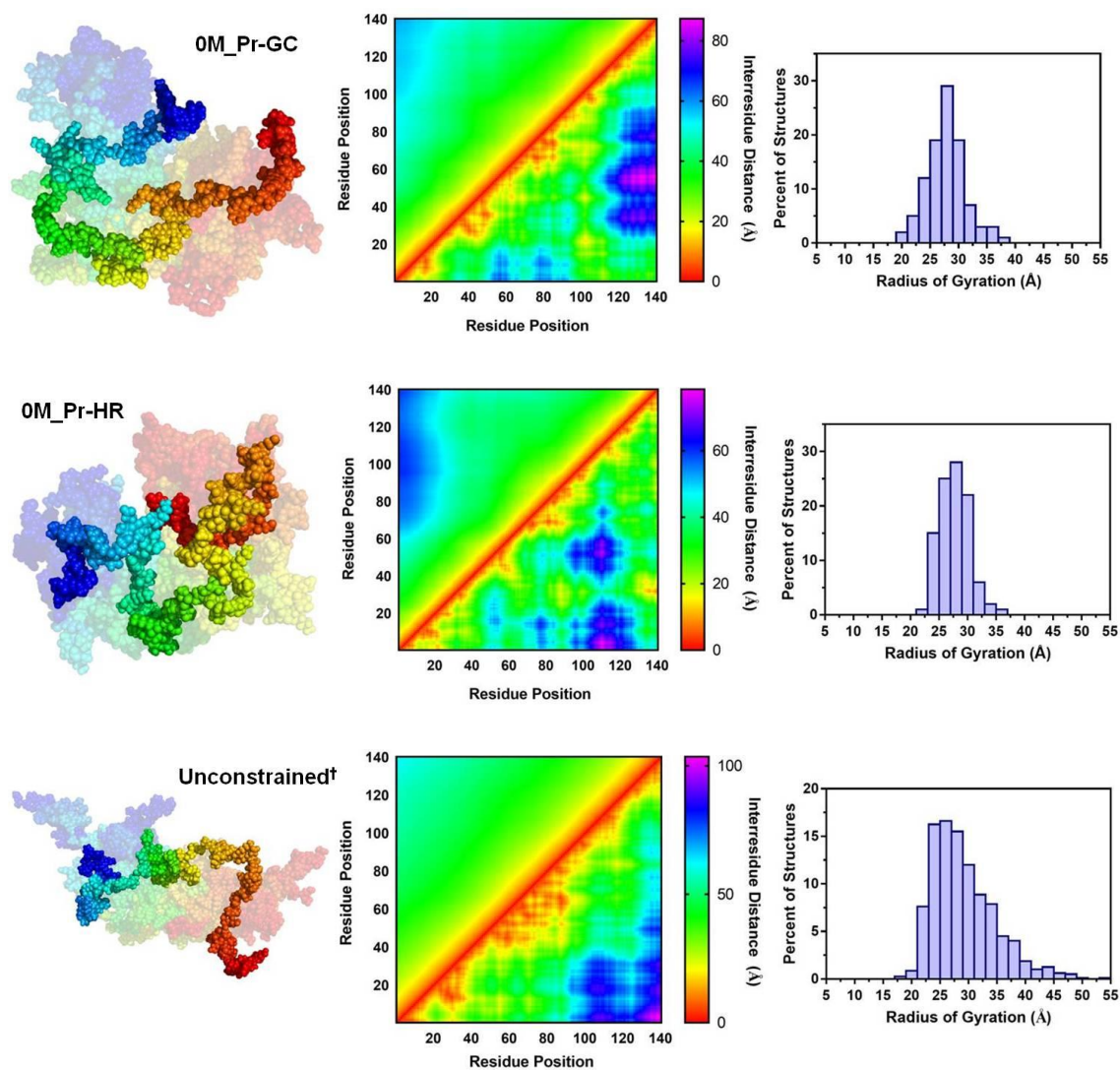


Figure 5.14b: Global structural summary for 0 M TMAO ensembles. Plots are derived from 0M_Pr-GC (top set), 0M_Pr-HR (middle set) and unconstrained† (bottom set) simulations. Simulation names represent the concentration of TMAO in which measurements were taken, the probability distribution used in the polymer-scaled Förster equation for obtaining distances (Gaussian-chain = GC or semi-empirical = Pr) and the constraining function employed (Gaussian-chain = GC or harmonic potential = HR), respectively. Structures (left) show the 10 lowest energy structures faded with the single lowest energy structure darkened (blue - red, N-term - C-term). Heat map (middle) shows the average inter-residue distances from all output structures (above diagonal) and the inter-residue distances from the single lowest energy structure (below diagonal) as a function of residue pair. Histogram (right) of the percent of output structures for a given radius of gyration with bin widths of 2 Å. Simulations marked with † indicate that the weight of the solvation term (f_{a_sol}) was set to 0.

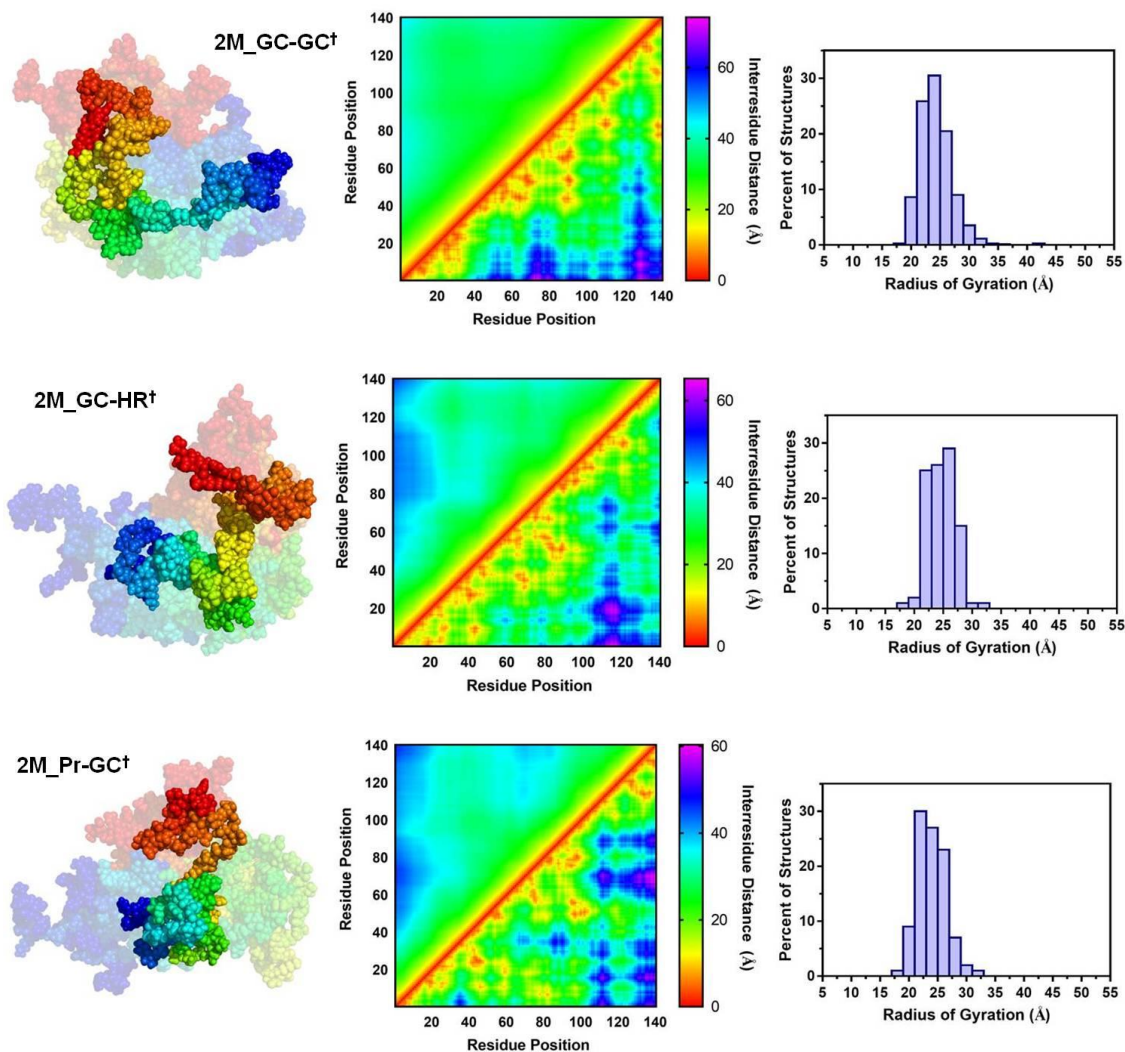


Figure 5.15a: Global structural summary for 2 M TMAO ensembles. Plots are derived from 2M_GC-GC† (top set), 2M_GC-HR† (middle set) and 2M_Pr-GC† (bottom set) simulations. Simulation names represent the concentration of TMAO in which measurements were taken, the probability distribution used in the polymer-scaled Förster equation for obtaining distances (Gaussian-chain = GC or semi-empirical = Pr) and the constraining function employed (Gaussian-chain = GC or harmonic potential = HR), respectively. Structures (left) show the 10 lowest energy structures faded with the single lowest energy structure darkened (blue - red, N-term - C-term). Heat map (middle) shows the average inter-residue distances from all output structures (above diagonal) and the inter-residue distances from the single lowest energy structure (below diagonal) as a function of residue pair. Histogram (right) of the percent of output structures for a given radius of gyration with bin widths of 2 Å. Simulations marked with † indicate that the weight of the solvation term (fa_{sol}) was set to 0.

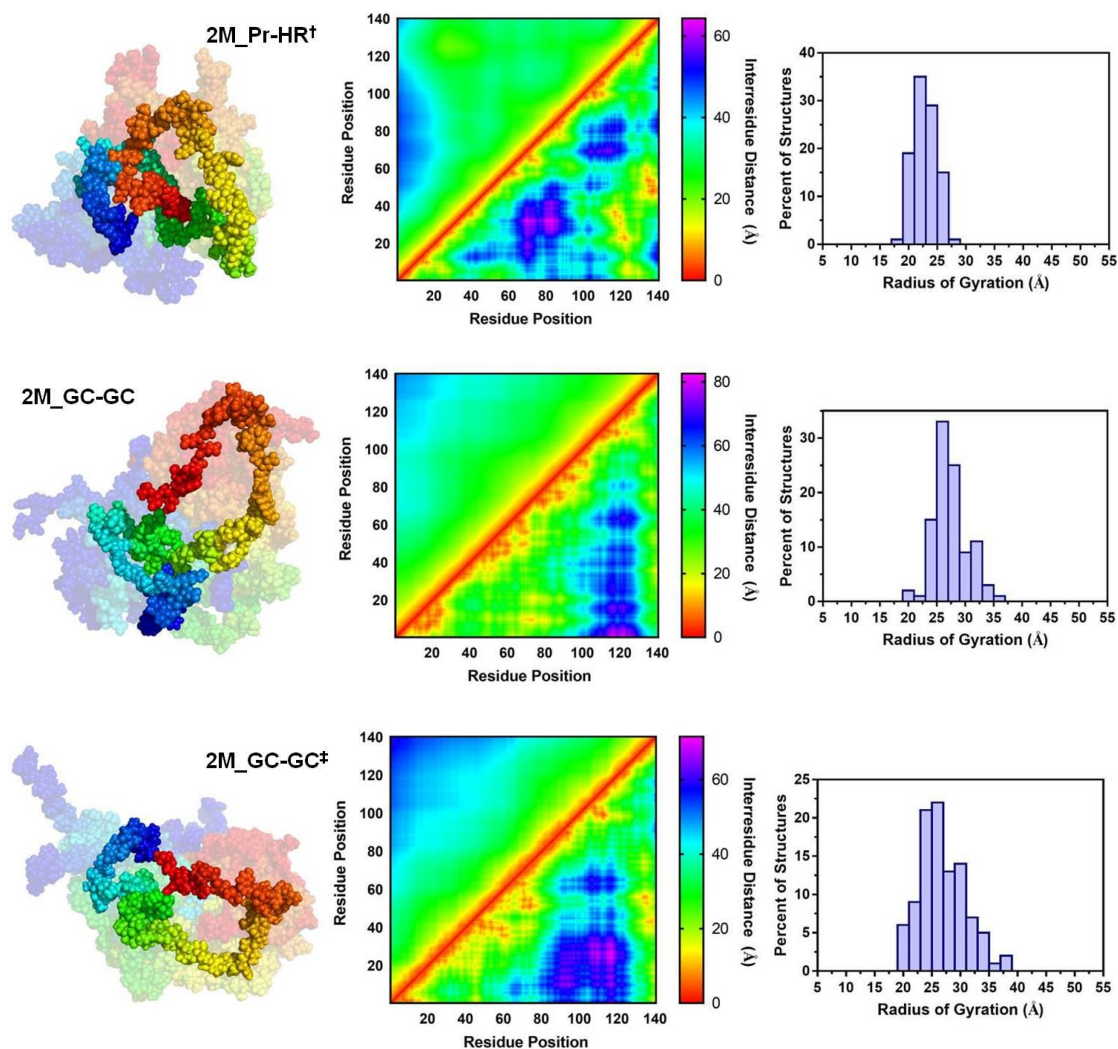


Figure 5.15b: Global structural summary for 2 M TMAO ensembles. Plots are derived from 2M_Pr-HR[†] (top set), 2M_GC-GC (middle set) and 2M_GC-GC[‡] (bottom set) simulations. Simulation names represent the concentration of TMAO in which measurements were taken, the probability distribution used in the polymer-scaled Förster equation for obtaining distances (Gaussian-chain = GC or semi-empirical = Pr) and the constraining function employed (Gaussian-chain = GC or harmonic potential = HR), respectively. Structures (left) show the 10 lowest energy structures faded with the single lowest energy structure darkened (blue - red, N-term - C-term). Heat map (middle) shows the average inter-residue distances from all output structures (above diagonal) and the inter-residue distances from the single lowest energy structure (below diagonal) as a function of residue pair. Histogram (right) of the percent of output structures for a given radius of gyration with bin widths of 2 Å. Simulations marked with [†] indicate that the weight of the solvation term (*fa_sol*) was set to 0 while simulations marked with [‡] indicated that fragment insertion was incorporated within the search.

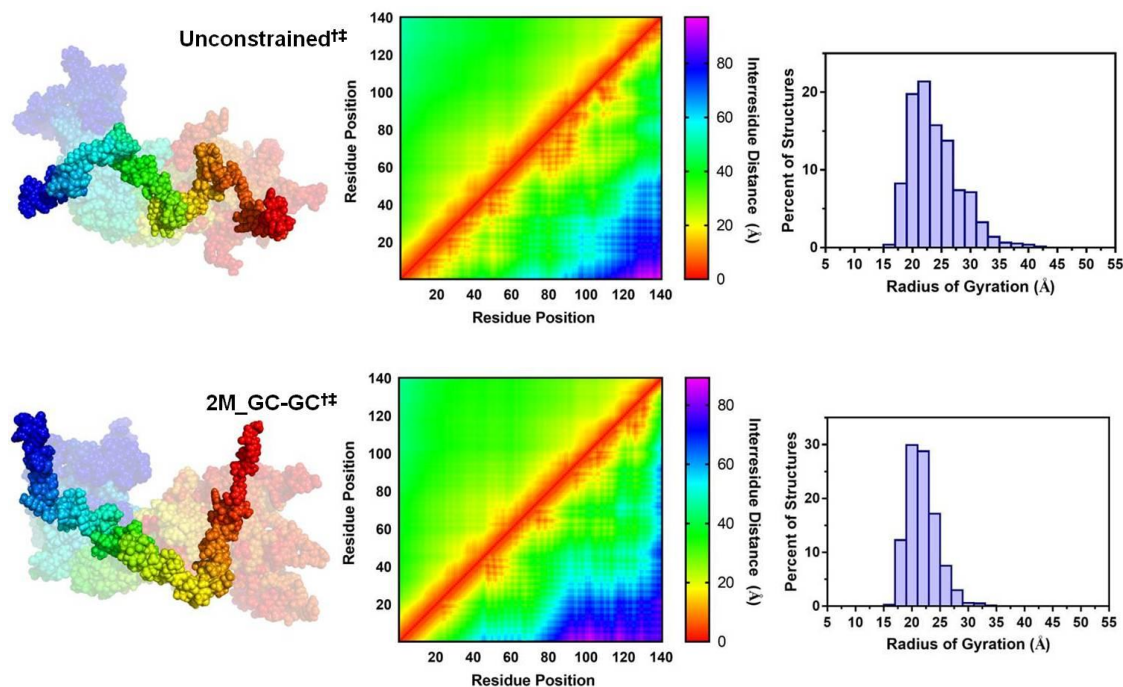


Figure 5.15c: Global structural summary for 2 M TMAO ensembles. Plots are derived from unconstrained†‡ (top set) and 2M_GC-GC†‡ (bottom set) simulations. Simulation names represent the concentration of TMAO in which measurements were taken, the probability distribution used in the polymer-scaled Förster equation for obtaining distances (Gaussian-chain = GC or semi-empirical = Pr) and the constraining function employed (Gaussian-chain = GC or harmonic potential = HR), respectively. Structures (left) show the 10 lowest energy structures faded with the single lowest energy structure darkened (blue - red, N-term - C-term). Heat map (middle) shows the average inter-residue distances from all output structures (above diagonal) and the inter-residue distances from the single lowest energy structure (below diagonal) as a function of residue pair. Histogram (right) of the percent of output structures for a given radius of gyration with bin widths of 2 Å. Simulations marked with † indicate that the weight of the solvation term (fa_{sol}) was set to 0 while simulations marked with ‡ indicated that fragment insertion was incorporated within the search.

Moreover, for simulations of the 2 M ensemble, fragment insertion was incorporated within the MC search to increase the amount of resultant secondary structure which has been observed by circular dichroism studies of the TMAO-induced conformation¹³³. Although this significantly increased the number of

helices, there was no marked improvement in the match of simulated structures with experimental FRET data (Fig. 5.4-5.7).

Initially, we compared the constrained simulations to the FRET data using Flory scaling plots, where the average interresidue distance (between α -carbons) as a function of sequence separation was plotted for the ensemble of 1000 lowest energy structures with experimental FRET data overlaid as discrete points (Fig. 5.10, *left*). All experimental data were close to the average simulation distance or within the standard deviation for both 0 and 2 M ensembles, with the notable exception of the Fam₉-Raz₁₃₆ FRET data. It is notable that the distance determined for this FRET pair using the harmonic potential (Fig. 5.14) does not show such dramatic discord with the other distances, potentially suggesting a limitation of the Gaussian chain model for very long distance ranges. We also predicted FRET values for all of the pairs of label sites for comparison to the experimental FRET data (Fig. 5.10, *right*). Distances between the α -carbons of the amino acids at the label sites were extracted from each structure in a simulated structural ensemble and converted to E_{FRET} values using the classical Förster equation (Eq. 5.12, where $P(r,x) = 1$). Since the variations in conformation in the ensemble explicitly capture the distribution of interresidue distances, the average E_{FRET} was computed as a simple average of the values extracted from each structure with no further correction for polymer scaling of the distance distribution. We found good agreement, with an average absolute E_{FRET} difference of 0.09 and all experimental values falling within the standard deviation of the simulated values. As discussed below, the close match between these values helps to validate our choice of a

polymer-scaled distance distribution function in interpreting E_{FRET} values to input constraints in the simulations.

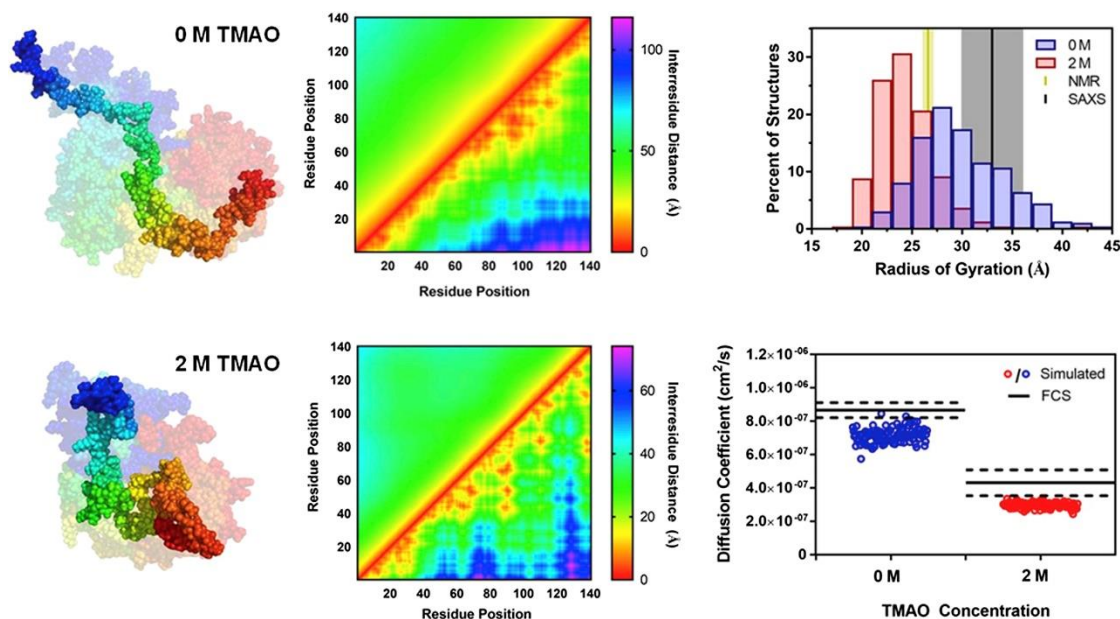


Figure 5.16: Analysis of α S structural ensembles. Left: Representative structures from 0 and 2 M simulations. The darkened structure in the foreground is the lowest energy structure and the faded ensemble in the background includes the 10 lowest energy structures. Middle: Heat maps showing the average interresidue distances in the simulated ensembles. Each heat map shows the average distance for the full simulated ensemble of structures above the diagonal and distances from the single lowest energy structure below the diagonal. Top right: Histograms of radii of gyration of structures from simulations, plotted with literature values of R_g determined from SAXS data and the hydrodynamic radius from NMR studies. Bottom right: Diffusion coefficients from FCS and simulated structures.

Fig. 5.16 shows the 10 lowest energy structures from the 0 and 2 M simulations, aligned and with a single structure darkened for clarity. One can see significant compaction of the 2 M ensemble, which can be quantified using histograms of the R_g for each structure in the ensemble. The average radius of gyration in 0 M TMAO is $29.3 \pm 4.6 \text{ \AA}$, which shrinks to $23.5 \pm 3.4 \text{ \AA}$ in 2 M TMAO.

Interestingly, despite this compaction, α S remains disordered in 2 M TMAO. The disorder can be visualized using a plot of the distribution of interresidue distances as a two-dimensional “heat map,” with interresidue distances for the entire ensemble plotted above the diagonal and interresidue distances for the single darkened structure below the diagonal. As seen in the middle plots in Fig. 5.16, in both 0 and 2 M TMAO, the average interresidue distance scales with primary sequence separation, despite the fact that long-range contacts occur in individual structures within the ensembles (Fig. 5.17).

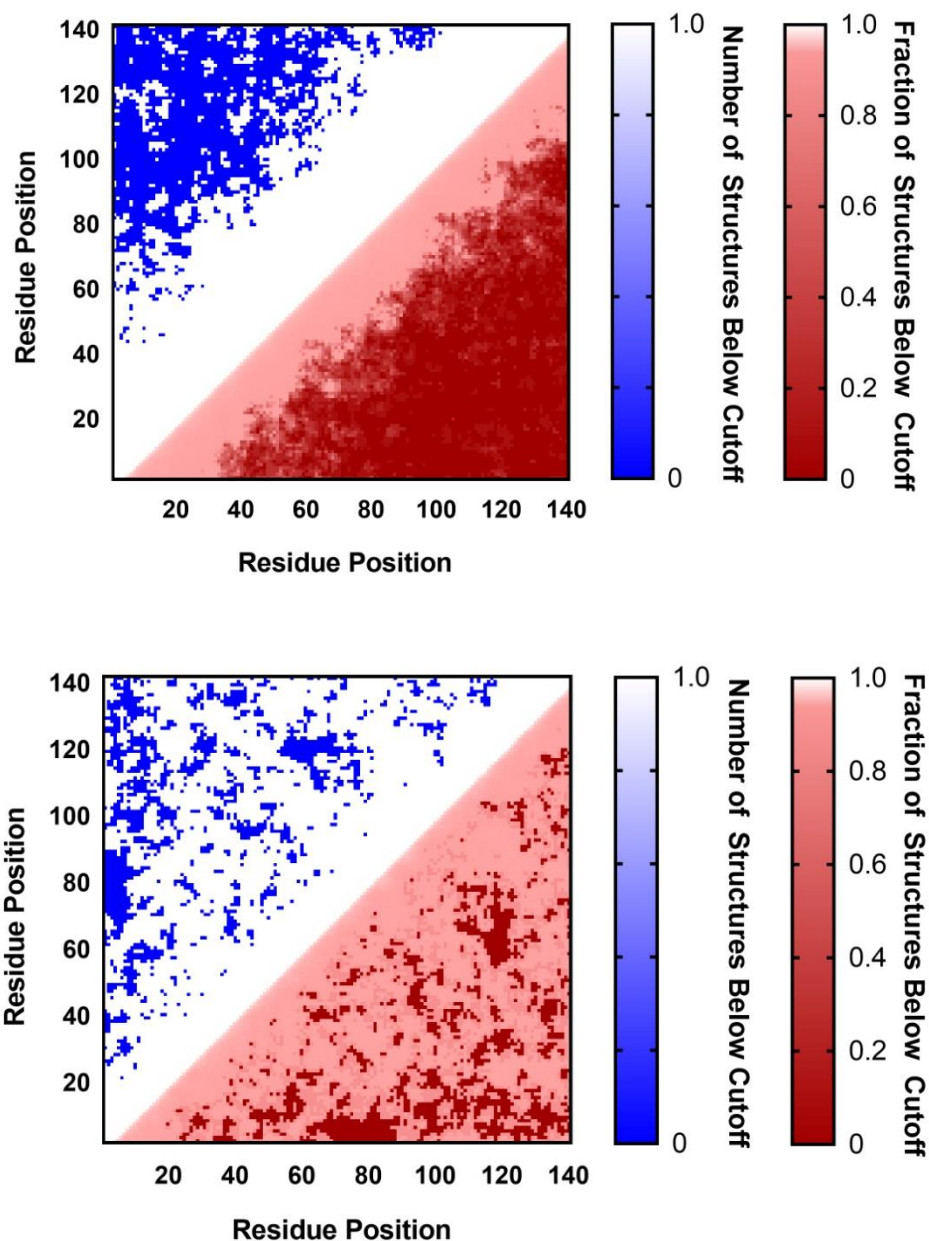


Figure 5.17: Contact maps from experimentally constrained ensembles. Contact maps from the 0M_GC-GC (top) and 2M_GC-GC† (bottom) ensembles. A cutoff distance of 10 Å was set as a contact. Each half of each heat map are analyses of the same ensemble and show all residue pairs which make contact in one or more structures from the ensemble as white while residue pairs which do not make contact are shown in blue (top left). The fraction of structures of the ensemble which make contact are shown over the full fractional scale with a red/white gradient (bottom right).

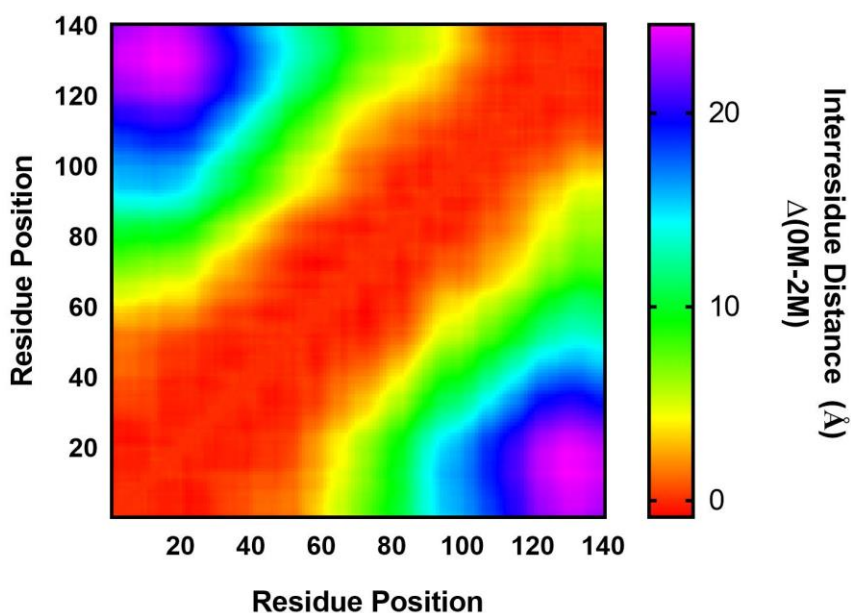


Figure 5.18: Average interresidue distance heat map. Heat map showing the average difference in interresidue distance between the 0M_GC-GC and 2M_GC-GC† ensembles.

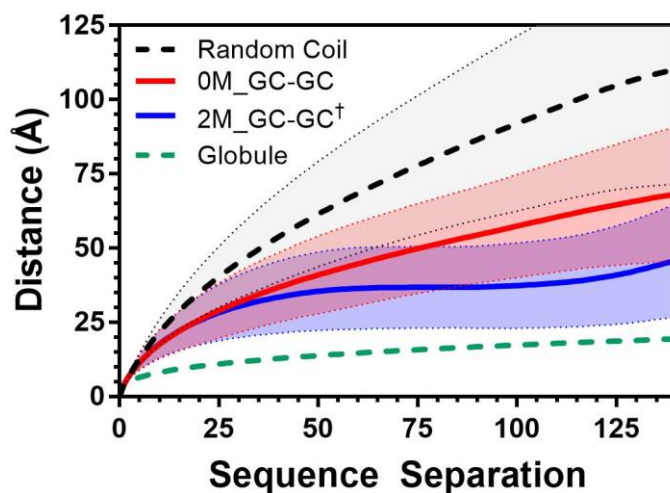


Figure 5.19: Experimentally constrained ensemble Flory scaling plot. Flory scaling plots of 0M_GC-GC (red) and 2M_GC-GC† (blue) simulated ensembles along with the random coil simulation (black) used for the construction of $P2(r,x)$ and a curve representing the scaling of a globule protein (green) ⁶⁸.

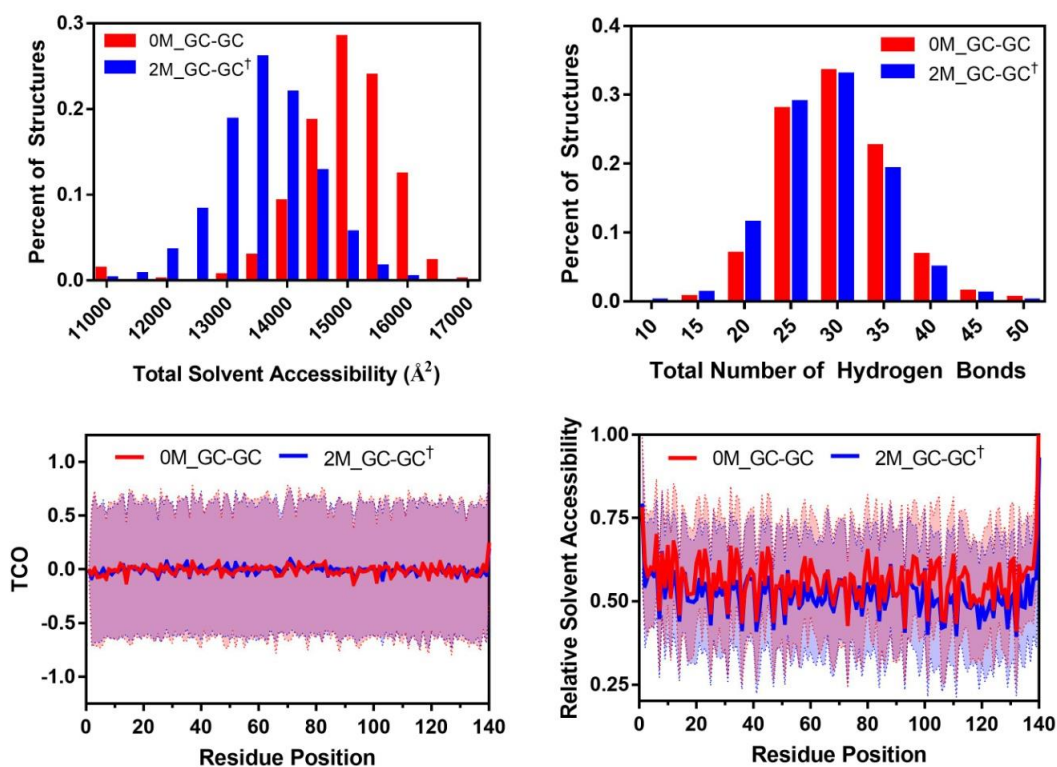


Figure 5.20: Experimentally constrained ensemble DSSP analysis. DSSP analysis of 0M_GC-GC (red) and 2M_GC-GC† (blue) ensembles for comparison. (Top Left) Histogram showing the percent of structures binned based on the total solvent accessible surface area of a given structure (Å²). (Top Right) Histogram showing the percent of structures binned based on the total number of intramolecular hydrogen bonds present in a given structure. (Bottom Left) Plot of the average value of the cosine of the angle between the backbone carbonyl of the current residue and the carbonyl of the previous residue (TCO) for each residue for each structure within a given ensemble. (Bottom Right) Plot of the average value relative solvent accessibility for each residue for each structure within a given ensemble.

This combined information indicated that, in the presence of 2 M TMAO, α S populates a compacted disordered ensemble that still maintains a high degree of structural heterogeneity. To confirm that the structure ensembles were consistent with other experimental observables, we calculated the diffusion coefficient for each structure of the 0 and 2 M ensembles using HydroPro²²⁶. As seen in

Fig. 5.16, the calculated diffusion coefficients are slightly (16%) smaller, but in reasonable agreement with those measured in our FCS experiments.

Comparison of structure ensembles to data from literature

We compared our structural models in the absence of TMAO to data from the literature as well as other published models. The average R_g of our 0 M ensemble matched well with reported values of R_g from SAXS experiments and R_h (typically 30% larger than R_g) from NMR experiments^{205, 228}. Furthermore, our models agreed remarkably well with both the primary NMR PRE data (Fig. 5.21) and Flory scaling plots of the resulting models (Fig. 5.22) reported by Allison *et al.*

126

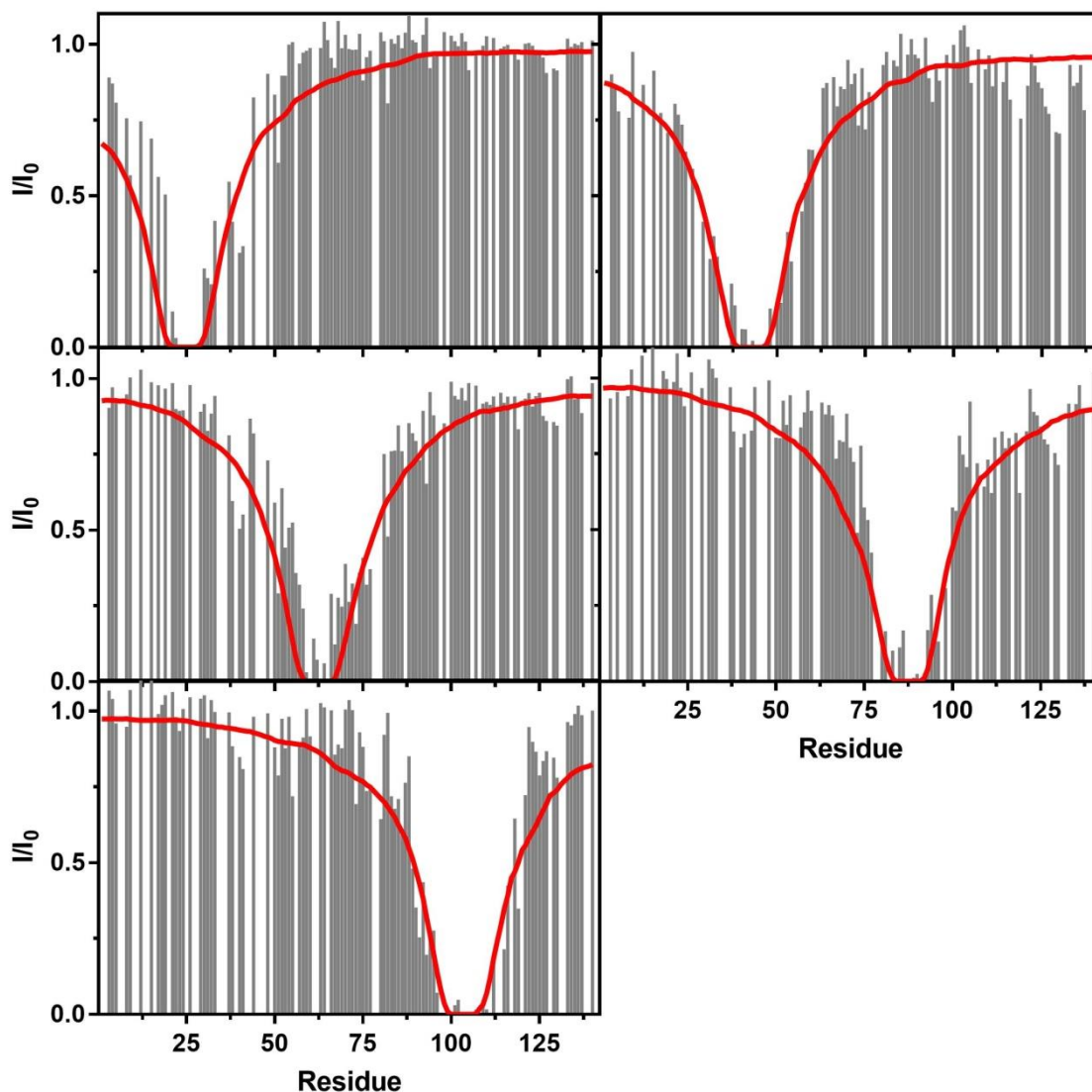


Figure 5.21: Intensity ratios from PRE measurements. Each plot shows the measured intensity ratio from PRE data obtained by Allison et. al. (grey bars) and the calculated average intensity ratio from the constrained simulation (red lines) were derived from data acquired in 0 M TMAO and interpreted using the Gaussian-chain polymer scaled Förster equation ¹²⁶. Plots show data for spin-label placement at residues 24 (top left), 42 (top right), 62 (middle left), 87 (middle right) and 103 (bottom right).

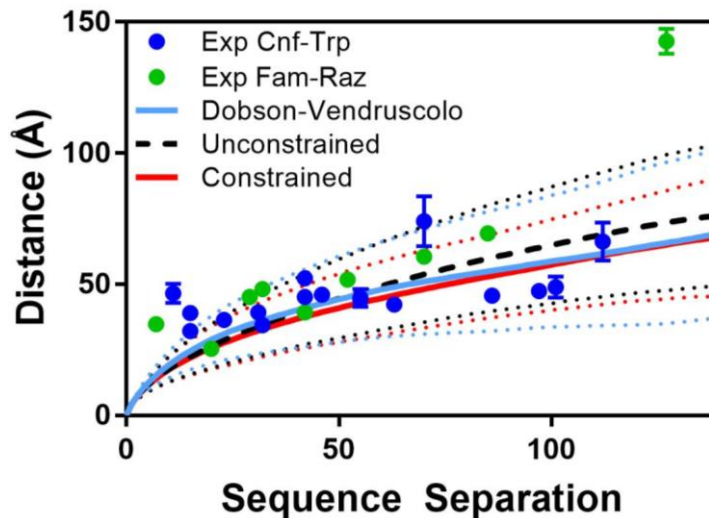


Figure 5.22: Flory scaling plot comparison to previously published ensemble. Plot shows the average distance as a function of sequence separation for simulated ensembles (PED9AAC) and unconstrained and constrained simulations reported herein¹²⁵⁻¹²⁶. Distance constraints were derived from data acquired in 0 M TMAO and interpreted using the Gaussian-chain polymer scaled Förster equation. Constraints were implemented using the Gaussian-chain constraint function.

We do observe rare contacts between regions of the protein that are distant in primary sequence (Fig. 5.17), consistent with reports of transient C-terminal contacts with the NAC and N-terminus²⁰⁰.

Although there is limited information on the structure of α S in 2 M TMAO, we find that our models agree with the available information. Similar to previous studies, our circular dichroism measurements made in 0 or 2 M TMAO show little change in α S helicity (Fig. 5.23), consistent with our models which show that the 2 M ensembles are still highly disordered.

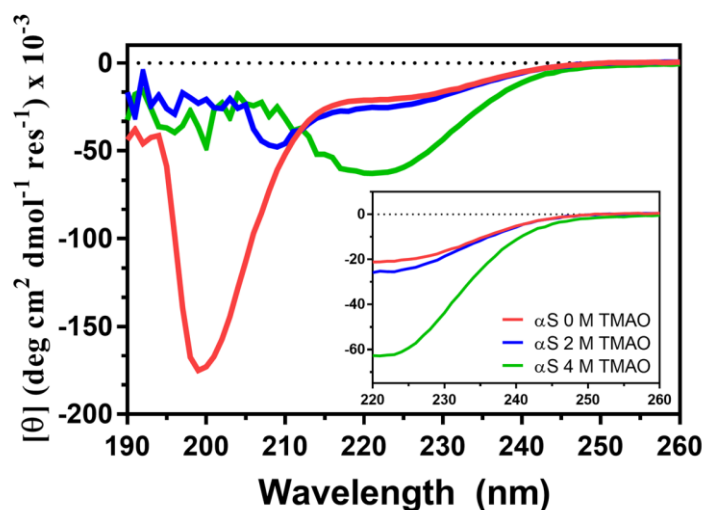


Figure 5.23: CD spectra of α S in varying concentrations of TMAO. Full plot and insert show the same data set, where the insert is adjusted to show difference in spectra above 220 nm. Measurements were taken on an Aviv model 410 circular dichroism spectrometer in 1 mm path length quartz cuvettes and were performed with 20 μ M wild-type α S in 20 mM Tris 100 mM NaCl pH 7.5. TMAO was added to each buffer and the pH was readjusted at 25 $^{\circ}$ C. Scans from 190-260 nm were performed with a 1 nm bandwidth, 1 nm step size and an averaging time of 10 seconds.

Previously, Uversky and coworkers monitored fibrillization of α S in TMAO using thioflavin T (ThT) fluorescence, and found that aggregation was accelerated, but that the final level of ThT fluorescence was lower. Our own studies with Congo Red (Fig. 5.24) corroborated this observation, although, intriguingly, changes in fluorescence polarization (Fig. 5.25) were slower in 2 M TMAO, possibly indicating that the ThT and Congo Red spectroscopic properties are changed in TMAO.

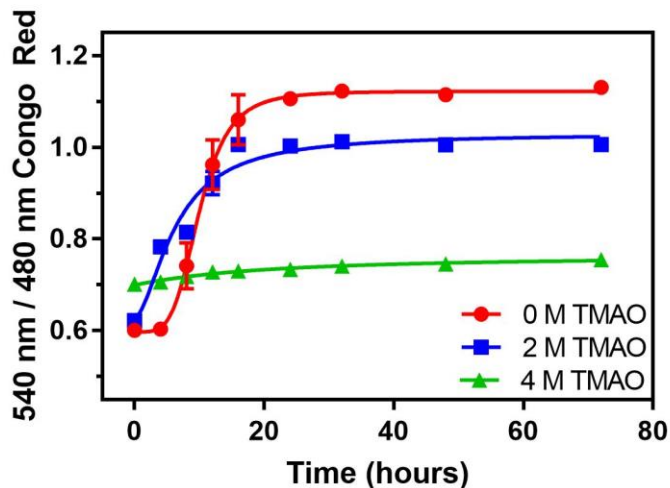


Figure 5.24: Aggregation kinetics of α S monitored by Congo Red. Aggregation was performed with 100 μ M wild-type α S in 20 mM Tris 100 mM NaCl pH 7.5. TMAO was added to each buffer and the pH was readjusted. Samples were agitated in an orbital shaker at 1500 rpm at 37 $^{\circ}$ C. At each timepoint, 10 μ L of sample was removed and added to 140 μ L of 20 μ M Congo Red in water and incubated for 20 minutes at room temperature prior to measurement. Absorption spectra (230-700 nm) were measured in a 96-well black CoStar clear bottom plate on a Tecan M1000 plate reader. Kinetics were fit to Eq. 5.26 in Prism with the detailed values in Table 5.7.

$$Y = A + (B-A)/(1 + (C/x)^D) \quad (\text{Eq. 5.26})$$

Table 5.7: Values from fits of Congo Red aggregation kinetics.

	0 M TMAO	2 M TMAO	4 M TMAO
A	0.5977	0.6289	0.7003
B	1.123	1.03	0.7656
C	9.989	6.397	19.72
D	4.357	1.708	1.104
R²	0.9855	0.9526	0.9062

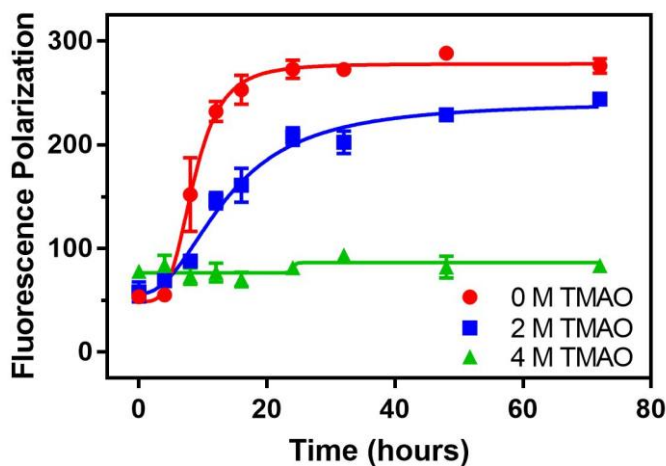


Figure 5.25: Aggregation kinetics of α S monitored by fluorescence polarization. Aggregation was performed with 100 μ M α S-Fam136 in 20 mM Tris 100 mM NaCl pH 7.5. TMAO was added to each buffer and the pH was readjusted. Samples were agitated in an orbital shaker at 1500 rpm at 37 $^{\circ}$ C. At each timepoint, 10 μ L of sample was removed and added to 90 μ L of 20 mM Tris 100 mM NaCl pH 7.5. Fluorescence polarization measurements were taken in a 96-well black CoStar clear bottom plate on a Tecan F200 plate reader. Kinetics were fit to Eq. 5.26 in Prism with the detailed values in Table 5.8.

Table 5.8: Values from fits of fluorescence polarization aggregation kinetics.

	0 M TMAO	2 M TMAO	4 M TMAO
A	48.65	58.28	76.27
B	278.1	240.9	86.22
C	8.495	13.74	23.99
D	3.82	2.248	80.41
R²	0.9806	0.9734	0.2733

Thus, although compacted, α S remains disordered and able to sample conformations that lead to fibril formation. This can be seen in Figs. 5.19 and 5.20 where, despite compaction observed by changes in solvent-accessible surface area, metrics such as numbers of intramolecular hydrogen bonds and backbone dihedral angles do not indicate the presence of persistent structure.

§ 5.7 Discussion

Our previous study of α S structures using MC simulations with only a repulsive Lennard-Jones potential and harmonic constraints based on single-molecule FRET data gave structural ensembles with global properties that matched well to experimental measurements such as R_g . This modeling protocol, with constrained simulations utilizing exclusively a repulsive van der Waals potential, was extremely efficient in its simplicity, but the resulting models lacked atomic-scale details that could be used in generating hypotheses or interpreting

mechanisms of conformational change. We and others have also performed unconstrained simulations which included the amino acids using single-sphere “centroid” representations of the side chains, as well as all-atom molecular dynamics simulations of α S, which do provide such detailed information, but are much more computationally intensive and generally limited in the conformational space explored. We wished to find an intermediate level of simulation wherein we could include side chains in MC simulations and maintain efficient sampling. However, we reasoned that the previous long-range constraints (>30 Å) derived from single-molecule FRET data would not provide sufficient information on short-range interactions to properly direct these simulations. Thus, we here included data from the short-range Cnf-Trp pair along with data from the Fam-Raz pair, which has a comparable FRET range to the previous single-molecule FRET probes. These short-range probe pairs may be crucial for identifying contacts or collapsed regions that can exist under varied solution conditions or in the presence of allosteric molecules.

Assessing the validity of $P(r)$

We reasoned at the outset that the Gaussian chain $P(r)$ was more accurate because it led to greater consistency between the distances obtained from the Cnf-Trp and Fam-Raz libraries. However, this method of analysis was complicated by the limited numbers of intramolecular distances in a range that could be accurately captured by both pairs, as these probes were selected specifically for their efficacy across different distance ranges. Furthermore, one can observe that the Cnf-Trp

distances are largely invariant after reaching a sequences separation of ~ 50 residues due to a working range that extends to only ~ 35 Å. Therefore, our assignment of the Gaussian chain $P(r)$ as the more accurate distribution function arises from visual inspection of the Flory scaling plots in Fig. 5.14 in the 25–45 Å range, which encompass the upper and lower bounds of the Cnf-Trp and Fam-Raz probe pairs, respectively. It is important to note that at very short (<15 residues) and very long (>115 residues) sequence separations, the calculated distances are likely unable to be accommodated by any conformation of the protein. Furthermore, the functional form of the distribution may also be dependent on the number of residues between two probes (e.g., short sequence separations cannot have truly polymer-like behavior and may need to be treated with classical FRET equations) or heterogeneity not accounted for by polymer-scaling behaviors²²⁹.

Differences between constraint methods

Since IDPs have relatively flat energetic landscapes, improper introduction of constraints can easily result in ensembles where the resultant conformations are not sufficiently diverse, especially proximal to constrained sites. This arises primarily from constraining potentials that are too deep or too narrow, but can also occur when conformational space is not effectively sampled. The latter problem could be caused by poor parameterization of the Metropolis criterion (such as selection of a kT value that is too low) or by not having an appropriate score function to produce the multiple local minima present within a single pair-wise constraint potential. By expanding our score function from a single repulsive van

der Waals term to the current optimal score function utilized by Rosetta, we assure that output states populate local minima that allow us to extract high-resolution information and provide an effective buffer for constraint introduction. Achieving unconstrained simulations of this quality required significant adjustment of PyRosetta scripts, which normally are used to model folded proteins rather than IDPs. These optimization efforts will be reported in detail elsewhere. In addition to improved modeling scripts, we hypothesized that issues related to the constraint functional form could be circumvented by directly employing appropriate distance distributions for disordered ensembles to analyze the FRET data.

Nonetheless, we performed simulations with all four combinations of the two $P(r)$ functions used to obtain distances and the two corresponding constraining functions. This allowed us to consider how the ensembles were influenced by the method of constraining in addition to the sets of distances used as constraints. Figs. 5.4-5.7, 5.14a-5.15c demonstrate that neither the method of constraint nor the $P(r)$ utilized for the distance determination through Eq. 5.12 had a dramatic impact on the resulting structures. This likely arises from the fact that distances from the unconstrained simulation are already very close to the distances obtained from the experiments. Despite the relative agreement between the ensembles produced across all constraining methods for a given concentration of TMAO, there are several noteworthy observations. Figs. 5.14a-5.15c show that in all cases in which a harmonic potential was employed, the structural diversity, most clearly observable in the dispersion of R_g values, is visibly decreased. Moreover, harmonic constraints consistently produced deviations in the scaling behavior,

where the heat maps (Figs. 5,14a-5,15c) reveal distances between the N-terminal region and the NAC domain that are, on average, longer than those between the C-terminal region and the NAC. This observation is independent of the input data, as these nuances are observed for both data sets, with constraints from data obtained utilizing $P_2(r,x)$ further enhancing these trends. These observations demonstrate that there are significant differences between these ensembles which result exclusively from the functional form of the constraining potential.

The relative impacts of changes in the score function or in the constraints are evident in comparisons between 2 M simulations performed using Gaussian chain derived data with or without constraints or solvation term modification (Figs. 5.4, 5.14a-5.15c). FRET constraints alone compacted the 2 M ensemble (2 M GC-GC) relative to the 0 M ensemble (0 M GC-GC), but were not sufficient to generate ensembles that were consistent with the 2 M experimental data. However, it is important to note the marked differences between the 0 and 2 M ensembles, as these demonstrate that simulations in PyRosetta can be significantly influenced by constraints, which was not evident from the 0 M simulations alone. The fact that these constraints alone were unable to produce ensembles in full agreement with the 2 M data suggests that α S in the presence of molar quantities of TMAO populates a subset of conformations that are not accessible with the standard score function. With the solvation term removed, comparison of the 2 M constrained (2M GC-GC[†], see Fig. 5.15a) and unconstrained (Unconstrained[†], see Fig. 5.14b) simulations demonstrate that the constraints serve to exclude extended structures. This further suggests that the application of constraints does not

produce new conformations not present in the unconstrained population, but can effectively remove unfavorable structures from an ensemble based on experimental data. Overall, we see this as a favorable compromise as the necessity to modulate the unconstrained population through changes to the PyRosetta scoring is likely driven by the dramatic change in buffer conditions, which would not be present in many other applications. Systems where dramatic changes in solvent conditions occur (as in this study) can be effectively represented by adjusting the score function, whereas intramolecular (or intermolecular) contacts can be accounted for by changing the constraint functional form to accurately represent experimental data.

Comparison of experimental and simulated data

Initial comparisons between the experimental and simulated distances in Figs. 5.10 and 5.4 demonstrate that most of the obtained distances fall within 1 SD of the average interresidue distance for the given sequence separation. The observed disagreements at high and low sequence separation are likely due to inaccuracies in the determination of the experimental distance arising from either the working range of the probe or the polymer-scaling function, as previously discussed. Moreover, agreement in simulated and experimental FRET efficiencies in Figs. 5.10, 5.5, and 5.6 demonstrate that the underlying conformation dispersion is accurately captured, since appropriate efficiencies are obtained for both probe pairs. The FCS measurements herein also demonstrate that the simulated ensembles are of approximately the correct overall size and that the degree of

compaction is qualitatively accurate. The quantitative disagreement between the simulated diffusion coefficients observed in Fig. 5.16, where the simulated diffusion coefficients are systematically lower than the experimental values, could arise from several different factors. Given that the overall size agrees with previously published NMR and SAXS data, it is possible that we encountered small, systematic inaccuracies in determining diffusion coefficients resulting from a combination of the treatment of the disordered ensemble in HydroPro, the need for an empirical conversion factor, or accounting for the change in viscosity induced by TMAO. Importantly, we have demonstrated in Figs. 5.21 and 5.22, respectively, that our 0 M ensembles agree with previously published PRE data and the structural ensembles generated from those data ^{104, 126}.

It is worth noting that in our effort to establish an efficient, intermediate level of simulation, we chose not to pursue some elements that could have further improved the accuracy or tested the boundaries of our simulations. For example, we elected not to incorporate probe/linker spatial exploration. Although we acknowledge that simulations would be made more accurate by including representations of the probes, with FRET based on distances between fluorophore transition dipoles instead of α -carbons, this would have significantly increased the computational time, undermining our intention of creating an efficient approach. Additionally, we envision that coupling our method with the strategy previously employed by Dobson and Vendruscolo, where constraining functions were iteratively updated, could produce a more refined, but computationally intensive, version of our modeling protocol ¹⁰⁴. Moreover, this method circumvents the need

to assume a probability distribution function when interpreting the FRET data, as the constraints would be applied directly as FRET efficiencies. Finally, since introduction of a new constraint functional form is as simple as writing a new function in Python, we hypothesize that our method of constraint introduction, where assumptions regarding underlying distributions are directly converted into a potential, is not only useful for applying FRET data from disordered systems, but would also allow for efficient incorporation of other types of experimental data such as PRE data ^{107-108, 126} .

§ 5.8 Conclusions

We have developed a Rosetta modeling protocol using explicit protein side chains and sophisticated score functions in combination with appropriately weighted distance constraints to generate models of IDPs. By performing simulations that were constrained with experimental FRET data from two libraries containing different FRET pairs, we were able to model the ensemble of α S in buffer and in the presence of 2 M TMAO. Our models agreed well with independent measurements of α S structure from FCS, NMR, PRE, and SAXS data, and were computationally less taxing than traditional molecular dynamics simulations. In future work, we intend to explore the degree to which ensembles generated in PyRosetta capture residual secondary structure and the accuracy of values computed from these ensembles, such as chemical shifts or J-couplings in comparison with NMR experiments. The modifications made to our Rosetta modeling protocol allowed us to produce reasonable α S starting models in the

absence of constraints. The quality of these unconstrained models gives us greater confidence in interpreting the interactions observed in the constrained structural ensembles, and makes the simulations more robust to the inclusion of an inconsistent constraint. After careful consideration of constraint function shape and weight, we have found a form that allows them to influence the structural ensemble without overconstraining. Future investigations could employ different weights for different FRET pairs, or different functional forms for different distance ranges.

Our modeling protocol for IDPs incorporates atomic detail relevant to the study of chemical-, ligand-, or environment-induced conformational changes, and yet is sufficiently rapid both in data collection and simulation time to be applied in a moderate throughput fashion. Once libraries of labeled proteins have been generated, FRET measurements of the type reported here could easily be acquired under a variety of conditions, potentially even in a multiwell format. The collection of FRET constraints could then be used to generate structural ensembles to explore hypotheses for mechanisms of conformational change or to rationalize trends among molecules and modifications that modulate IDP conformation. For example, Kakish et al.²³⁰ have recently shown that bis-heterocycles linked by a flexible tether are able to bind to α S and induce conformational changes in the monomer that inhibit its propensity to aggregate. The procedures used here to study the effects of TMAO on α S structure could be applied to study such molecules with therapeutic potential. Furthermore, recent work has shown that modifications such as serine glycosylation and tyrosine phosphorylation affect aggregation and membrane binding respectively²³¹⁻²³².

Subsequent modification of the current labeled library, or production of a new library, would allow one to utilize the method reported herein to visualize the effect of these and other posttranslational modifications on the disordered ensemble. Although certain classes of molecules or solution conditions may spectroscopically interfere with our FRET probes, it is important to note that we can easily vary these probes using essentially the same labeling strategies. For example, we have recently reported a methoxycoumarin-acridonylalanine FRET pair that can be introduced by a combination of cysteine modification and unnatural amino acid mutagenesis⁵¹. Finally, it is important to note that our methods are not restricted to pure IDPs like monomeric α S, but can be applied to disordered regions of folded proteins or ordered aggregates such as the N- and C-terminal regions of fibrillar α S^{47, 233}.

**CHAPTER 6: A UNIFIED DE NOVO APPROACH FOR PREDICTING THE
STRUCTURES OF ORDERED AND DISORDERED PROTEINS.**

§ 6.1 Introduction

Intrinsically disordered proteins (IDPs) and intrinsically disordered regions (IDRs) of proteins have garnered increasing attention due to growing recognition of their roles in cellular health and disease.^{2, 234} These highly dynamic systems play crucial roles in signaling pathways, the function of the nuclear pore, lipid transport, membraneless organelles, and several pathologies.^{2, 20, 234} These functions are often related to structural transitions, where protein-protein interactions or ligand binding events facilitate conversion from a broad conformational ensemble to a much smaller number of states.² The ability to model both the conformationally diverse states as well as the induced ordered structures of IDPs is essential for mechanistic understanding of their activity and for potential therapeutic development for IDP-related proteinopathies.

As interest in IDPs and IDRs has continued to grow, experimental and theoretical techniques originally developed for structured proteins have been adapted to accommodate these dynamic systems. Although NMR, fluorescence and scattering-based methods have all been effectively utilized to characterize unstructured systems, the production of representative structural ensembles for IDPs using computational methods has proven far more challenging.^{53, 60-61, 68, 104, 107-108, 200-202, 219} Many approaches have constructed ensembles by using experimental data to filter sets of potential structures generated from random-coil or Protein Data Bank (PDB) fragment libraries.^{107-108, 199} Molecular dynamics (MD) based methods have been developed and optimized to predict IDP ensembles.^{97-100, 105} Monte Carlo (MC) methods, such as those employed in CAMPARI, have

demonstrated efficacy in modeling structural ensembles with and without data restraints, but protocols developed to date cannot generally accommodate both folded proteins and IDPs.^{53, 68, 235-236}

Our goal was to develop an architecture which effectively handles both ordered and disordered proteins and can generate accurate structural representations of both, *de novo*. Furthermore, we want to develop algorithms that do not suffer the hardware and computational time burdens associated with MD simulations. We posit that MC sampling, which allows for broad sampling of conformational space, is more efficient than the relatively local sampling associated with MD. However, one major hurdle with MC methods is that the identification of true minima is often challenging. With these ideas in mind, we have turned our attention towards the Rosetta Modeling Suite. This platform has demonstrated extensive success in predicting structures for ordered systems through the use of a simulated annealing MC approach with limited gradient-based minimization.²³⁷ Simulated annealing methods effectively reduce simulation time by utilizing coarse grained representations during early stages of sampling and can greatly favor the identification of minima. In addition, Rosetta has been utilized successfully to model IDRs and IDPs with the application of experimental constraints.^{53, 68, 236}

Here, we focus on adapting two algorithms: AbInitio, which can predict folded protein structure from sequence, and FloppyTail, which can produce structural ensembles of IDRs.^{236, 238-240} For the first time, to our knowledge, we tested these method's ability to produce an accurate structural ensemble of an

IDP, α -synuclein (α S), for which there exists a battery of experimental data. α S is fully disordered in solution and is a key player in the pathogenesis of Parkinson's Disease and other proteinopathies and is one of the most well-studied IDPs.¹⁹⁻²⁰ After demonstrating the limitations of both algorithms, we developed a generalized simulation format in PyRosetta that allowed us to test the efficacy of the underlying energy term and sampling methods that exist in Rosetta.¹¹⁴ After identifying issues with Rosetta's knowledge-based terms and fragment sampling approach, we demonstrate methods for improving the energy term accuracy and fragment selection process. Although the solutions posited here are applied exclusively to Rosetta, both the issues themselves and the solutions are finding that are likely generalizable for other Monte-Carlo based methods.¹¹⁰ Lastly, we demonstrate that the incorporation of these improvements, along with additional improvements to sampling schemes, affords two improved algorithms, AbInitioVO and FastFloppyTail, which in concert allow for the accurate prediction of structural representations of proteins across all degrees of foldedness.

When considering a unified architecture, we noted that Rosetta-based algorithms such as AbInitio, designed for folded structures, and FloppyTail, designed for disordered structures, vary dramatically in their approaches.^{236, 238, 240} AbInitio utilized simulated annealing, coarse-graining and knowledge-based scoring terms to sample nine-residue (9-mer) and three-residue (3-mer) fragments culled from the PDB.²³⁸⁻²⁴⁰ With each simulation stage more score terms are added to the MC score function and the number of fragment insertions tested between score events is reduced. Lastly, the simulated structures are processed via the

Relax algorithm, which performs a series of gradient based minimizations with varied weightings of the repulsive van der Waals term.¹⁹⁰ This allows the protein to pulse and wiggle towards an energetically minimized conformation. By comparison, FloppyTail also utilized knowledge-based scoring and coarse-graining to perform initial 3-mer sampling.²³⁶ Following initial simulation under a coarse-grained representation, the structure is converted into an all atom representation and 3-mer sampling proceed under the “score12” all-atom score function used by Rosetta. Throughout, fragment insertion steps are coupled to gradient based minimization and, after the algorithm switches to an all-atom representation, sidechain rotamer sampling is also added.

§ 6.2 Methods

Computational Resources

All simulations were performed on the University of Pennsylvania School of Arts and Sciences General Purpose Cluster (GPC). Each of the four GPC compute nodes contains 24 cores at 2.6 Ghz, 256 GB RAM, 1 GbE Networking, 56 Gb Infiniband and 3.6 TB OS/Scratch storage. All times reported herein were from 140-residue α S. On average, Full-atom Generalized Simulations required ~3 hours per structure while simulations using exclusively Centroid course-graining required ~30 minutes per structure. Simulations using both course-grained and all-atom molecular representations (SimAnn) took ~ 1.5 hours per structure and the previously reported FloppyTail simulation method took ~30 minutes per structure whereas FastFloppyTail simulation method took ~ 3 minutes per structure. The

AbInitio and AbInitioVO algorithms both required ~ 45 minutes per structure and Relaxes took ~ 3 minutes a structure.

Generalized Simulation Format:

We chose to sample the effects of full-atom and centroid AtomTypes on the simulation, in an effort to assess the possibility of using the course-grained method to speed up the simulation during initial sampling steps as is frequently done in Rosetta. Additionally, we assessed the utility of different score terms and focused on six different score weights (VDW: exclusively repulsive van der Waals term "fa_rep", CenStd: "cen_std" score term weights, CenStdExt: "cen_std" score term weights adding "rama", "cenpack", "hbond_lr_bb" and "hbond_sr_bb" all at weights of 1.0, Beta: "beta_nov15" score term weights, SimAnn: simulated annealing where additional score terms were included as sampling progressed utilizing "score0", "score1", "score2", "score3" and "beta_nov15", CenNath: "vdw", "rama", "pair", "env", "hbond_lr_bb" and "hbond_sr_bb" all at weights of 1.0, and CenOpt: identical to the SimAnn centroid score functions with "rg" at a weight of 0 and "hbond_lr_bb" and "hbond_sr_bb" at weights of 1.0).⁶⁸ Finally, we utilized three different types of Movers within each of the different AtomType/ScoreFunction combinations (PP: ϕ/ψ torsion angle changes using Small/Shear Movers, FI: fragment insertion, and SC: side-chain rotamer optimization using the PackRotamersMover). The simulation nomenclature is in the form AtomType_Scoring_Sampling. For example, FA_Beta_PPF1 utilizes the full-atom AtomType, the "beta_nov15" scoring term weights for the ScoreFunction and

application of both phi/psi sampling and fragment insertion. To assess the impact of each of these parameters on the resultant ensemble, a base script was devised which could be easily altered to analyze variable effects while keeping the number of Movers applied to the protein backbone over the course of the simulation constant. In brief, the method consisted primarily of RandomMovers, which select and apply a single move from a detailed set of movers comprised of SmallMovers and ShearMovers (apply ϕ/ψ torsion angle changes), as well as MinMovers (perform gradient-descent minimization to locate local minima). Fragment insertion in the form of the ClassicFragmentMover was added directly to RandomMovers for both 9-mer and 3-mer fragments. Side-chain sampling was not added directly to the RandomMover, but was applied after a move from the RandomMover. The number of moves applied by the RandomMover prior to application of the Metropolis-Hastings acceptance criteria were decreased over the course of the simulation. The temperature (specified kT) value was also decreased. Lastly, the structure was set to the lowest energy structure found after each sampling stage and at the end of the simulation. For each simulation the resultant ensembles consisted of ~1000 structures.

Construction of Fragment Libraries:

The Robetta server was used to generate the initial fragment library for testing all variants of the “Generalized Simulation” along with “FloppyTail” and “FloppyTail_ref2015”.²²⁵ Production of the custom fragment libraries used were prepared using the FragmentPicker application in Rosetta.²⁴¹ The

"FloppyTail_Quota" library was prepared using secondary structure probability predictions from the primary sequence of α -synuclein (α S) using the Jufo, PsiPred and RaptorX servers.²⁴²⁻²⁴⁴ The "FloppyTail_Loop" library was prepared by exchanging the secondary structure predictions with a single manually crafted prediction input that contains 1.0 loop probabilities at every residue of the sequence.

All "FloppyTail" simulations not previously addressed, including all FastFloppyTail simulations of non- α S disordered proteins, utilized a quota-style fragment library using the same inputs as detailed above, where the disordered probability prediction was used to re-weight loop contributions as described in the main text. Additionally, AbInitio and AbInitioVO simulations were generated using the FragmentPicker following the best protocol, using only PsiPred predictions as inputs. AbInitioVO PsiPred predictions were loop reweighted as described in the main text. All fragment libraries contained 200 fragments.

§ 6.3 Calculation of Data from Ensembles

FRET Data:

FRET efficiencies (E_{FRET}) for each residue pair were computed for each individual structure and averaged across all structures in a given ensemble to determine the average E_{FRET} of the ensemble. This was performed through application of the Förster Equation:

$$E_{FRET} = \frac{1}{1+(r/R_0)^6} \quad (\text{Eqn. 6.1})^{57}$$

which is dependent on the interfluorophore distance, r , and the Förster distance (R_0). For expedience, we approximated the interfluorophore distance as the distance between the C α atoms of the labeled residues.

Distance Data

Distances for each residue pair were determined from the distances between the C α atoms of each residue for each structure in a given ensemble and were averaged over all members of the ensemble.

PRE Data:

Paramagnetic relaxation enhancement (PRE) values (Γ_2) were computed and converted to the experimentally observed value (I_{ox}/I_{red}) from each structure in the ensemble using the formulas:

$$\Gamma_2 = \left[\frac{K}{r^6} \left(4\tau_c + \frac{3\tau_c}{1+\omega^2\tau_c^2} \right) \right] \quad (\text{Eqn. 6.2})$$

$$\frac{I_{ox}}{I_{red}} = \frac{R2_{red} \exp(-\Gamma_2 t)}{R2_{red} + \Gamma_2} \quad (\text{Eqn. 6.3})$$

previously detailed by Sung *et al.*²⁰¹ Here, K is a constant that describes the spin properties for the nitroxide radical ($1.23 \times 10^{-32} \text{ cm}^6 \text{ s}^{-2}$), τ_c is the correlation time for the electron-nuclear interaction vector (4 ns) and $\omega/2\pi$ is the Larmour frequency of an amide proton (computed in all cases for a 700 MHz field), $R2_{red}$ is the transverse relaxation rate in the diamagnetic state (set to 4 s^{-1}) and t is the total INEPT evolution time of the HSQC experiment, which was 10 ms for comparisons

to data from Dedmon *et al.*¹⁰⁴ (as previously done by Piana *et al.*⁹⁷) and 4 ms for comparisons to data from Sung *et al.*²⁰¹

Chemical Shift Data:

The amide proton (H), amide nitrogen (N), carbonyl carbon (C), α -carbon ($C\alpha$) and β -carbon ($C\beta$) chemical shifts were all computed using the SPARTA+ package developed by Shen and Bax.²⁴⁵ The chemical shift values were computed for each structure and averages were computed uniformly across a given ensemble.

Residual Dipolar Couplings:

All residual dipolar couplings (RDCs) were computed using PALES.²⁴⁶ For each structure in a given ensemble, the RDC for a given residue was computed considering only a 15 residue segment where the residue of interest occupied the central position, as previously done by Piana *et al.*⁹⁷ For the first and last seven residues in a given protein, the N- and C-terminal 15 residue segments were used. The averages were computed for the 1000 lowest energy structures in each ensemble.

J-Couplings:

The NMR *J*-couplings were calculated using fitted Karplus equations previously utilized by Shen and Bax from backbone ϕ/ψ dihedral angles.²⁴⁷ For each structure, ϕ and ψ were computed in DSSP.²⁴⁸ As with other values, the *J*-

coupling was computed for each residue in each structure and averaged across all members of the ensemble. The Karplus equations used to compute each J -coupling are detailed below:

$${}^3J_{HN-HA} = 7.97 \cos^2(\varphi - 60^\circ) - 1.26 \cos(\varphi - 60^\circ) + 0.63 \quad (\text{Eqn 6.4})^{249}$$

$${}^1J_{C\alpha-H\alpha} = A_{RC} + 1.4 \sin(\psi + 138^\circ) - 4.1 \cos(2[\psi + 138^\circ]) + 1.7 \cos(2[\varphi + 60^\circ]) \quad (\text{Eqn 6.5})^{250}$$

$${}^1J_{C\alpha-N} = 9.5098 - 0.9799 \cos(\psi) + 1.7040 \cos^2(\psi) \quad (\text{Eqn 6.6})^{251}$$

$${}^2J_{C\alpha-N} = C - 1.5176 \cos(\psi) - 0.2047 \cos^2(\psi) \quad (\text{Eqn 6.7})^{247}$$

$${}^3J_{C'-C'} = 0.46 - 0.95 \cos(\psi) + 1.78 \cos^2(\psi) \quad (\text{Eqn 6.7})^{252}$$

For both the ${}^1J_{C\alpha-H\alpha}$ and the ${}^2J_{C\alpha-N}$ couplings, the value of the constants A_{RC} and C are amino acid dependent. For Val, Thr, Ile, and Ser $C = 7.65$ while for all other amino acids, $C = 8.15$.²⁴⁷ The constant A_{RC} values are listed in Table S1 below and originate from the random coil ${}^1J_{C\alpha-H\alpha}$ values.²⁵⁰

Table 6.1: Values for Computing J-Couplings.

Residue	A_{RC}	Residue	A_{RC}	Residue	A_{RC}	Residue	A_{RC}
Ala	143.7	Glu	141.9	Met	142.2	Trp	143.0
Arg	141.5	His	143.8	Phe	142.9	Tyr	143.0
Asn	141.5	Ile	141.3	Pro	148.4	Val	141.3
Asp	142.5	Leu	141.1	Ser	142.1	Other	140.3
Gln	141.1	Lys	141.5	Thr	141.4		

§ 6.3 Results

To determine the effective capacity of each method, structural predictions were performed using the *AbInitio* and *FloppyTail* methods, and were compared to 1) Förster resonance energy transfer (FRET) data from four different probe pairs, 2) inter-residue distances, determined from FRET, 3) electron transfer experiments, and 4) paramagnetic relaxation enhancement (PRE) measurements, to assess global ensemble accuracy.^{51, 53, 68, 104, 201, 253-254} Furthermore, the accuracy of residue level information for α S ensembles was analyzed through comparison to 5) chemical shift (δ) data, 6) residual dipolar couplings (RDCs) and 7) NMR J-couplings.^{201, 247}

Of the two methods, *FloppyTail* significantly outperformed the *AbInitio* in generating accurate disordered ensembles of α S, as might be expected (referred to as *FloppyTail_score12* and *AbInitio*, respectively, Tables 6.2-4). The *FloppyTail* ensemble demonstrated an impressive degree of agreement with

global descriptors of the proteins overall topology including radius of gyration (R_g), EFRET, distance and PRE data (Figs. 6.3b, 6.4b, 6.5b and 6.22). However, upon comparison to residue-level information, we noted that the predicted C and C β chemical shifts (Fig. 6.52) as well as $^2J_{NC\alpha}$ J-couplings (Fig. 6.84) deviate from the experimental data (Tables 6.3-4). These deviations are likely attributable to overpopulations of helical structure and is clearly observed when plotting per residue helical propensity across all members of the resultant ensemble (Fig. 6.2b). Indeed, overpopulations of helical architectures was not isolated to the FloppyTail simulation and was more pronounced in the AbInitio simulated ensemble (Fig. 6.2b). In contrast with the FloppyTail output, the AbInitio algorithm also presented severe discrepancies in the overall R_g and other global parameters compared to experimental values (Figs 6.3b, 6.4b, 6.5b and 6.33). Overall, both methods overpopulate helical architecture, while the AbInitio method additionally produces overcompacted structures. In comparison to other simulation approaches previously explored in the literature, overcompaction and overpopulation of helices are common problems in both MD and MC-based methods which have been optimized for ordered proteins.^{97, 99-100, 109-110}

Employment of a Generalized Simulation Format

To overcome these barriers, a new method was drafted in PyRosetta allowing for a comparison of 17 different combinations of score functions, sampling schemes and atomic representations.¹¹⁴ PyRosetta is a python-wrapped version of Rosetta, that allows for new algorithm development in a relatively easy coding

language without the need to compile. The nomenclature and specifics of each simulation are detailed in the Supporting Information. In short, the format consists of six sampling stages where the temperature used for assessing the Metropolis criteria is decreased at each stage. Sampling consists of ϕ/ψ backbone torsion angle sampling, fragment insertion and/or sidechain rotamer sampling (indicated by PP, FI and SC suffixes in the generalized simulation names respectively). Additionally, simulations were tested under different score function and atomic representations utilizing the standard centroid scoring procedure in Rosetta, the standard full-atom score function in Rosetta, or the scoring procedure utilized by AbInitio (*CenStd*, *Beta*, and *SimAnn* prefixes respectively).^{194, 238-240} Additionally, polymer-like self-avoid walk simulations were also performed where the score function was exclusively comprised of a repulsive van der Waals energy term. (*VDW* prefix).¹⁹⁴

We observed that over-compactness was most severe when using the centroid coarse-grained representation (*CenStd_PP* and *CenStd_PPFI* simulations, Table 6.2, Figs. 6.3a, 6.4a, 6.5a, 6.10 and 6.11). The “centroid” coarse-graining approach represents the backbone atoms explicitly, while reducing sidechain atoms to a single pseudo-atom located on the gamma-carbon, with appropriate parameters to approximate the side-chain.²⁴⁰ This representation utilizes several knowledge-based score terms in place of the physics-based energy terms which are utilized in the explicit, all-atom representations in Rosetta.¹⁹⁴ Although we observed overcompactness when employing centroid coarse graining, we observe that either subsequent all-atom sampling (all *SimAnn* simulations,

Figs. 6.15-18) or sampling utilizing the specific centroid score function utilized by the FloppyTail algorithm (all *CenStd_Ext* simulations, Figs. 6.12-13) alleviated compaction (Table 6.2 and Figs. 6.3a, 6.4a and 6.5a).²³⁶ Therefore, we were able to determine that the “rg” score term, which is excluded from the score terms utilized in the centroid phase of the FloppyTail algorithm, was the likely source of overcompaction.

Unlike overcompaction, overpopulation of helical structures appeared to occur independent of the atomic representation or score function and was solely an artifact of sampling. Indeed, overpopulation occurred in all cases where fragments were sampled (all *PPFI* and *PPFISC* simulations, chemical shifts: Figs. 6.37, 6.38, 6.41, 6.43, 6.46 and 6.47 J-couplings: Figs. 6.68, 6.69, 6.72, 6.74 and 6.78) except when the protein was reduced to self-avoiding polymer (*VDW_PPFI* and *VDW_PPFISC* simulations, chemical shifts: Figs. 6.50 and 6.51 and J-couplings: Figs. 6.81 and 6.82) as observed in Figure 6.2. Therefore, we hypothesized that fragment selection, as opposed to the fragment sampling condition, was the likely source of error.

Lastly, we observed that sidechain sampling had a relatively minimal impact in the absence of fragment sampling. Inspecting the all-atom simulations and simulated annealing simulations (*Beta*, *SimAnn* and *VDW* simulations) we observe that the impact of sidechain sampling is relatively small compared to the impact of sampling fragments (Tables 6.2-4). Since sidechain packing in Rosetta accounts for a significant portion of the simulation time, we rationalize that simulation speed might be enhanced by reducing the frequency of sidechain sampling events.

Increased Accuracy through Improved Fragment Selection

Fragment libraries in Rosetta are typically generated from 3-state secondary structure predictions based on protein primary sequences.²⁴¹ For the initial assessment of the performance of FloppyTail, a fragment library generated from the Robetta server was employed.²²⁵ We further corroborated our hypothesis that the observed overpopulation of helical structures originated from sampling, and not scoring, by confirming that the FloppyTail algorithm yielded a similar output when using an up-to-date all-atom score function, “ref2015”.¹⁹⁴ This was indeed the case with no apparent differences in comparisons to all experimental data (*FloppyTail_ref2015* simulation, Figs. 6.2b, 6.3b, 6.4b, 6.5b, 6.23, 6.53, 6.66d and 6.84) Therefore, we wanted to assess the impact of sampling a broader array of input fragments through manual construction of a new fragment library. To do this we employed the FragmentPicker protocol.²⁴¹ This method utilizes secondary structure inputs from servers such as PsiPred, Jufo and RaptorX, which predict the probability of each residue existing in a helix, sheet or loop for a supplied sequence.²⁴²⁻²⁴⁴ Traditionally, structural predictions in Rosetta utilize a fragment selection scheme where only fragments that match the highest probability architecture for each residue are chosen.²⁴¹ This is referred to as a “best” protocol, which can be exchanged for a “quota” fragment selection protocol, which probabilistically collects fragments such that the frequency with which each secondary structural element is sampled for a given residues matches the probabilities in the input prediction. This library, though sampling a broader array

of secondary structural elements, provided a very minor improvement (Tables 6.2-4) and still presented regions with a significant helical content (*FloppyTail_Quota* simulation, Figs. 6.2b, 6.3b, 6.4b, 6.5b, 6.24, 6.54, 6.66d and 6.85). This was unsurprising as all three predictions used contained a significant helical probability corresponding to the region of α S that organizes into a helices when bound to membranes.²³⁻²⁴ Therefore, we attempted to remove all fragment sampling from the protocol (*FloppyTail_NoFrag*s simulation), exclusively employing pseudo-random ϕ/ψ torsional sampling. The resulting simulation provided very good agreement with all data (Tables 6.2-4, Figs. 6.2b, 6.3b, 6.4b, 6.5b, 6.25, 6.55, 6.66d and 6.86), however we suspected that this approach would be insufficient for IDPs which transit more frequently between folded and unfolded architectures.

Recent studies by Bax and coworkers demonstrated that loops from PDB structures populated a similar Ramachandran backbone dihedral space to that of IDPs.²⁵⁵ Therefore, we hypothesized that a fragment library composed exclusively of loops might allow us to re-introduce fragment insertion. Replacement of the initial fragment library with an all-loop library in the FloppyTail algorithm resulted in an ensemble which was devoid of extended secondary structure (*FloppyTail_Loops* simulation, Tables 6.2-4, Figs. 6.2b, 6.3b, 6.4b, 6.5b, 6.26, 6.56, 6.66e and 6.87). Although this fragment library was sufficient to demonstrate that fragment sampling can be employed in an efficacious manner, this strategy for fragment library construction would require prior knowledge of a protein's degree of order. Therefore, we sought to identify a surrogate for this knowledge, identifying disordered probability predictions from primary sequence as a

potentially useful candidate. Along with secondary structure predictions, many servers provide disordered probability predictions.²⁴³⁻²⁴⁴ Although many servers incorrectly characterized α S as being probabilistically ordered in at least a single stretch of residues, the RaptorX server provided a correct prediction that the protein was entirely disordered.²⁴³⁻²⁴⁴ Therefore, we envisioned that disordered probability predictions from RaptorX could be used to reweight the secondary structure predictions by adjusting the loop probability as follows:

$$P'_{res}(loop) = P_{res}(disorder) \quad (\text{Eq. 6.9})$$

$$P'_{res}(sheet) = (1 - P_{res}(disorder)) \times P_{res}(sheet) \quad (\text{Eq. 6.10})$$

$$P'_{res}(helix) = (1 - P_{res}(disorder)) \times P_{res}(helix) \quad (\text{Eq. 6.11})$$

Here, $P_{res}(disorder)$ represents the predicted probability of a given residue to be disordered, while $P_{res}(loop)$, $P_{res}(helix)$, $P_{res}(sheet)$ represent the probability of the residue being a loop, helix or sheet. Additionally, primed and non-primed variables represent the reweighted and original predictions respectively. Gratifyingly, the resultant fragment library constructed using these reweighted loop predictions as inputs for the previously described “quota” fragment library construction approach resulted in a highly accurate FloppyTail output ensemble (*FloppyTail* simulation, Tables 6.2-4, Figs. 6.2b, 6.3b, 6.4b, 6.5b, 6.27, 6.57, 6.66e and 6.88) comprised of very few helices.

Acceleration and Improved Accuracy via Adjustment to Sampling

As previously stated, results from the generalized simulation suggest that a reduction of sidechain sampling could afford an accelerated simulation with comparable accuracy. Therefore, we removed sidechain sampling from one of the interior loops of the FloppyTail protocol, reducing the sampling frequency 15-fold. This resulted in a substantial ~10-fold reduction in compute time and produced an ensemble (*FloppyTail_Rot* simulation, Tables 6.2-4, Figs. 6.2b, 6.3b, 6.4b, 6.5b, 6.29, 6.61, 6.66f and 6.92) that was nearly identical across all experimental comparisons. Additionally, unlike a traditional MC approach, at several stages of the FloppyTail simulation, the structure is returned to the lowest energy structure encountered in the search up until that point.²³⁶ This helps to restrict sampling to states near minima, while allowing limited sampling following Metropolis criteria acceptance to avoid trapping in local minima. Serendipitously, we discovered that reducing the amount of sampling between instances where the structure is returned to the lowest energy conformer further improved agreement with the experimental data, resulting in the algorithm now called FastFloppyTail (Tables 6.2-4, Figs. 6.2b, 6.3b, 6.4b, 6.5b, 6.31, 6.59, 6.66e and 6.90). Finally, we employed the Relax algorithm, commonly used after the AbIntio algorithm, to further minimize outputs from the FastFloppyTail simulation, and again were met with an additional improvement in the overall agreement with experimental data (Tables 6.2-4, Figs. 6.2b, 6.3b, 6.4b and 6.5b, *FloppyTail_Relax*: 6.28, 6.58, 6.66e and 6.89, *FloppyTail_Rot_Relax*: 6.30, 6.62, 6.66f, 6.93, and *FastFloppyTail_Relax*: 6.32, 6.60, 6.66f, 6.91.^{194, 238-239} This final coupled

algorithm, FastFloppyTail-Relax, produces α S ensembles with comparable agreement with ensembles generated from state-of-the-art MD (Tables 6.2-4).

Re-defining Knowledge Based Terms

Although improved fragment selection was sufficient for delivering the improved FastFloppyTail method for generating disordered protein structural ensembles, the reliance of knowledge-based terms for the prediction of folded proteins required additional attention. The final stages of the AbInitio simulation is reliant on the “rg” score term, whose energetic penalty is equal to the R_g of the structure.²⁴⁰ Although this facilitates the compaction necessary for the adoption of well packed folded proteins, this produces severe over competition in IDPs and IDRs. Therefore, we propose replacement of this term with a score term that compares the R_g of the current simulated protein structure to that expected from polymer-scaling laws. Employing the following equations from Schuler and co-workers, we can predict an expected R_g from the hydrophobic or charge content of a protein from a given sequence:

$$\overline{R_g} = \sqrt{\frac{2l_p^*b}{(2v+1)(2v+2)}} N^v \quad (\text{Eq. 6.12})^{256}$$

Here, $\overline{R_g}$ is the predicted mean R_g value determined from the persistence length ($l_p^* = 0.53$ nm) and the average distance between two C α atoms ($b = 0.38$ nm), which have been previously determined, along with the number of residues

in the sequence, N , and the scaling factor, ν . The polymer scaling term, ν , can be computed using Eqs. 6.13 or 6.14 depending on whether net charge or hydrophobicity serves as the dominating characteristic.

$$\nu(Q) = \frac{1}{3} + a \left[1 + \frac{\exp(x_0 - Q)}{z} \right]^{-1} \quad (\text{Eq. 6.13})^{256}$$

$$\nu(H) = \frac{1}{3} + a \left[1 + \frac{\exp(x_0 + cH - d)}{z} \right]^{-1} \quad (\text{Eq. 6.14})^{256}$$

In Eqs. 6.13 and 6.14, constants have been previously determined through fitting to experimental data ($a = 0.394$, $z = 0.09$, $x_0 = 0.114$, $c = 1.72$ and $d = 0.9$) while variable Q and H represent the net charge and the mean hydrophobicity respectively.²⁵⁶ To select between the scaling factors computed by the net charge and hydrophobicity when computing an expected radius of gyration, the polyampholyte theory was applied to determine the effect of the net charge on the excluded volume of the system by computing the excess volume:

$$\nu^* = \frac{4\pi l_B (f-g)^2}{\kappa^2} - \frac{\pi l_B^2 (f+g)^2}{\kappa} \quad (\text{Eq. 6.15})^{256}$$

Here, f and g are the fraction of positively (Arg/Lys) and negatively (Glu/Asp) charged residues, respectively, κ is the Debye length defined by $\kappa^{-1} = 0.304/\sqrt{I}$, where I is the ionic strength and l_B is the Bjerrum length defined by:

$$l_B = e^2 / (4\pi\epsilon_0\epsilon_r k_B T) \quad (\text{Eq. 6.16})^{256}$$

In the above equation, e is the elementary charge (1.602×10^{-19}), ϵ_0 is the dielectric constant (8.854×10^{-12}), ϵ_r is the permittivity of water (78.7) and k_B is the Boltzmann constant. In the general case for the calculation of v^* , the ionic strength and the temperature are set to 0.15 M and 298 K respectively. Since values of v^* greater than zero correspond to net electrostatic repulsion, when v^* is greater than zero, the scaling term computed using the net charge is applied when predicting the radius of gyration. Conversely, when v^* is less than zero, there is a net electrostatic attraction, forcing the scaling based on the hydrophobicity to dominate. In cases where both the fractions of positively and negatively charged residues are zero, the scaling is determined using the net hydrophobicity. Lastly, in the cases where either the fraction of positively or negatively charged residues equals zero, the scaling due to the net charge is selected.

Upon further inspection of the above method proposed by Hofmann et al., we noted that when testing this function for lysozyme (PDB: 2LZM) the predicted scaling value ($v = 0.519$) is significantly higher than the expected value for a folded protein.²⁵⁶ Since the fitting was performed on denatured proteins, the work by Hofmann et al. shows that when determining the radius of gyration for a suite of proteins from the Protein Data Bank, the minimum observed scaling factor is not equal to 0.33, which is the value for well-folded proteins, but instead is ~ 0.4 .²⁵⁶ Therefore, we utilized the per residue disordered probability of the protein to

determine if a given segment of the protein of interest was folded or unfolded. For predictions by the RaptorX server we used the suggested cutoff, where residues with a disordered probability >0.5 were determined to be disordered.²⁴⁴ Segments were defined as any stretch of residues greater than 10 in which all residues were of the same classification, ordered or disordered. Segments deemed ordered were assigned a scaling value of $v = 0.33$ while the scaling value for disordered regions was computed using the above equations. Therefore, we intend to predict R_g values for each ordered/disordered segment and compute the associated score on a per segment bases, with the overall score simply representing the sum of the scores across all segments.

To incorporate the predicted R_g within an energy term we elected to craft a potential from a general version of a self-avoiding walk (SAW) probability distribution, previously employed by Zheng, Best, Schuler and coworkers, that accommodates changes in the scaling exponent.⁶⁵

$$P(r) = A \frac{4\pi}{R_g} \left(\frac{r_g}{R_g}\right)^{2+g} \exp\left[-\alpha \left(\frac{r_g}{R_g}\right)^\delta\right] \quad (\text{Eq. 6.17})^{65}$$

$$g = (\gamma - 1)/v \text{ and } \delta = 1/(1 - v) \quad (\text{Eq. 6.18 and 6.19})^{65}$$

Above, r_g represents the R_g of the structure being assess by the scoring function and again $\overline{R_g}$ represents the predicted mean R_g . The constants g and δ in Eq. 6.17 are defined in the subsequent equations Eq. 6.18 and 6.19. The constant

γ (1.1615) has been previously estimated for proteins, while the constants A and α are determined for a given \overline{R}_g and v pair based on the normalization conditions:

$$\int_0^{\infty} P(r_g) dr_g = 1 \text{ and } \int_0^{\infty} P(r_g) r_g^2 dr_g = \overline{R}_g^2 \quad (\text{Eq. 6.20 and 6.21})^{65}$$

This distribution provides a broad distribution of radii of gyration while at the same time preventing structures from becoming overly compact. The per segment energy potential is defined as:

$$E_{\overline{R}_g}(r_g) = 1 - P(r_g)/\max[P(r_g)] \quad (\text{Eq. 6.22})^{65}$$

Through this form, the energy value scales from zero to one for any average R_g allowing the weight of this score term within our overall score function to determine the depth of the potential. To determine the optimal weight of the new R_g score term, α S simulations were performed under the AbInitio protocol using a range of weighting values. We identified the optimal weighting value as that which maximized the impact of the score term on the average radius of gyration while minimizing the restriction on the conformational diversity (Fig. 6.1). Combination of this novel scoring approach with the previously described fragment selection strategy results in the new simulation termed AbInitioVO, or AbInitio Variable Order.

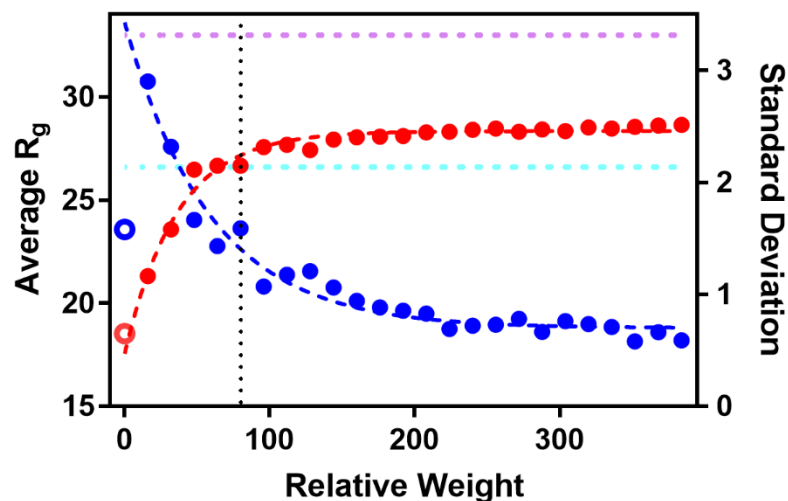


Figure 6.1: Determination of optimal new R_g score term weight. Plot shows the average (filled blue circles) and standard deviation (filled red circles) radius of gyration resultant from simulations with various relative weights of the new R_g score term compared to the simulation lacking the traditional rg term (open red and blue circles) and experimental data from SAXS²²⁸ (pink dashed line) and NMR²⁰⁵ (cyan dashed line). The vertical black dashed line illustrates the chosen optimal score term weight, as this pushes the average R_g closest to the experimental value without over-constraining the sampling. Therefore, the weights of all terms in the new score3 term is equivalent to the traditional score3 term, with the canonical rg term set to weight a zero and the weight of this new term set to 84.

Accurate De Novo Prediction of Disordered and Ordered Proteins

To determine whether this modified AbInitioVO protocol outperforms the original AbInitio method, a set of 25 proteins were selected containing a variety of different secondary structural elements and spanning varying levels of order/disorder. For each AbInitioVO simulation, disordered probability predictions were performed based on the input sequence using RaptorX and fragment libraries were assembled using a best fragment selection protocol based on a loop-reweighted secondary structure prediction from PsiPred, using Eqs. 6.9-6.11. The resultant structural predictions from AbInitioVO were compared to structures

generated using the standard AbInitio protocol (Table 6.5). For ordered and partially-ordered proteins, prediction accuracy was assessed by comparing the resultant structures to x-ray crystal structures and NMR structures deposited in the PDB by computing C α -RMSDs. Disordered proteins were inspected for agreement with the experimentally determined R $_g$ of each protein (Table 6.5). Overall, we observed that across all ordered and partially-ordered proteins tested, that AbInitioVO performs comparably to AbInitio for the prediction of folded regions across all secondary structures (Table 6.5, Folding Funnels: Figs. 6.100-117, Structural Overlays: Figs. 6.124-141). Although there are some cases where AbInitioVO appears to be outperformed by AbInitio (ex: 1b3a, 1bq9, 1ejf, 1lwm) there are at least as many cases where AbInitioVO outperforms AbInitio (ex: 1shf, 1bk2, 1ghc, 1ubi). For partially-ordered proteins, as expected, the average R $_g$ from the AbInitioVO simulations is larger than the for AbInitio simulations (Fig 6.98). Curiously, we observed that ordered proteins were on average slightly more compact in AbInitioVO compared to AbInitio (Fig. 6.97). Lastly, we find that the resultant average R $_g$ values for disordered protein from AbInitioVO are more in-line with the experimental R $_g$ s than those from AbInitio (Table 6.5, Fig 6.99). These results support the notion that AbInitioVO is able to correctly predict the folded portions of ordered and partially-ordered proteins, while identifying disordered regions with the correct secondary structure and overall size. These disordered regions can then be simulated using the FastFloppyTail approach to create accurate structural ensemble of disordered and partially-ordered proteins.

Table 6.2: Comparison of Simulated α S Ensembles to Global Experimental Data.

Name	R_g (Å)	E_{FRET} RMSD	I_o/I_r RMSD	Dist. RMSD
Generalized Simulation				
VDW_PP	43.7 + 8.0	0.18	0.23	17.5
VDW_PPF1	39.7 + 6.8	0.15	0.21	13.17
VDW_PPFISC	40.0 + 7.0	0.15	0.21	14.0
CenStd_PP	19.6 + 2.5	0.28	0.20	11.9
CenStd_PPF1	19.2 + 2.5	0.28	0.22	12.5
CenNath	33.4 + 7.9	0.11	0.18	8.09
CenStd_Ext_PP	32.7 + 8.7	0.11	0.17	7.02
CenStd_Ext_PPF1	25.6 + 5.6	0.17	0.16	6.69
Beta_PP	30.6 + 5.5	0.13	0.17	8.23
Beta_PPF1	36.4 + 6.9	0.12	0.19	11.7
Beta_PPFISC	40.5 + 6.7	0.18	0.20	15.2
Beta_PPSC	30.3 + 5.3	0.13	0.17	7.74
SimAnn_PP	33.9 + 7.0	0.10	0.18	8.54
SimAnn_PPF1	33.4 + 6.8	0.11	0.18	9.02
SimAnn_PPFISC	25.5 + 4.0	0.18	0.17	8.14
SimAnn_PPSC	31.7 + 6.3	0.12	0.17	7.38
FloppyTail				
FloppyTail_score12	32.8 + 6.0	0.12	0.18	8.36
FloppyTail_ref2015	33.9 + 6.4	0.12	0.18	9.39
FloppyTail_Quota	34.2 + 6.1	0.12	0.18	9.02
FloppyTail_Loops	35.3 + 6.7	0.12	0.18	9.75
FloppyTail_NoFrag	34.8 + 6.3	0.12	0.18	9.00
FloppyTail-Relax	27.7 + 8.1	0.15	0.15	7.03
FloppyTail	34.0 + 7.7	0.11	0.18	7.94
FastFloppyTail-Relax	31.0 + 9.0	0.11	0.17	5.37
FastFloppyTail	38.0 + 9.4	0.12	0.20	10.1
FloppyTail_Rot	37.2 + 8.6	0.12	0.19	10.4
FloppyTail_Rot-Relax	29.2 + 8.7	0.13	0.15	6.25
AbInitio				
DeNovoIDP	28.4 + 5.1	0.15	0.17	9.63
AbInitio	17.3 + 1.0	0.32	0.31	12.2
AbInitioVO	22.7 + 2.6	0.24	0.22	12.9
Robustelli et. al.⁹⁹				
a99SB-disp	36.73	-	0.17	-

*Atom type abbreviations: CEN = centroid, FA = full-atom.

** R_g values can be compared to experimentally determined values of 33.0 ± 3.0 Å and 26.6 ± 0.5 Å from SAXS²²⁸ and NMR²⁰⁵ data, respectively. E_{FRET} RMSD values were computed from data from Ferrie *et. al.*, Ferrie *et. al.* and Nath *et. al.*^{51, 53, 68} Distance RMSD values were computed from Grupi *et. al.* and Lee *et al.*²⁵³⁻²⁵⁴ PRE RMSD values were computed from data from Sung *et. al.* and Dedmon *et. al.*^{104, 201}

Table 6.3: Comparison of Simulated α S Ensembles to Chemical Shift and RDC Data.

Simulation Name	N	H	C	C α	C β	All CS	RDC D.E	RDC M.Z.
Generalized Simulation								
VDW_PP	1.19	0.88	0.84	0.76	1.05	0.87	0.65	0.55
VDW_PPFI	1.52	0.17	0.55	0.65	1.06	0.91	0.63	0.55
VDW_PPFI SC	1.83	0.19	0.49	0.50	1.08	1.01	0.61	0.52
CenStd_PP	3.42	0.25	0.70	0.65	1.28	1.69	0.64	0.55
CenStd_PPFI	3.66	0.20	0.66	1.07	1.31	1.83	0.60	0.44
CenNath	2.10	0.22	0.40	0.50	1.37	1.16	0.64	0.51
CenStd_Ext_PP	2.07	0.22	0.40	0.49	1.35	1.15	0.64	0.51
CenStd_Ext_PPFI	3.60	0.28	1.28	1.52	1.42	1.95	0.61	0.36
Beta_PP	0.95	1.06	0.57	2.67	0.14	1.38	0.64	0.54
Beta_PPFI	3.37	0.31	1.18	1.58	1.18	1.83	0.62	0.39
Beta_PPFI SC	2.93	0.26	1.68	2.60	1.23	1.99	0.61	0.32
Beta_PPSC	1.59	0.13	1.00	0.41	1.11	0.99	0.66	0.57
SimAnn_PP	2.74	0.14	0.65	0.49	1.06	1.37	0.63	0.52
SimAnn_PPFI	3.63	0.31	0.83	1.10	1.14	1.81	0.60	0.43
SimAnn_PPFI SC	3.10	0.27	1.20	1.81	1.08	1.77	0.59	0.37
SimAnn_PPSC	2.17	0.12	0.64	0.43	1.03	1.13	0.63	0.51
FloppyTail								
FloppyTail_score12	2.03	0.18	0.39	0.58	1.02	1.07	0.63	0.48
FloppyTail_ref2015	1.93	0.17	0.39	0.58	1.02	1.02	0.63	0.48
FloppyTail_Quota	1.75	0.18	0.38	0.46	1.03	0.95	0.64	0.49
FloppyTail_Loops	1.52	0.13	0.40	0.39	1.02	0.86	0.65	0.55
FloppyTail_NoFrag s	1.52	0.13	0.41	0.38	1.01	0.86	0.64	0.55
FloppyTail-Relax	1.79	0.13	0.38	0.34	0.98	0.94	0.63	0.44
FloppyTail	1.69	0.13	0.36	0.35	0.98	0.90	0.62	0.44
FastFloppyTail-Relax	1.62	0.11	0.47	0.36	1.02	0.90	0.63	0.42
FastFloppyTail	1.55	0.12	0.42	0.33	1.01	0.86	0.62	0.41
FloppyTail_Rot	1.66	0.13	0.36	0.33	0.98	0.89	0.62	0.44
FloppyTail_Rot-Relax	1.77	0.12	0.38	0.34	0.97	0.93	0.63	0.44
AbInitio								
DeNovoIDP	1.67	0.22	0.39	0.48	1.04	0.93	0.67	0.52
AbInitio	2.68	0.25	1.37	2.08	1.23	1.73	0.60	0.38
AbInitioVO	2.24	0.24	0.61	0.72	1.03	1.19	0.63	0.47
Robustelli et. al.								
a99SB-disp	1.46	0.14	0.31	0.51	1.04	0.85	-	0.41

*All chemical shift data are reported as RMSD values from computed from data from Sung *et al.*²⁰¹ RDC values are computed as Q-values, as described in Zweckstetter *et. al.*²⁴⁶, based on data from Sung *et. al.* (RDC D.E.)²⁰¹ and Bertocini *et. al.* (RDC M.Z.)²⁰⁰

Table 6.4: Comparison of Simulated α S Ensembles to J-Coupling Data

Simulation Name	$^3J_{\text{HNH}\alpha}$	$^1J_{\text{CaH}\alpha}$	$^1J_{\text{NC}\alpha}$	$^2J_{\text{NC}\alpha}$	$^3J_{\text{C}'\text{C}'}$	All J
Generalized Simulation						
VDW_PP	2.38	1.31	0.63	0.42	0.69	1.30
VDW_PPF1	2.63	1.10	0.72	0.74	0.57	1.38
VDW_PPFISC	2.39	1.51	0.94	0.93	0.52	1.42
CenStd_PP	1.41	1.37	0.66	0.46	0.61	0.99
CenStd_PPF1	1.21	1.68	0.98	1.25	0.14	1.17
CenNath	1.67	1.46	0.25	0.45	0.41	1.03
CenStd_Ext_PP	1.61	1.43	0.25	0.45	0.39	1.00
CenStd_Ext_PPF1	1.85	3.15	0.89	1.55	0.42	1.83
Beta_PP	0.89	1.39	0.43	0.40	0.73	0.85
Beta_PPF1	1.22	2.81	0.91	1.29	0.38	1.55
Beta_PPFISC	2.41	4.31	1.15	1.48	0.40	2.37
Beta_PPSC	0.82	1.28	0.38	0.40	0.74	0.79
SimAnn_PP	0.67	1.68	0.55	0.45	0.58	0.91
SimAnn_PPF1	0.84	2.08	0.91	1.37	0.24	1.25
SimAnn_PPFISC	1.50	3.21	0.99	1.39	0.34	1.77
SimAnn_PPSC	0.43	1.39	0.60	0.49	0.50	0.76
FloppyTail						
FloppyTail_score12	0.48	1.64	0.54	0.70	0.16	0.88
FloppyTail_ref2015	0.52	1.68	0.52	0.79	0.17	0.88
FloppyTail_Quota	0.62	1.55	0.49	0.74	0.19	0.85
FloppyTail_Loops	0.51	1.10	0.35	0.43	0.10	0.60
FloppyTail_NoFrag	0.53	1.11	0.35	0.43	0.10	0.60
FloppyTail-Relax	0.59	1.25	0.41	0.54	0.16	0.69
FloppyTail	0.64	1.45	0.40	0.53	0.16	0.77
FastFloppyTail-Relax	0.65	1.14	0.36	0.39	0.26	0.64
FastFloppyTail	0.56	1.22	0.35	0.39	0.24	0.65
FloppyTail_Rot	0.61	1.41	0.40	0.51	0.16	0.75
FloppyTail_Rot-Relax	0.60	1.23	0.40	0.52	0.17	0.68
AbInitio						
DeNovoIDP	0.74	1.55	0.72	0.98	0.28	0.95
AbInitio	2.15	3.49	0.99	1.32	0.37	1.98
AbInitioVO	1.16	1.82	0.71	1.12	0.25	1.13
Robustelli et. al						
a99SB-disp	1.11	-	-	-	0.18	-

*All data are reported as RMSD values computed from data from Mantsyzov *et al.* and Lee *et al.*^{247, 252}

Analysis of Per Residue Percent Helicity for α -Synuclein Simulations

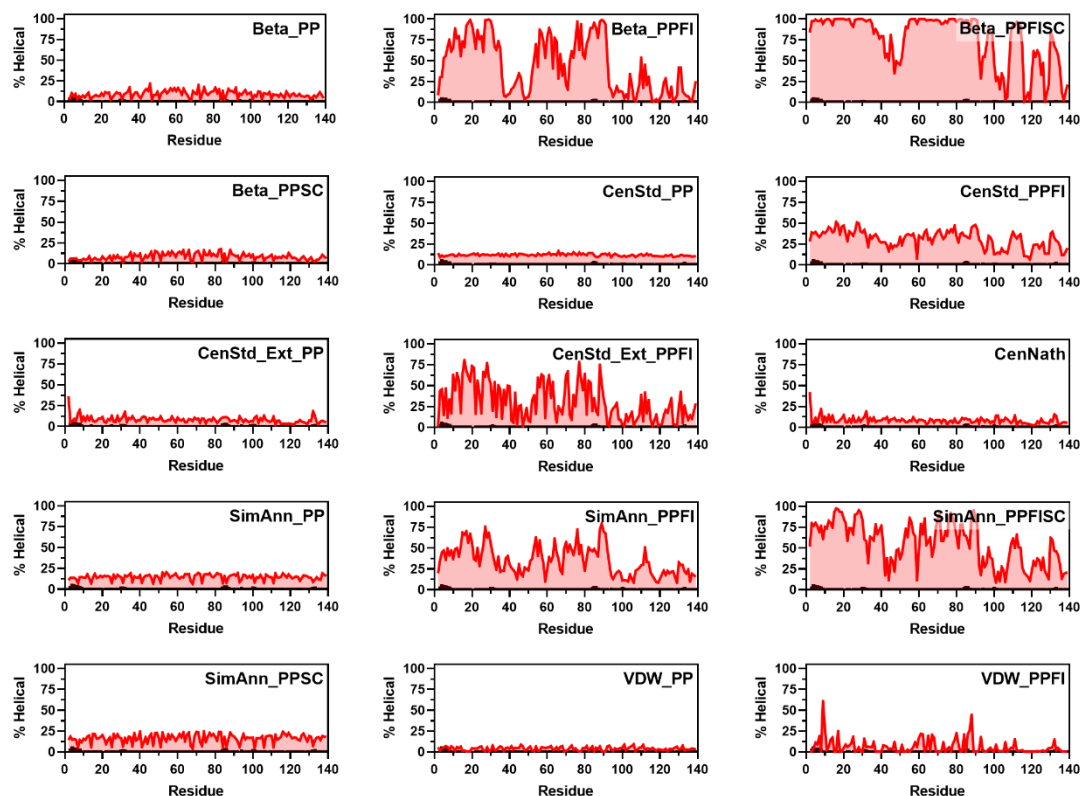


Figure 6.2a: Plots of Percent Helicity. Plots of the percentage of structures containing helices on a per residue basis from Beta_PP (Row 1 Left), Beta_PPF1 (Row 1 Middle), Beta_PPFISC (Row 1 Right), Beta_PPSC (Row 2 Left), CenStd_PP (Row 2 Middle), CenStd_PPF1 (Row 2 Right), CenStd_Ext_PP (Row 3 Left), CenStd_Ext_PPF1 (Row 3 Middle), CenNath (Row 3 Right), SimAnn_PP (Row 4 Left), SimAnn_PPF1 (Row 4 Middle), SimAnn_PPFISC (Row 4 Right), SimAnn_PPSC (Row 5 Left), VDW_PP (Row 5 Middle) and VDW_PPF1 (Row 5 Right) ensembles compared to percentages computed from Sung et. al. chemical shift data using the D2D method from Camilloni et. al.^{201, 257}

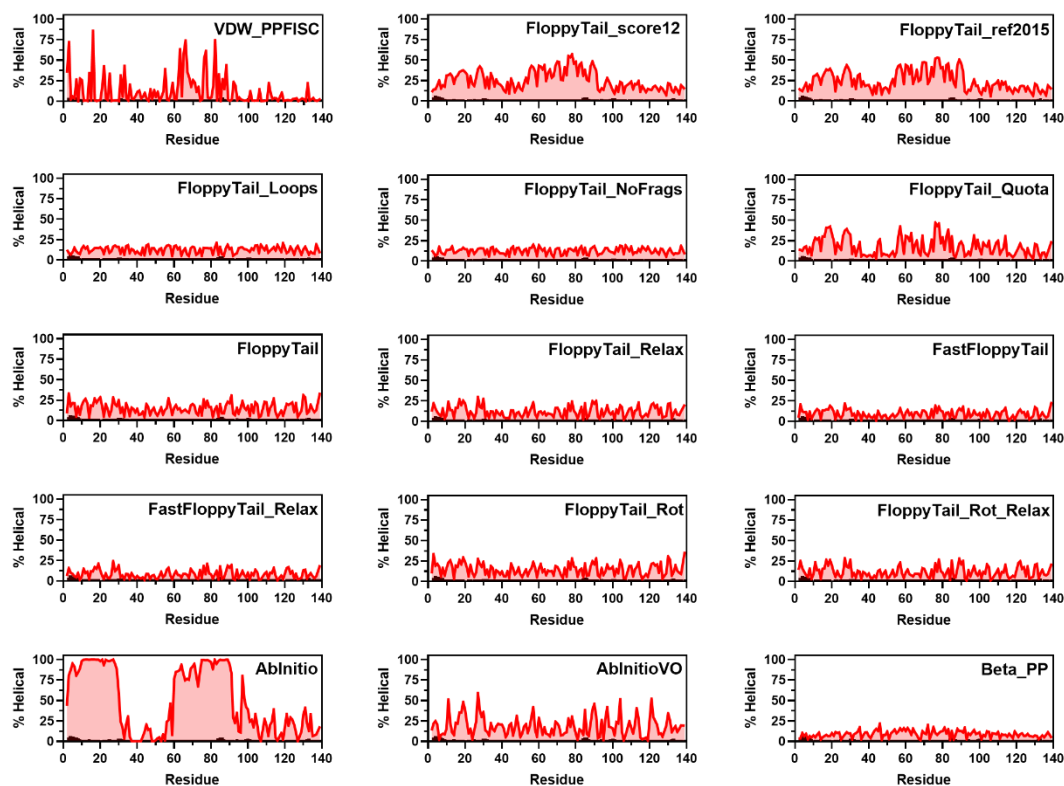


Figure 6.2b: Plots of Percent Helicity. Plots of the percentage of structures containing helices on a per residue basis from VDW_PPFISC (Row 1 Left), FloppyTail_score12 (Row 1 Middle), FloppyTail_ref2015 (Row 1 Right), FloppyTail_Quota (Row 2 Left), FloppyTail_NoFrag (Row 2 Middle), FloppyTail_Loops (Row 2 Right), FloppyTail (Row 3 Left), FloppyTail_Relax (Row 3 Middle), FastFloppyTail (Row 3 Right), FastFloppyTail_Relax (Row 4 Left), FloppyTail_Rot (Row 4 Middle), FloppyTail_Rot_Relax (Row 4 Right), AbInitio (Row 5 Left), AbInitioVO (Row 5 Middle) and DeNovoIDP (Row 5 Right) ensembles compared to percentages computed from Sung et. al. chemical shift data using the D2D method from Camilloni et al.^{201, 257}

Radii of Gyration for α -Synuclein Simulations

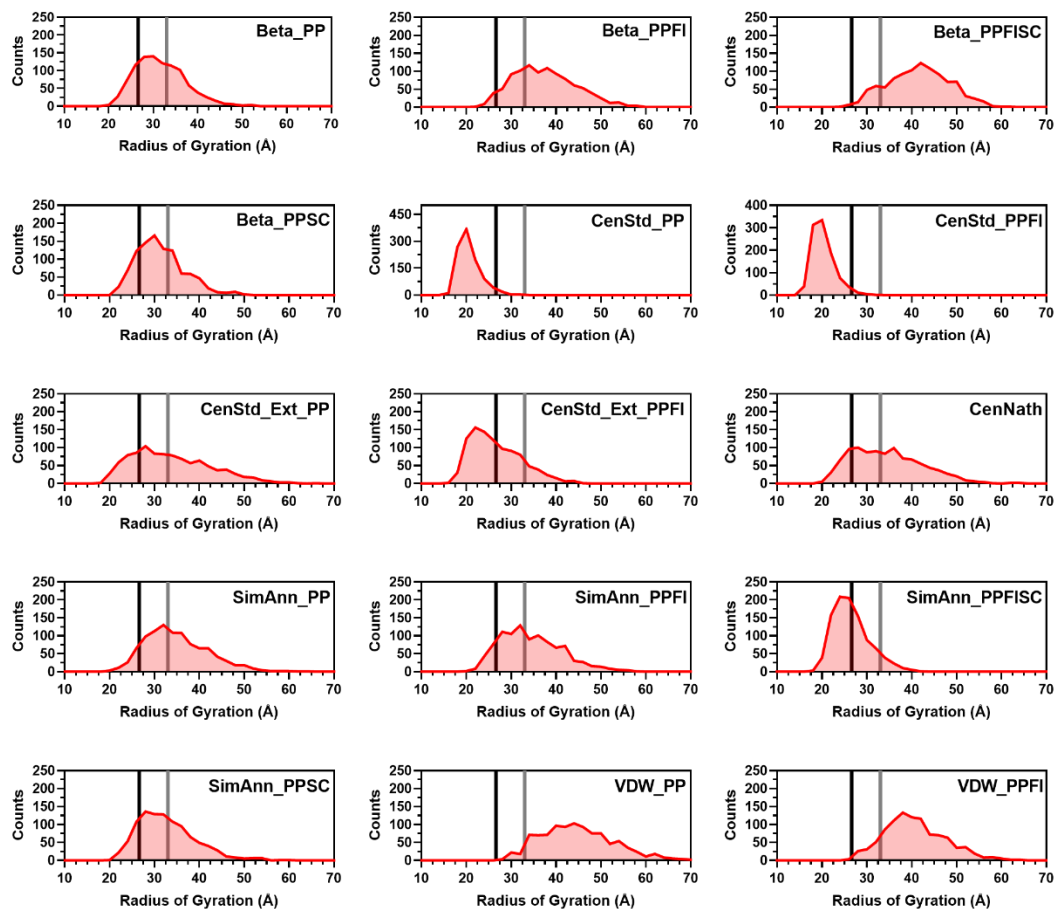


Figure 6.3a: Histograms of Radii of Gyration. Histograms of the radius of gyration from Beta_PP (Row 1 Left), Beta_PPFI (Row 1 Middle), Beta_PPFI SC (Row 1 Right), Beta_PPSC (Row 2 Left), CenStd_PP (Row 2 Middle), CenStd_PPFI (Row 2 Right), CenStd_Ext_PP (Row 3 Left), CenStd_Ext_PPFI (Row 3 Middle), CenNath (Row 3 Right), SimAnn_PP (Row 4 Left), SimAnn_PPFI (Row 4 Middle), SimAnn_PPFI SC (Row 4 Right), SimAnn_PPSC (Row 5 Left), VDW_PP (Row 5 Middle) and VDW_PPFI (Row 5 Right) ensembles compared to experimental values from SAXS²²⁸ (grey) and NMR²⁰⁵ (black).

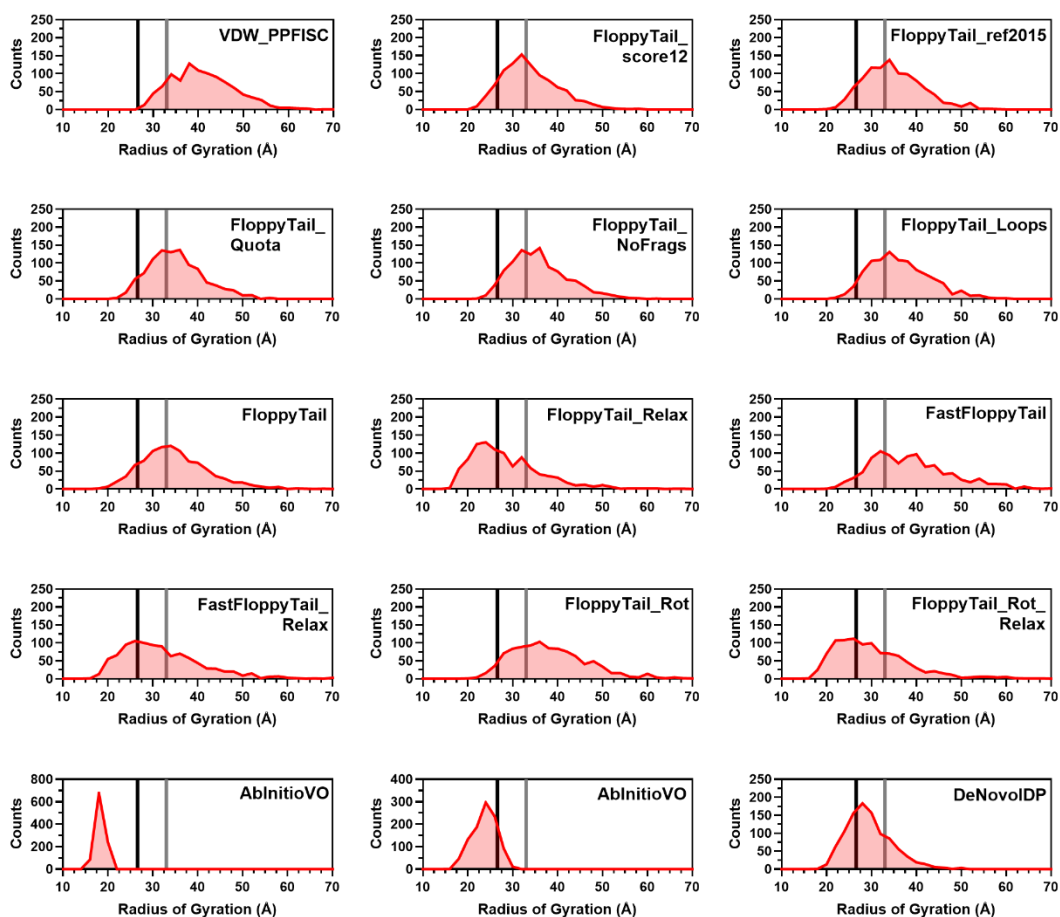


Figure 6.3b: Histograms of Radii of Gyration. Histograms of the radius of gyration from VDW_PPFISC (Row 1 Left), FloppyTail_score12 (Row 1 Middle), FloppyTail_ref2015 (Row 1 Right), FloppyTail_Quota (Row 2 Left), FloppyTail_NoFrag (Row 2 Middle), FloppyTail_Loops (Row 2 Right), FloppyTail (Row 3 Left), FloppyTail_Relax (Row 3 Middle), FastFloppyTail (Row 3 Right), FastFloppyTail_Relax (Row 4 Left), FloppyTail_Rot (Row 4 Middle), FloppyTail_Rot_Relax (Row 4 Right), AbInitio (Row 5 Left), AbInitioVO (Row 5 Middle) and DeNovoIDP (Row 5 Right) ensembles compared to experimental values from SAXS²²⁸ (grey) and NMR²⁰⁵ (black).

Comparison of EFRET Data from α -Synuclein Simulations

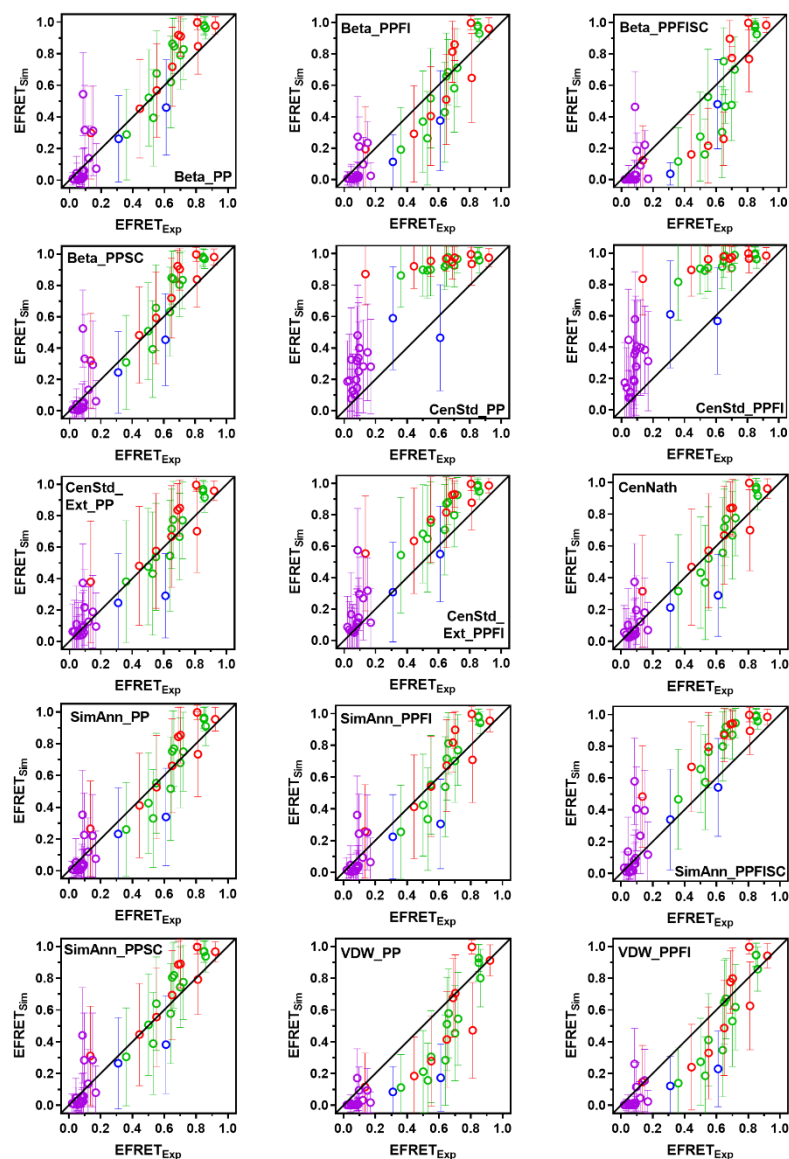


Figure 6.4a: Comparison of Simulated EFRET with Experimental EFRET. Simulated E_{FRET} s for Beta_PP (Row 1 Left), Beta_PPF (Row 1 Center), Beta_PPFISC (Row 1 Right), Beta_PPSC (Row 2 Left), CenStd_PP (Row 2 Center), CenStd_PPF (Row 2 Right), CenStd_Ext_PP (Row 3 Left), CenStd_Ext_PPF (Row 3 Center), CenNath (Row 3 Right), SimAnn_PP (Row 4 Left), SimAnn_PPF (Row 4 Center), SimAnn_PPFISC (Row 4 Right), SimAnn_PPSC (Row 5 Left), VDW_PP (Row 5 Middle), VDW_PPFISC (Row 5 Right) with data from Ferrie *et al.* Cnf-Trp⁵³ (Purple) and Fam-Raz⁵³ (Red) Pairs, Ferrie *et al.* Mcm-Acd pair⁵¹ (Blue), and Nath *et al.*⁶⁸ (Green).

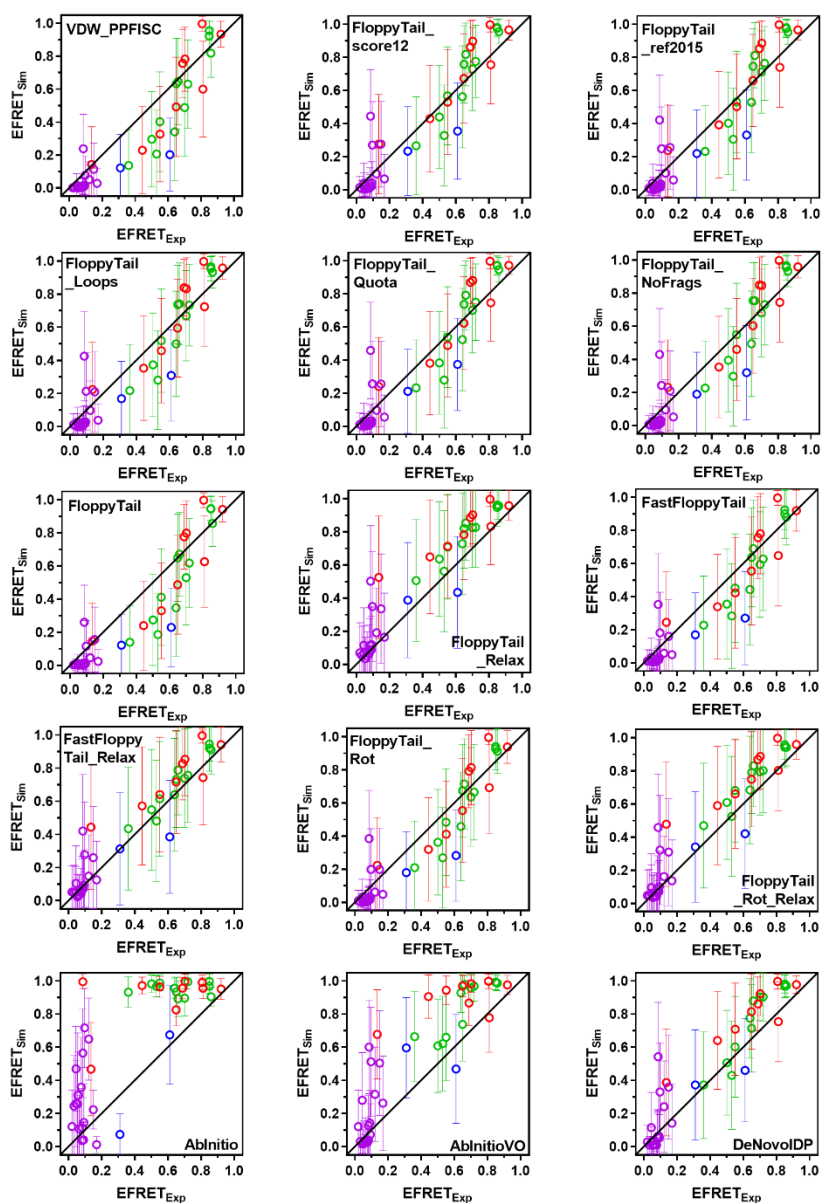


Figure 6.4b: Comparison of Simulated EFRET with Experimental EFRET. Simulated E_{FRET} s for VDW_PPFISC (Row 1 Left), FloppyTail_score12 (Row 1 Center), FloppyTail_ref2015 (Row 1 Right), FloppyTail_Loops (Row 2 Left), FloppyTail_Quota (Row 2 Center), FloppyTail_NoFrag (Row 2 Right), FloppyTail (Row 3 Left), FloppyTail_Relax (Row 3 Center), FastFloppyTail (Row 3 Right), FastFloppyTail_Relax (Row 4 Left), FloppyTail_Rot (Row 4 Center), FloppyTail_Rot_Relax (Row 4 Right), AbInitio (Row 5 Left), AbInitioVO (Row 5 Middle), DeNovoDP (Row 5 Right) with data from Ferrie *et al.* Cnf-Trp⁵³ (Purple) and Fam-Raz⁵³ (Red) Pairs, Ferrie *et al.* Mcm-Acd pair⁵¹ (Blue), and Nath *et al.*⁶⁸ (Green).

Comparison of Distance from α -Synuclein Simulations

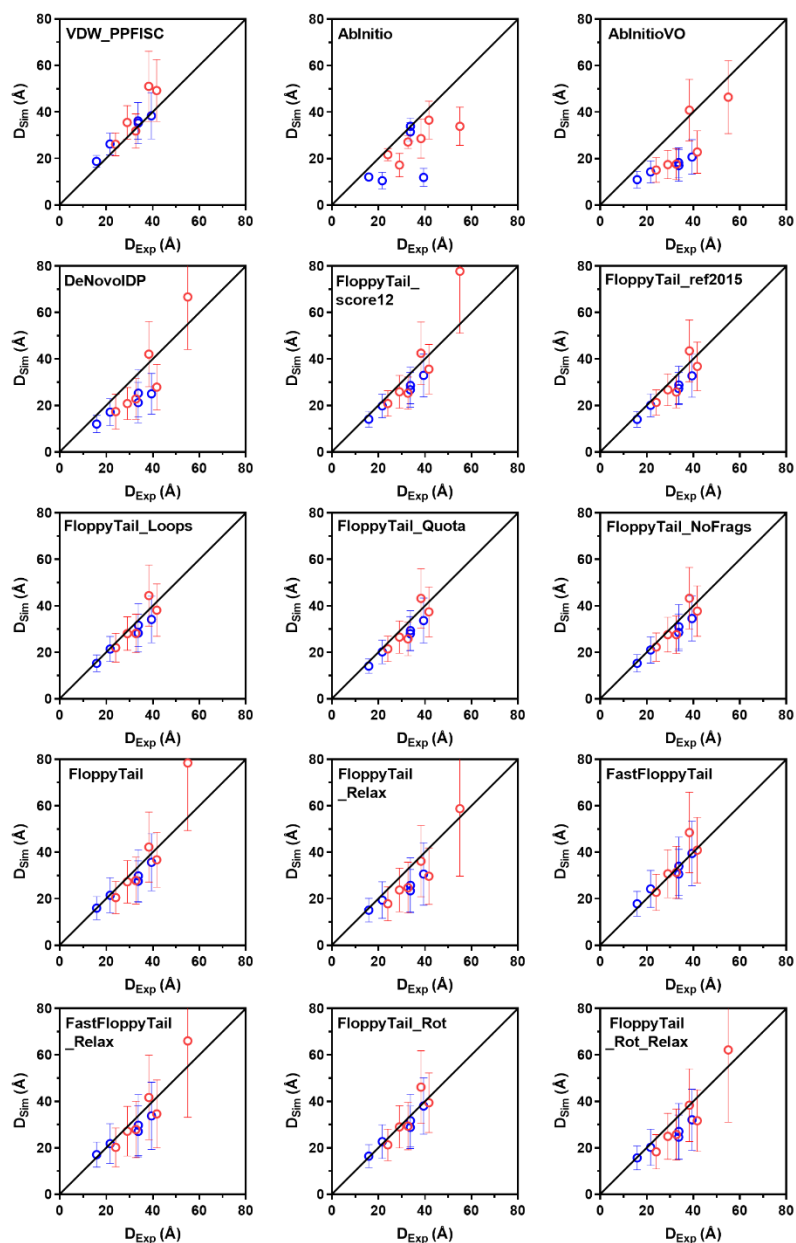


Figure 6.5a: Comparison of Simulated Distance with Experimental Distances. Simulated E_{FRET} for Beta_PP (Row 1 Left), Beta_PPF1 (Row 1 Center), Beta_PPFISC (Row 1 Right), Beta_PPSC (Row 2 Left), CenStd_PP (Row 2 Center), CenStd_PPF1 (Row 2 Right), CenStd_Ext_PP (Row 3 Left), CenStd_Ext_PPF1 (Row 3 Center), CenNath (Row 3 Right), SimAnn_PP (Row 4 Left), SimAnn_PPF1 (Row 4 Center), SimAnn_PPFISC (Row 4 Right), SimAnn_PPSC (Row 5 Left), VDW_PP (Row 5 Middle), VDW_PPFISC (Row 5 Right) with data from Lee *et al.*²⁵³ (Red) and Grupi *et al.*²⁵⁴ (Blue).

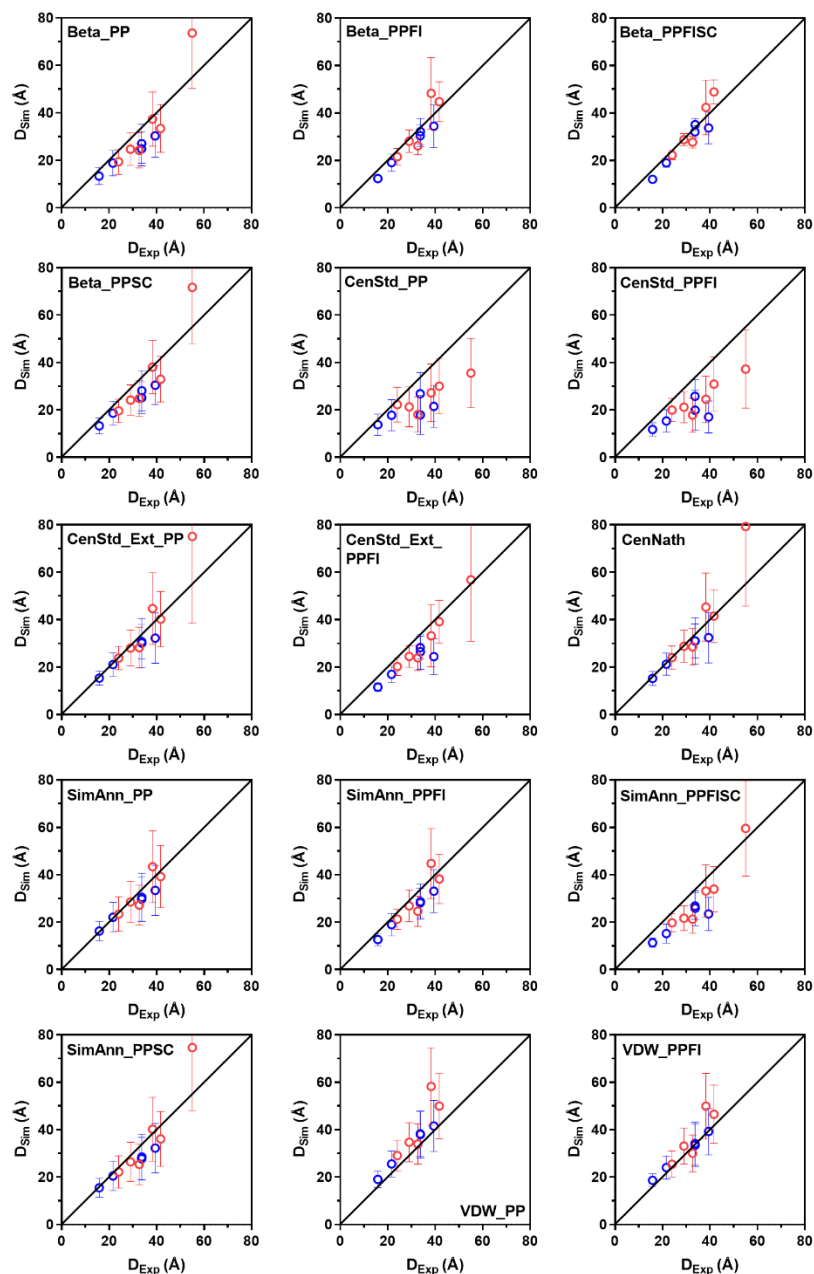


Figure 6.5b: Comparison of Simulated Distance with Experimental Distances. Simulated E_{FRETs} for VDW_PPFI SC (Row 1 Left), FloppyTail_score12 (Row 1 Center), FloppyTail_ref2015 (Row 1 Right), FloppyTail_Loops (Row 2 Left), FloppyTail_Quota (Row 2 Center), FloppyTail_NoFrag (Row 2 Right), FloppyTail (Row 3 Left), FloppyTail_Relax (Row 3 Center), FastFloppyTail (Row 3 Right), FastFloppyTail_Relax (Row 4 Left), FloppyTail_Rot (Row 4 Center), FloppyTail_Rot_Relax (Row 4 Right), AbInitio (Row 5 Left), AbInitioVO (Row 5 Middle), DeNovoIDP (Row 5 Right) with data from Lee *et al.*²⁵³ (Red) and Grupi *et al.*²⁵⁴ (Blue).

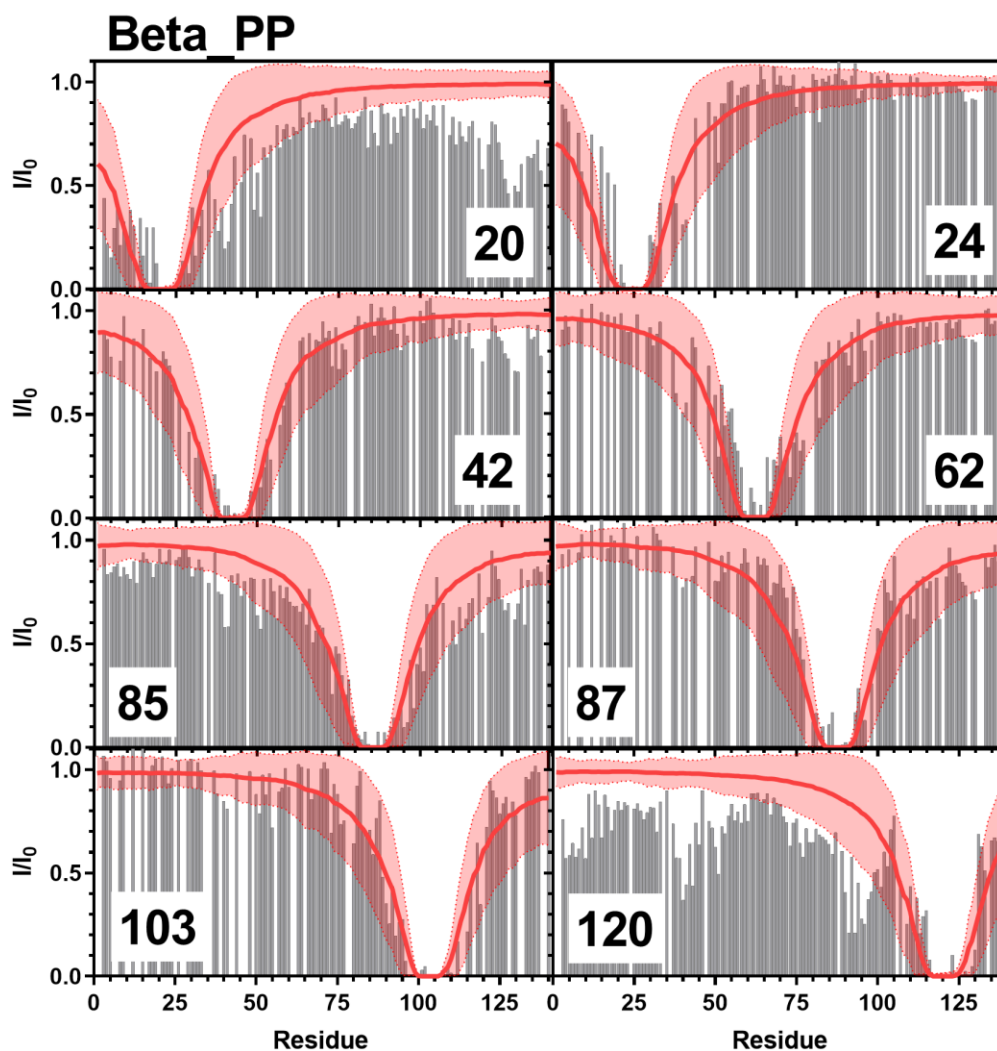


Figure 6.6: Comparison of Simulated Beta_PP PRE Values and Experimental PRE Values. Simulated PRE values from Beta_PP (red line) overlaid on top of experimental data (grey bars) from positions 20 (Top Left), 24 (Top Right), 42 (Upper Middle Left), 62 (Upper Middle Right), 85 (Lower Middle Left), 87 (Lower Middle Right), 103 (Bottom Left), 120 (Bottom Right). Experimental data for positions 20, 85, and 120 are from Sung *et al.*²⁰¹ and data for positions 24, 42, 62, 87, and 103 are from Dedmon *et al.*¹⁰⁴

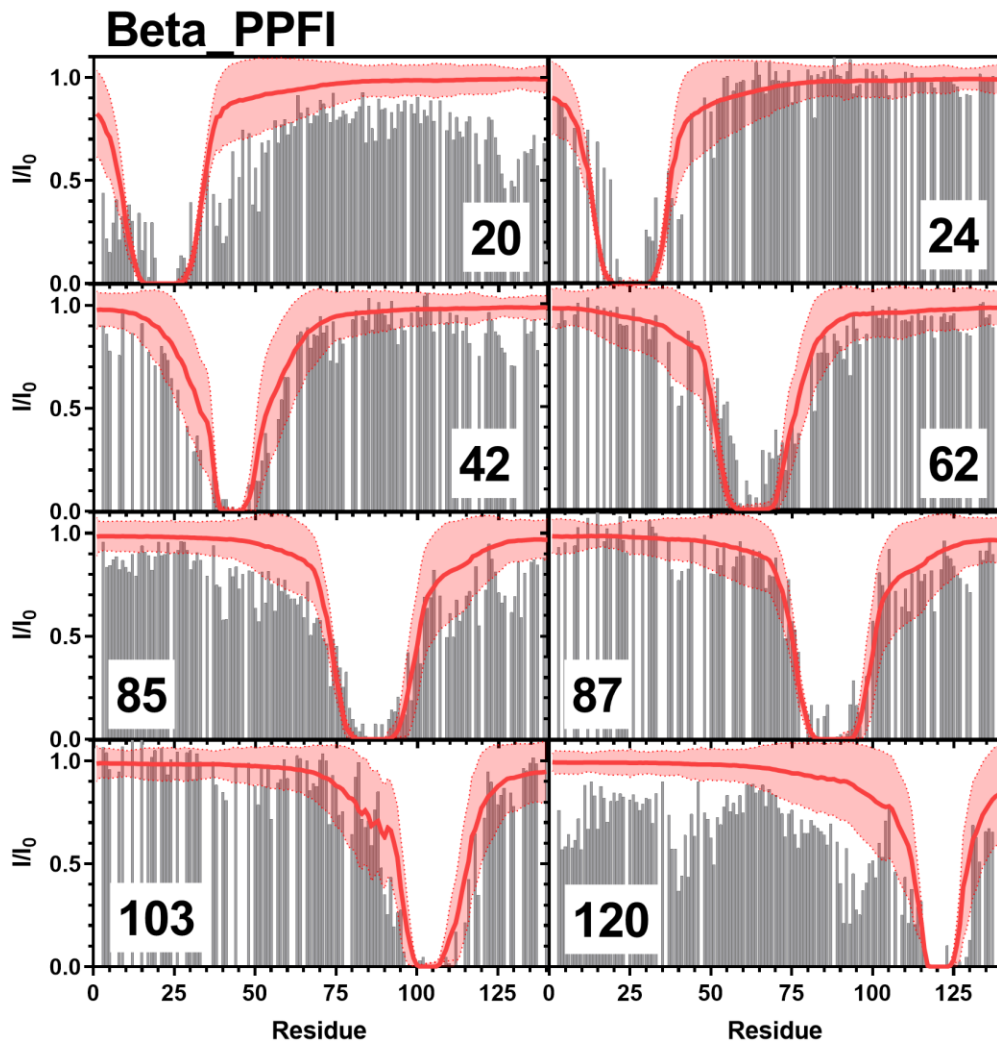


Figure 6.7: Comparison of Simulated Beta_PPF1 PRE Values and Experimental PRE Values. Simulated PRE values from Beta_PPF1 (red line) overlaid on top of experimental data (grey bars) from positions 20 (Top Left), 24 (Top Right), 42 (Upper Middle Left), 62 (Upper Middle Right), 85 (Lower Middle Left), 87 (Lower Middle Right), 103 (Bottom Left), 120 (Bottom Right). Experimental data for positions 20, 85, and 120 are from Sung *et al.*²⁰¹ and data for positions 24, 42, 62, 87, and 103 are from Dedmon *et al.*¹⁰⁴

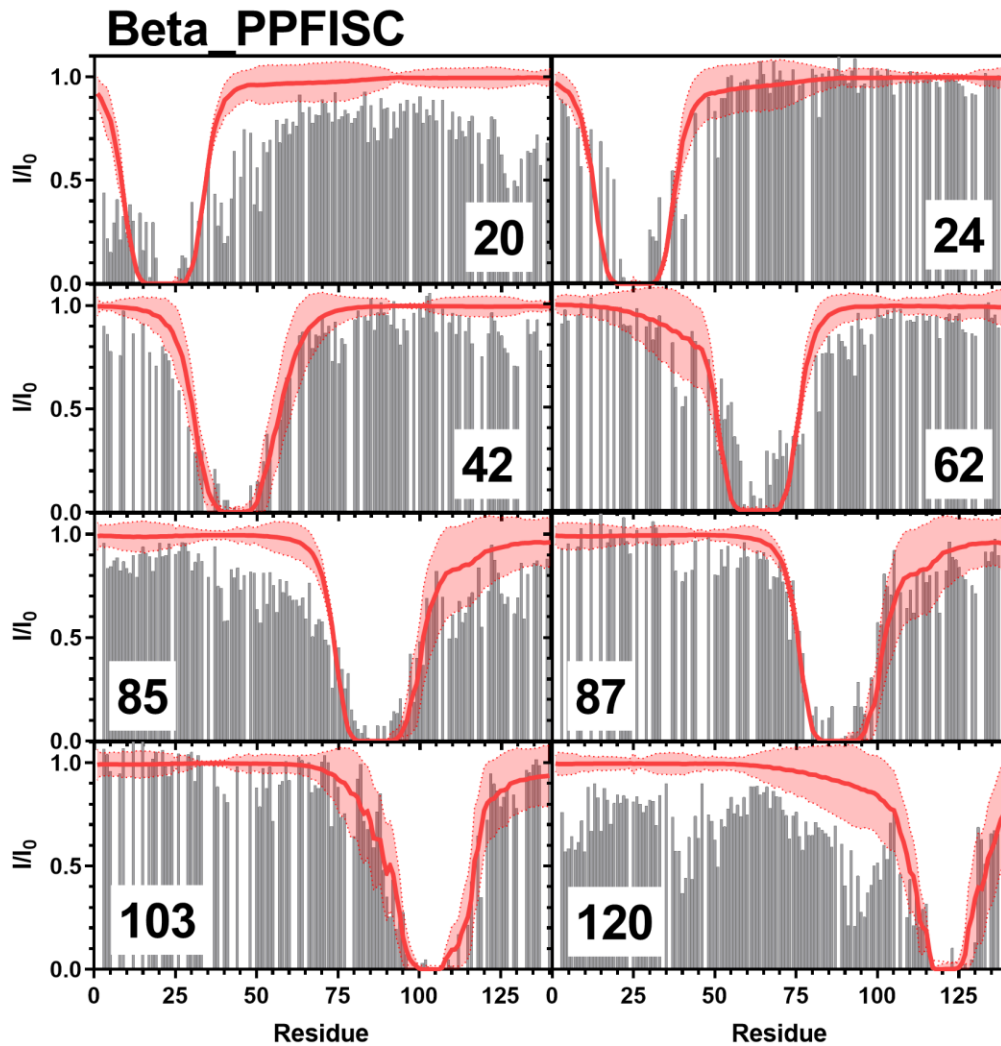


Figure 6.8: Comparison of Simulated Beta_PPFISC PRE Values and Experimental PRE Values. Simulated PRE values from Beta_PPFISC (red line) overlaid on top of experimental data (grey bars) from positions 20 (Top Left), 24 (Top Right), 42 (Upper Middle Left), 62 (Upper Middle Right), 85 (Lower Middle Left), 87 (Lower Middle Right), 103 (Bottom Left), 120 (Bottom Right). Experimental data for positions 20, 85, and 120 are from Sung *et al.*²⁰¹ and data for positions 24, 42, 62, 87, and 103 are from Dedmon *et al.*¹⁰⁴

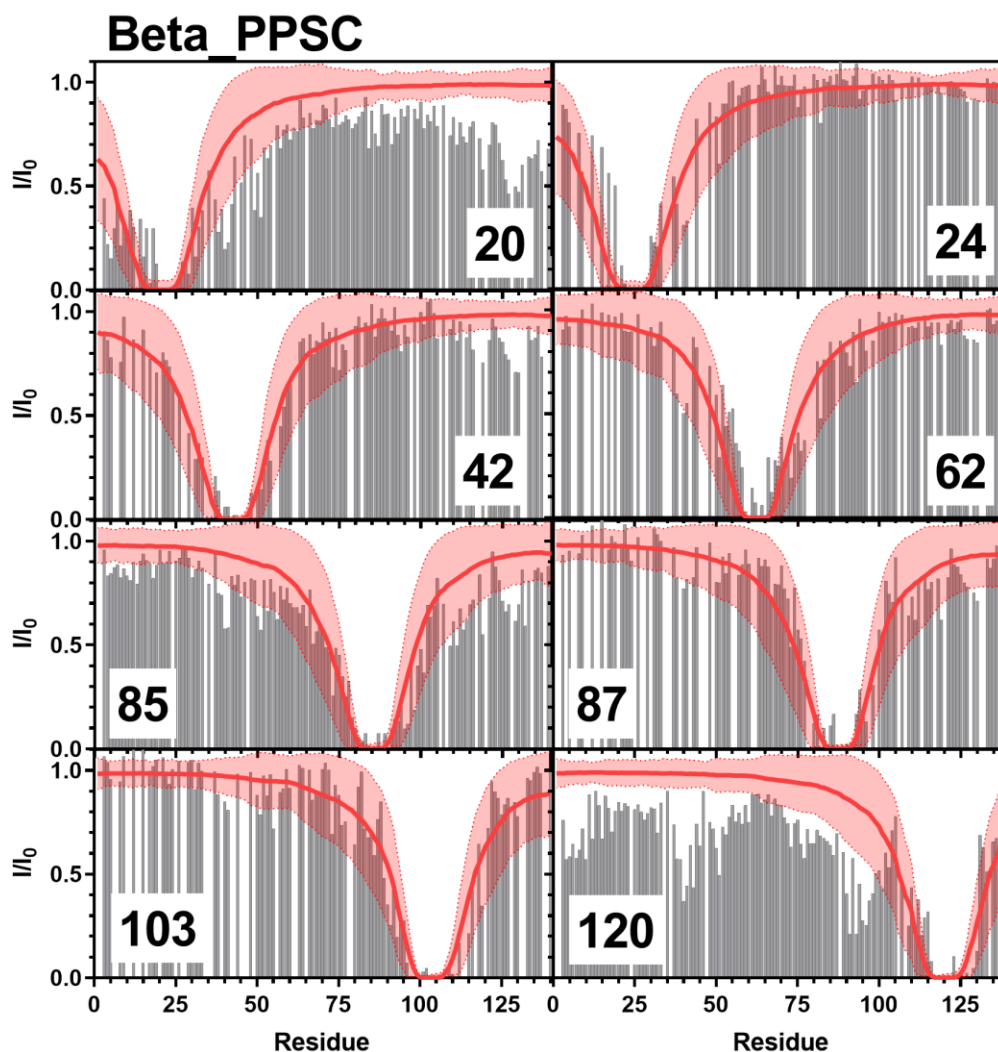


Figure 6.9: Comparison of Simulated Beta_PPSC PRE Values and Experimental PRE Values. Simulated PRE values from Beta_PPSC (red line) overlaid on top of experimental data (grey bars) from positions 20 (Top Left), 24 (Top Right), 42 (Upper Middle Left), 62 (Upper Middle Right), 85 (Lower Middle Left), 87 (Lower Middle Right), 103 (Bottom Left), 120 (Bottom Right). Experimental data for positions 20, 85, and 120 are from Sung *et al.*²⁰¹ and data for positions 24, 42, 62, 87, and 103 are from Dedmon *et al.*¹⁰⁴

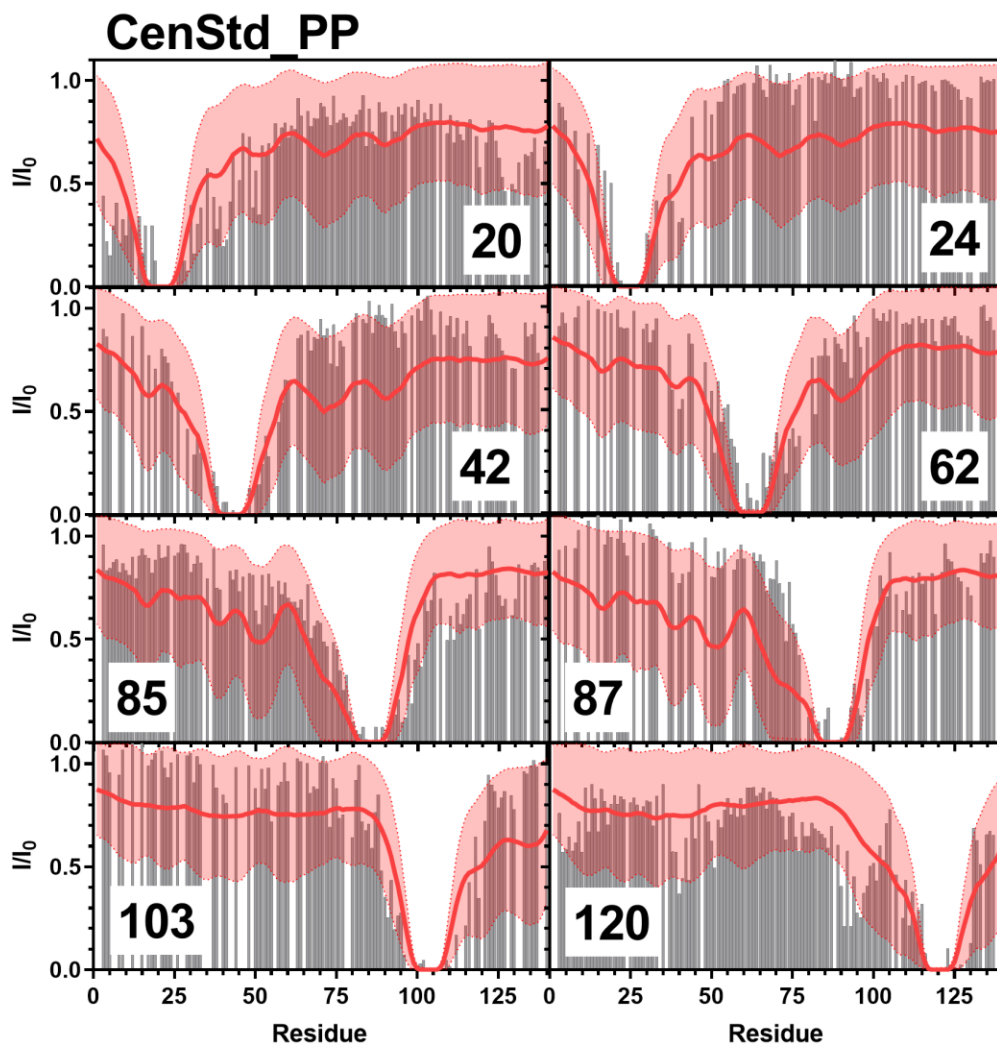


Figure 6.10: Comparison of Simulated CenStd_PP PRE Values and Experimental PRE Values. Simulated PRE values from CenStd_PP (red line) overlaid on top of experimental data (grey bars) from positions 20 (Top Left), 24 (Top Right), 42 (Upper Middle Left), 62 (Upper Middle Right), 85 (Lower Middle Left), 87 (Lower Middle Right), 103 (Bottom Left), 120 (Bottom Right). Experimental data for positions 20, 85, and 120 are from Sung *et al.*²⁰¹ and data for positions 24, 42, 62, 87, and 103 are from Dedmon *et al.*¹⁰⁴

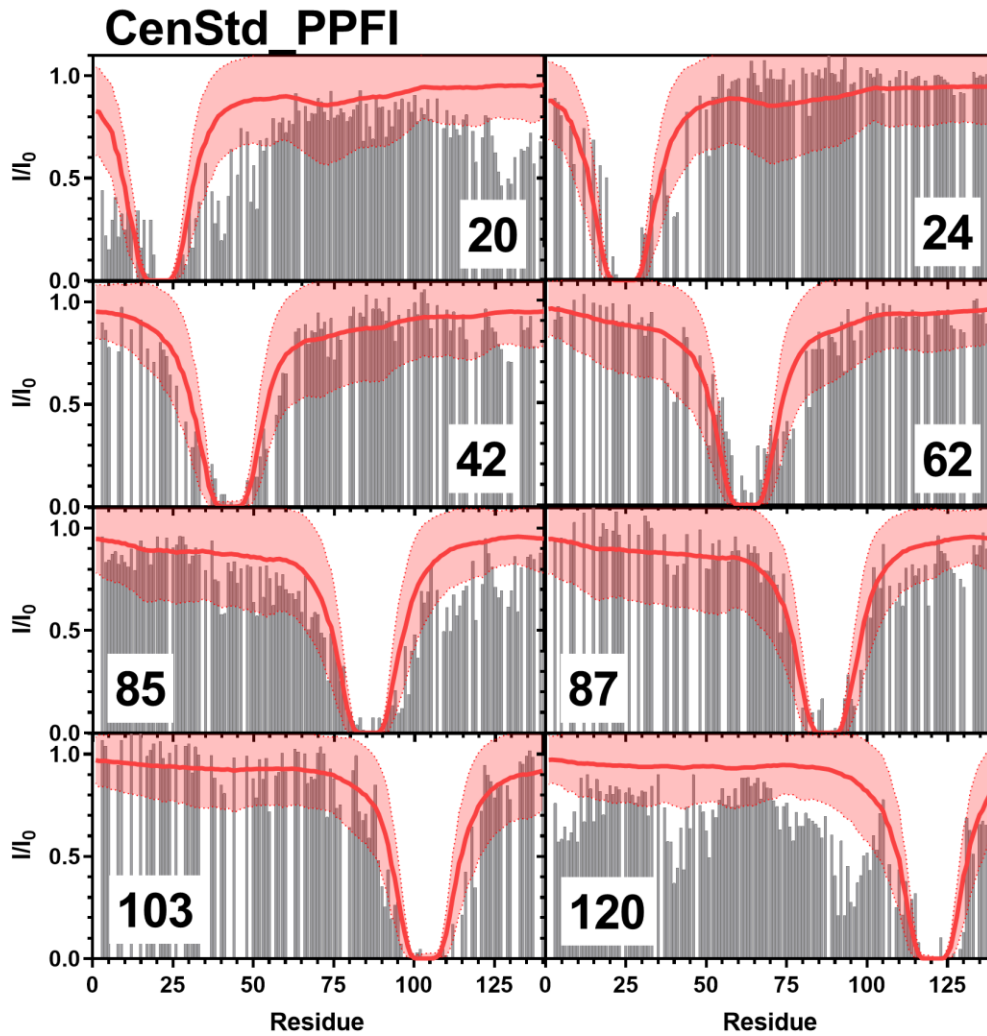


Figure 6.11: Comparison of Simulated CenStd_PPFI PRE Values and Experimental PRE Values. Simulated PRE values from CenStd_PPFI (red line) overlaid on top of experimental data (grey bars) from positions 20 (Top Left), 24 (Top Right), 42 (Upper Middle Left), 62 (Upper Middle Right), 85 (Lower Middle Left), 87 (Lower Middle Right), 103 (Bottom Left), 120 (Bottom Right). Experimental data for positions 20, 85, and 120 are from Sung *et al.*²⁰¹ and data for positions 24, 42, 62, 87, and 103 are from Dedmon *et al.*¹⁰⁴

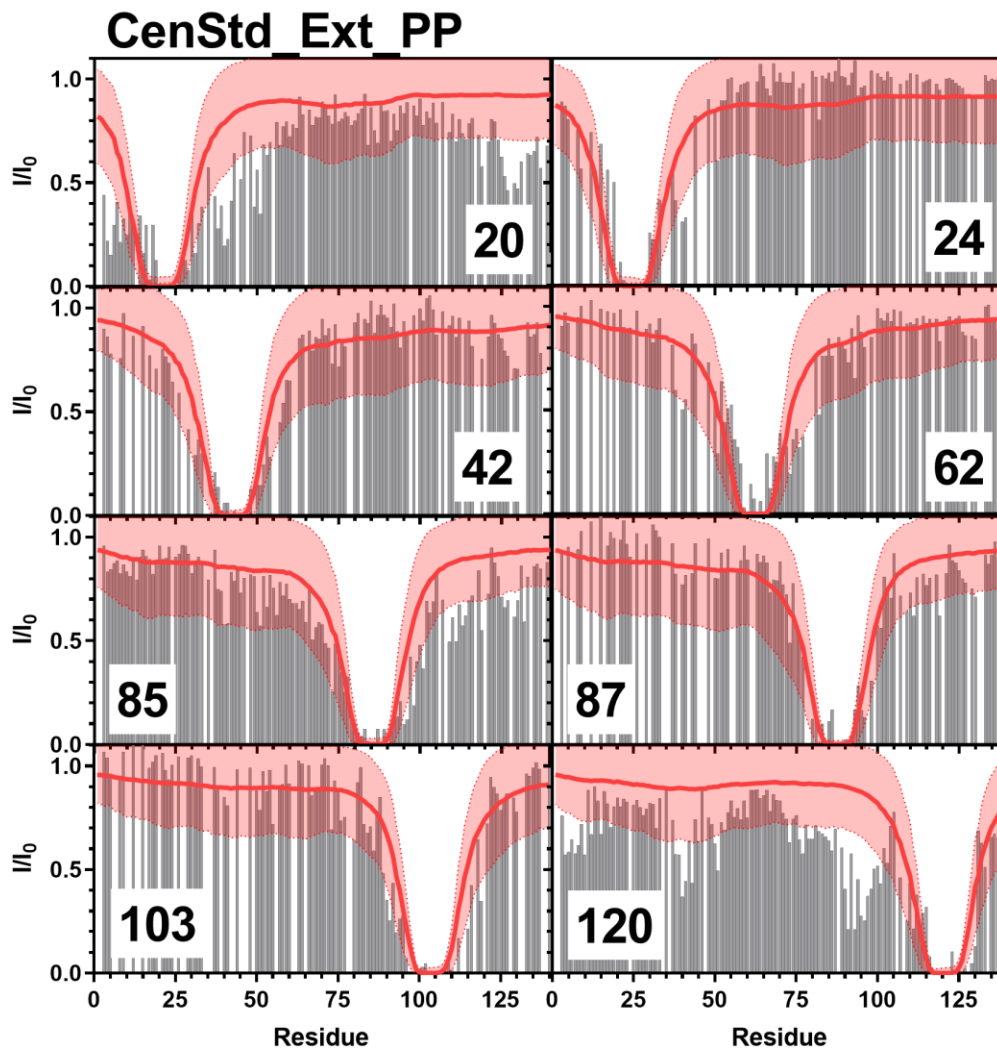


Figure 6.12: Comparison of Simulated CenStd_Ext_PP PRE Values and Experimental PRE Values. Simulated PRE values from CenStd_Ext_PP (red line) overlaid on top of experimental data (grey bars) from positions 20 (Top Left), 24 (Top Right), 42 (Upper Middle Left), 62 (Upper Middle Right), 85 (Lower Middle Left), 87 (Lower Middle Right), 103 (Bottom Left), 120 (Bottom Right). Experimental data for positions 20, 85, and 120 are from Sung *et al.*²⁰¹ and data for positions 24, 42, 62, 87, and 103 are from Dedmon *et al.*¹⁰⁴

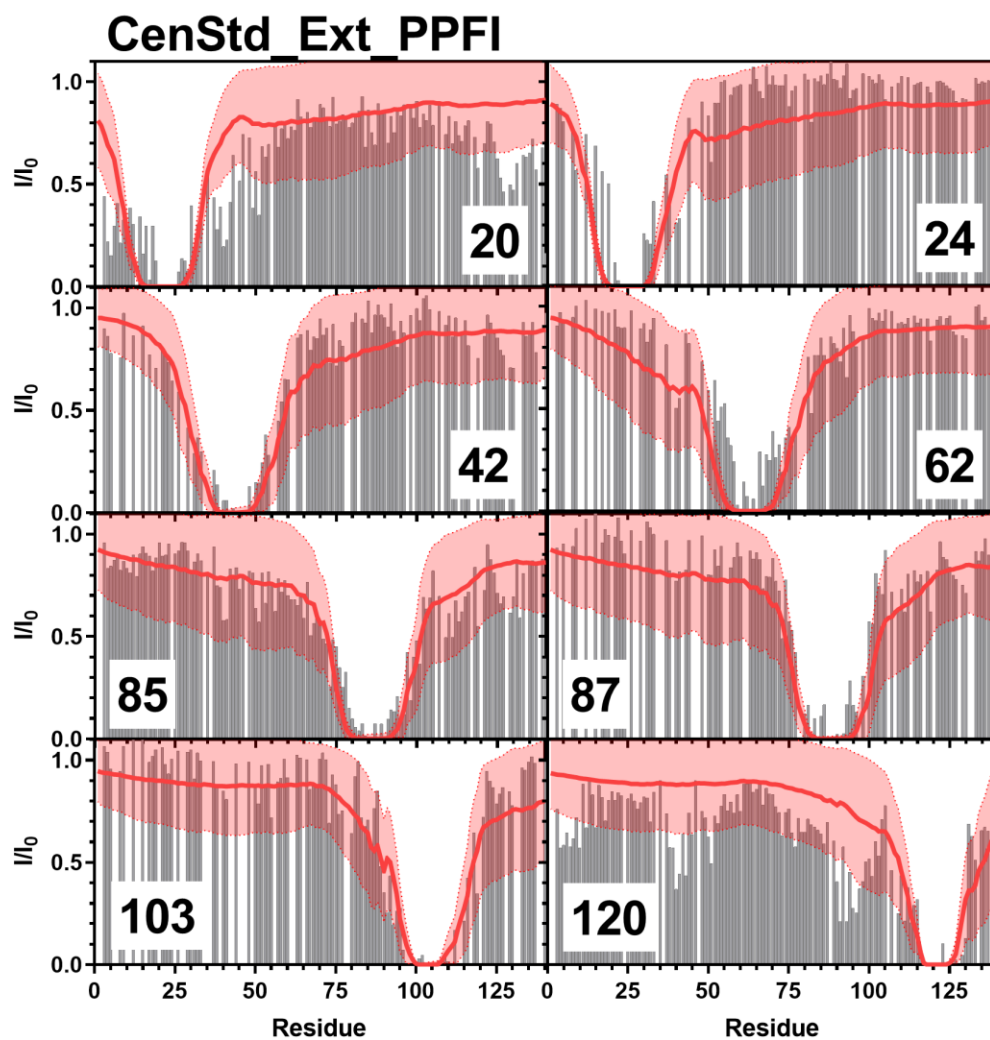


Figure 6.13: Comparison of Simulated CenStd_Ext_PPFi PRE Values and Experimental PRE Values. Simulated PRE values from CenStd_Ext_PPFi (red line) overlaid on top of experimental data (grey bars) from positions 20 (Top Left), 24 (Top Right), 42 (Upper Middle Left), 62 (Upper Middle Right), 85 (Lower Middle Left), 87 (Lower Middle Right), 103 (Bottom Left), 120 (Bottom Right). Experimental data for positions 20, 85, and 120 are from Sung *et al.*²⁰¹ and data for positions 24, 42, 62, 87, and 103 are from Dedmon *et al.*¹⁰⁴

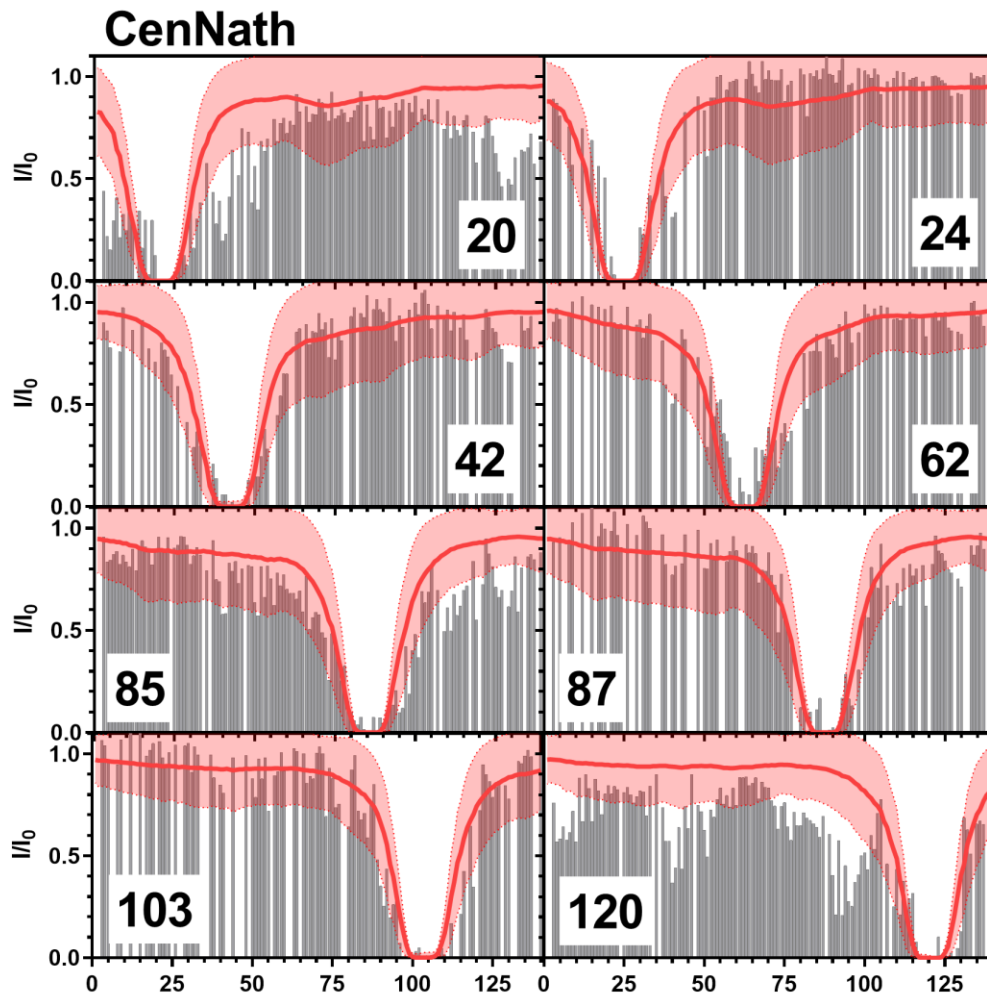


Figure 6.14: Comparison of Simulated CenNath PRE Values and Experimental PRE Values. Simulated PRE values from CenNath (red line) overlaid on top of experimental data (grey bars) from positions 20 (Top Left), 24 (Top Right), 42 (Upper Middle Left), 62 (Upper Middle Right), 85 (Lower Middle Left), 87 (Lower Middle Right), 103 (Bottom Left), 120 (Bottom Right). Experimental data for positions 20, 85, and 120 are from Sung *et al.*²⁰¹ and data for positions 24, 42, 62, 87, and 103 are from Dedmon *et al.*¹⁰⁴

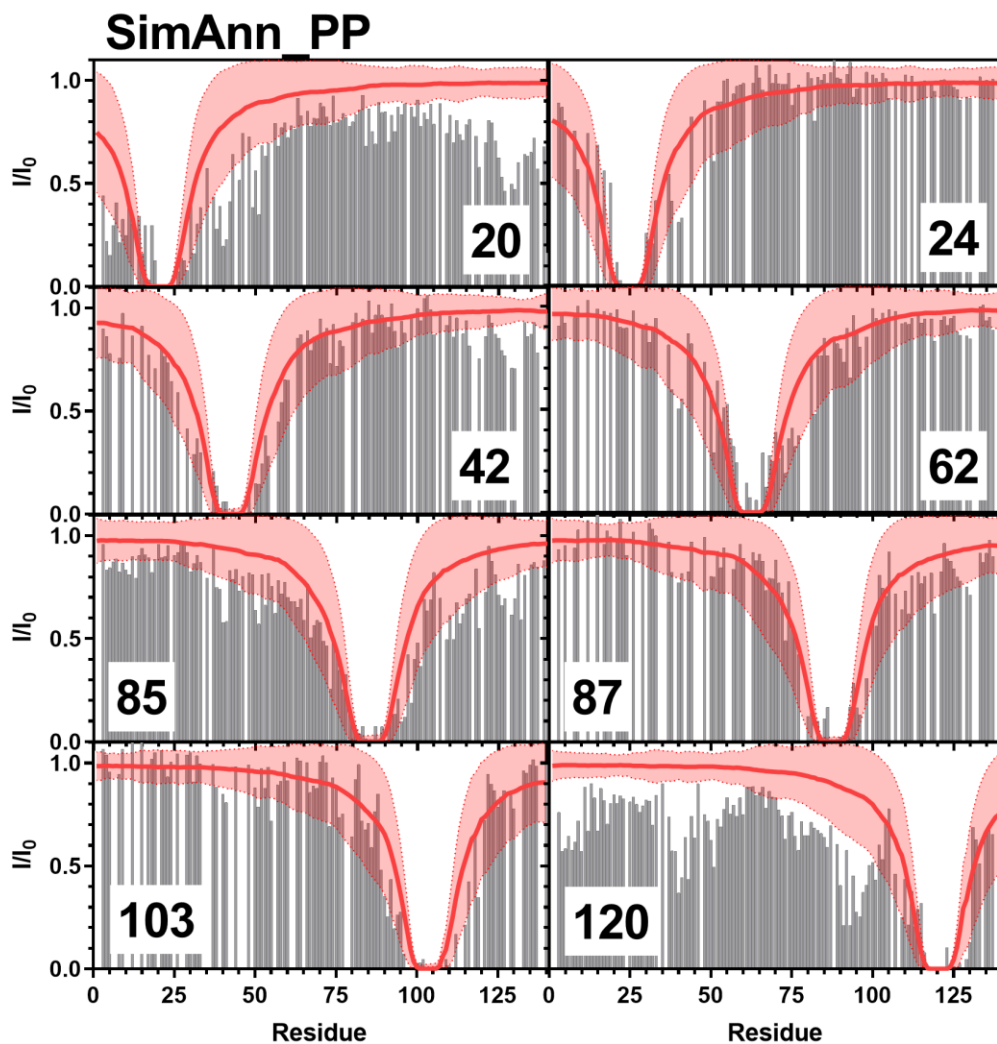


Figure 6.15: Comparison of Simulated SimAnn_PP PRE Values and Experimental PRE Values. Simulated PRE values from SimAnn_PP (red line) overlaid on top of experimental data (grey bars) from positions 20 (Top Left), 24 (Top Right), 42 (Upper Middle Left), 62 (Upper Middle Right), 85 (Lower Middle Left), 87 (Lower Middle Right), 103 (Bottom Left), 120 (Bottom Right). Experimental data for positions 20, 85, and 120 are from Sung *et al.*²⁰¹ and data for positions 24, 42, 62, 87, and 103 are from Dedmon *et al.*¹⁰⁴

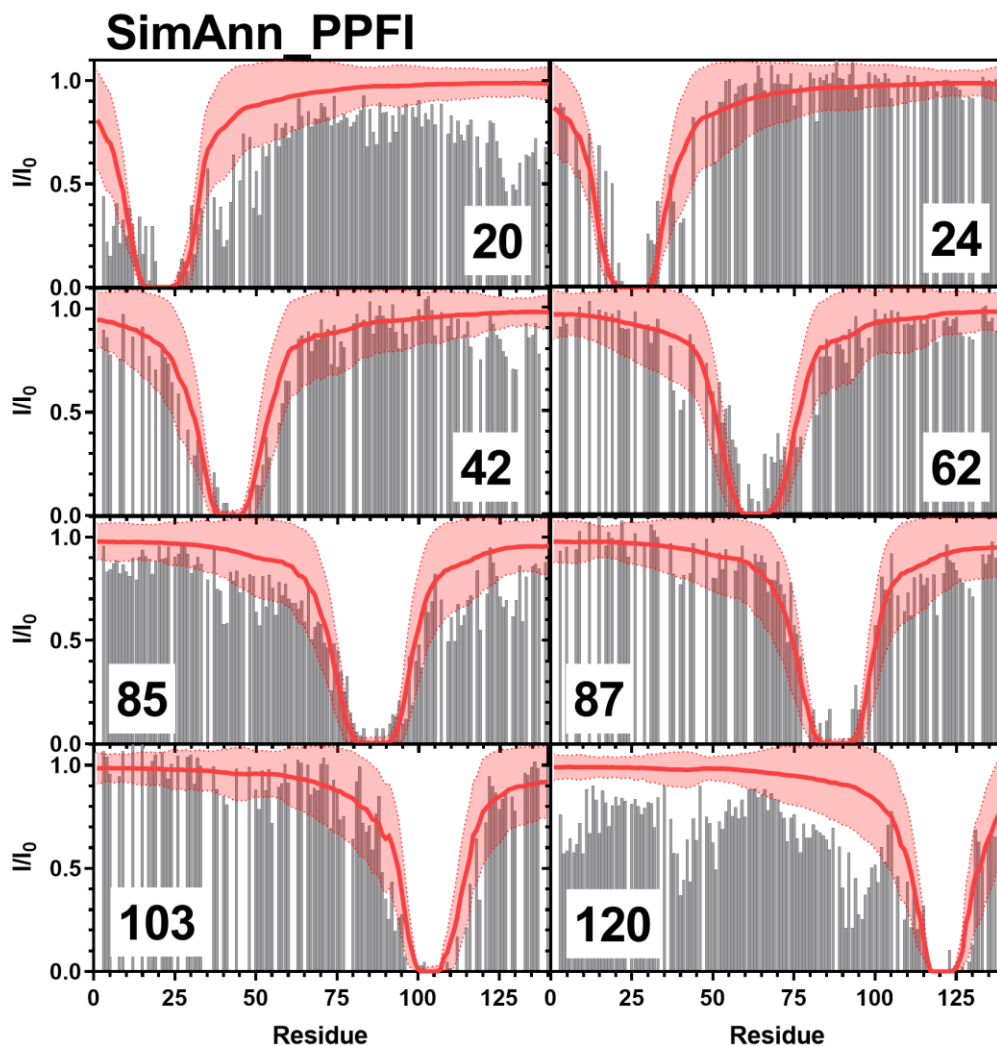


Figure 6.16: Comparison of Simulated SimAnn_PPFi PRE Values and Experimental PRE Values. Simulated PRE values from SimAnn_PPFi (red line) overlaid on top of experimental data (grey bars) from positions 20 (Top Left), 24 (Top Right), 42 (Upper Middle Left), 62 (Upper Middle Right), 85 (Lower Middle Left), 87 (Lower Middle Right), 103 (Bottom Left), 120 (Bottom Right). Experimental data for positions 20, 85, and 120 are from Sung *et al.*²⁰¹ and data for positions 24, 42, 62, 87, and 103 are from Dedmon *et al.*¹⁰⁴

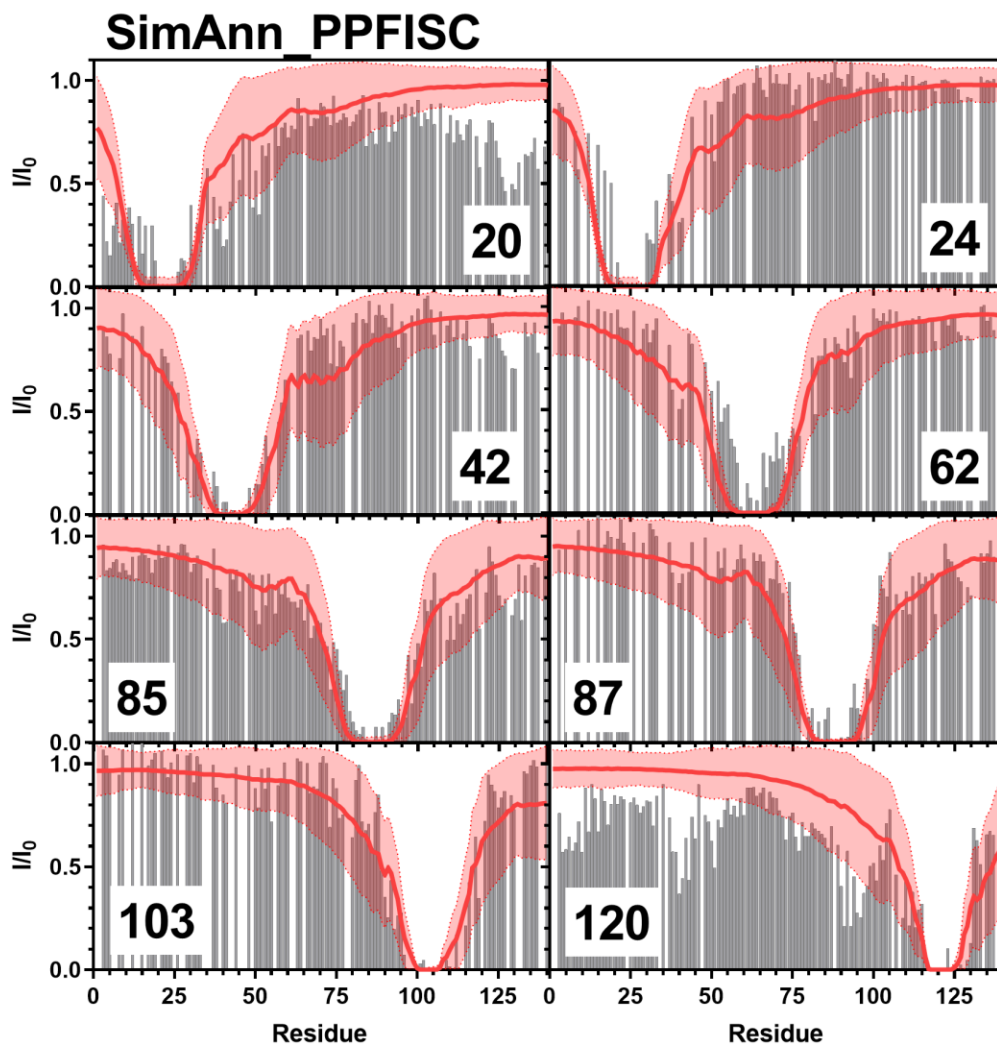


Figure 6.17: Comparison of Simulated SimAnn_PPFISC PRE Values and Experimental PRE Values. Simulated PRE values from SimAnn_PPFISC (red line) overlaid on top of experimental data (grey bars) from positions 20 (Top Left), 24 (Top Right), 42 (Upper Middle Left), 62 (Upper Middle Right), 85 (Lower Middle Left), 87 (Lower Middle Right), 103 (Bottom Left), 120 (Bottom Right). Experimental data for positions 20, 85, and 120 are from Sung *et al.*²⁰¹ and data for positions 24, 42, 62, 87, and 103 are from Dedmon *et al.*¹⁰⁴

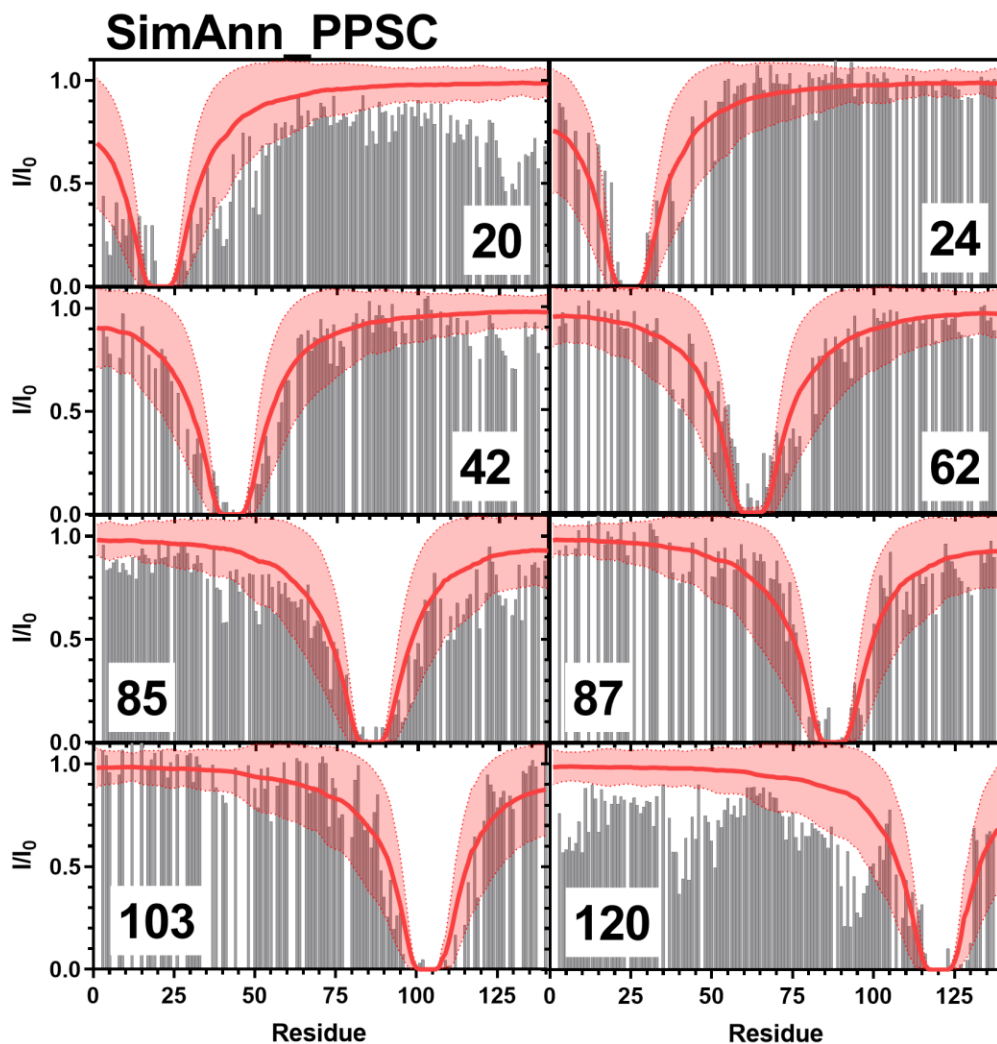


Figure 6.18: Comparison of Simulated SimAnn_PPSC PRE Values and Experimental PRE Values. Simulated PRE values from SimAnn_PPSC (red line) overlaid on top of experimental data (grey bars) from positions 20 (Top Left), 24 (Top Right), 42 (Upper Middle Left), 62 (Upper Middle Right), 85 (Lower Middle Left), 87 (Lower Middle Right), 103 (Bottom Left), 120 (Bottom Right). Experimental data for positions 20, 85, and 120 are from Sung *et al.*²⁰¹ and data for positions 24, 42, 62, 87, and 103 are from Dedmon *et al.*¹⁰⁴

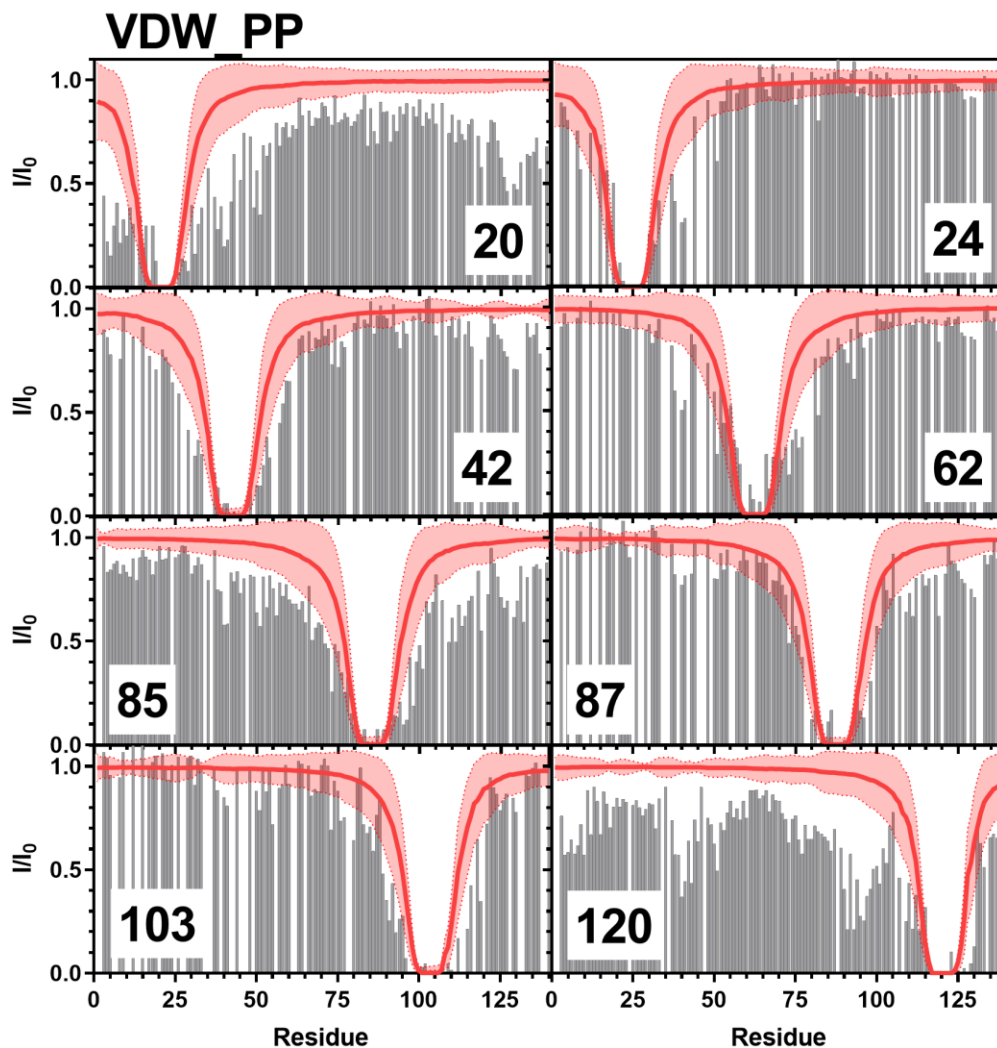


Figure 6.19: Comparison of Simulated VDW_PP PRE Values and Experimental PRE Values. Simulated PRE values from VDW_PP (red line) overlaid on top of experimental data (grey bars) from positions 20 (Top Left), 24 (Top Right), 42 (Upper Middle Left), 62 (Upper Middle Right), 85 (Lower Middle Left), 87 (Lower Middle Right), 103 (Bottom Left), 120 (Bottom Right). Experimental data for positions 20, 85, and 120 are from Sung *et al.*²⁰¹ and data for positions 24, 42, 62, 87, and 103 are from Dedmon *et al.*¹⁰⁴

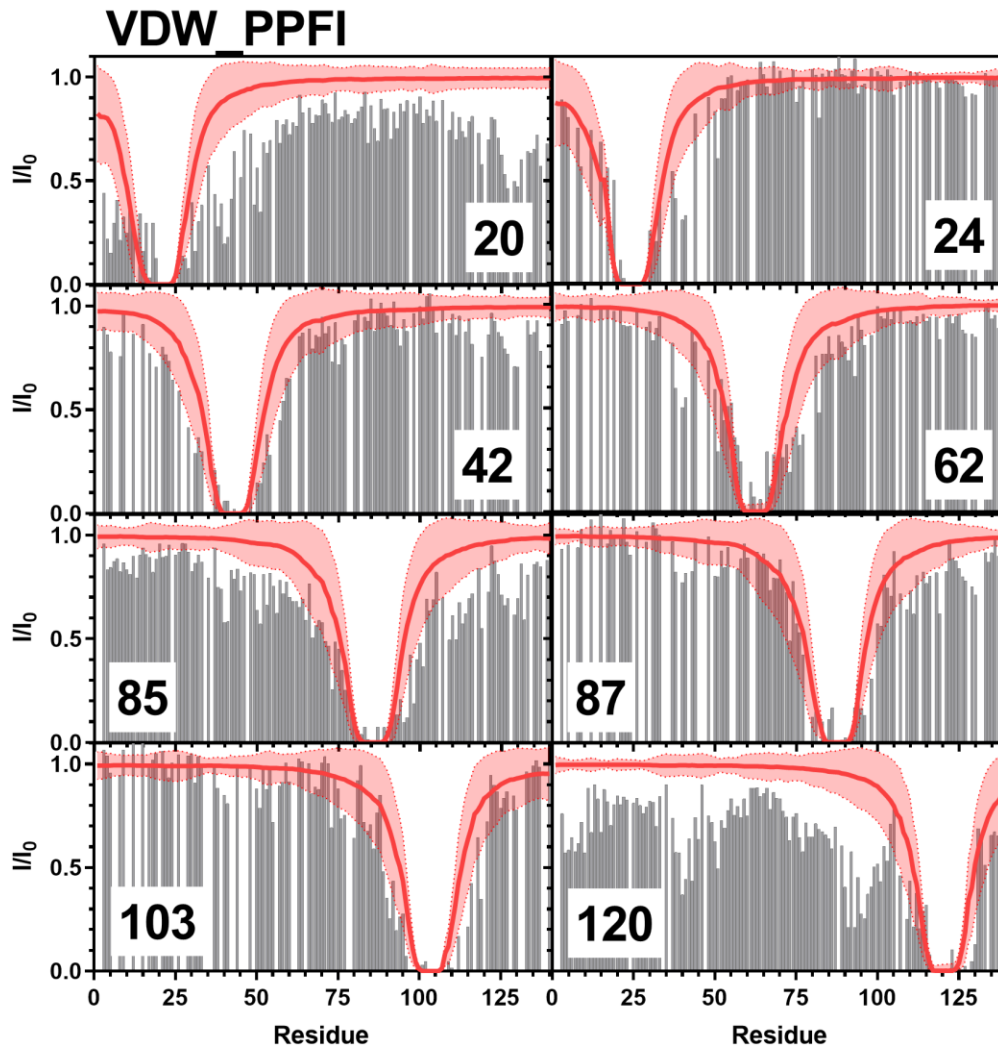


Figure 6.20: Comparison of Simulated VDW_PPFI PRE Values and Experimental PRE Values. Simulated PRE values from VDW_PPFI (red line) overlaid on top of experimental data (grey bars) from positions 20 (Top Left), 24 (Top Right), 42 (Upper Middle Left), 62 (Upper Middle Right), 85 (Lower Middle Left), 87 (Lower Middle Right), 103 (Bottom Left), 120 (Bottom Right). Experimental data for positions 20, 85, and 120 are from Sung *et al.*²⁰¹ and data for positions 24, 42, 62, 87, and 103 are from Dedmon *et al.*¹⁰⁴

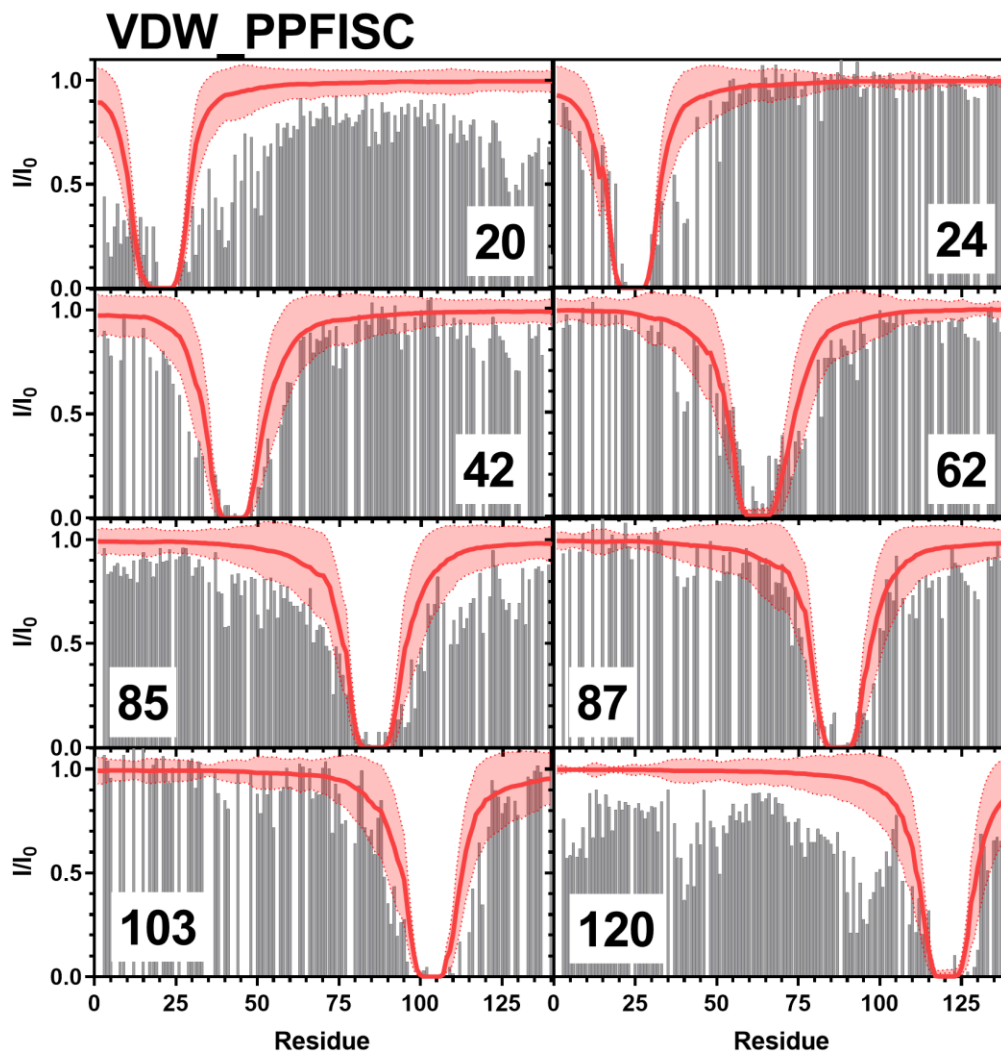


Figure 6.21: Comparison of Simulated VDW_PPFISC PRE Values and Experimental PRE Values. Simulated PRE values from VDW_PPFISC (red line) overlaid on top of experimental data (grey bars) from positions 20 (Top Left), 24 (Top Right), 42 (Upper Middle Left), 62 (Upper Middle Right), 85 (Lower Middle Left), 87 (Lower Middle Right), 103 (Bottom Left), 120 (Bottom Right). Experimental data for positions 20, 85, and 120 are from Sung *et al.*²⁰¹ and data for positions 24, 42, 62, 87, and 103 are from Dedmon *et al.*¹⁰⁴

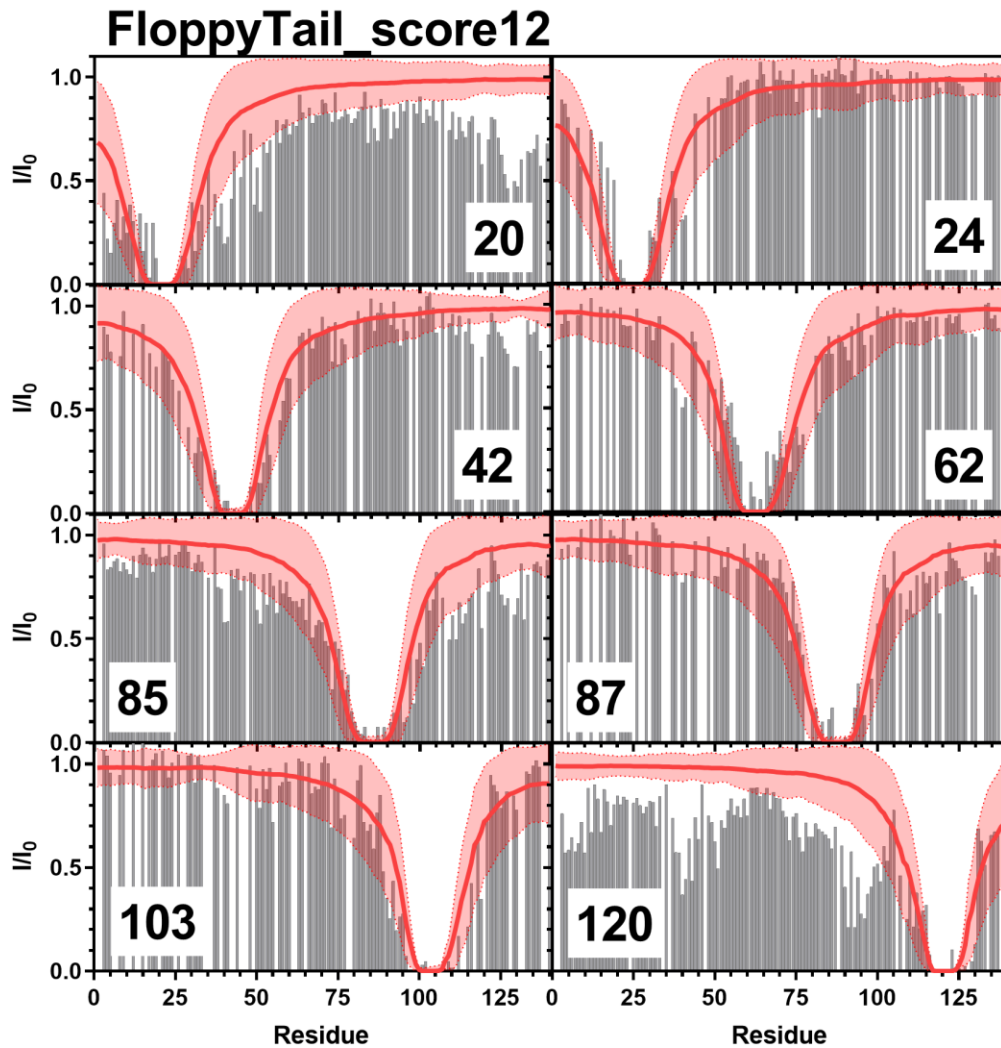


Figure 6.22: Comparison of Simulated FloppyTail_score12 PRE Values and Experimental PRE Values. Simulated PRE values from FloppyTail_score12 (red line) overlaid on top of experimental data (grey bars) from positions 20 (Top Left), 24 (Top Right), 42 (Upper Middle Left), 62 (Upper Middle Right), 85 (Lower Middle Left), 87 (Lower Middle Right), 103 (Bottom Left), 120 (Bottom Right). Experimental data for positions 20, 85, and 120 are from Sung *et al.*²⁰¹ and data for positions 24, 42, 62, 87, and 103 are from Dedmon *et al.*¹⁰⁴

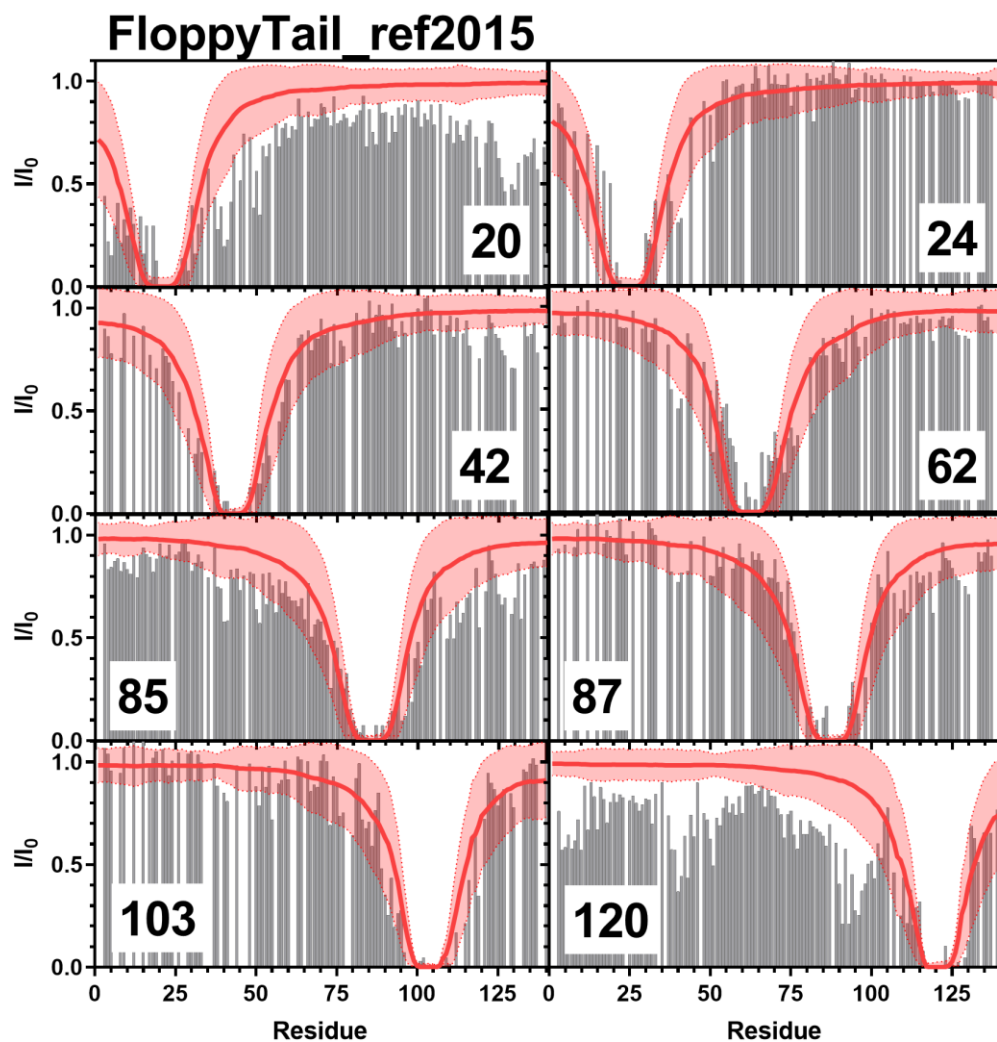


Figure 6.23: Comparison of Simulated FloppyTail_ref2015 PRE Values and Experimental PRE Values. Simulated PRE values from FloppyTail_ref2015 (red line) overlaid on top of experimental data (grey bars) from positions 20 (Top Left), 24 (Top Right), 42 (Upper Middle Left), 62 (Upper Middle Right), 85 (Lower Middle Left), 87 (Lower Middle Right), 103 (Bottom Left), 120 (Bottom Right). Experimental data for positions 20, 85, and 120 are from Sung *et al.*²⁰¹ and data for positions 24, 42, 62, 87, and 103 are from Dedmon *et al.*¹⁰⁴

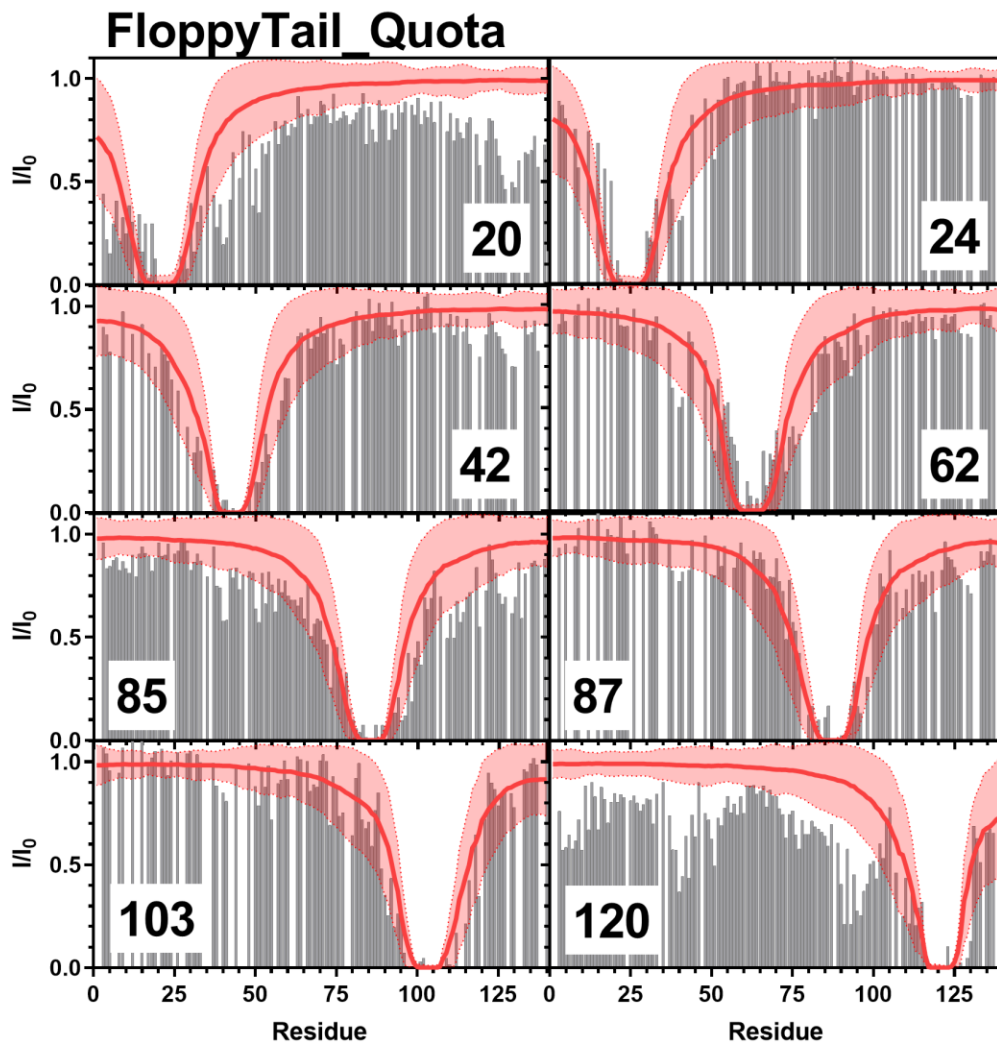


Figure 6.24: Comparison of Simulated FloppyTail_Quota PRE Values and Experimental PRE Values. Simulated PRE values from FloppyTail_Quota (red line) overlaid on top of experimental data (grey bars) from positions 20 (Top Left), 24 (Top Right), 42 (Upper Middle Left), 62 (Upper Middle Right), 85 (Lower Middle Left), 87 (Lower Middle Right), 103 (Bottom Left), 120 (Bottom Right). Experimental data for positions 20, 85, and 120 are from Sung *et al.*²⁰¹ and data for positions 24, 42, 62, 87, and 103 are from Dedmon *et al.*¹⁰⁴

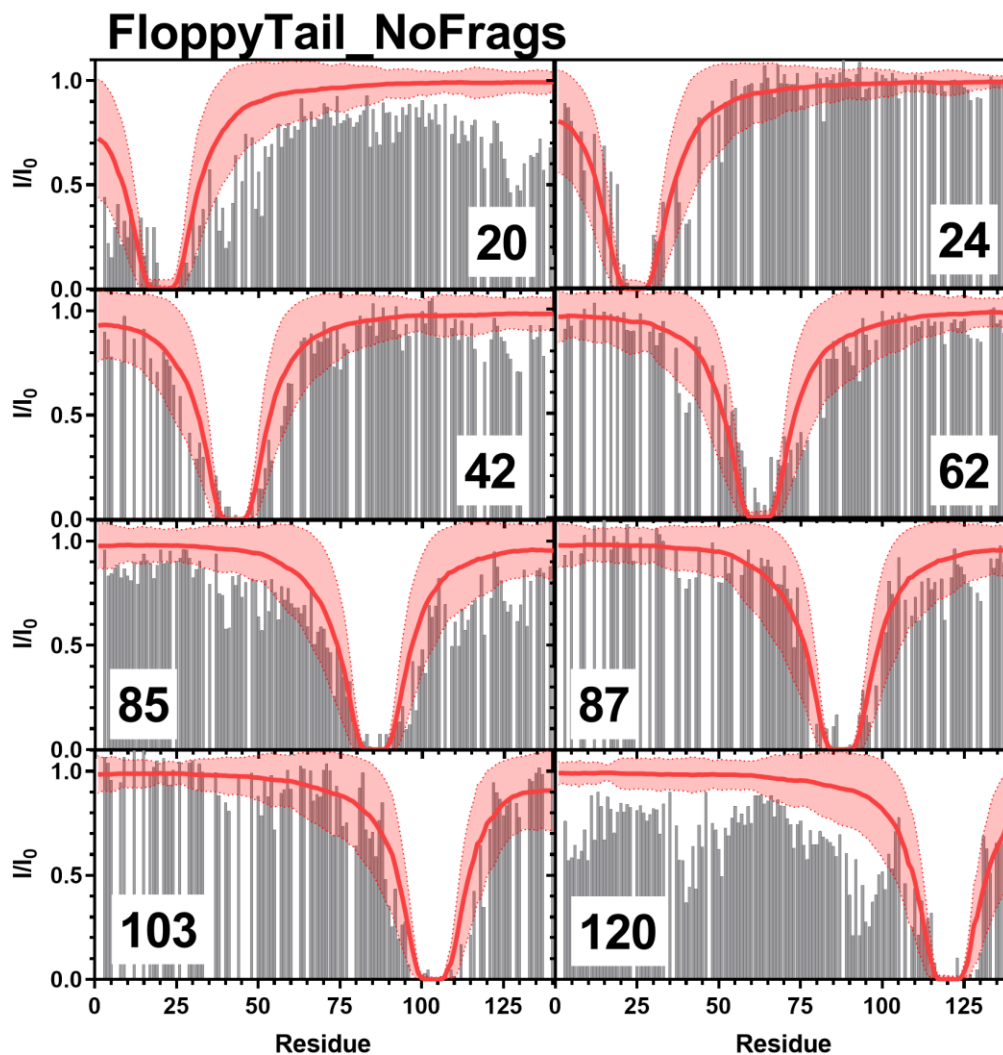


Figure 6.25: Comparison of Simulated FloppyTail_NoFrag PRE Values and Experimental PRE Values. Simulated PRE values from FloppyTail_NoFrag (red line) overlaid on top of experimental data (grey bars) from positions 20 (Top Left), 24 (Top Right), 42 (Upper Middle Left), 62 (Upper Middle Right), 85 (Lower Middle Left), 87 (Lower Middle Right), 103 (Bottom Left), 120 (Bottom Right). Experimental data for positions 20, 85, and 120 are from Sung *et al.*²⁰¹ and data for positions 24, 42, 62, 87, and 103 are from Dedmon *et al.*¹⁰⁴

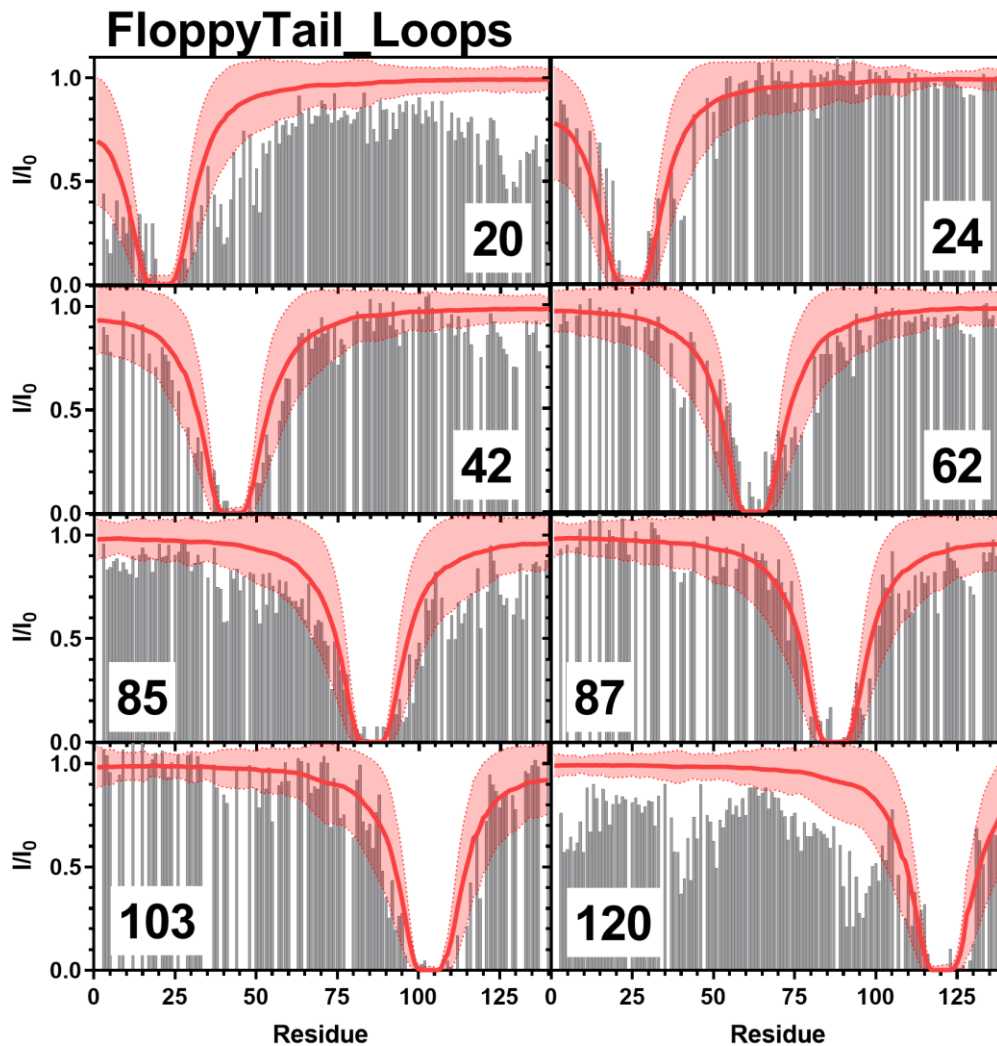


Figure 6.26: Comparison of Simulated FloppyTail_Loops PRE Values and Experimental PRE Values. Simulated PRE values from FloppyTail_Loops (red line) overlaid on top of experimental data (grey bars) from positions 20 (Top Left), 24 (Top Right), 42 (Upper Middle Left), 62 (Upper Middle Right), 85 (Lower Middle Left), 87 (Lower Middle Right), 103 (Bottom Left), 120 (Bottom Right). Experimental data for positions 20, 85, and 120 are from Sung *et al.*²⁰¹ and data for positions 24, 42, 62, 87, and 103 are from Dedmon *et al.*¹⁰⁴

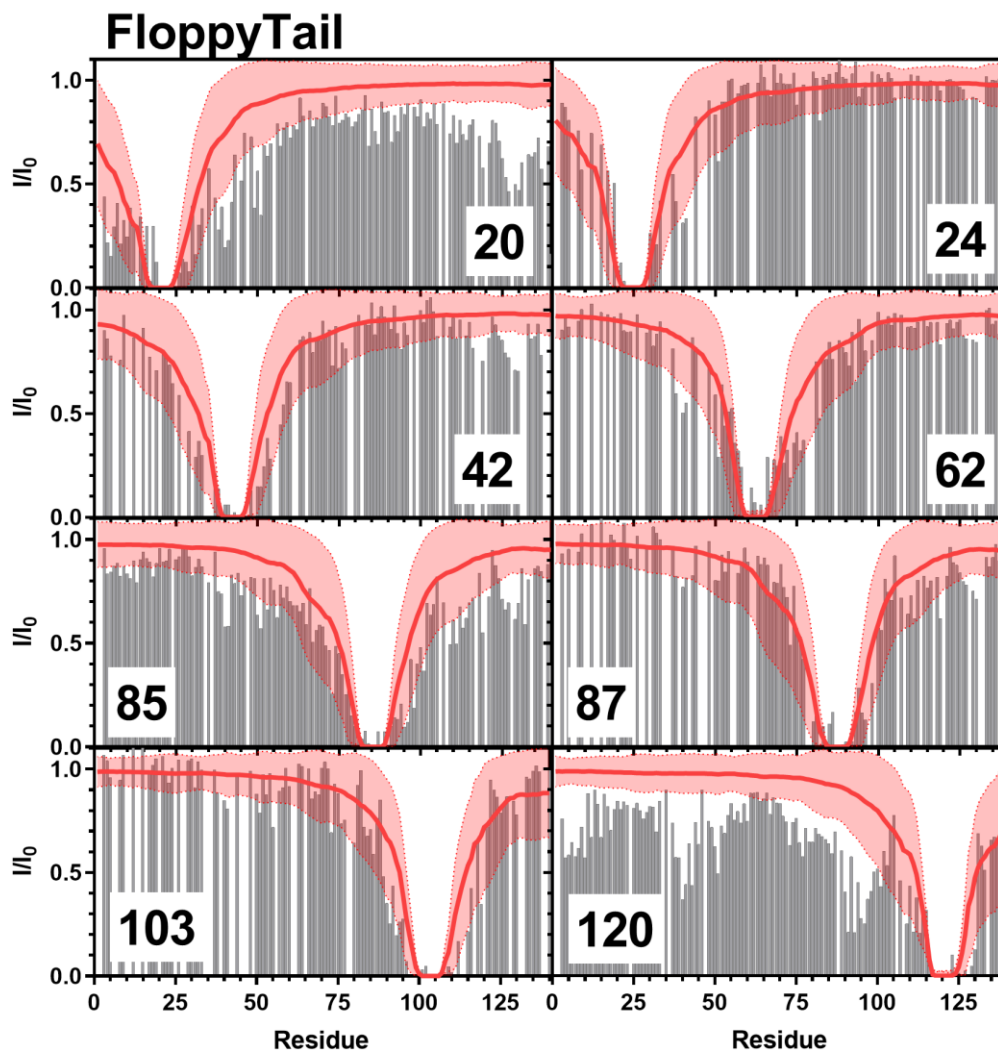


Figure 6.27: Comparison of Simulated FloppyTail PRE Values and Experimental PRE Values. Simulated PRE values from FloppyTail (red line) overlaid on top of experimental data (grey bars) from positions 20 (Top Left), 24 (Top Right), 42 (Upper Middle Left), 62 (Upper Middle Right), 85 (Lower Middle Left), 87 (Lower Middle Right), 103 (Bottom Left), 120 (Bottom Right). Experimental data for positions 20, 85, and 120 are from Sung *et al.*²⁰¹ and data for positions 24, 42, 62, 87, and 103 are from Dedmon *et al.*¹⁰⁴

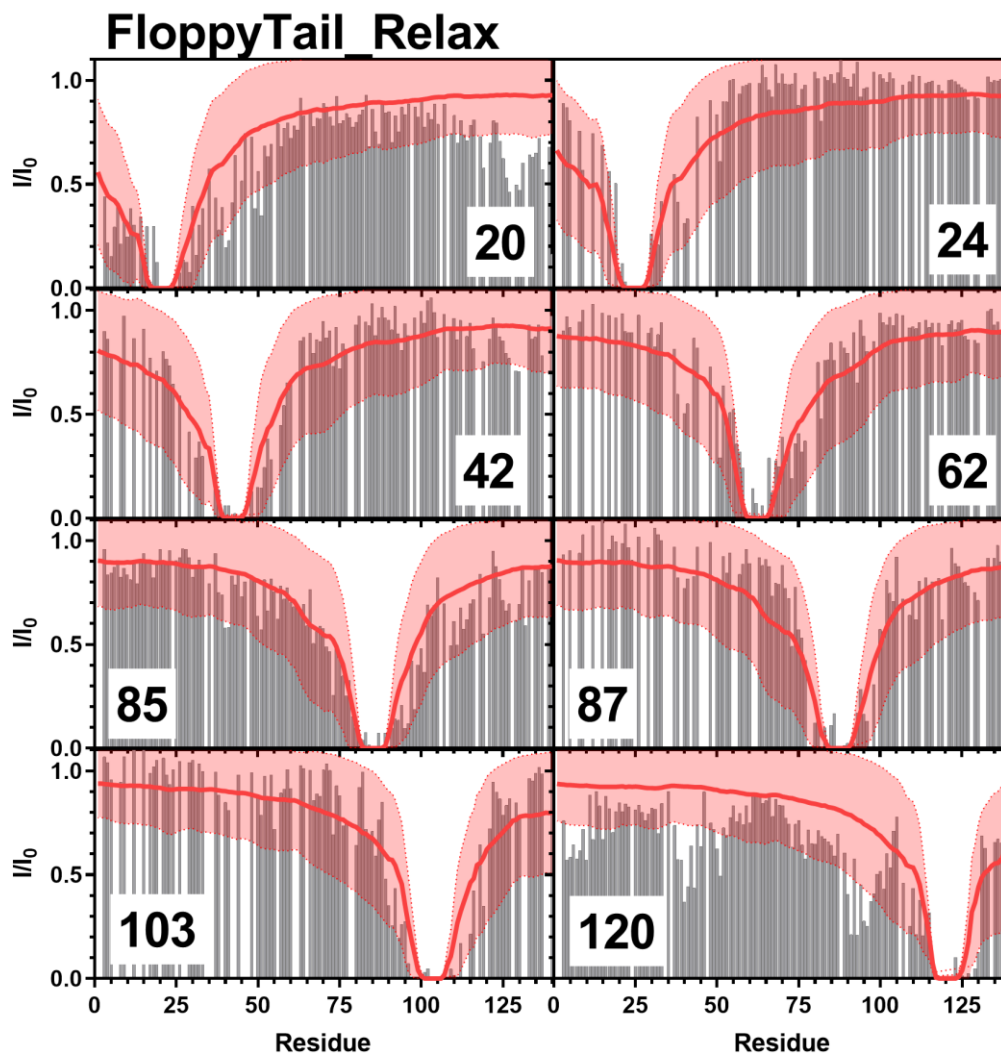


Figure 6.28: Comparison of Simulated FloppyTail_Relax PRE Values and Experimental PRE Values. Simulated PRE values from FloppyTail_Relax (red line) overlaid on top of experimental data (grey bars) from positions 20 (Top Left), 24 (Top Right), 42 (Upper Middle Left), 62 (Upper Middle Right), 85 (Lower Middle Left), 87 (Lower Middle Right), 103 (Bottom Left), 120 (Bottom Right). Experimental data for positions 20, 85, and 120 are from Sung *et al.*²⁰¹ and data for positions 24, 42, 62, 87, and 103 are from Dedmon *et al.*¹⁰⁴

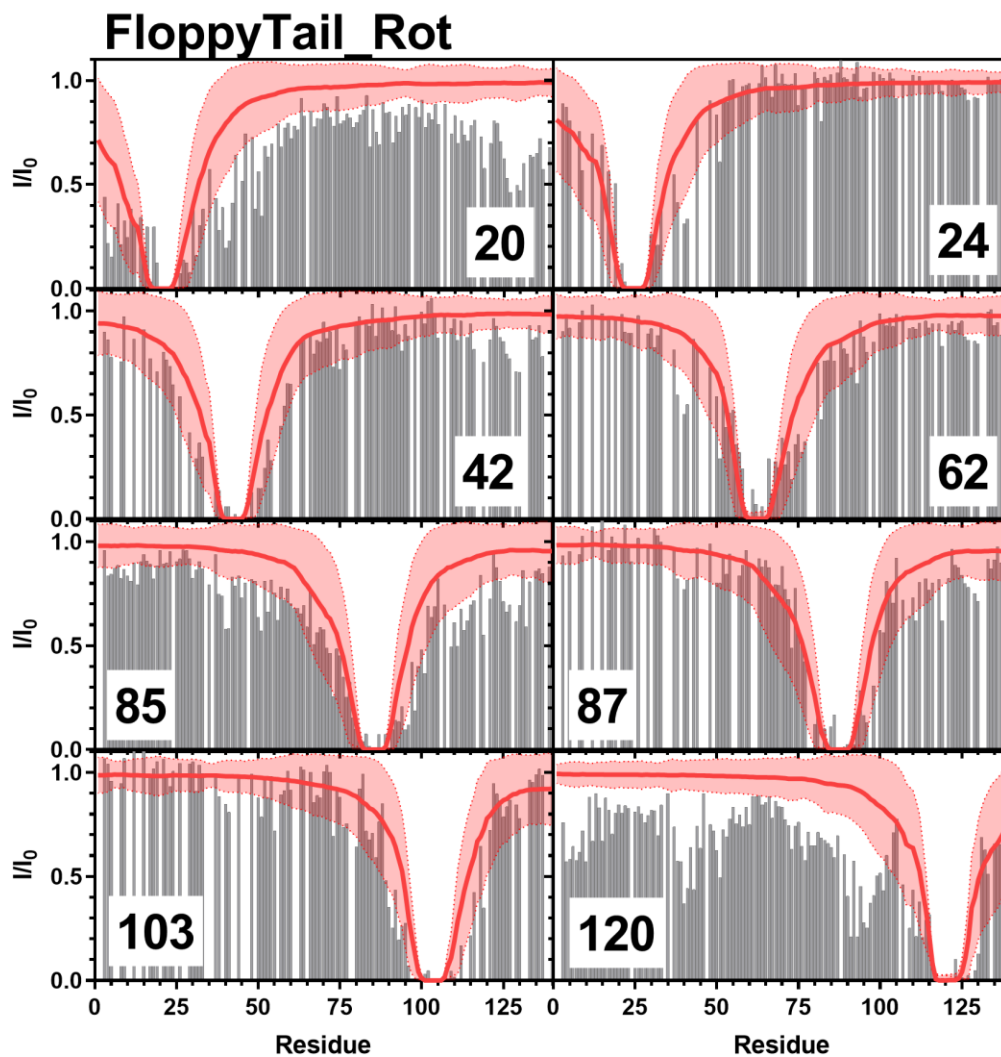


Figure 6.29: Comparison of Simulated FloppyTail_Rot PRE Values and Experimental PRE Values. Simulated PRE values from FloppyTail_Rot (red line) overlaid on top of experimental data (grey bars) from positions 20 (Top Left), 24 (Top Right), 42 (Upper Middle Left), 62 (Upper Middle Right), 85 (Lower Middle Left), 87 (Lower Middle Right), 103 (Bottom Left), 120 (Bottom Right). Experimental data for positions 20, 85, and 120 are from Sung *et al.*²⁰¹ and data for positions 24, 42, 62, 87, and 103 are from Dedmon *et al.*¹⁰⁴

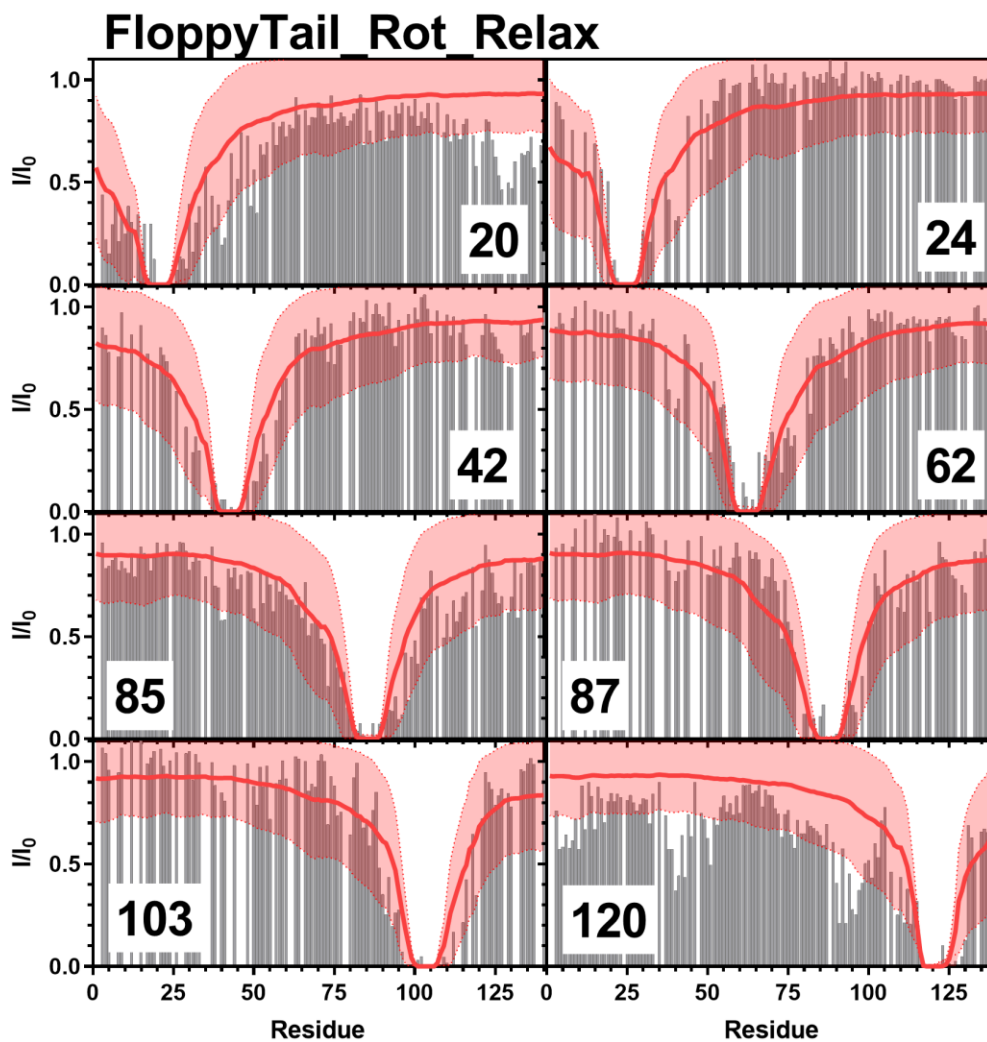


Figure 6.30: Comparison of Simulated FloppyTail_Rot_Relax PRE Values and Experimental PRE Values. Simulated PRE values from FloppyTail_Rot_Relax (red line) overlaid on top of experimental data (grey bars) from positions 20 (Top Left), 24 (Top Right), 42 (Upper Middle Left), 62 (Upper Middle Right), 85 (Lower Middle Left), 87 (Lower Middle Right), 103 (Bottom Left), 120 (Bottom Right). Experimental data for positions 20, 85, and 120 are from Sung *et al.*²⁰¹ and data for positions 24, 42, 62, 87, and 103 are from Dedmon *et al.*¹⁰⁴

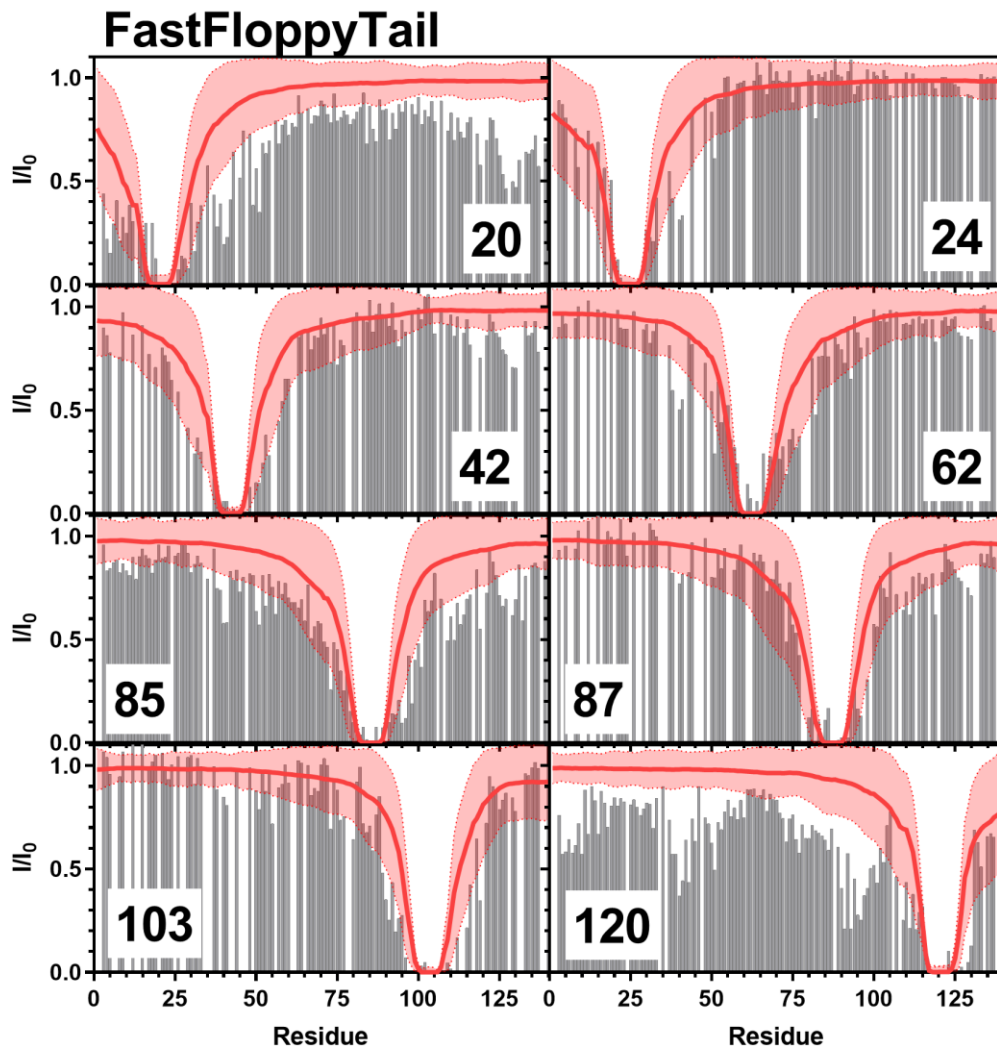


Figure 6.31: Comparison of Simulated FastFloppyTail PRE Values and Experimental PRE Values. Simulated PRE values from FastFloppyTail (red line) overlaid on top of experimental data (grey bars) from positions 20 (Top Left), 24 (Top Right), 42 (Upper Middle Left), 62 (Upper Middle Right), 85 (Lower Middle Left), 87 (Lower Middle Right), 103 (Bottom Left), 120 (Bottom Right). Experimental data for positions 20, 85, and 120 are from Sung *et al.*²⁰¹ and data for positions 24, 42, 62, 87, and 103 are from Dedmon *et al.*¹⁰⁴

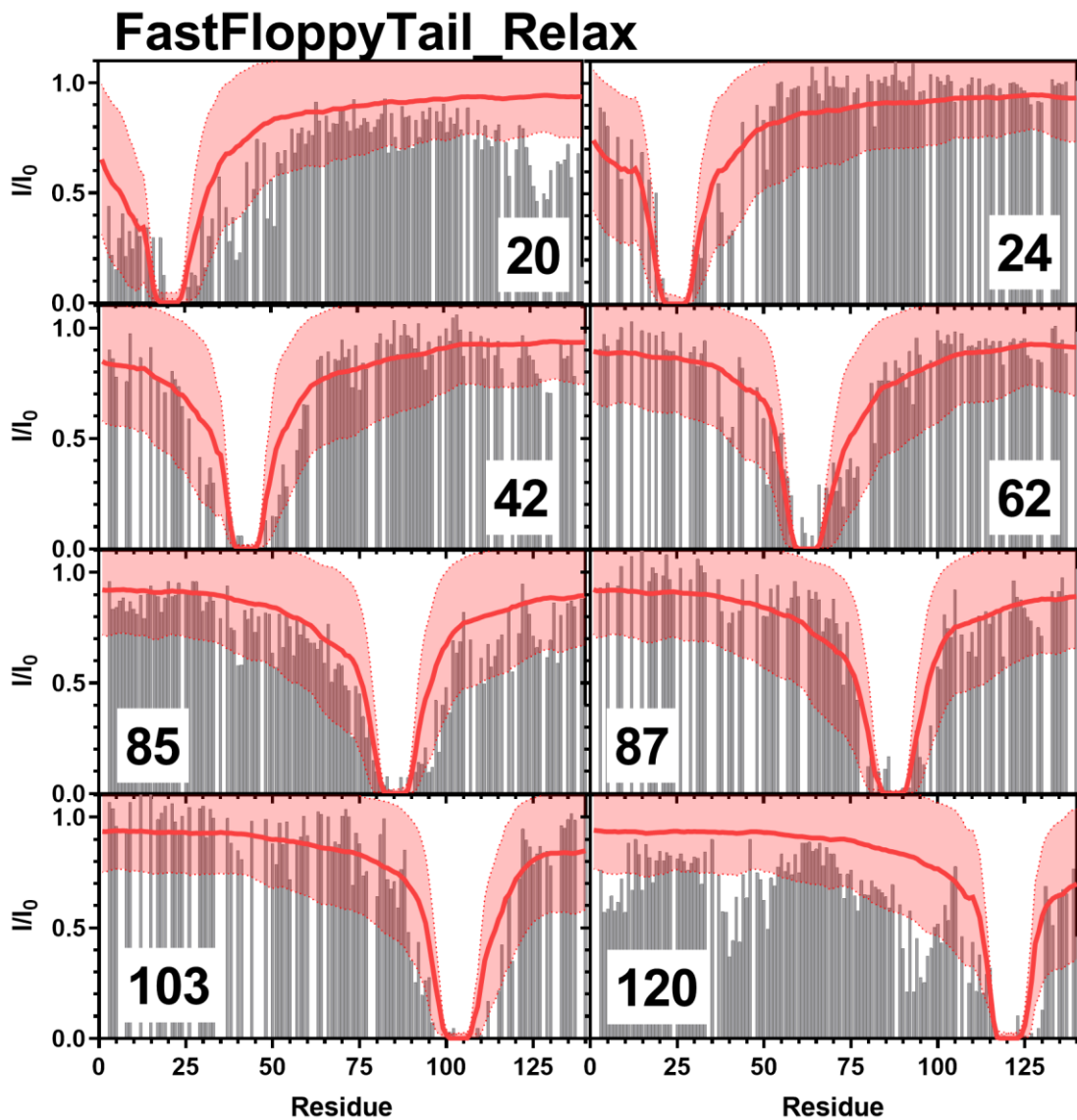


Figure 6.32: Comparison of Simulated FastFloppyTail PRE Values and Experimental PRE Values. Simulated PRE values from FastFloppyTail (red line) overlaid on top of experimental data (grey bars) from positions 20 (Top Left), 24 (Top Right), 42 (Upper Middle Left), 62 (Upper Middle Right), 85 (Lower Middle Left), 87 (Lower Middle Right), 103 (Bottom Left), 120 (Bottom Right). Experimental data for positions 20, 85, and 120 are from Sung *et al.*²⁰¹ and data for positions 24, 42, 62, 87, and 103 are from Dedmon *et al.*¹⁰⁴

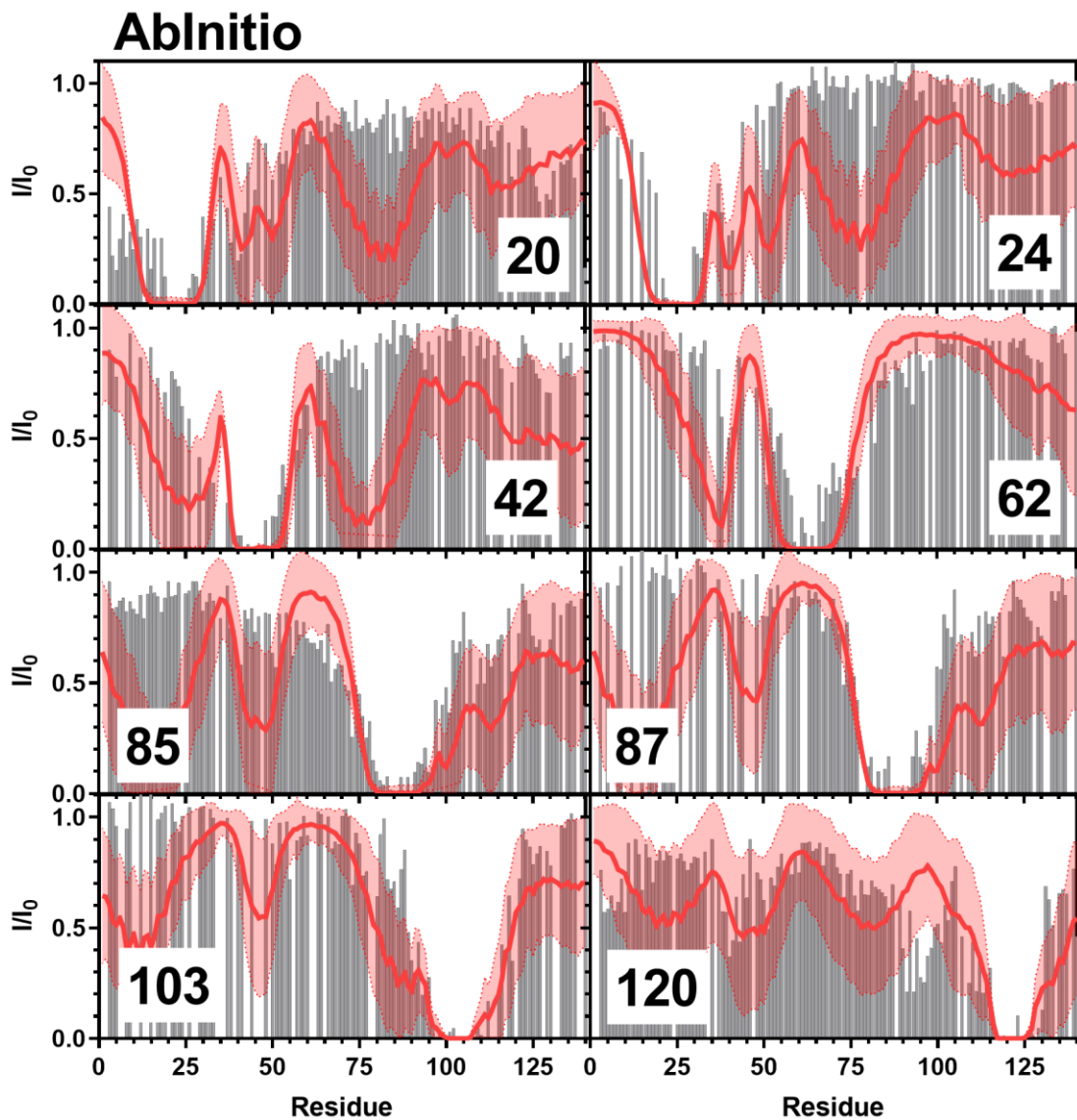


Figure 6.33: Comparison of Simulated AbInitio PRE Values and Experimental PRE Values. Simulated PRE values from AbInitio (red line) overlaid on top of experimental data (grey bars) from positions 20 (Top Left), 24 (Top Right), 42 (Upper Middle Left), 62 (Upper Middle Right), 85 (Lower Middle Left), 87 (Lower Middle Right), 103 (Bottom Left), 120 (Bottom Right). Experimental data for positions 20, 85, and 120 are from Sung *et al.*²⁰¹ and data for positions 24, 42, 62, 87, and 103 are from Dedmon *et al.*¹⁰⁴

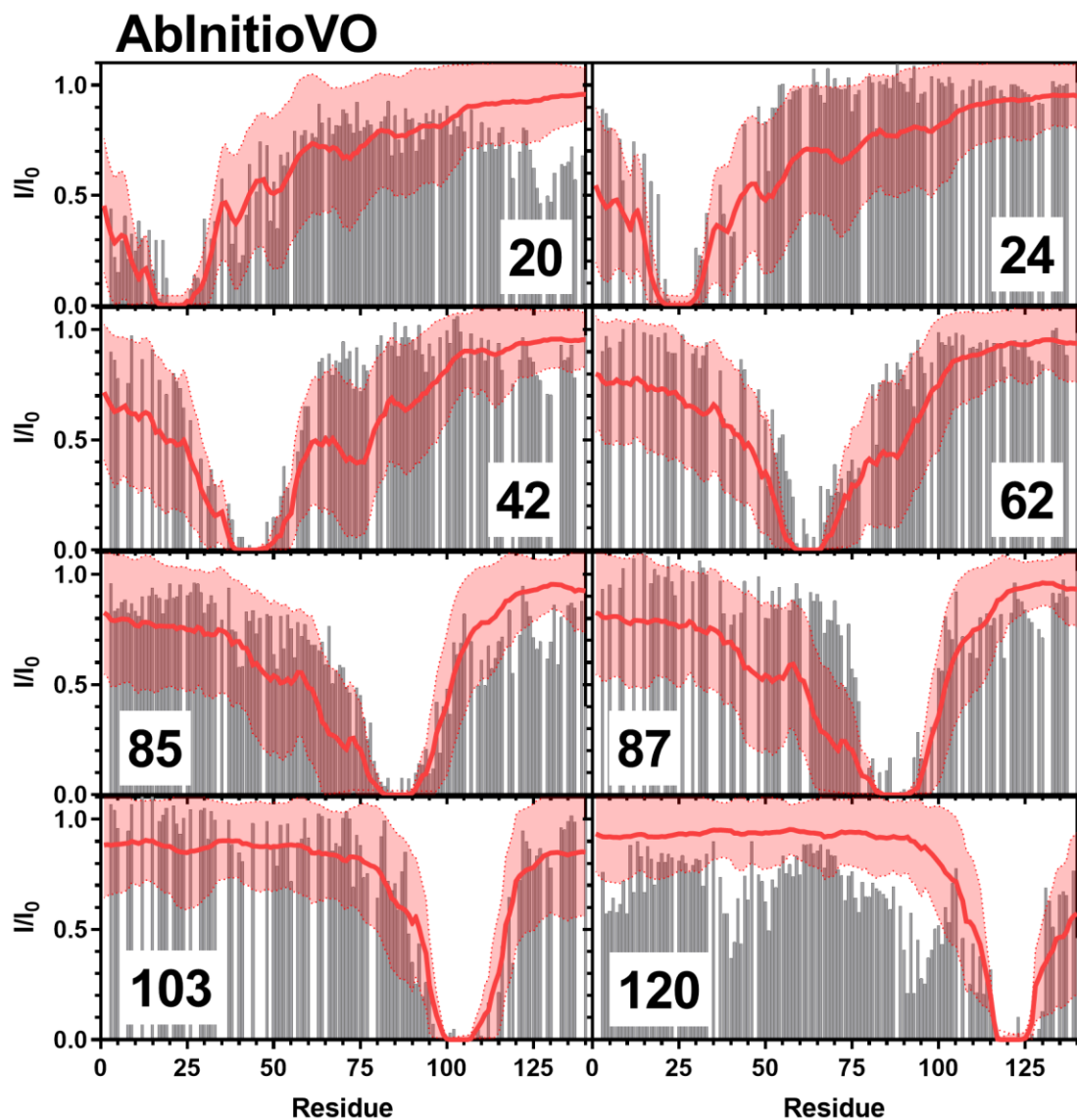


Figure 6.34: Comparison of Simulated AbInitioVO PRE Values and Experimental PRE Values. Simulated PRE values from AbInitioVO (red line) overlaid on top of experimental data (grey bars) from positions 20 (Top Left), 24 (Top Right), 42 (Upper Middle Left), 62 (Upper Middle Right), 85 (Lower Middle Left), 87 (Lower Middle Right), 103 (Bottom Left), 120 (Bottom Right). Experimental data for positions 20, 85, and 120 are from Sung *et al.*²⁰¹ and data for positions 24, 42, 62, 87, and 103 are from Dedmon *et al.*¹⁰⁴

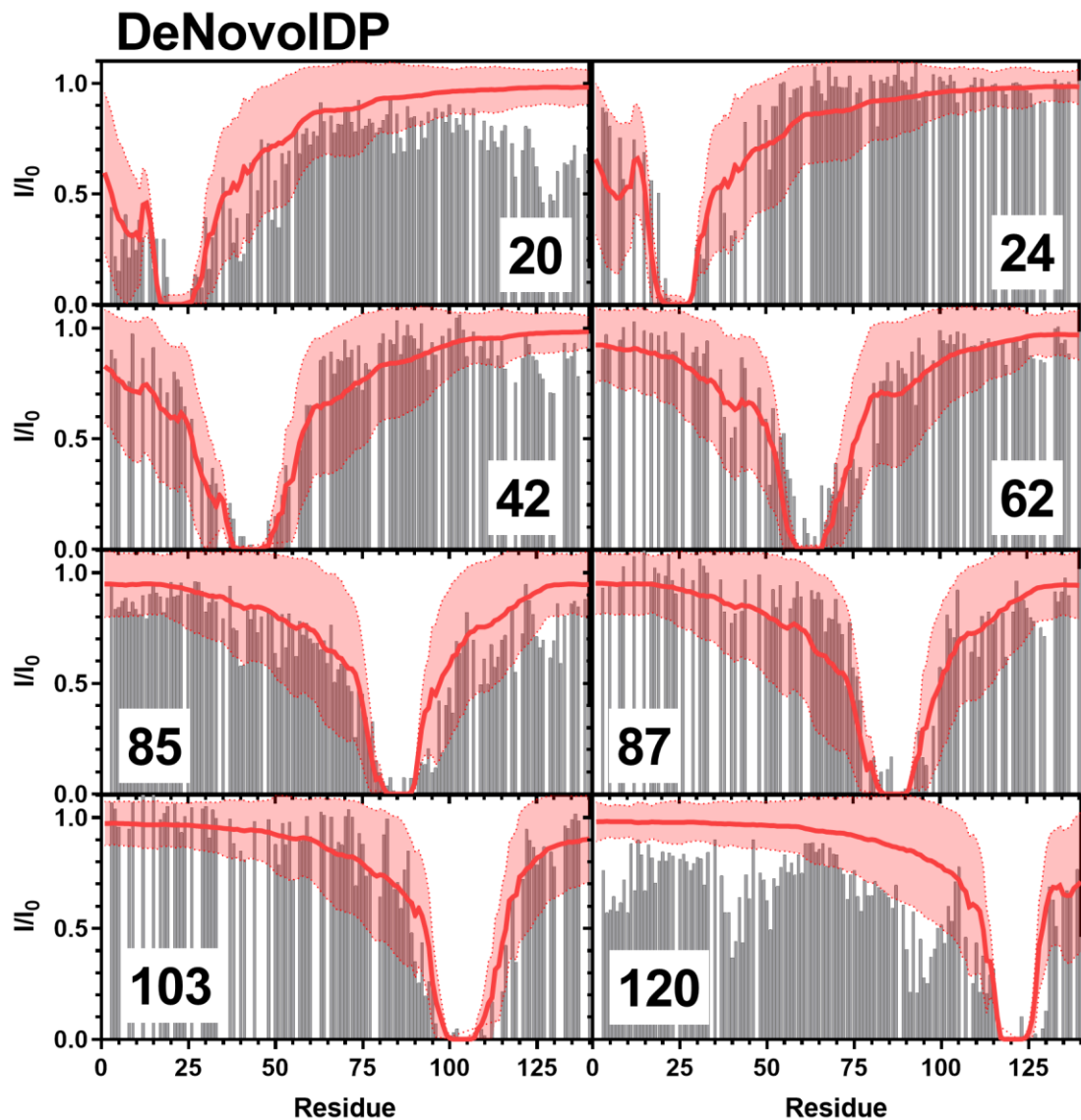


Figure 6.35: Comparison of Simulated DeNovoIDP PRE Values and Experimental PRE Values. Simulated PRE values from DeNovoIDP (red line) overlaid on top of experimental data (grey bars) from positions 20 (Top Left), 24 (Top Right), 42 (Upper Middle Left), 62 (Upper Middle Right), 85 (Lower Middle Left), 87 (Lower Middle Right), 103 (Bottom Left), 120 (Bottom Right). Experimental data for positions 20, 85, and 120 are from Sung *et al.*²⁰¹ and data for positions 24, 42, 62, 87, and 103 are from Dedmon *et al.*¹⁰⁴

Comparison with α -Synuclein Chemical Shift Data

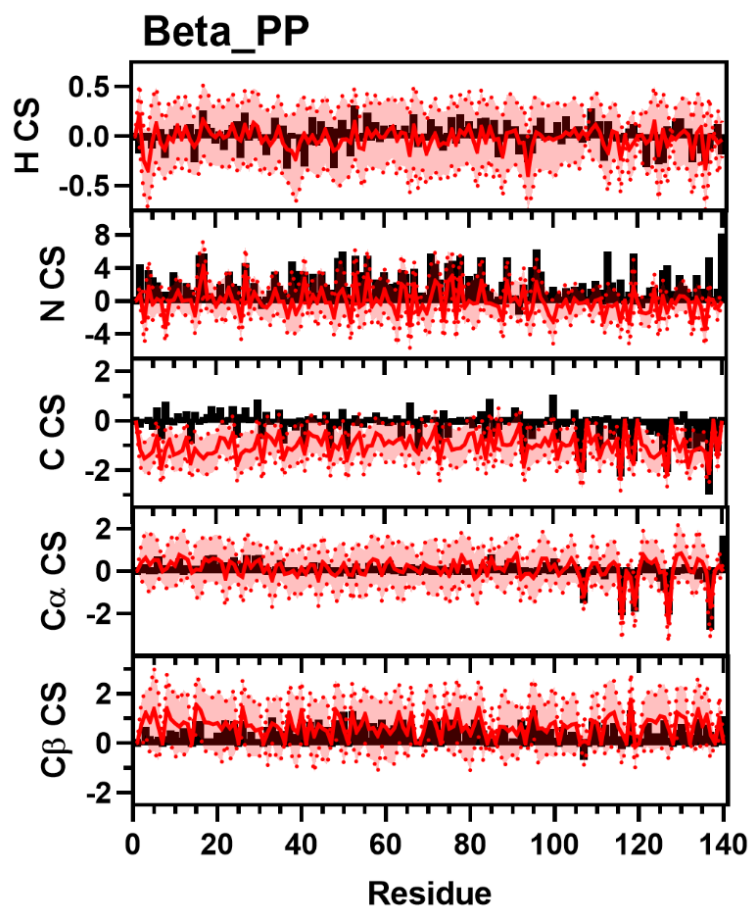


Figure 6.36: Comparison of Simulated Beta_PP and Experimental NMR Chemical Shift Data. Simulated Chemical Shift values from the Beta_PP ensemble (red) overlaid on experimental data (black bars) of N (Top), H (Upper Middle), C (Middle), C α (Lower Middle), and C β (Bottom) chemical shifts from Sung *et al.*²⁰¹ Neighbor corrected random coil chemical shift values have been subtracted from both simulated and experimental data.²⁴⁵

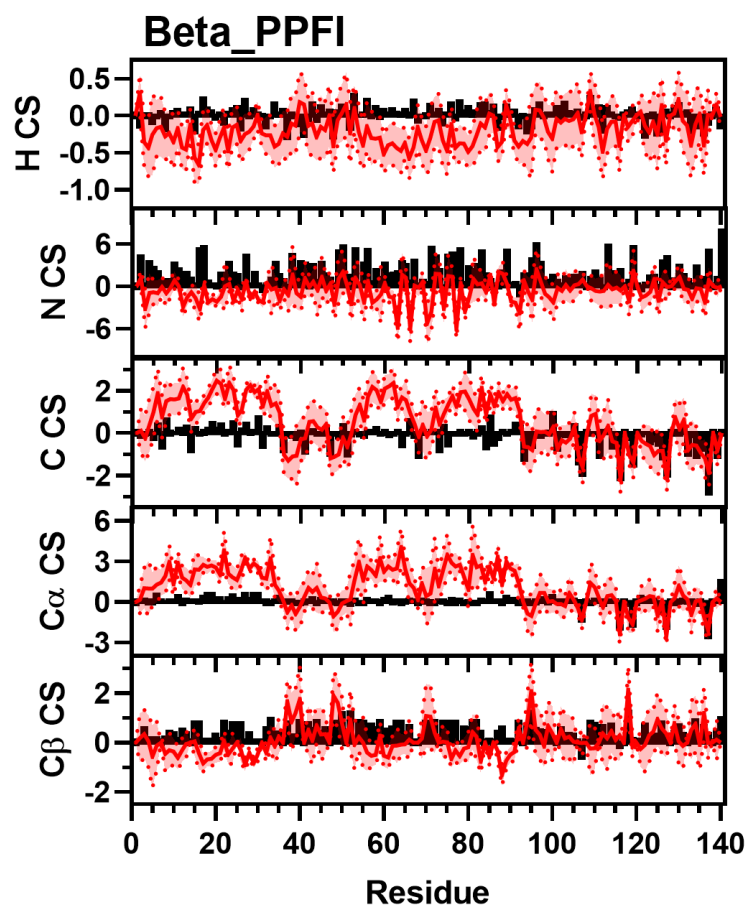


Figure 6.37: Comparison of Simulated Beta_PPF1 and Experimental NMR Chemical Shift Data. Simulated Chemical Shift values from the Beta_PPF1 ensemble (red) overlaid on experimental data (black bars) of N (Top), H (Upper Middle), C (Middle), C α (Lower Middle), and C β (Bottom) chemical shifts from Sung *et al.*²⁰¹ Neighbor corrected random coil chemical shift values have been subtracted from both simulated and experimental data.²⁴⁵

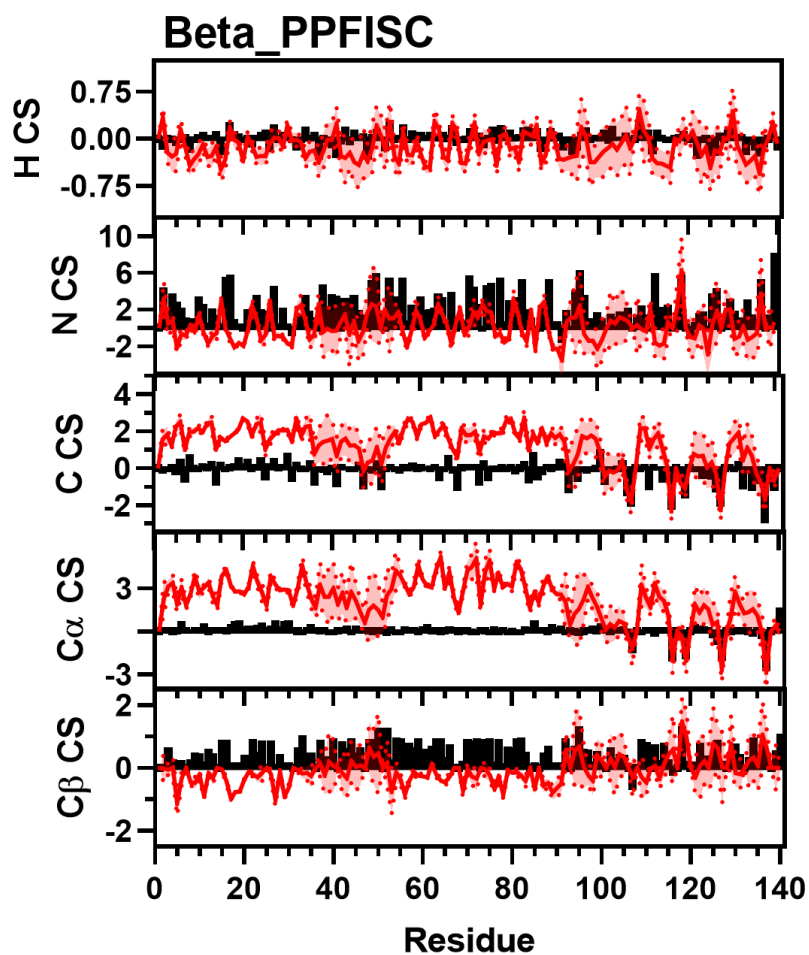


Figure 6.38: Comparison of Simulated Beta_PPFISC and Experimental NMR Chemical Shift Data. Simulated Chemical Shift values from the Beta_PPFISC ensemble (red) overlaid on experimental data (black bars) of N (Top), H (Upper Middle), C (Middle), C α (Lower Middle), and C β (Bottom) chemical shifts from Sung *et al.*²⁰¹ Neighbor corrected random coil chemical shift values have been subtracted from both simulated and experimental data.²⁴⁵

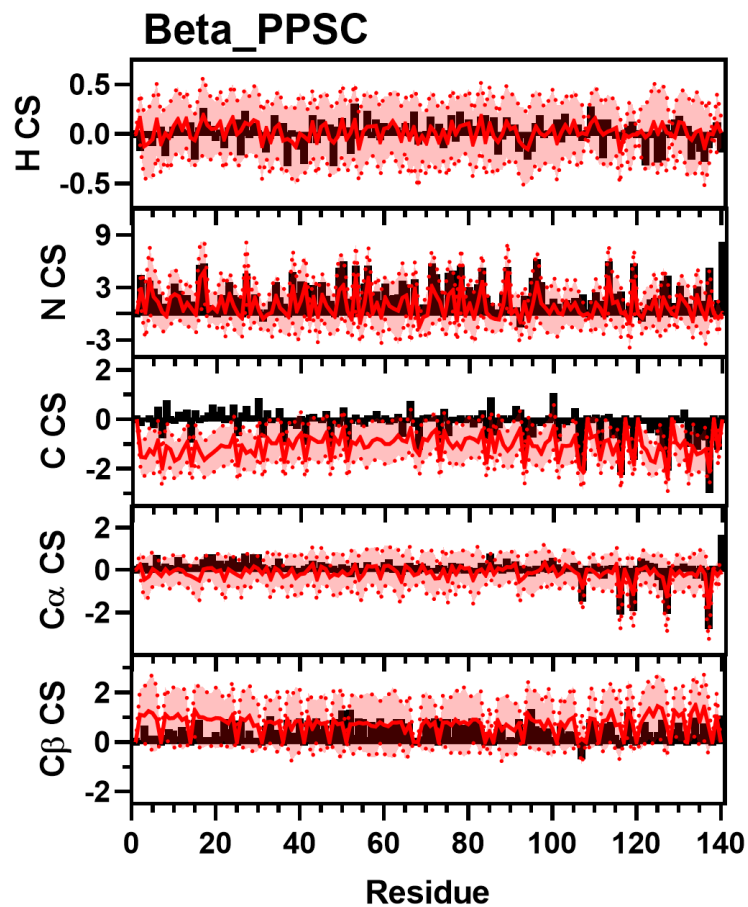


Figure 6.39: Comparison of Simulated Beta_PPSC and Experimental NMR Chemical Shift Data. Simulated Chemical Shift values from the Beta_PPSC ensemble (red) overlaid on experimental data (black bars) of N (Top), H (Upper Middle), C (Middle), C α (Lower Middle), and C β (Bottom) chemical shifts from Sung *et al.*²⁰¹ Neighbor corrected random coil chemical shift values have been subtracted from both simulated and experimental data.²⁴⁵

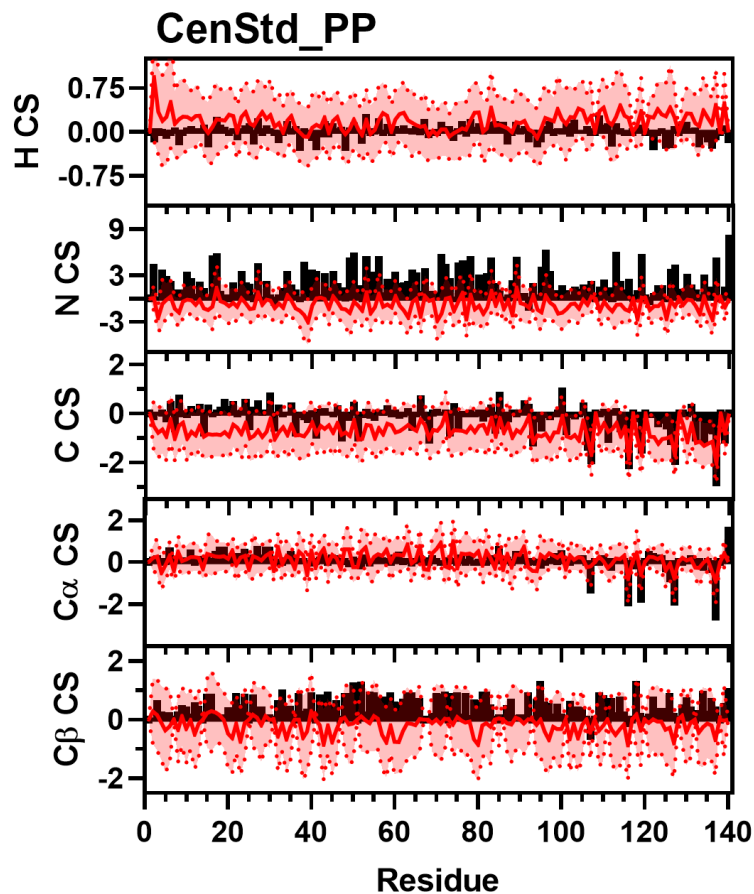


Figure 6.40: Comparison of Simulated CenStd_PP and Experimental NMR Chemical Shift Data. Simulated Chemical Shift values from the CenStd_PP ensemble (red) overlaid on experimental data (black bars) of N (Top), H (Upper Middle), C (Middle), C α (Lower Middle), and C β (Bottom) chemical shifts from Sung *et al.*²⁰¹ Neighbor corrected random coil chemical shift values have been subtracted from both simulated and experimental data.²⁴⁵

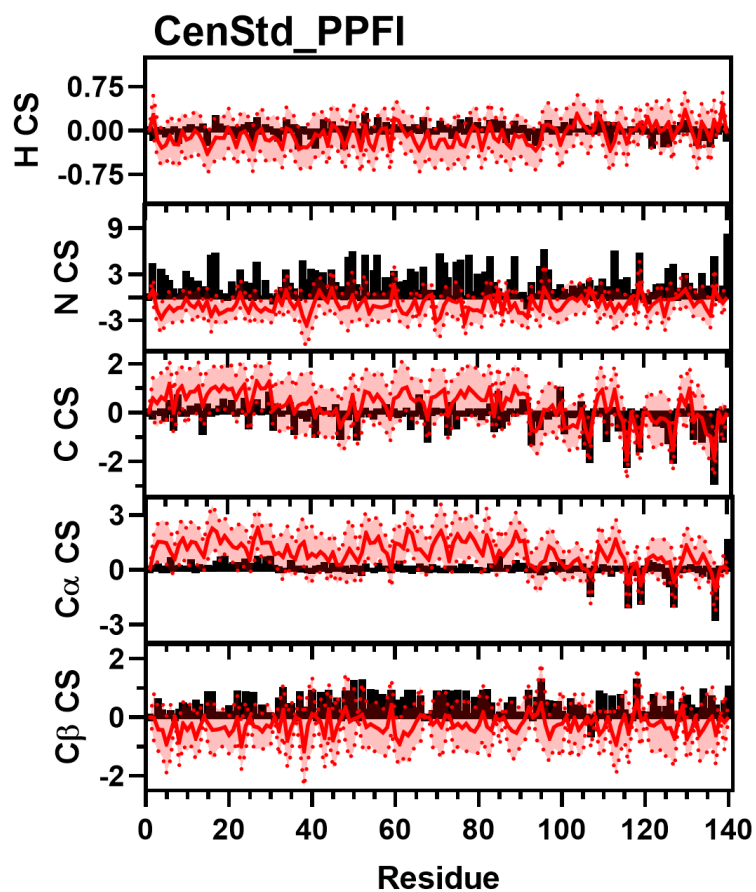


Figure 6.41: Comparison of Simulated CenStd_PPFI and Experimental NMR Chemical Shift Data. Simulated Chemical Shift values from the CenStd_PPFI ensemble (red) overlaid on experimental data (black bars) of N (Top), H (Upper Middle), C (Middle), C α (Lower Middle), and C β (Bottom) chemical shifts from Sung *et al.*²⁰¹ Neighbor corrected random coil chemical shift values have been subtracted from both simulated and experimental data.²⁴⁵

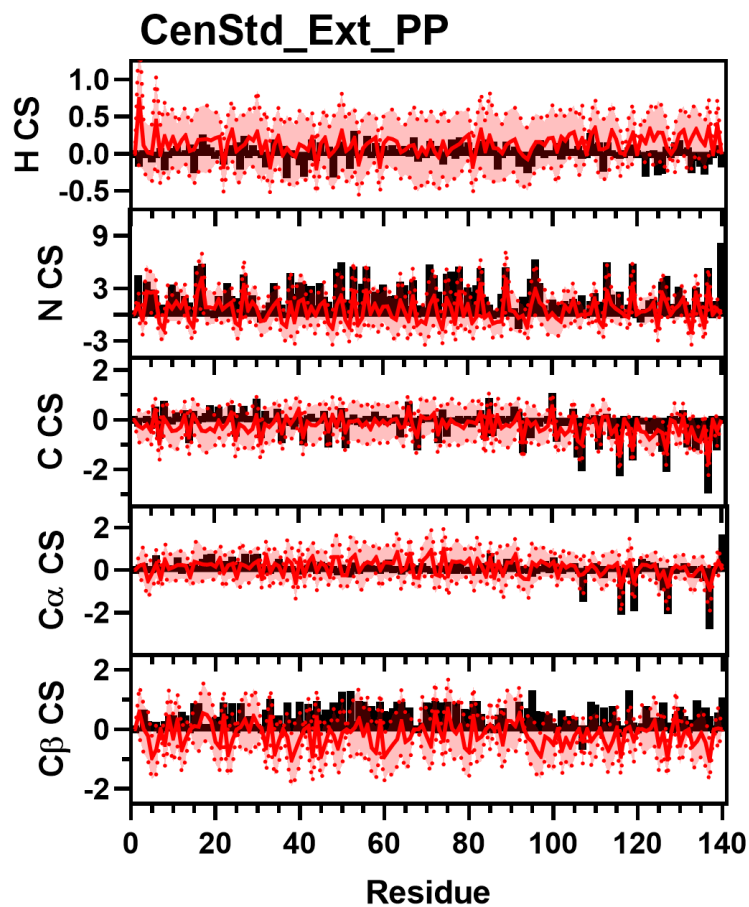


Figure 6.42: Comparison of Simulated CenStd_Ext_PP and Experimental NMR Chemical Shift Data. Simulated Chemical Shift values from the CenStd_Ext_PP ensemble (red) overlaid on experimental data (black bars) of N (Top), H (Upper Middle), C (Middle), C α (Lower Middle), and C β (Bottom) chemical shifts from Sung *et al.*²⁰¹ Neighbor corrected random coil chemical shift values have been subtracted from both simulated and experimental data.²⁴⁵

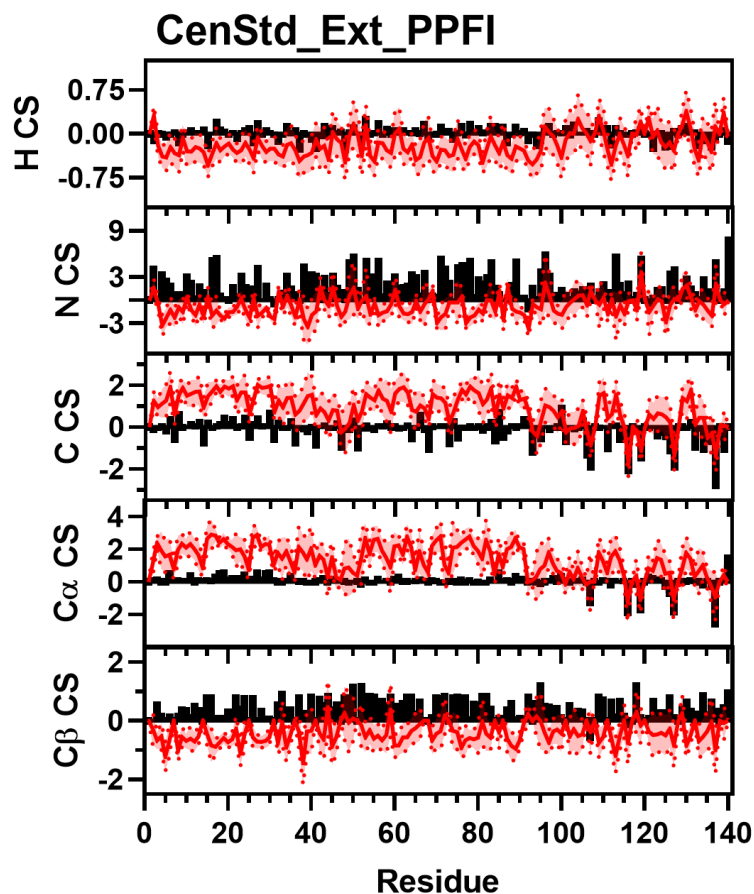


Figure 6.43: Comparison of Simulated CenStd_Ext_PPFI and Experimental NMR Chemical Shift Data. Simulated Chemical Shift values from the CenStd_Ext_PPFI ensemble (red) overlaid on experimental data (black bars) of N (Top), H (Upper Middle), C (Middle), C α (Lower Middle), and C β (Bottom) chemical shifts from Sung *et al.*²⁰¹ Neighbor corrected random coil chemical shift values have been subtracted from both simulated and experimental data.²⁴⁵

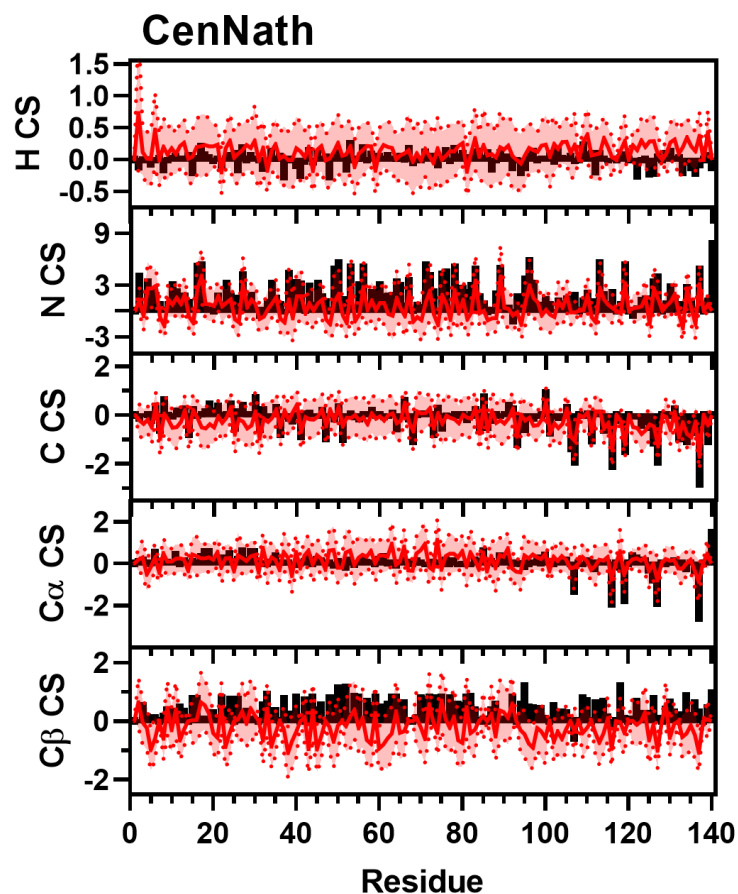


Figure 6.44: Comparison of Simulated CenNath and Experimental NMR Chemical Shift Data. Simulated Chemical Shift values from the CenNath ensemble (red) overlaid on experimental data (black bars) of N (Top), H (Upper Middle), C (Middle), C α (Lower Middle), and C β (Bottom) chemical shifts from Sung *et al.*²⁰¹ Neighbor corrected random coil chemical shift values have been subtracted from both simulated and experimental data.²⁴⁵

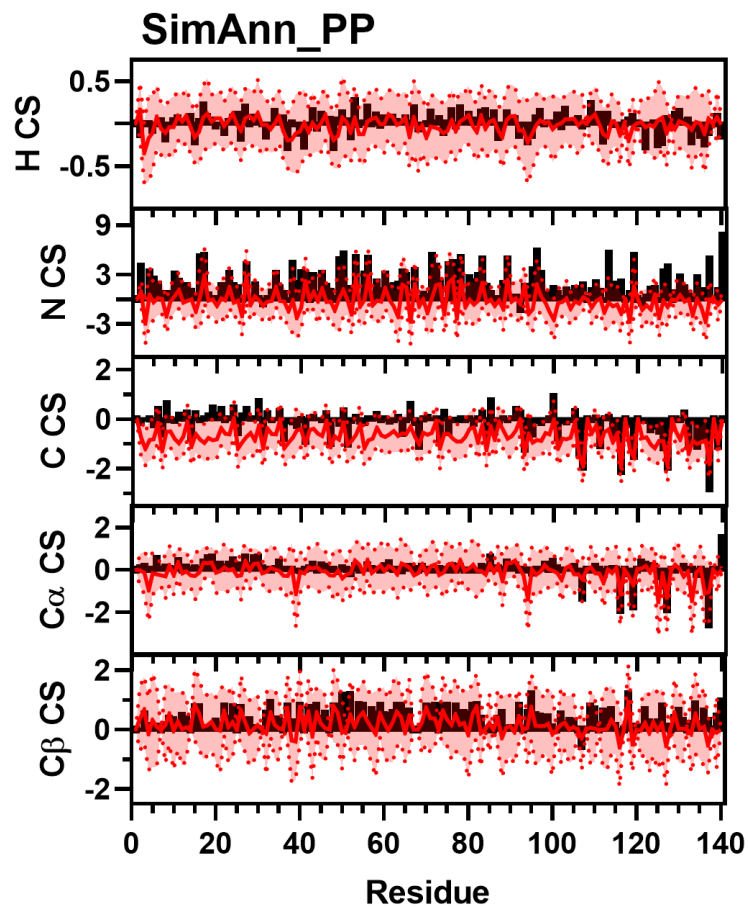


Figure 6.45: Comparison of Simulated SimAnn_PP and Experimental NMR Chemical Shift Data. Simulated Chemical Shift values from the SimAnn_PP ensemble (red) overlaid on experimental data (black bars) of N (Top), H (Upper Middle), C (Middle), C α (Lower Middle), and C β (Bottom) chemical shifts from Sung *et al.*²⁰¹ Neighbor corrected random coil chemical shift values have been subtracted from both simulated and experimental data.²⁴⁵

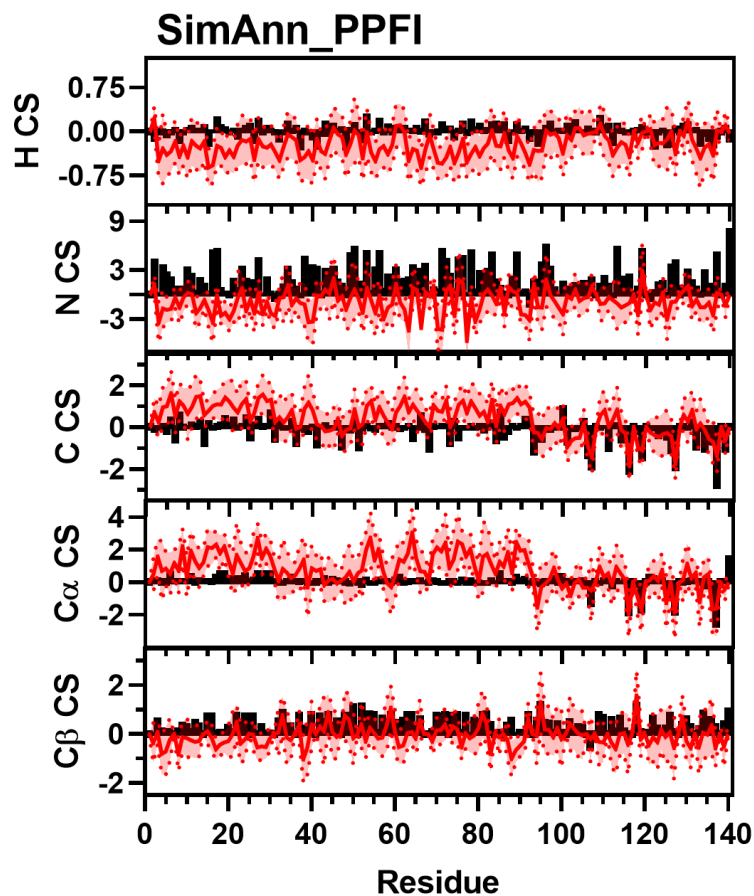


Figure 6.46: Comparison of Simulated SimAnn_PPFI and Experimental NMR Chemical Shift Data. Simulated Chemical Shift values from the SimAnn_PPFI ensemble (red) overlaid on experimental data (black bars) of N (Top), H (Upper Middle), C (Middle), C α (Lower Middle), and C β (Bottom) chemical shifts from Sung *et al.*²⁰¹ Neighbor corrected random coil chemical shift values have been subtracted from both simulated and experimental data.²⁴⁵

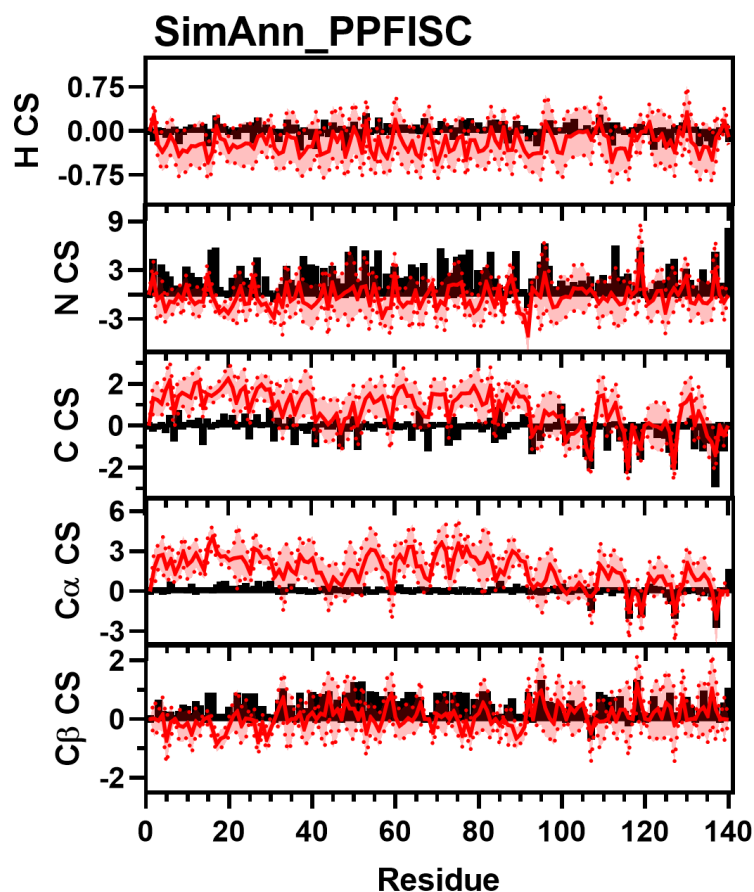


Figure 6.47: Comparison of Simulated SimAnn_PPFISC and Experimental NMR Chemical Shift Data. Simulated Chemical Shift values from the SimAnn_PPFISC ensemble (red) overlaid on experimental data (black bars) of N (Top), H (Upper Middle), C (Middle), C α (Lower Middle), and C β (Bottom) chemical shifts from Sung *et al.*²⁰¹ Neighbor corrected random coil chemical shift values have been subtracted from both simulated and experimental data.²⁴⁵

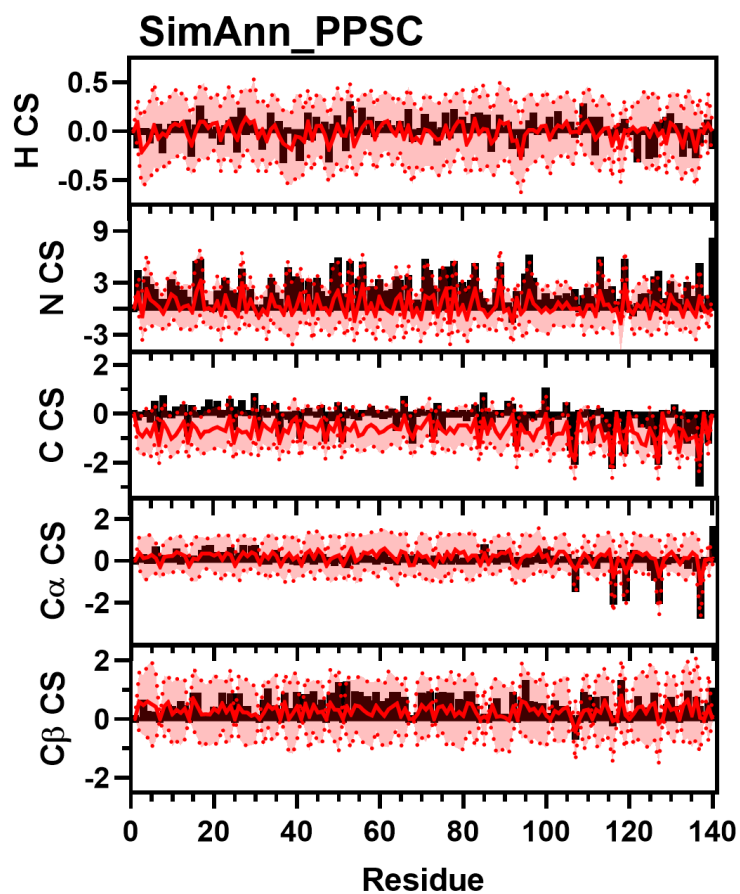


Figure 6.48: Comparison of Simulated SimAnn_PPSC and Experimental NMR Chemical Shift Data. Simulated Chemical Shift values from the SimAnn_PPSC ensemble (red) overlaid on experimental data (black bars) of N (Top), H (Upper Middle), C (Middle), C α (Lower Middle), and C β (Bottom) chemical shifts from Sung *et al.*²⁰¹ Neighbor corrected random coil chemical shift values have been subtracted from both simulated and experimental data.²⁴⁵

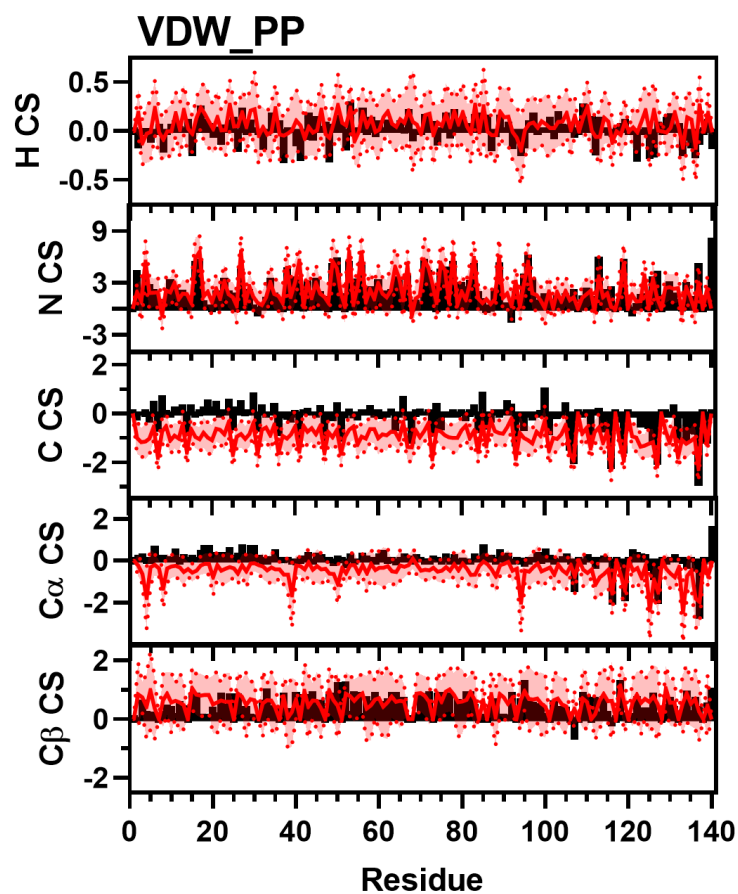


Figure 6.49: Comparison of Simulated VDW_PP and Experimental NMR Chemical Shift Data. Simulated Chemical Shift values from the VDW_PP ensemble (red) overlaid on experimental data (black bars) of N (Top), H (Upper Middle), C (Middle), C α (Lower Middle), and C β (Bottom) chemical shifts from Sung *et al.*²⁰¹ Neighbor corrected random coil chemical shift values have been subtracted from both simulated and experimental data.²⁴⁵

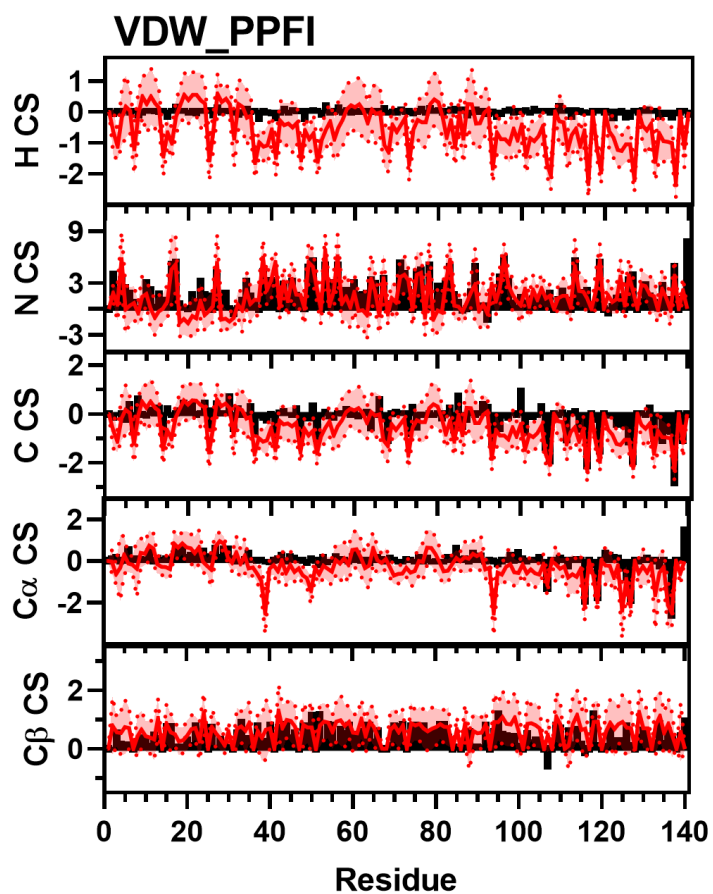


Figure 6.50: Comparison of Simulated VDW_PPFI and Experimental NMR Chemical Shift Data. Simulated Chemical Shift values from the VDW_PPFI ensemble (red) overlaid on experimental data (black bars) of N (Top), H (Upper Middle), C (Middle), C α (Lower Middle), and C β (Bottom) chemical shifts from Sung *et al.*²⁰¹ Neighbor corrected random coil chemical shift values have been subtracted from both simulated and experimental data.²⁴⁵

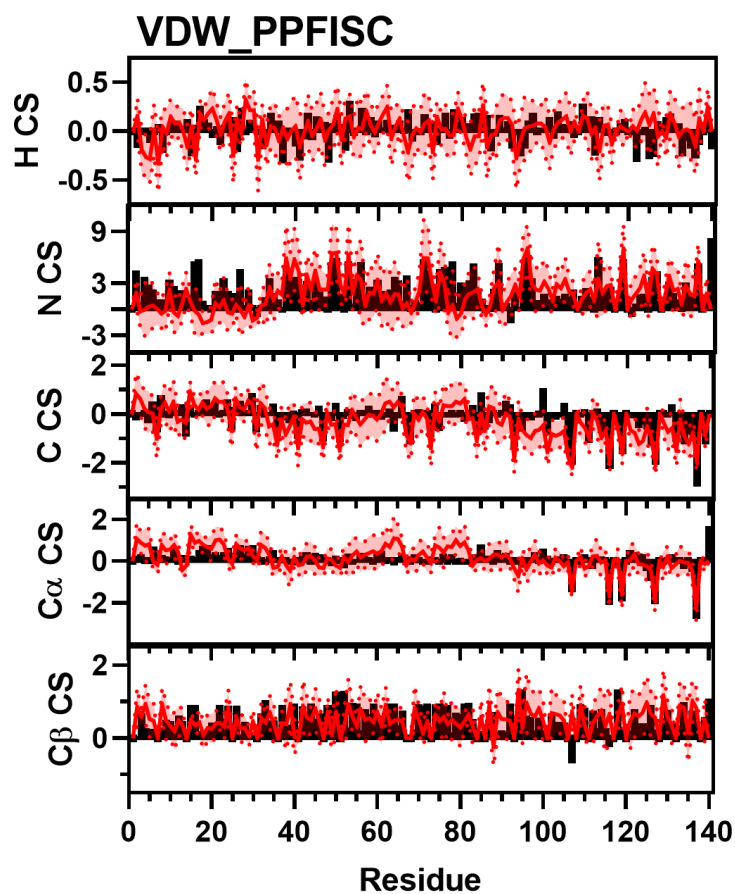


Figure 6.51: Comparison of Simulated VDW_PPFISC and Experimental NMR Chemical Shift Data. Simulated Chemical Shift values from the VDW_PPFISC ensemble (red) overlaid on experimental data (black bars) of N (Top), H (Upper Middle), C (Middle), C α (Lower Middle), and C β (Bottom) chemical shifts from Sung *et al.*²⁰¹ Neighbor corrected random coil chemical shift values have been subtracted from both simulated and experimental data.²⁴⁵

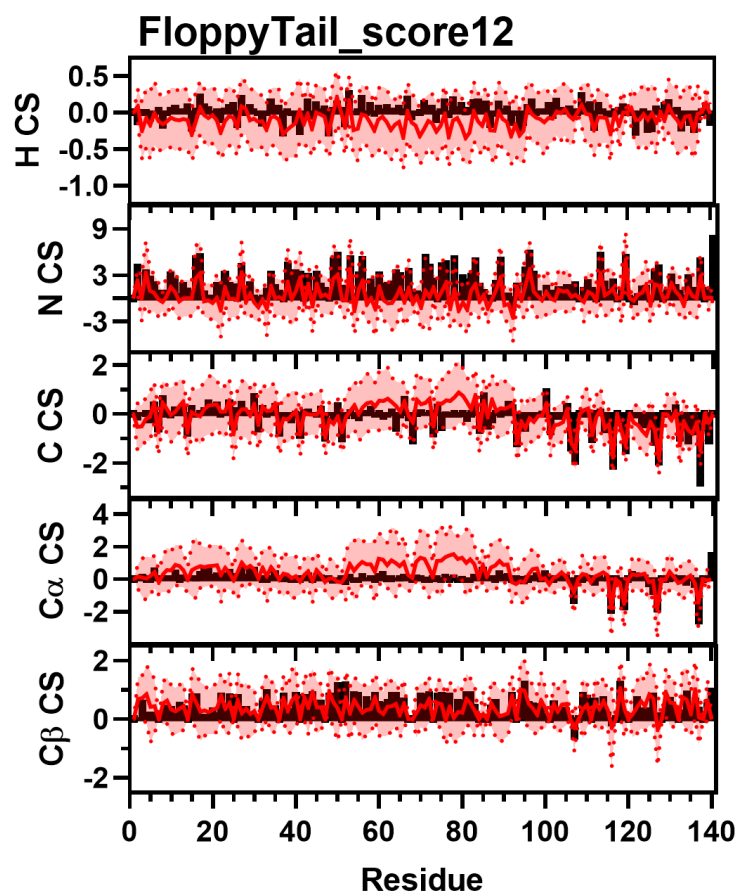


Figure 6.52: Comparison of Simulated FloppyTail_score12 and Experimental NMR Chemical Shift Data. Simulated Chemical Shift values from the FloppyTail_score12 ensemble (red) overlaid on experimental data (black bars) of N (Top), H (Upper Middle), C (Middle), $C\alpha$ (Lower Middle), and $C\beta$ (Bottom) chemical shifts from Sung *et al.*²⁰¹ Neighbor corrected random coil chemical shift values have been subtracted from both simulated and experimental data.²⁴⁵

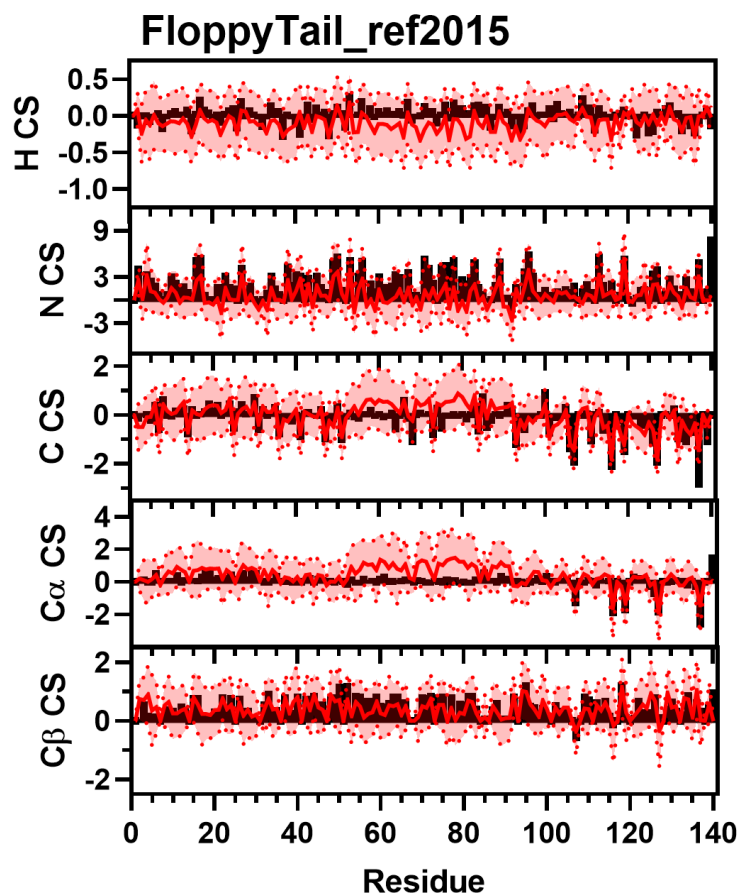


Figure 6.53: Comparison of Simulated FloppyTail_ref2015 and Experimental NMR Chemical Shift Data. Simulated Chemical Shift values from the FloppyTail_ref2015 ensemble (red) overlaid on experimental data (black bars) of N (Top), H (Upper Middle), C (Middle), C α (Lower Middle), and C β (Bottom) chemical shifts from Sung *et al.*²⁰¹ Neighbor corrected random coil chemical shift values have been subtracted from both simulated and experimental data.²⁴⁵

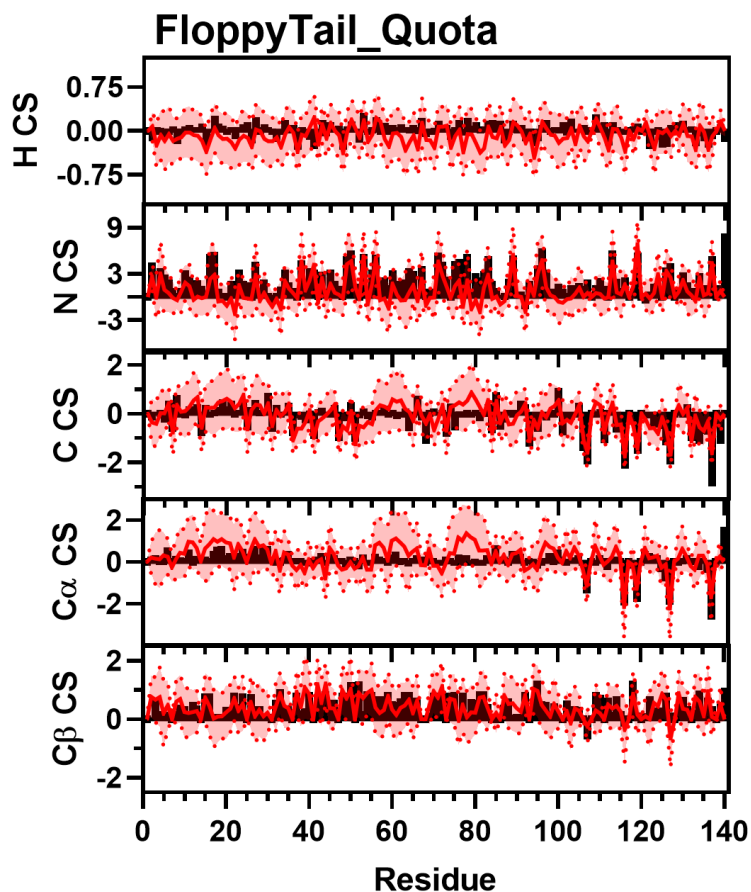


Figure 6.54: Comparison of Simulated FloppyTail_Quota and Experimental NMR Chemical Shift Data. Simulated Chemical Shift values from the FloppyTail_Quota ensemble (red) overlaid on experimental data (black bars) of N (Top), H (Upper Middle), C (Middle), C α (Lower Middle), and C β (Bottom) chemical shifts from Sung *et al.*²⁰¹ Neighbor corrected random coil chemical shift values have been subtracted from both simulated and experimental data.²⁴⁵

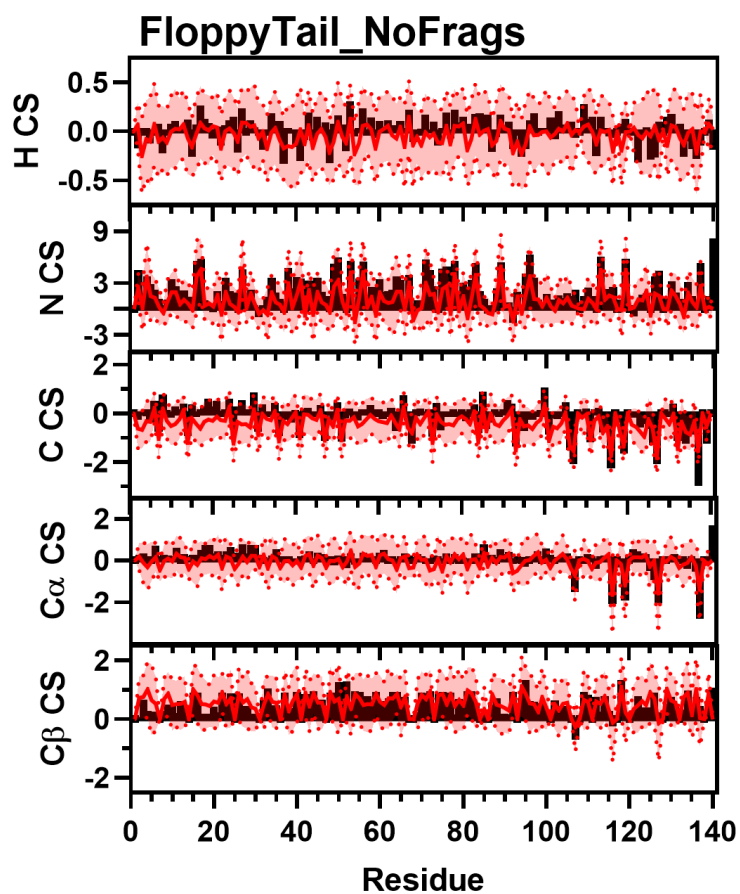


Figure 6.55: Comparison of Simulated FloppyTail_NoFrag and Experimental NMR Chemical Shift Data. Simulated Chemical Shift values from the FloppyTail_NoFrag ensemble (red) overlaid on experimental data (black bars) of N (Top), H (Upper Middle), C (Middle), C α (Lower Middle), and C β (Bottom) chemical shifts from Sung *et al.*²⁰¹ Neighbor corrected random coil chemical shift values have been subtracted from both simulated and experimental data.²⁴⁵

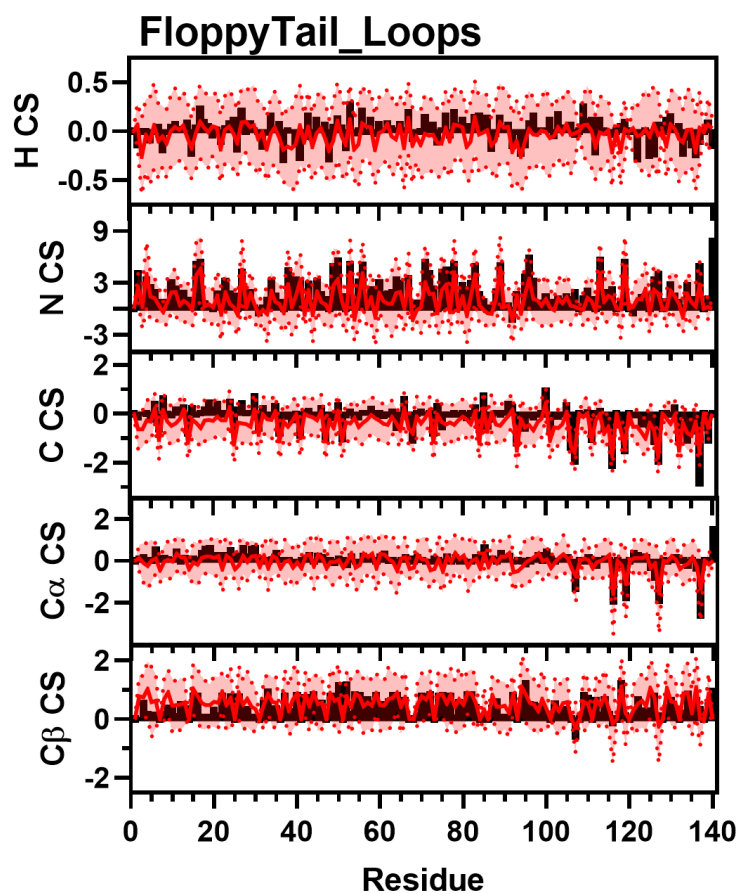


Figure 6.56: Comparison of Simulated FloppyTail_Loops and Experimental NMR Chemical Shift Data. Simulated Chemical Shift values from FloppyTail_Loops ensemble (red) overlaid on experimental data (black bars) of N (Top), H (Upper Middle), C (Middle), C α (Lower Middle), and C β (Bottom) chemical shifts from Sung *et al.*²⁰¹ Neighbor corrected random coil chemical shift values have been subtracted from both simulated and experimental data.²⁴⁵

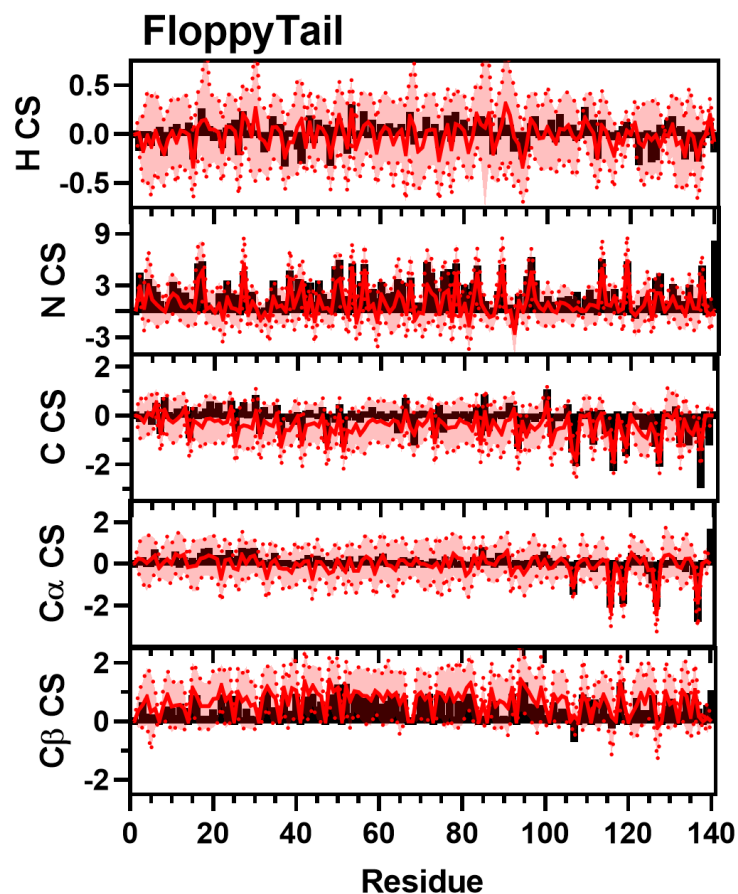


Figure 6.57: Comparison of Simulated FloppyTail and Experimental NMR Chemical Shift Data. Simulated Chemical Shift values from the FloppyTail ensemble (red) overlaid on experimental data (black bars) of N (Top), H (Upper Middle), C (Middle), C α (Lower Middle), and C β (Bottom) chemical shifts from Sung *et al.*²⁰¹ Neighbor corrected random coil chemical shift values have been subtracted from both simulated and experimental data.²⁴⁵

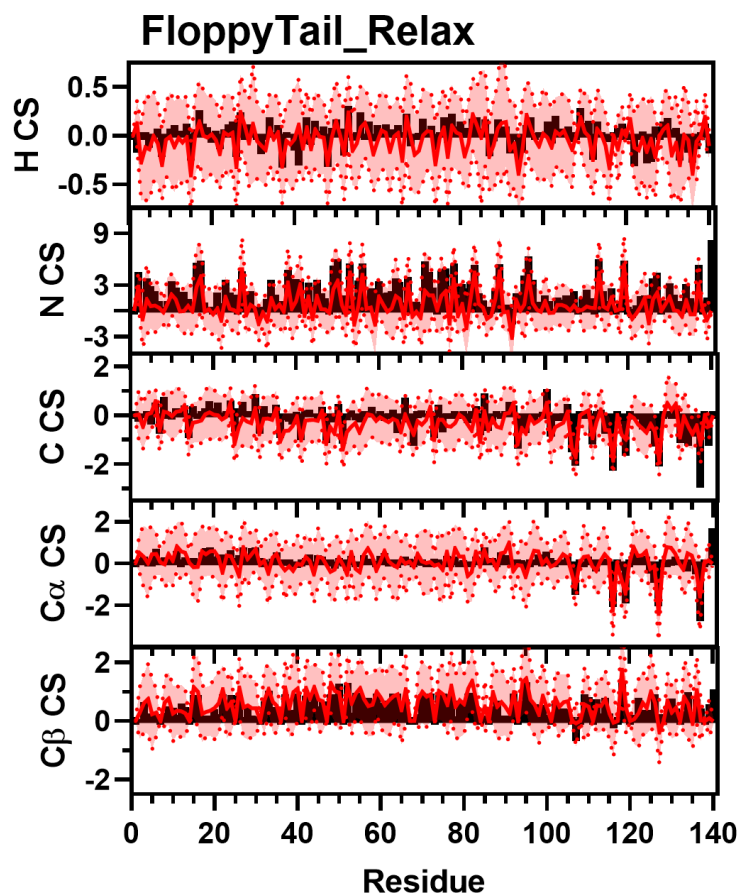


Figure 6.58: Comparison of Simulated FloppyTail_Relax and Experimental NMR Chemical Shift Data. Simulated Chemical Shift values from the FloppyTail_Relax ensemble (red) overlaid on experimental data (black bars) of N (Top), H (Upper Middle), C (Middle), C α (Lower Middle), and C β (Bottom) chemical shifts from Sung *et al.*²⁰¹ Neighbor corrected random coil chemical shift values have been subtracted from both simulated and experimental data.²⁴⁵

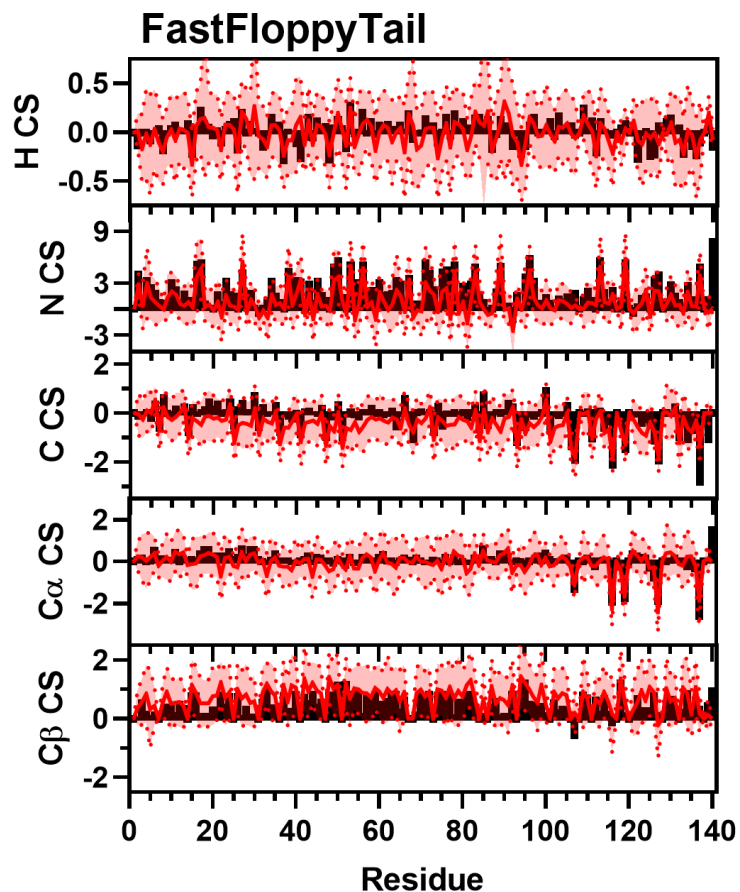


Figure 6.59: Comparison of Simulated FastFloppyTail and Experimental NMR Chemical Shift Data. Simulated Chemical Shift values from the FastFloppyTail ensemble (red) overlaid on experimental data (black bars) of N (Top), H (Upper Middle), C (Middle), C α (Lower Middle), and C β (Bottom) chemical shifts from Sung *et al.*²⁰¹ Neighbor corrected random coil chemical shift values have been subtracted from both simulated and experimental data.²⁴⁵

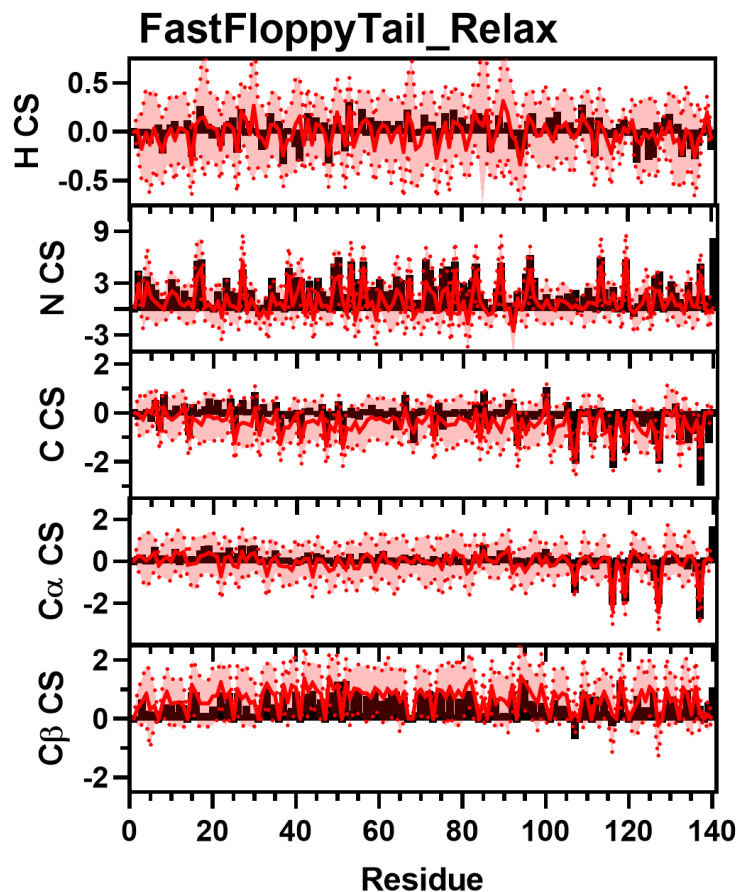


Figure 6.60: Comparison of Simulated FastFloppyTail_Relax and Experimental NMR Chemical Shift Data. Simulated Chemical Shift values from the FastFloppyTail_Relax ensemble (red) overlaid on experimental data (black bars) of N (Top), H (Upper Middle), C (Middle), C α (Lower Middle), and C β (Bottom) chemical shifts from Sung *et al.*²⁰¹ Neighbor corrected random coil chemical shift values have been subtracted from both simulated and experimental data.²⁴⁵

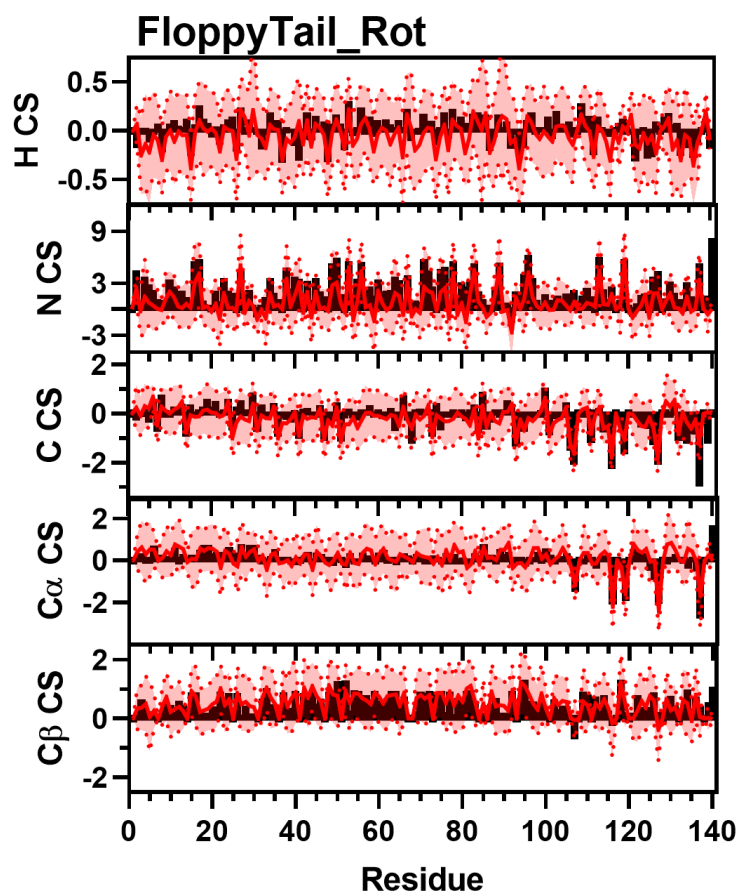


Figure 6.61: Comparison of Simulated FloppyTail_Rot and Experimental NMR Chemical Shift Data. Simulated Chemical Shift values from the FloppyTail_Rot ensemble (red) overlaid on experimental data (black bars) of N (Top), H (Upper Middle), C (Middle), C α (Lower Middle), and C β (Bottom) chemical shifts from Sung *et al.*²⁰¹ Neighbor corrected random coil chemical shift values have been subtracted from both simulated and experimental data.²⁴⁵

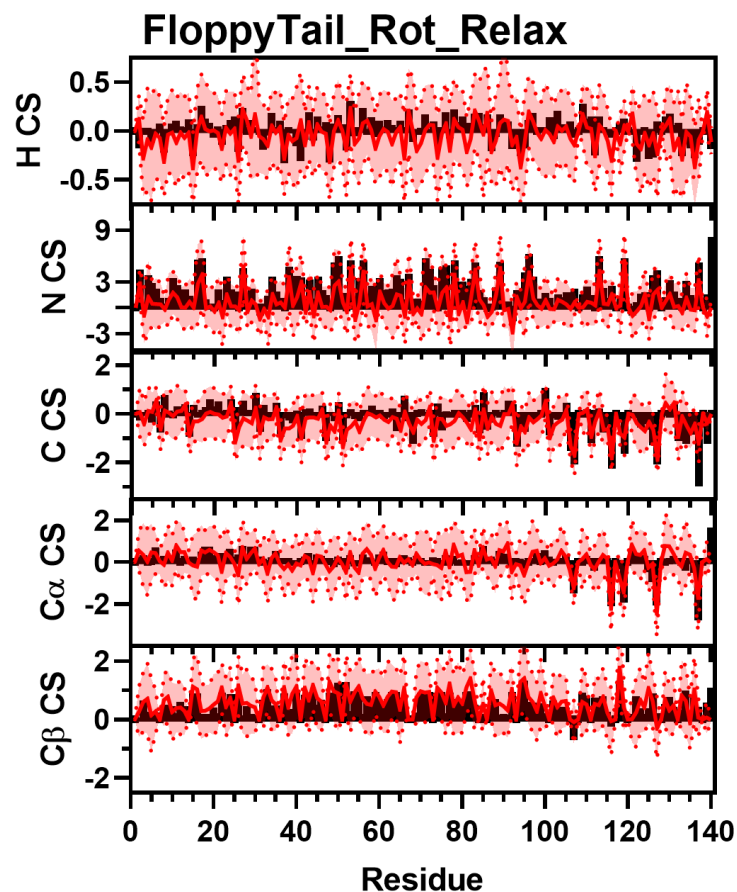


Figure 6.62: Comparison of Simulated FloppyTail_Rot_Relax and Experimental NMR Chemical Shift Data. Simulated Chemical Shift values from the FloppyTail_Rot_Relax ensemble (red) overlaid on experimental data (black bars) of N (Top), H (Upper Middle), C (Middle), C α (Lower Middle), and C β (Bottom) chemical shifts from Sung *et al.*²⁰¹ Neighbor corrected random coil chemical shift values have been subtracted from both simulated and experimental data.²⁴⁵

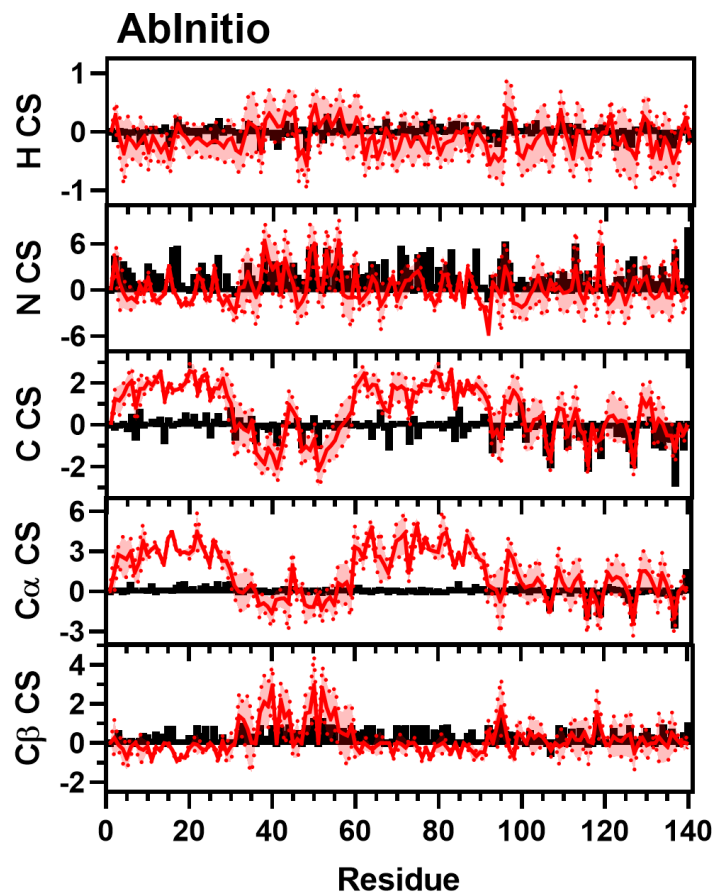


Figure 6.63: Comparison of Simulated AbInitio and Experimental NMR Chemical Shift Data. Simulated Chemical Shift values from the AbInitio ensemble (red) overlaid on experimental data (black bars) of N (Top), H (Upper Middle), C (Middle), C α (Lower Middle), and C β (Bottom) chemical shifts from Sung *et al.*²⁰¹ Neighbor corrected random coil chemical shift values have been subtracted from both simulated and experimental data.²⁴⁵

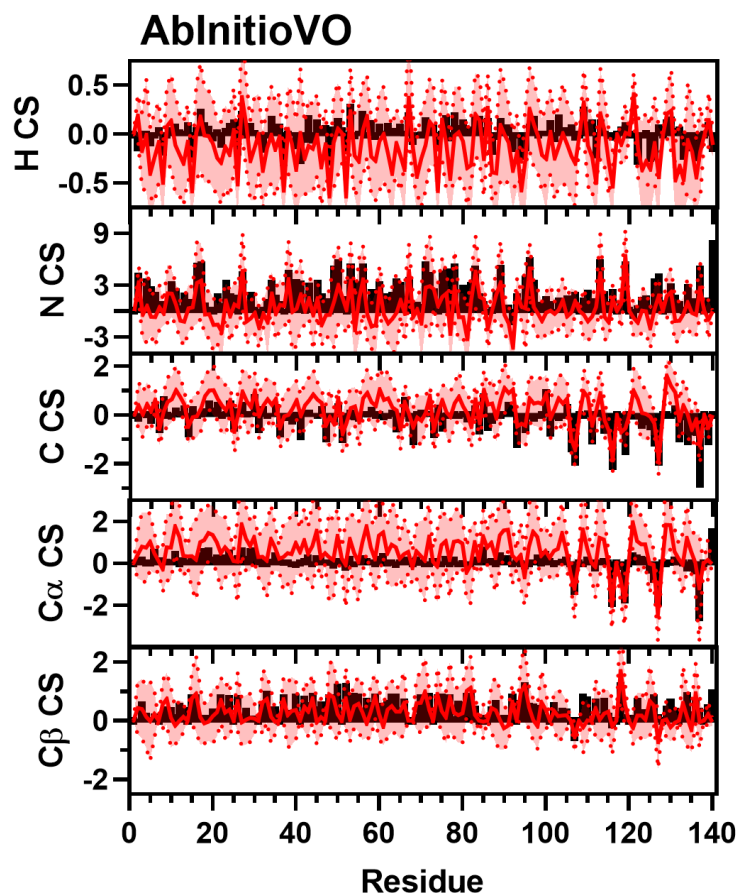


Figure 6.64: Comparison of Simulated AbInitioVO and Experimental NMR Chemical Shift Data. Simulated Chemical Shift values from the AbInitioVO ensemble (red) overlaid on experimental data (black bars) of N (Top), H (Upper Middle), C (Middle), C α (Lower Middle), and C β (Bottom) chemical shifts from Sung *et al.*²⁰¹ Neighbor corrected random coil chemical shift values have been subtracted from both simulated and experimental data.²⁴⁵

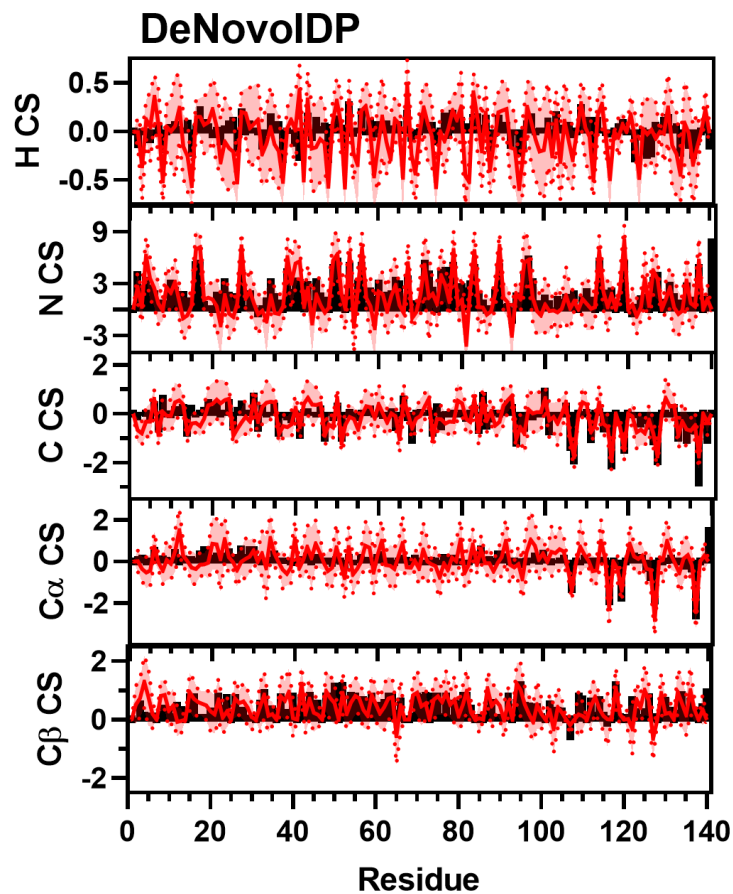


Figure 6.65: Comparison of Simulated DeNovoIDP and Experimental NMR Chemical Shift Data. Simulated Chemical Shift values from the DeNovoIDP ensemble (red) overlaid on experimental data (black bars) of N (Top), H (Upper Middle), C (Middle), C α (Lower Middle), and C β (Bottom) chemical shifts from Sung *et al.*²⁰¹ Neighbor corrected random coil chemical shift values have been subtracted from both simulated and experimental data.²⁴⁵

Comparison with α -Synuclein Residual Dipolar Coupling Data

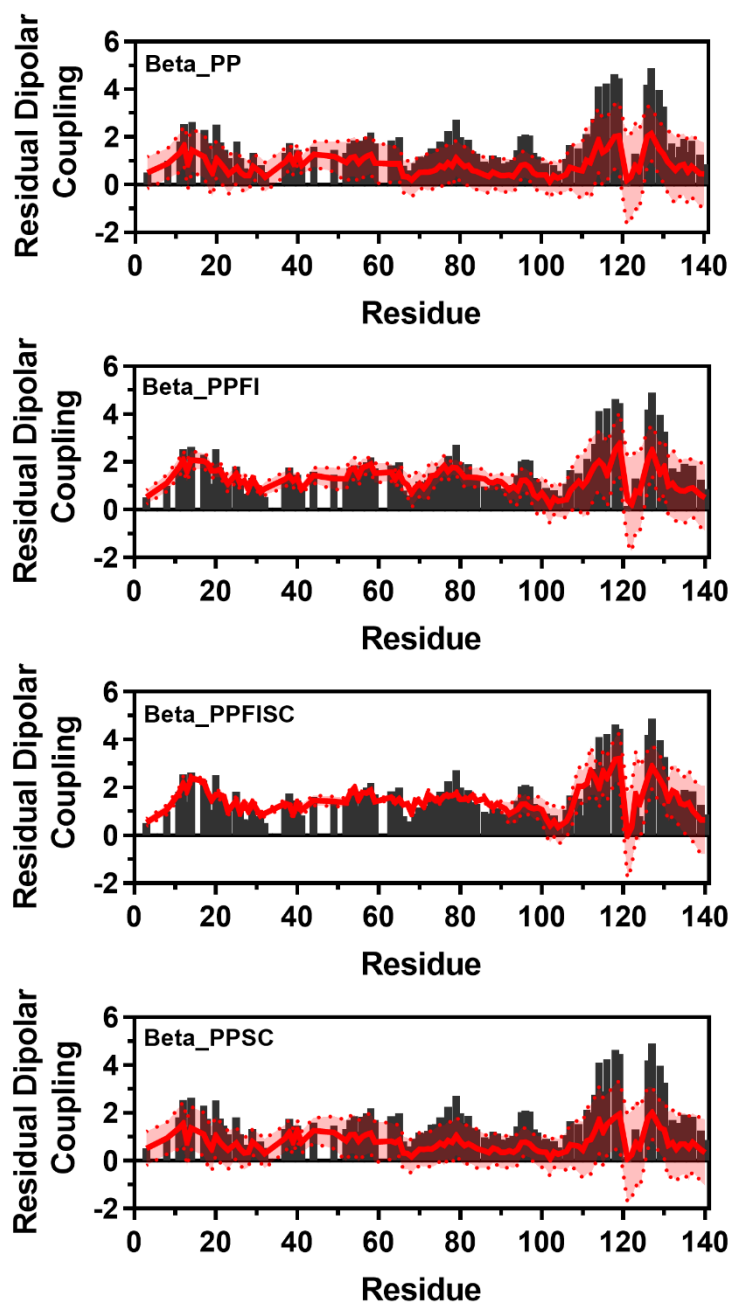


Figure 6.66a: Comparison of Simulated and Experimental Residual Dipolar Coupling Data. Simulated RDC values (red line) from Beta_PP (Top), Beta_PPF1 (Upper Middle), Beta_PPFISC (Lower Middle), and Beta_PPSC (Bottom) overlaid on experimental data (black bars) from Bertocini et. al.²⁰⁰

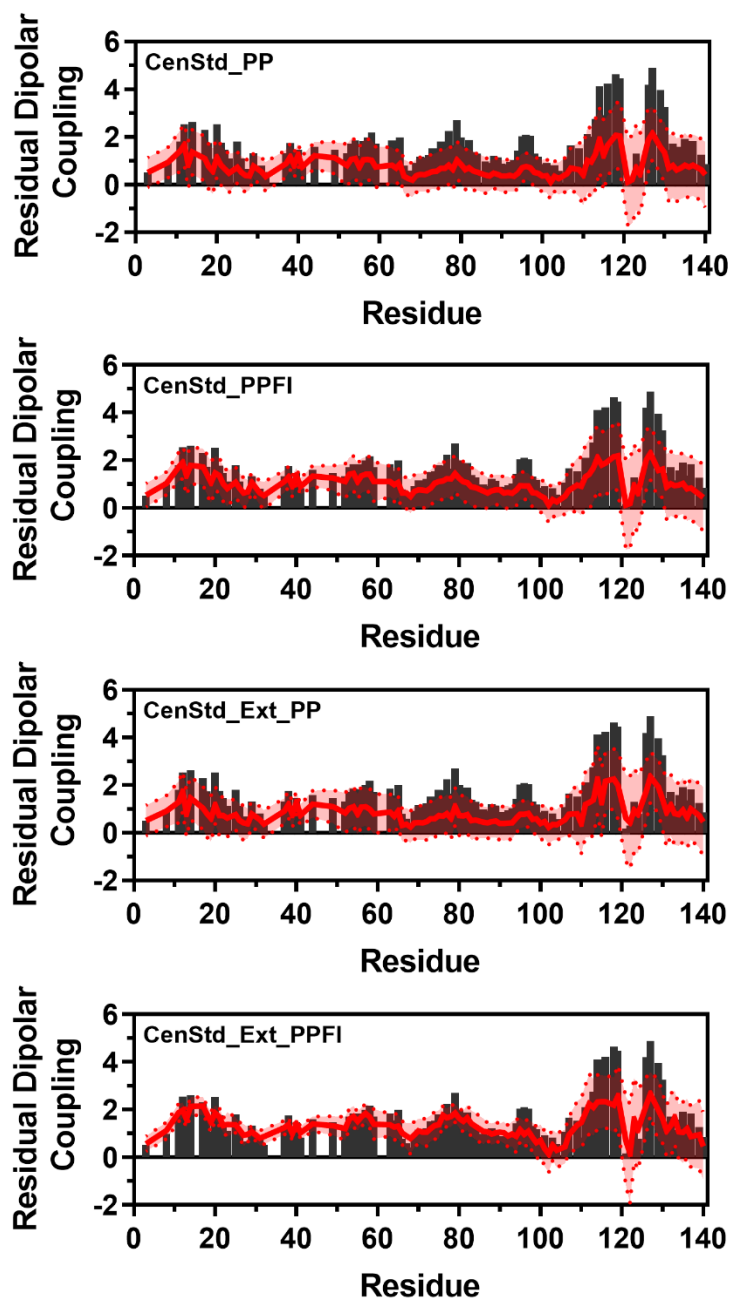


Figure 6.66b: Comparison of Simulated and Experimental Residual Dipolar Coupling Data. Simulated RDC values (red line) from CenStd_PP (Top), CenStd_PPFI (Upper Middle), CenStd_Ext_PP (Lower Middle), and CenStd_Ext_PPFI (Bottom) overlaid on experimental data (black bars) from Bertoncini et. al.²⁰⁰

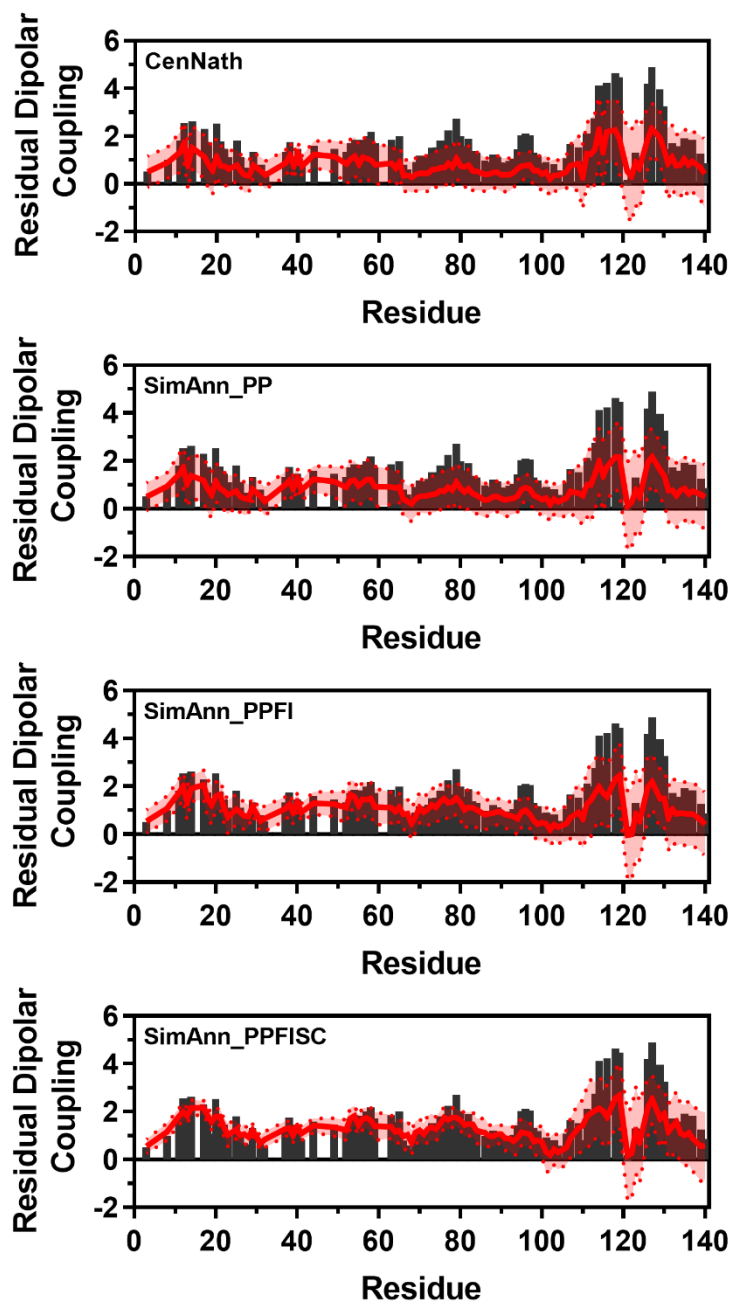


Figure 6.66c: Comparison of Simulated and Experimental Residual Dipolar Coupling Data. Simulated RDC values (red line) from SimAnn_PP (Top), SimAnn_PPFI (Upper Middle), SimAnn_PPFI SC (Lower Middle), and SimAnn_PPSC (Bottom) overlaid on experimental data (black bars) from Bertocini et. al.²⁰⁰

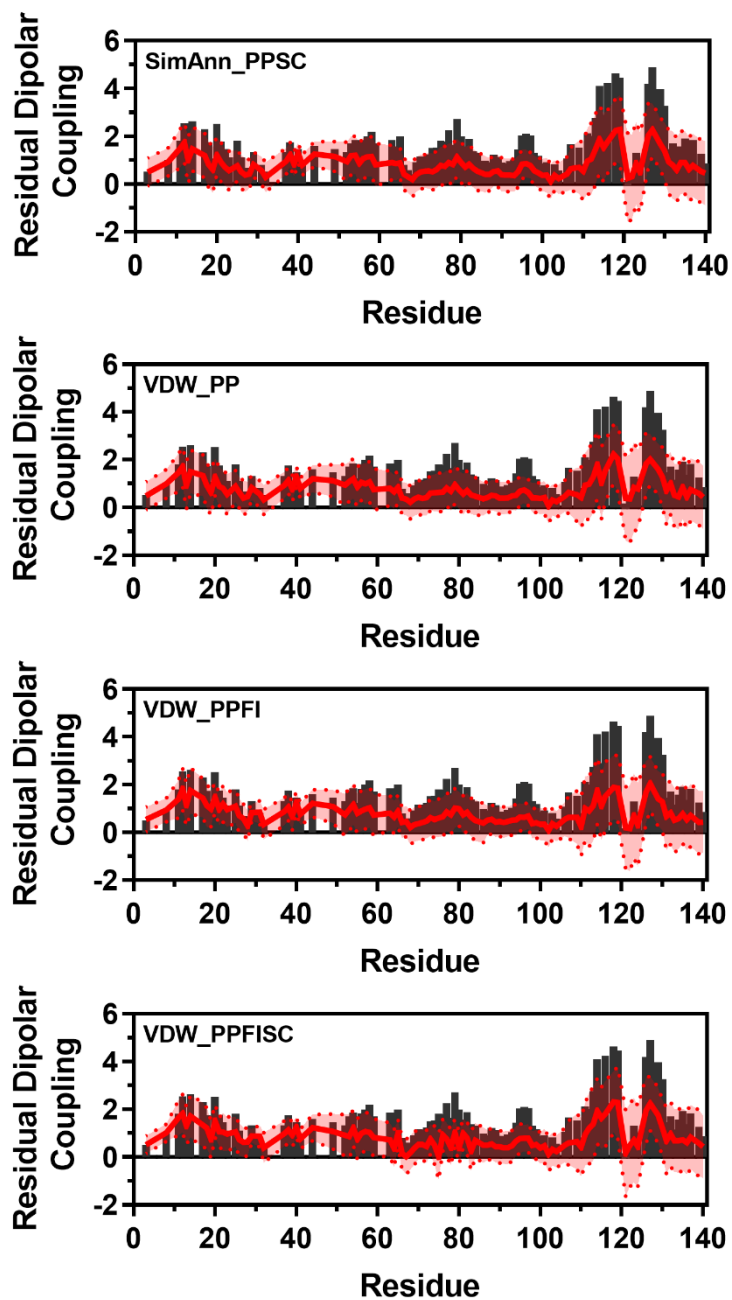


Figure 6.66d: Comparison of Simulated and Experimental Residual Dipolar Coupling Data. Simulated RDC values (red line) from CenNath (Top), VDW_PP (Upper Middle), VDW_PPFI (Lower Middle), and VDW_PPFI SC (Bottom) overlaid on experimental data (black bars) from Bertoncini et. al.²⁰⁰

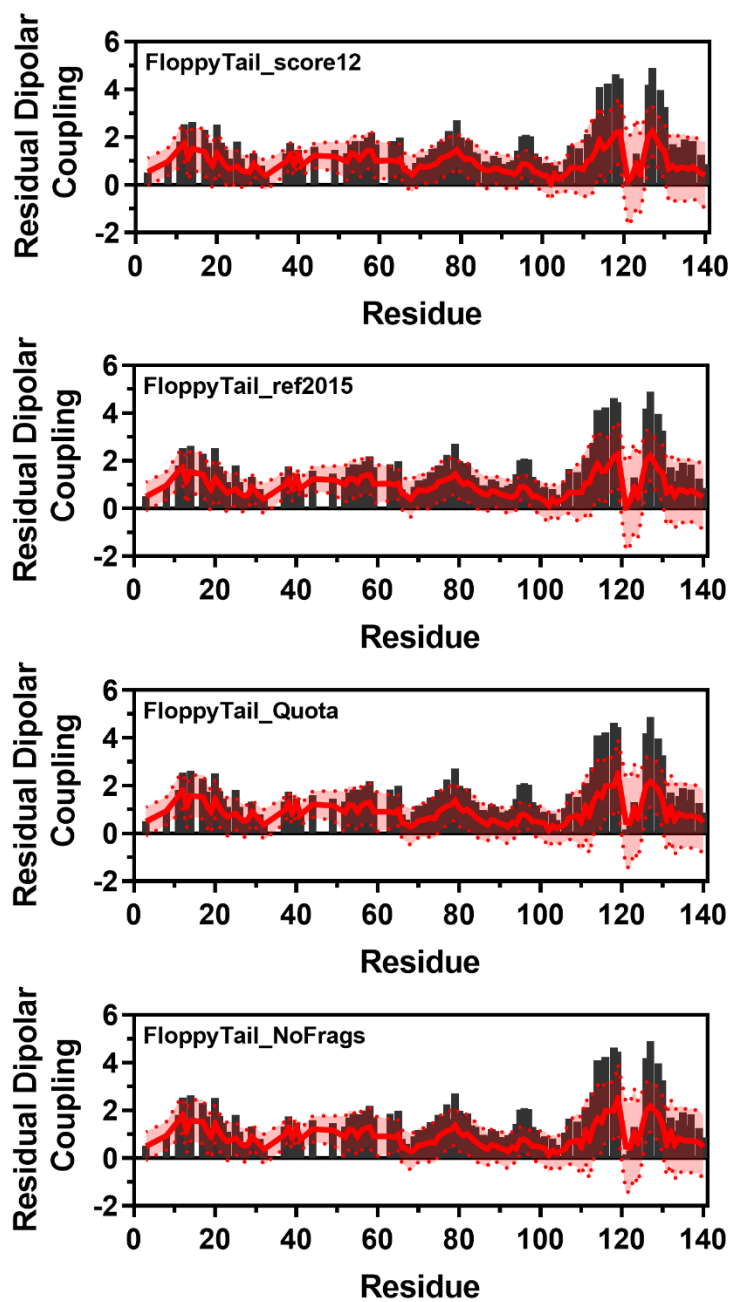


Figure 6.66d: Comparison of Simulated and Experimental Residual Dipolar Coupling Data. Simulated RDC values (red line) from FloppyTail_score12 (Top), FloppyTail_ref2015 (Upper Middle), FloppyTail_Quota (Lower Middle), and FloppyTail_NoFrag (Bottom) overlaid on experimental data (black bars) from Bertocini et. al.²⁰⁰

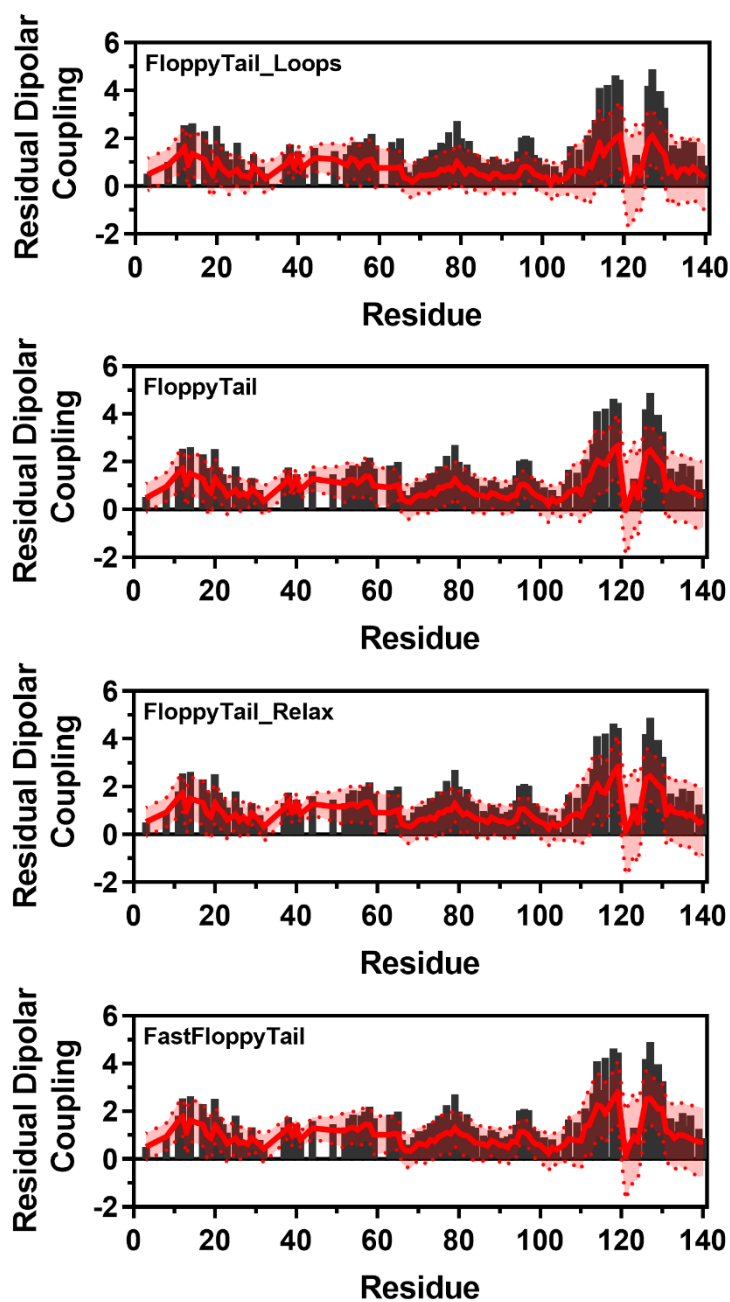


Figure 6.66e: Comparison of Simulated and Experimental Residual Dipolar Coupling Data. Simulated RDC values (red line) from FloppyTail_Loops (Top), FloppyTail (Upper Middle), FloppyTail_Relax (Lower Middle), and FastFloppyTail (Bottom) overlaid on experimental data (black bars) from Bertoncini et. al.²⁰⁰

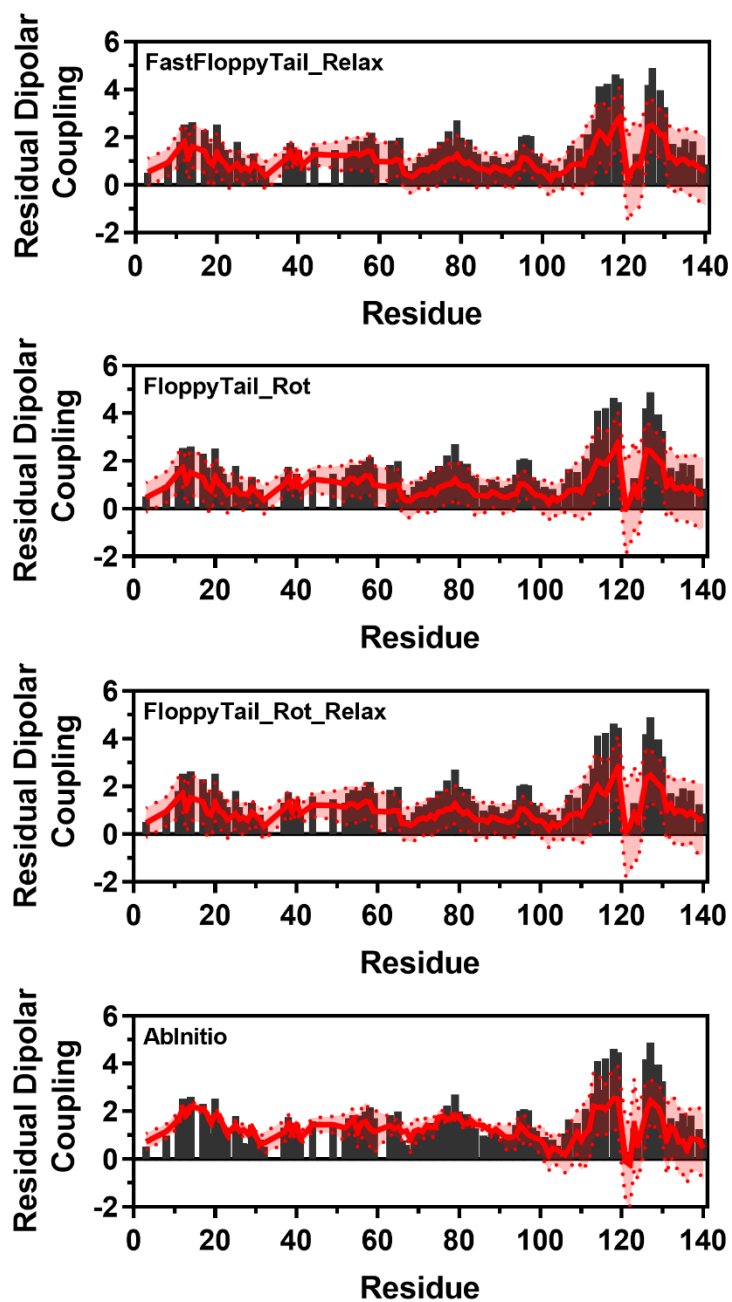


Figure 6.66f: Comparison of Simulated and Experimental Residual Dipolar Coupling Data. Simulated RDC values (red line) from FastFloppyTail_Relax (Top), FloppyTail_Rot (Upper Middle), FloppyTail_Rot_Relax (Lower Middle), and AbInitio (Bottom) overlaid on experimental data (black bars) from Bertoncini et. al.²⁰⁰

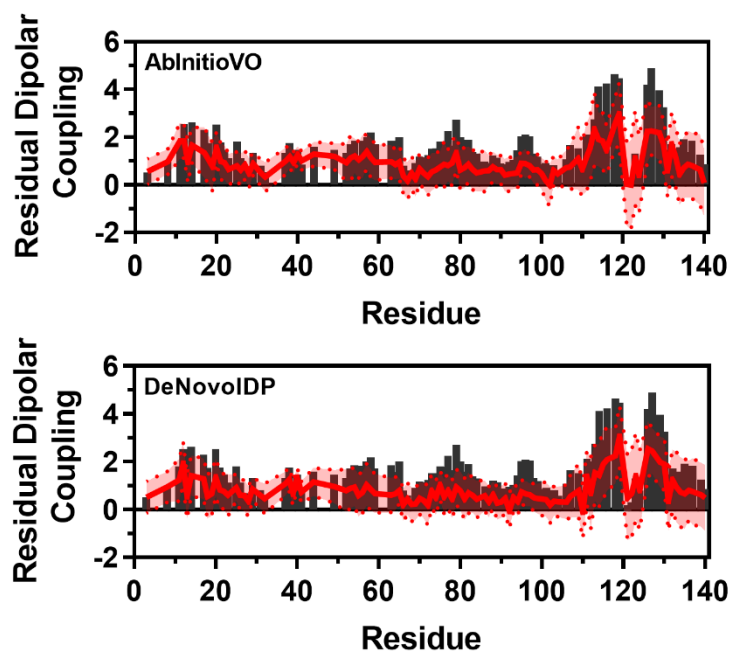


Figure 6.66g: Comparison of Simulated and Experimental Residual Dipolar Coupling Data. Simulated RDC values (red line) from AbInitioVO (Top) and DeNovoIDP (Bottom) overlaid on experimental data (black bars) from Bertoncini et. al.²⁰⁰

Comparison with α -Synuclein J-Coupling Data

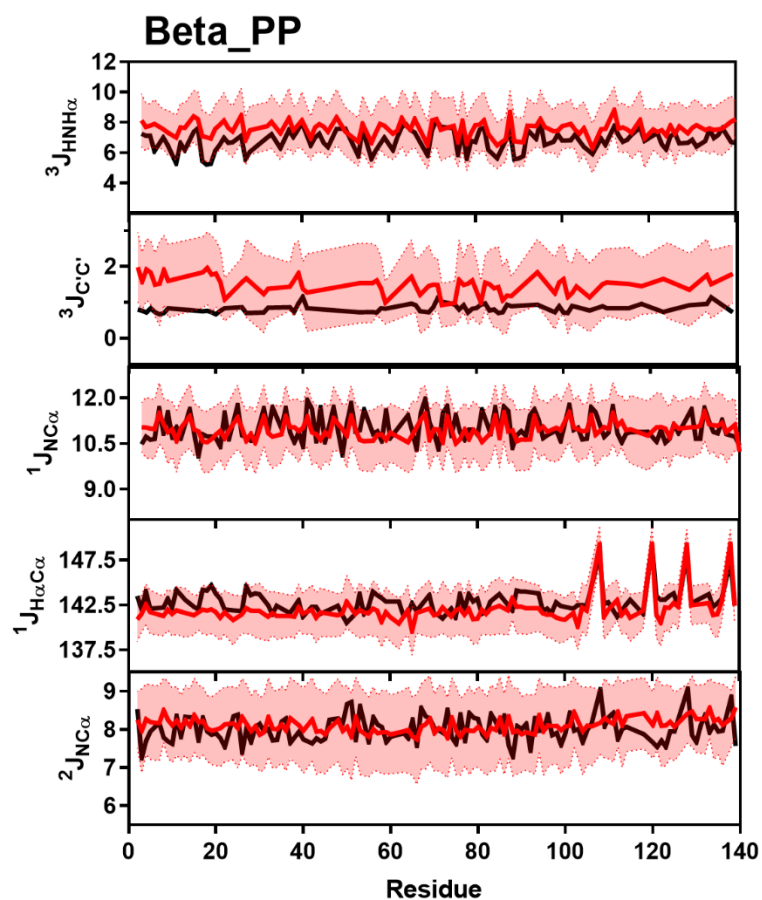


Figure 6.67: Comparison of Simulated Beta_PP and Experimental J-Coupling Data. Simulated Chemical Shift values from the Beta_PP ensemble (red) overlaid on experimental data (black) of $^3J_{\text{HNH}\alpha}$ (Top), $^3J_{\text{HNH}\alpha}$ (Upper Middle), $^1J_{\text{H}\alpha\text{C}\alpha}$ (Middle), $^1J_{\text{NC}\alpha}$ (Lower Middle) and $^2J_{\text{NC}\alpha}$ (Bottom) from Mantsyzov *et al.* and Lee *et. al.*^{247, 252}

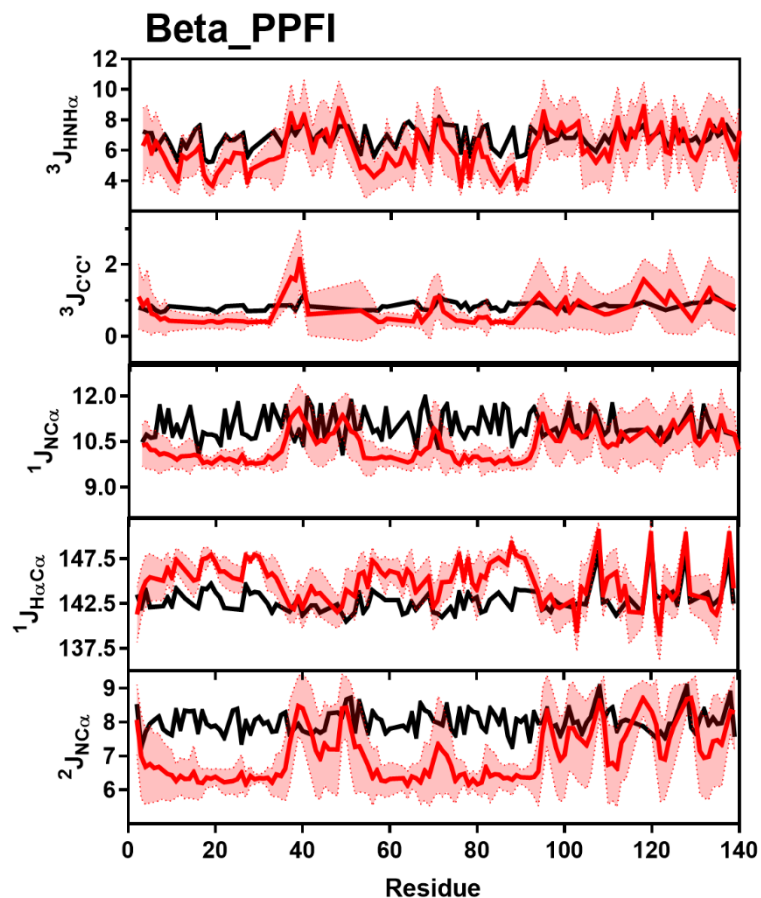


Figure 6.68: Comparison of Simulated Beta_PPFI and Experimental J-Coupling Data. Simulated Chemical Shift values from the Beta_PPFI ensemble (red) overlaid on experimental data (black) of $^3J_{\text{HNH}\alpha}$ (Top), $^3J_{\text{HNH}\alpha}$ (Upper Middle), $^1J_{\text{H}\alpha\text{C}\alpha}$ (Middle), $^1J_{\text{NC}\alpha}$ (Lower Middle) and $^2J_{\text{NC}\alpha}$ (Bottom) from Mantsyzov *et al.* and Lee *et al.*^{247, 252}

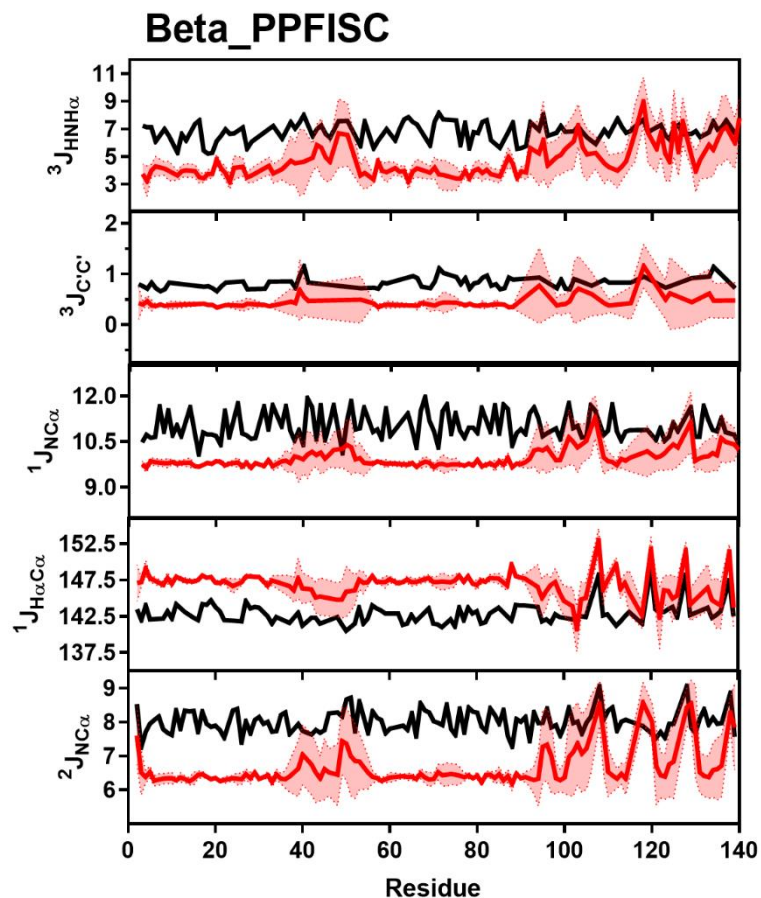


Figure 6.69: Comparison of Simulated Beta_PPFISC and Experimental J-Coupling Data. Simulated Chemical Shift values from the Beta_PPFISC ensemble (red) overlaid on experimental data (black) of $^3J_{\text{HNH}\alpha}$ (Top), $^3J_{\text{HNH}\alpha}$ (Upper Middle), $^1J_{\text{H}\alpha\text{C}\alpha}$ (Middle), $^1J_{\text{NC}\alpha}$ (Lower Middle) and $^2J_{\text{NC}\alpha}$ (Bottom) from Mantsyzov *et al.* and Lee *et al.*^{247, 252}

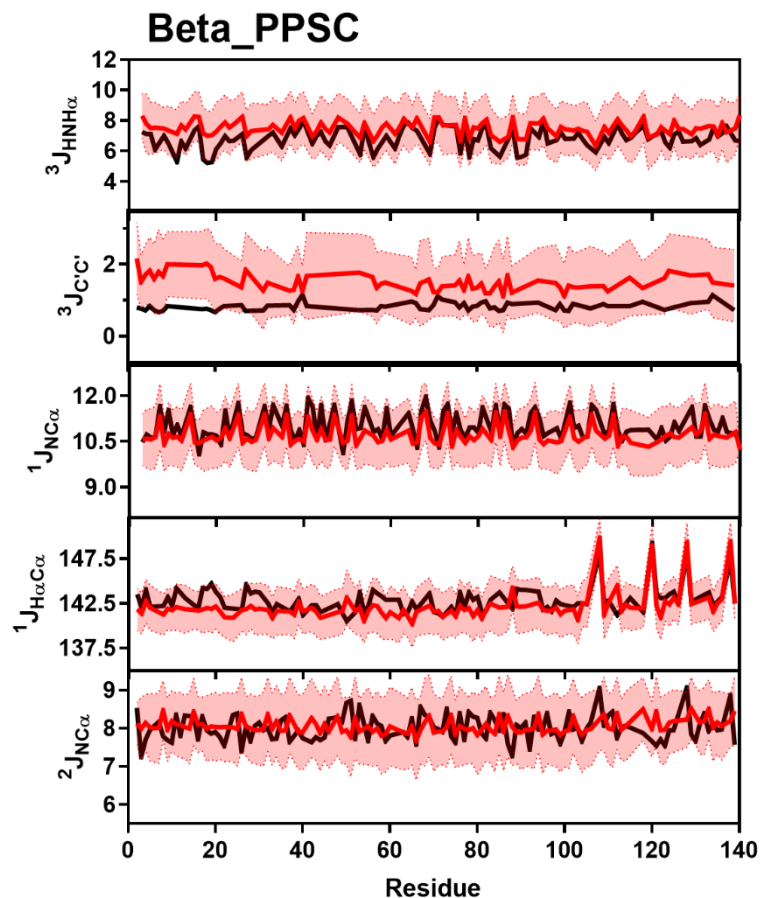


Figure 6.70: Comparison of Simulated Beta_PPSC and Experimental J-Coupling Data. Simulated Chemical Shift values from the Beta_PPSC ensemble (red) overlaid on experimental data (black) of $^3J_{\text{HNH}\alpha}$ (Top), $^3J_{\text{HNH}\alpha}$ (Upper Middle), $^1J_{\text{H}\alpha\text{C}\alpha}$ (Middle), $^1J_{\text{NC}\alpha}$ (Lower Middle) and $^2J_{\text{NC}\alpha}$ (Bottom) from Mantsyzov *et al.* and Lee *et al.*^{247, 252}

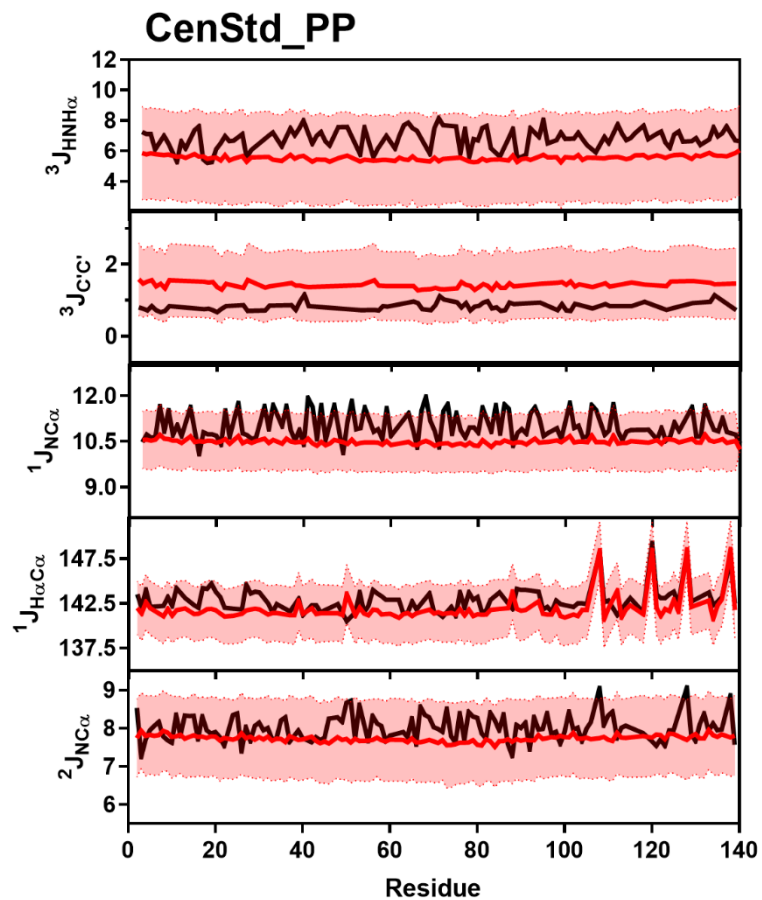


Figure 6.71: Comparison of Simulated CenStd_PP and Experimental J-Coupling Data. Simulated Chemical Shift values from the CenStd_PP ensemble (red) overlaid on experimental data (black) of ${}^3J_{\text{HNH}\alpha}$ (Top), ${}^3J_{\text{HNH}\alpha}$ (Upper Middle), ${}^1J_{\text{H}\alpha\text{C}\alpha}$ (Middle), ${}^1J_{\text{NC}\alpha}$ (Lower Middle) and ${}^2J_{\text{NC}\alpha}$ (Bottom) from Mantsyzov *et al.* and Lee *et al.*^{247, 252}

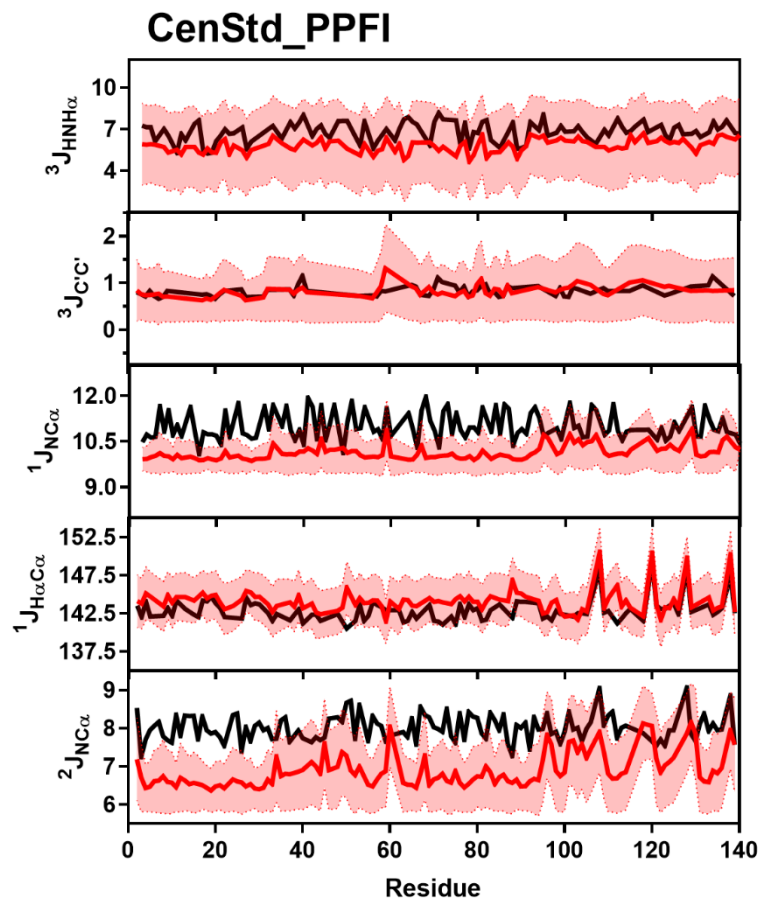


Figure 6.72: Comparison of Simulated CenStd_PPFI and Experimental J-Coupling Data. Simulated Chemical Shift values from the CenStd_PPFI ensemble (red) overlaid on experimental data (black) of $^3J_{\text{HNH}\alpha}$ (Top), $^3J_{\text{HNH}\alpha}$ (Upper Middle), $^1J_{\text{H}\alpha\text{C}\alpha}$ (Middle), $^1J_{\text{NC}\alpha}$ (Lower Middle) and $^2J_{\text{NC}\alpha}$ (Bottom) from Mantsyzov *et al.* and Lee *et. al.*^{247, 252}

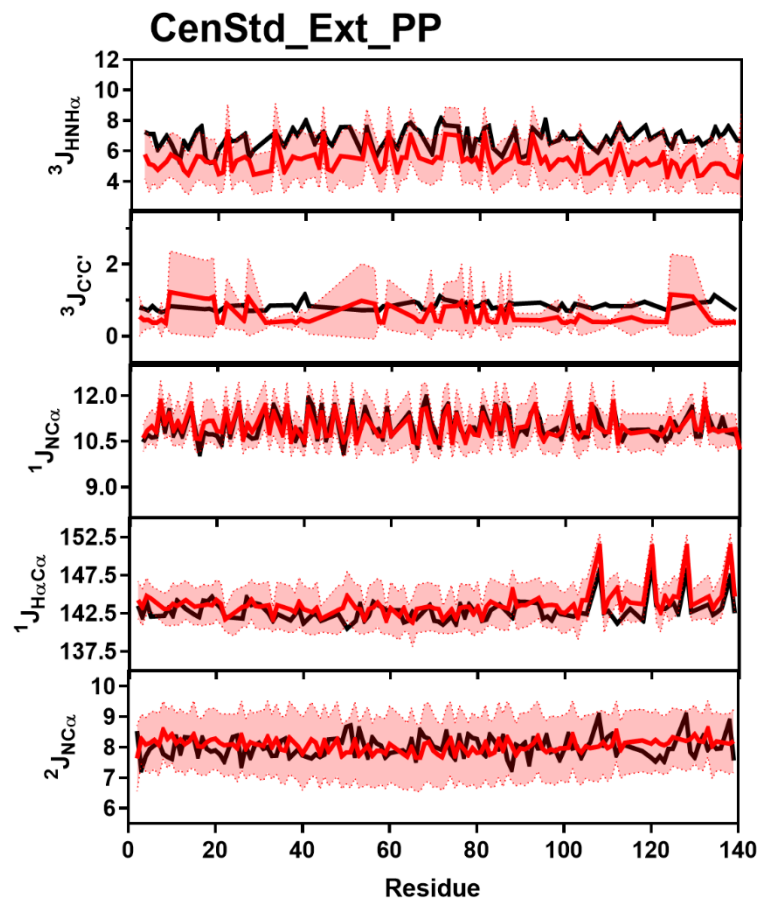


Figure 6.73: Comparison of Simulated CenStd_Ext_PP and Experimental J-Coupling Data. Simulated Chemical Shift values from the CenStd_Ext_PP ensemble (red) overlaid on experimental data (black) of $^3J_{\text{HNH}\alpha}$ (Top), $^3J_{\text{HNH}\alpha}$ (Upper Middle), $^1J_{\text{H}\alpha\text{C}\alpha}$ (Middle), $^1J_{\text{NC}\alpha}$ (Lower Middle) and $^2J_{\text{NC}\alpha}$ (Bottom) from Mantsyzov *et al.* and Lee *et al.*^{247, 252}

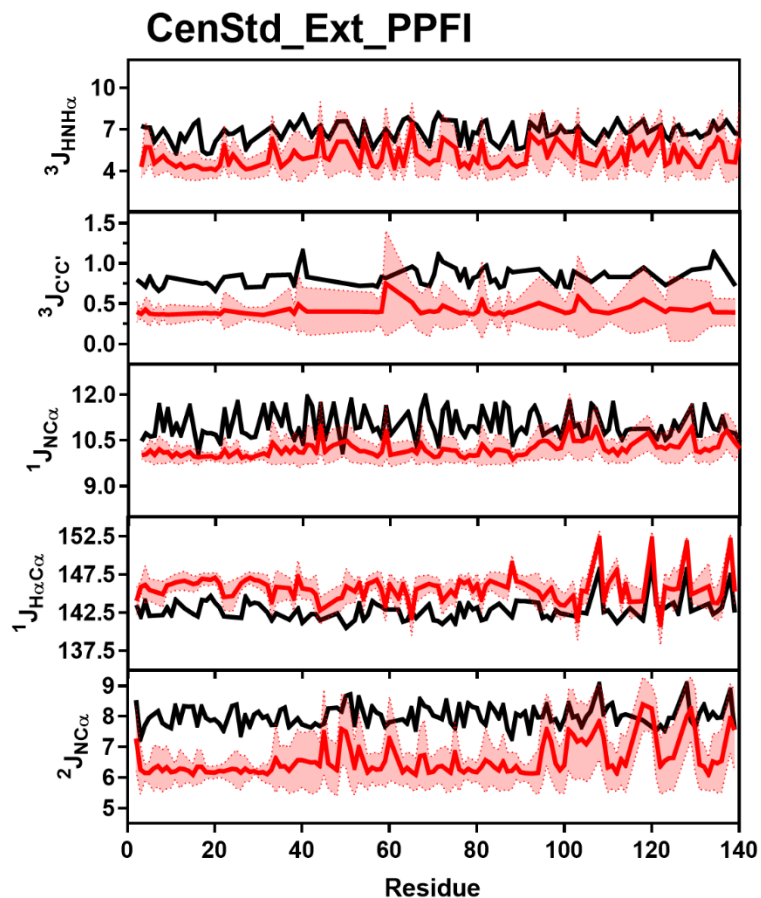


Figure 6.74: Comparison with α -Synuclein PRE Data Comparison of Simulated CenStd_Ext_PPFI and Experimental J-Coupling Data. Simulated Chemical Shift values from the CenStd_Ext_PPFI ensemble (red) overlaid on experimental data (black) of $^3J_{\text{HNH}\alpha}$ (Top), $^3J_{\text{HNH}\alpha}$ (Upper Middle), $^1J_{\text{H}\alpha\text{C}\alpha}$ (Middle), $^1J_{\text{NC}\alpha}$ (Lower Middle) and $^2J_{\text{NC}\alpha}$ (Bottom) from Mantsyzov *et al.* and Lee *et. al.*^{247, 252}

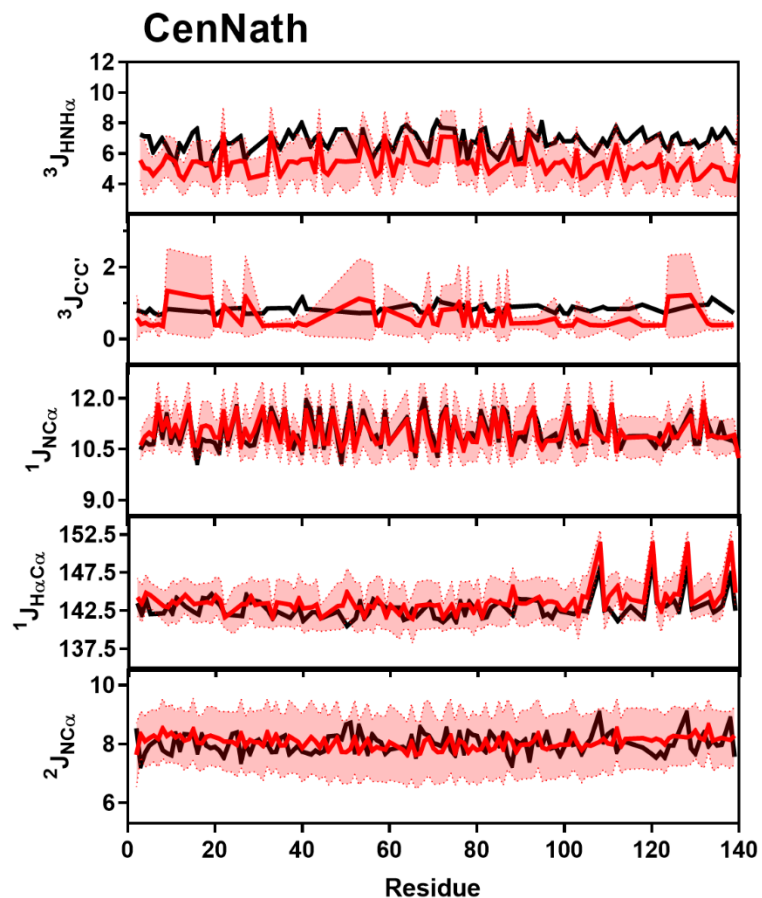


Figure 6.75: Comparison of Simulated CenNath and Experimental J-Coupling Data. Simulated Chemical Shift values from the CenNath ensemble (red) overlaid on experimental data (black) of $^3J_{\text{HNH}\alpha}$ (Top), $^3J_{\text{HNH}\alpha}$ (Upper Middle), $^1J_{\text{H}\alpha\text{C}\alpha}$ (Middle), $^1J_{\text{NC}\alpha}$ (Lower Middle) and $^2J_{\text{NC}\alpha}$ (Bottom) from Mantsyzov *et al.* and Lee *et al.*^{247, 252}

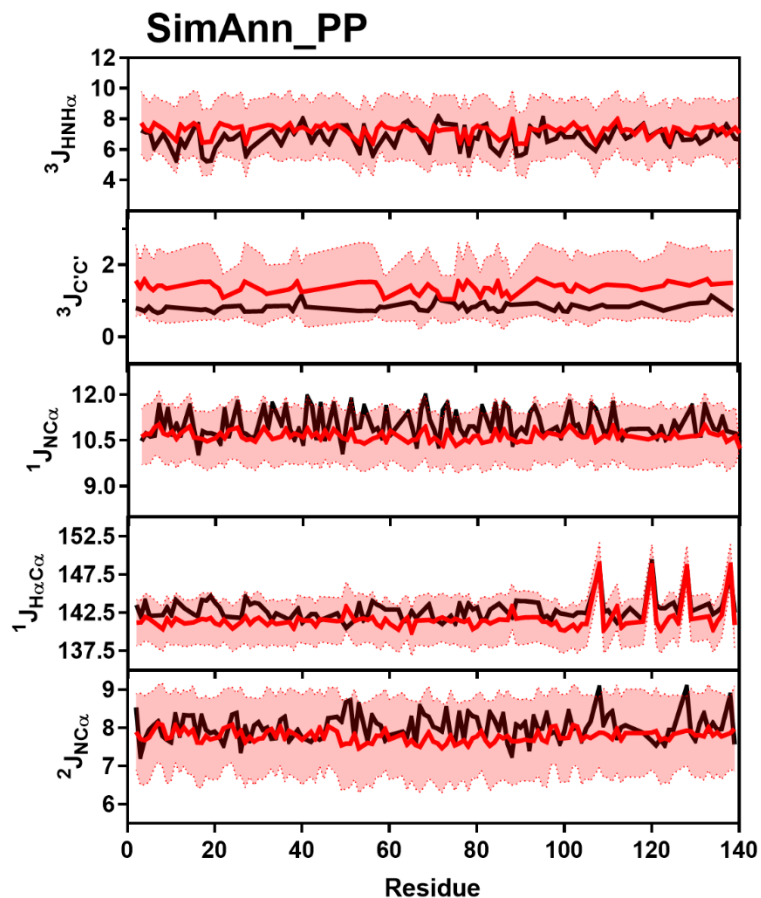


Figure 6.76: Comparison of Simulated SimAnn_PP and Experimental J-Coupling Data. Simulated Chemical Shift values from the SimAnn_PP ensemble (red) overlaid on experimental data (black) of $^3J_{\text{HNH}\alpha}$ (Top), $^3J_{\text{HNH}\alpha}$ (Upper Middle), $^1J_{\text{H}\alpha\text{C}\alpha}$ (Middle), $^1J_{\text{NC}\alpha}$ (Lower Middle) and $^2J_{\text{NC}\alpha}$ (Bottom) from Mantsyzov *et al.* and Lee *et al.*^{247, 252}

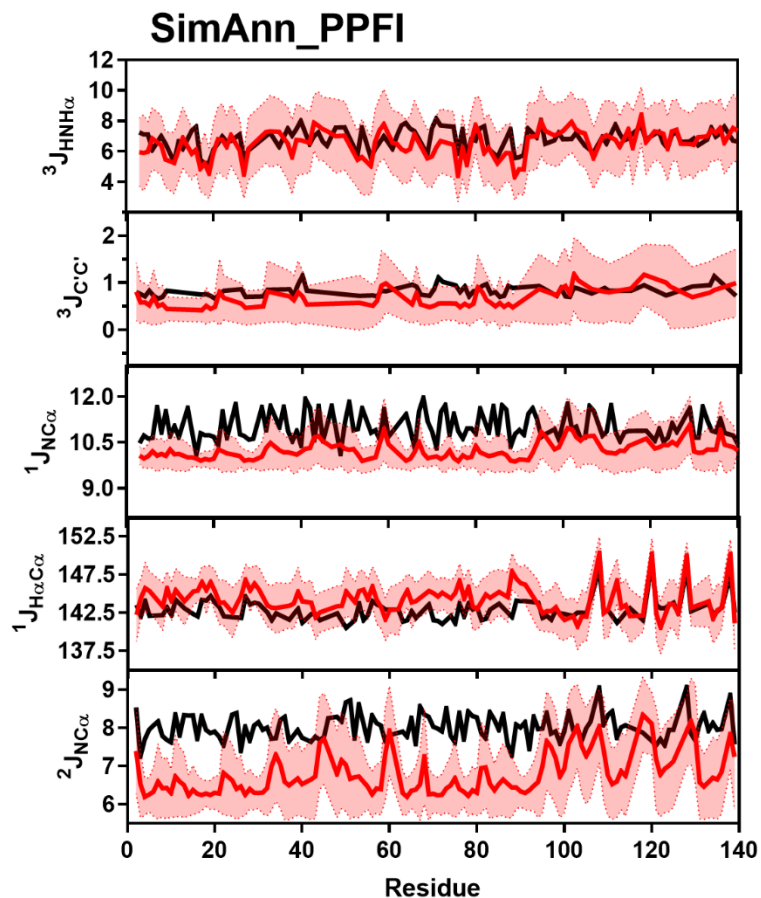


Figure 6.77: Comparison of Simulated SimAnn_PPFI and Experimental J-Coupling Data. Simulated Chemical Shift values from the SimAnn_PPFI ensemble (red) overlaid on experimental data (black) of $^3J_{\text{HNNH}\alpha}$ (Top), $^3J_{\text{HNNH}\alpha}$ (Upper Middle), $^1J_{\text{H}\alpha\text{C}\alpha}$ (Middle), $^1J_{\text{NC}\alpha}$ (Lower Middle) and $^2J_{\text{NC}\alpha}$ (Bottom) from Mantsyzov *et al.* and Lee *et. al.*^{247, 252}

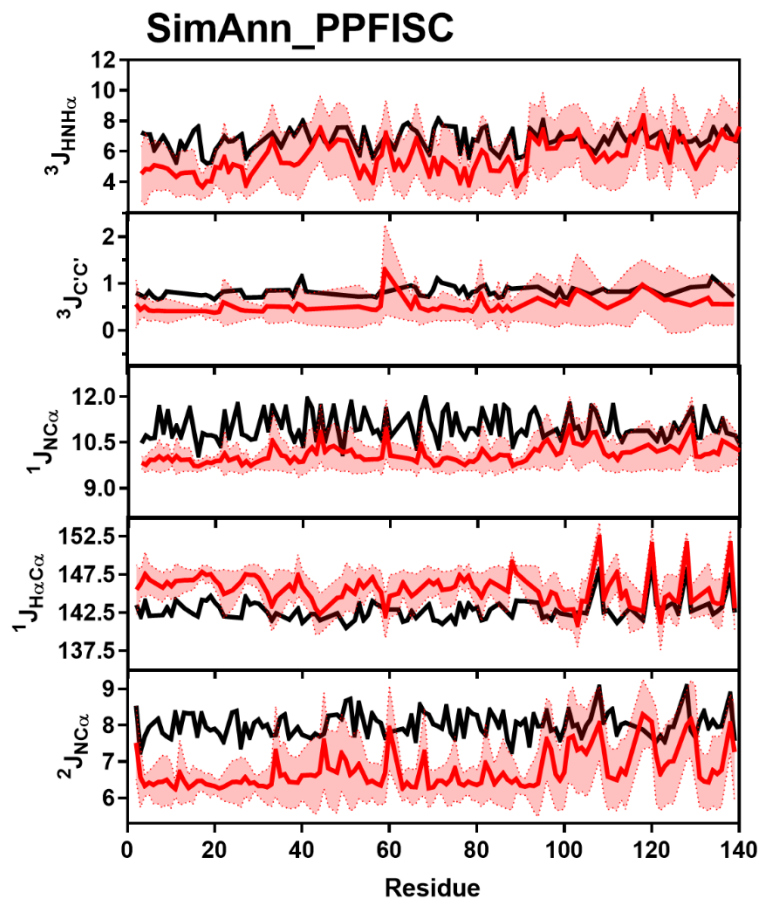


Figure 6.78: Comparison of Simulated SimAnn_PPFISC and Experimental J-Coupling Data. Simulated Chemical Shift values from the SimAnn_PPFISC ensemble (red) overlaid on experimental data (black) of $^3J_{\text{HNH}\alpha}$ (Top), $^3J_{\text{HNH}\alpha}$ (Upper Middle), $^1J_{\text{H}\alpha\text{C}\alpha}$ (Middle), $^1J_{\text{NC}\alpha}$ (Lower Middle) and $^2J_{\text{NC}\alpha}$ (Bottom) from Mantsyzov *et al.* and Lee *et. al.*^{247, 252}

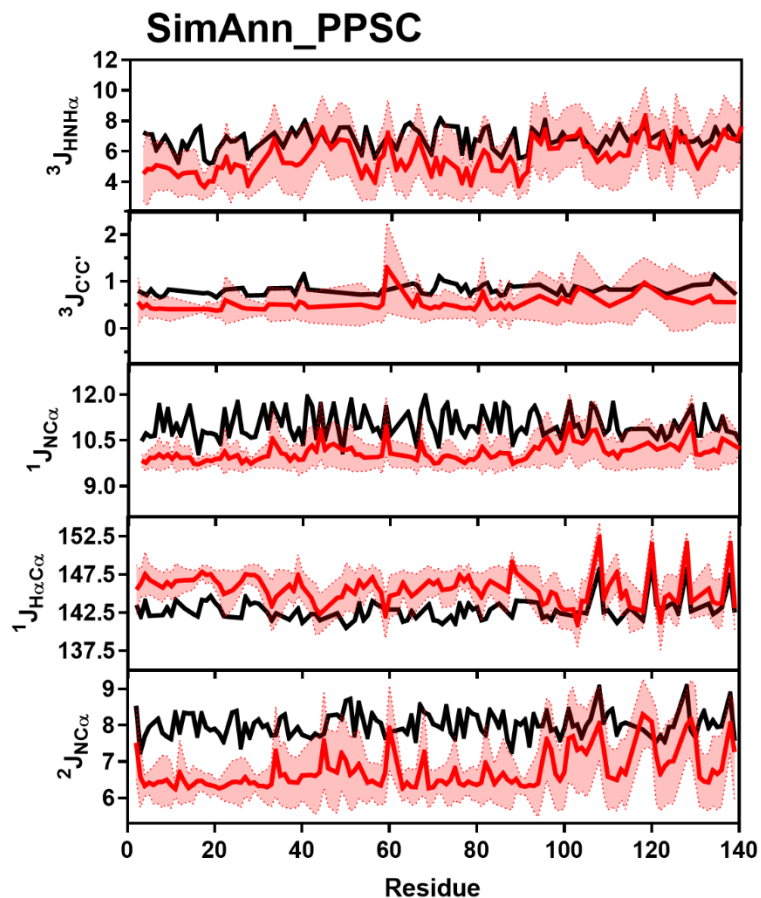


Figure 6.79: Comparison of Simulated SimAnn_PPSC and Experimental J-Coupling Data. Simulated Chemical Shift values from the SimAnn_PPSC ensemble (red) overlaid on experimental data (black) of $^3J_{\text{HNH}\alpha}$ (Top), $^3J_{\text{HNH}\alpha}$ (Upper Middle), $^1J_{\text{H}\alpha\text{C}\alpha}$ (Middle), $^1J_{\text{NC}\alpha}$ (Lower Middle) and $^2J_{\text{NC}\alpha}$ (Bottom) from Mantsyzov *et al.* and Lee *et. al.*^{247, 252}

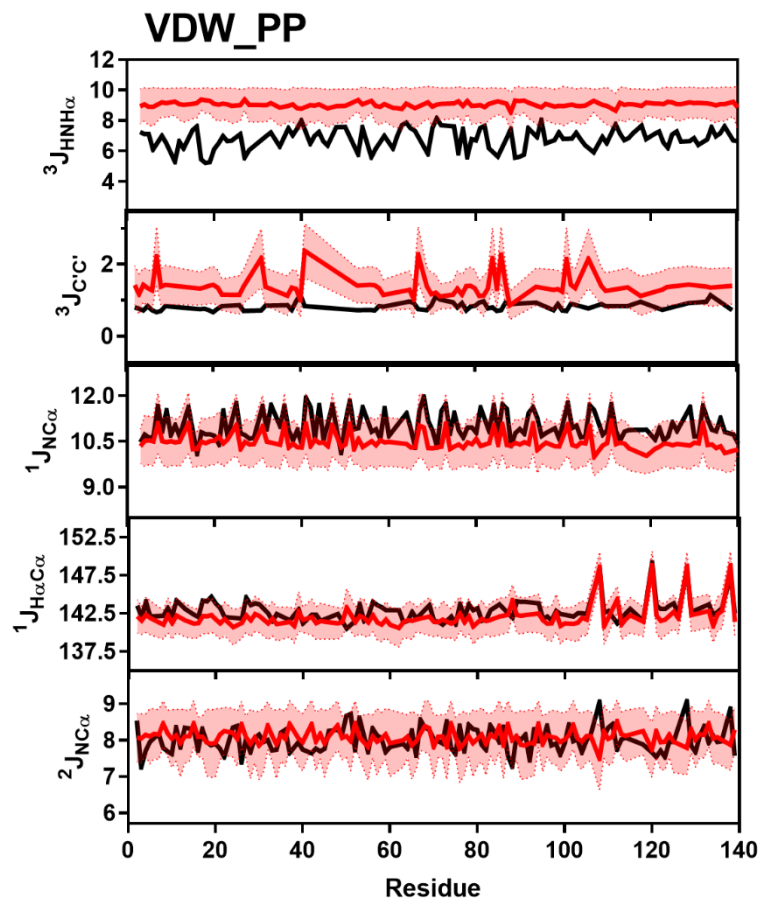


Figure 6.80: Comparison of Simulated VDW_PP and Experimental J-Coupling Data. Simulated Chemical Shift values from the VDW_PP ensemble (red) overlaid on experimental data (black) of $^3J_{\text{H}\alpha\text{H}\alpha}$ (Top), $^3J_{\text{H}\alpha\text{H}\alpha}$ (Upper Middle), $^1J_{\text{H}\alpha\text{C}\alpha}$ (Middle), $^1J_{\text{N}\alpha\text{C}\alpha}$ (Lower Middle) and $^2J_{\text{N}\alpha\text{C}\alpha}$ (Bottom) from Mantsyzov *et al.* and Lee *et al.*^{247, 252}

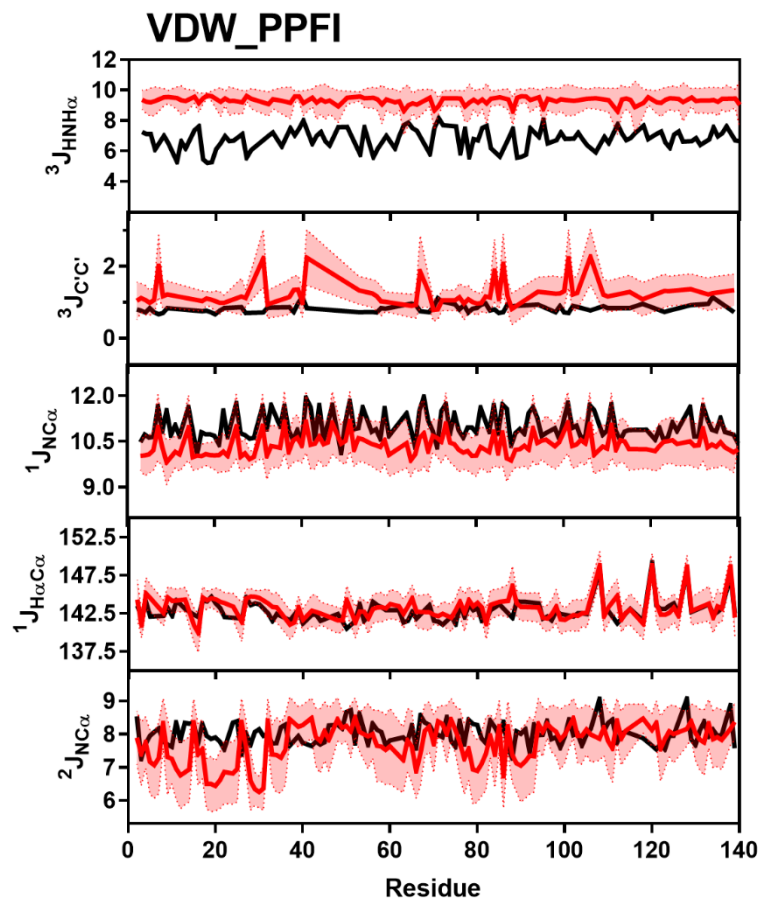


Figure 6.81: Comparison of Simulated VDW_PPFI and Experimental J-Coupling Data. Simulated Chemical Shift values from the VDW_PPFI ensemble (red) overlaid on experimental data (black) of $^3J_{\text{HNH}\alpha}$ (Top), $^3J_{\text{HNH}\alpha}$ (Upper Middle), $^1J_{\text{H}\alpha\text{C}\alpha}$ (Middle), $^1J_{\text{N}\text{C}\alpha}$ (Lower Middle) and $^2J_{\text{N}\text{C}\alpha}$ (Bottom) from Mantsyzov *et al.* and Lee *et al.*^{247, 252}

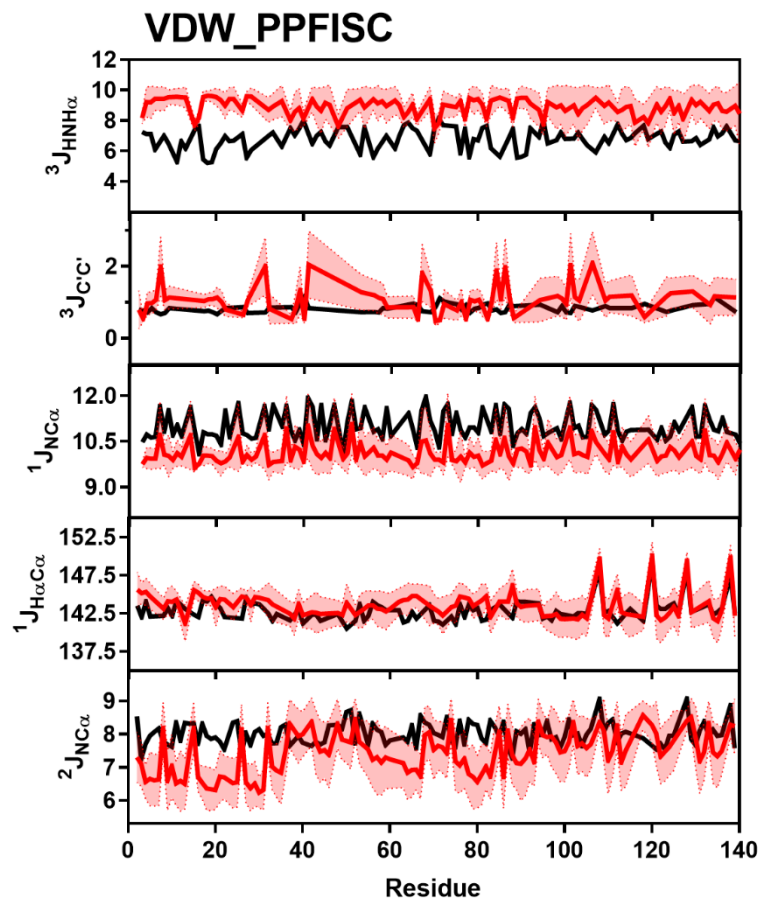


Figure 6.82: Comparison of Simulated VDW_PPFISC and Experimental J-Coupling Data. Simulated Chemical Shift values from the VDW_PPFISC ensemble (red) overlaid on experimental data (black) of $^3J_{\text{HNH}\alpha}$ (Top), $^3J_{\text{HNH}\alpha}$ (Upper Middle), $^1J_{\text{H}\alpha\text{C}\alpha}$ (Middle), $^1J_{\text{NC}\alpha}$ (Lower Middle) and $^2J_{\text{NC}\alpha}$ (Bottom) from Mantsyzov *et al.* and Lee *et al.*^{247, 252}

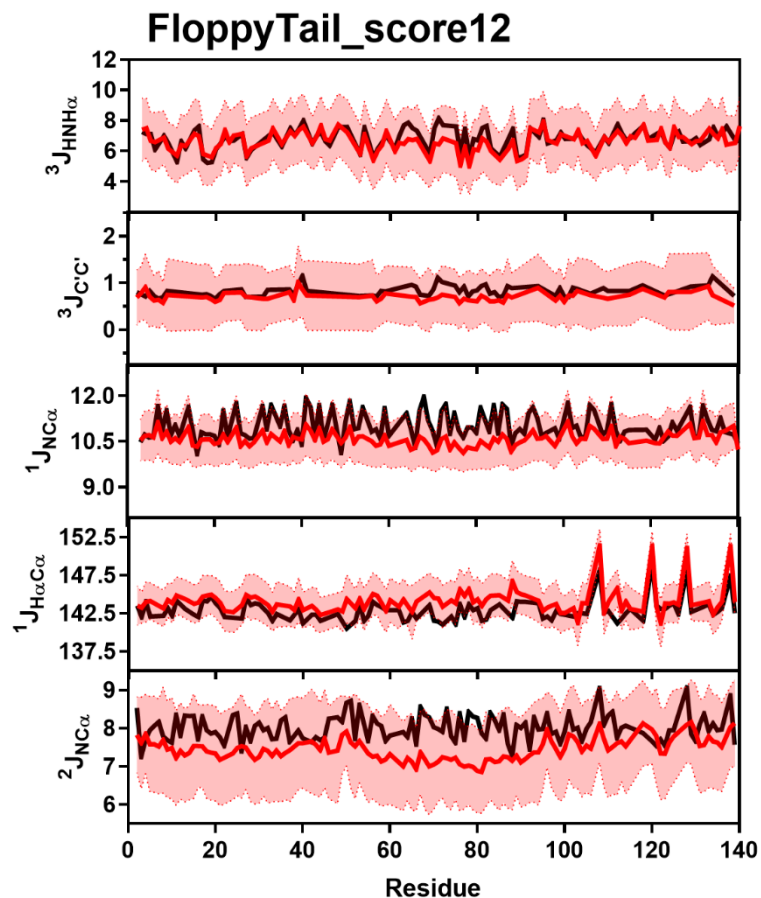


Figure 6.83: Comparison of Simulated FloppyTail_score12 and Experimental J-Coupling Data. Simulated Chemical Shift values from the FloppyTail_score12 ensemble (red) overlaid on experimental data (black) of $^3J_{\text{HNH}\alpha}$ (Top), $^3J_{\text{HNH}\alpha}$ Upper Middle), $^1J_{\text{H}\alpha\text{C}\alpha}$ (Middle), $^1J_{\text{NC}\alpha}$ (Lower Middle) and $^2J_{\text{NC}\alpha}$ (Bottom) from Mantsyzov *et al.* and Lee *et. al.*^{247, 252}

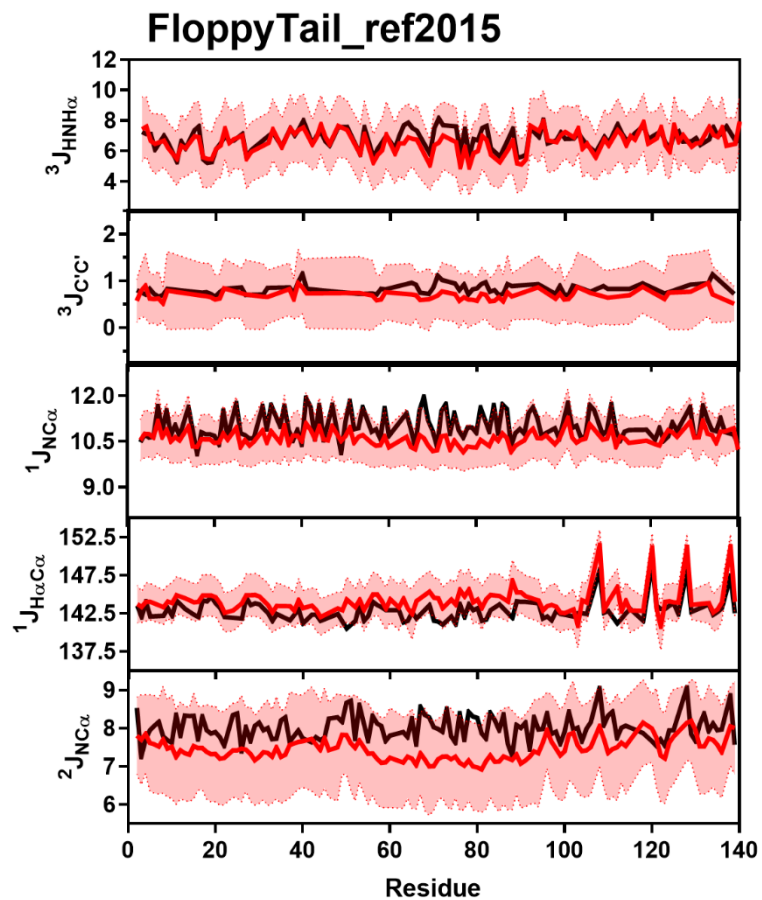


Figure 6.84: Comparison of Simulated FloppyTail_ref2015 and Experimental J-Coupling Data. Simulated Chemical Shift values from the FloppyTail_ref2015 ensemble (red) overlaid on experimental data (black) of $^3J_{\text{HNH}\alpha}$ (Top), $^3J_{\text{C}'\text{C}'}$ (Upper Middle), $^1J_{\text{H}\alpha\text{C}\alpha}$ (Middle), $^1J_{\text{NC}\alpha}$ (Lower Middle) and $^2J_{\text{NC}\alpha}$ (Bottom) from Mantsyzov *et al.* and Lee *et. al.*^{247, 252}

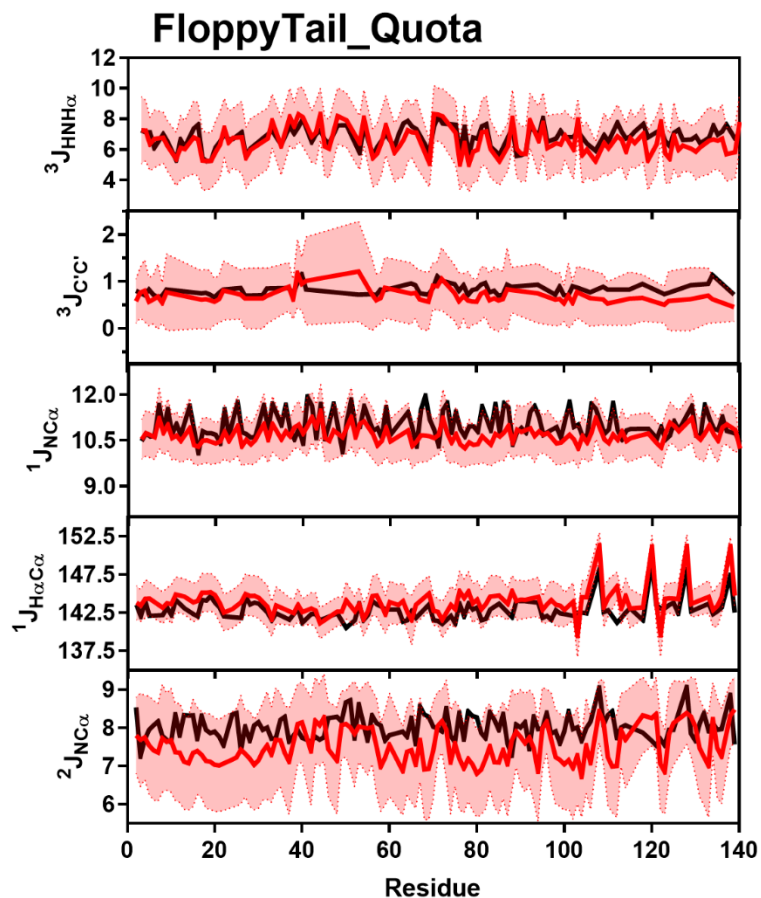


Figure 6.85: Comparison of Simulated FloppyTail_Quota and Experimental J-Coupling Data. Simulated Chemical Shift values from the FloppyTail_Quota ensemble (red) overlaid on experimental data (black) of $^3J_{\text{HNH}\alpha}$ (Top), $^3J_{\text{HNH}\alpha}$ (Upper Middle), $^1J_{\text{H}\alpha\text{C}\alpha}$ (Middle), $^1J_{\text{NC}\alpha}$ (Lower Middle) and $^2J_{\text{NC}\alpha}$ (Bottom) from Mantsyzov *et al.* and Lee *et. al.*^{247, 252}

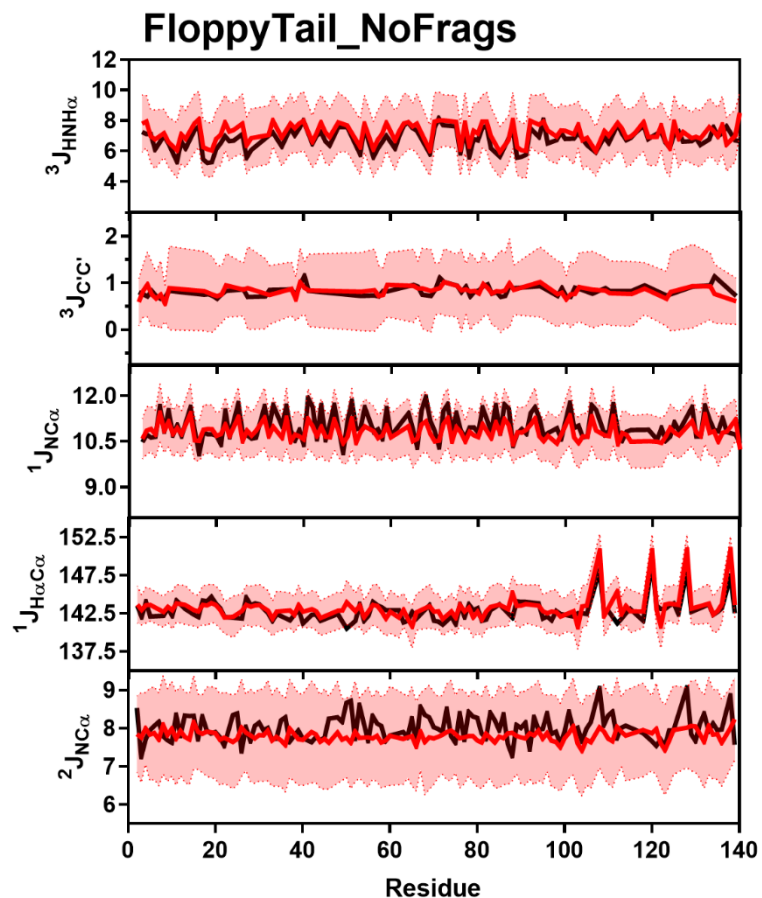


Figure 6.86: Comparison of Simulated FloppyTail_NoFrag and Experimental J-Coupling Data. Simulated Chemical Shift values from the FloppyTail_NoFrag ensemble (red) overlaid on experimental data (black) of $^3J_{\text{HNH}\alpha}$ (Top), $^3J_{\text{HNH}\alpha}$ (Upper Middle), $^1J_{\text{H}\alpha\text{C}\alpha}$ (Middle), $^1J_{\text{NC}\alpha}$ (Lower Middle) and $^2J_{\text{NC}\alpha}$ (Bottom) from Mantsyzov *et al.* and Lee *et. al.*^{247, 252}

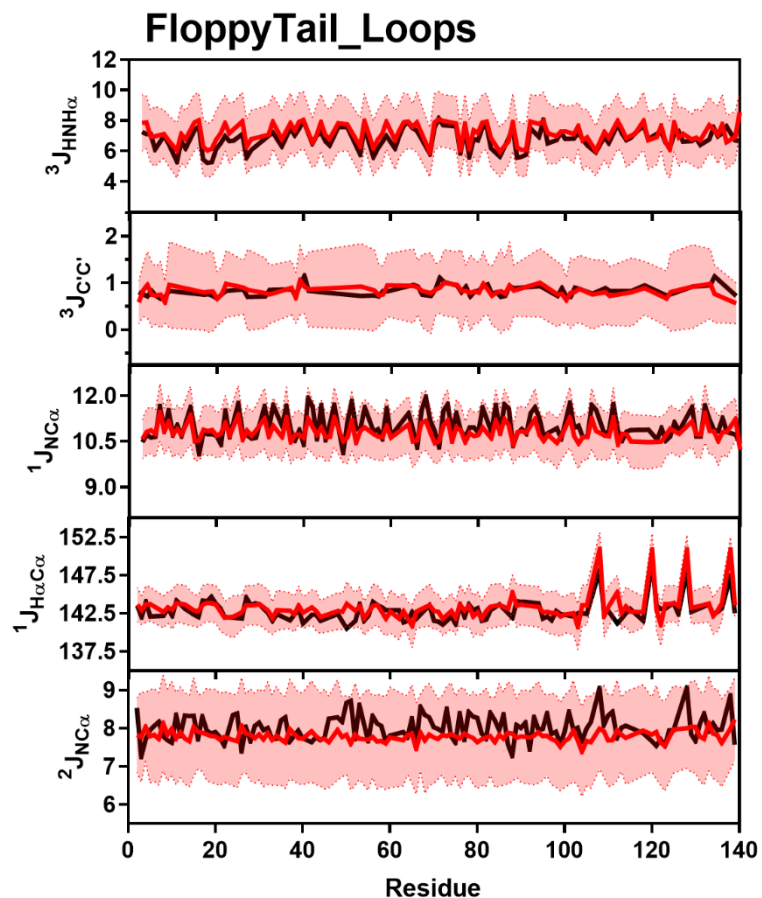


Figure 6.87: Comparison of Simulated FloppyTail_Loops and Experimental J-Coupling Data. Simulated Chemical Shift values from the FloppyTail_Loops ensemble (red) overlaid on experimental data (black) of $^3J_{\text{HNH}\alpha}$ (Top), $^3J_{\text{HNH}\alpha}$ (Upper Middle), $^1J_{\text{H}\alpha\text{C}\alpha}$ (Middle), $^1J_{\text{NC}\alpha}$ (Lower Middle) and $^2J_{\text{NC}\alpha}$ (Bottom) from Mantsyzov *et al.* and Lee *et. al.*^{247, 252}

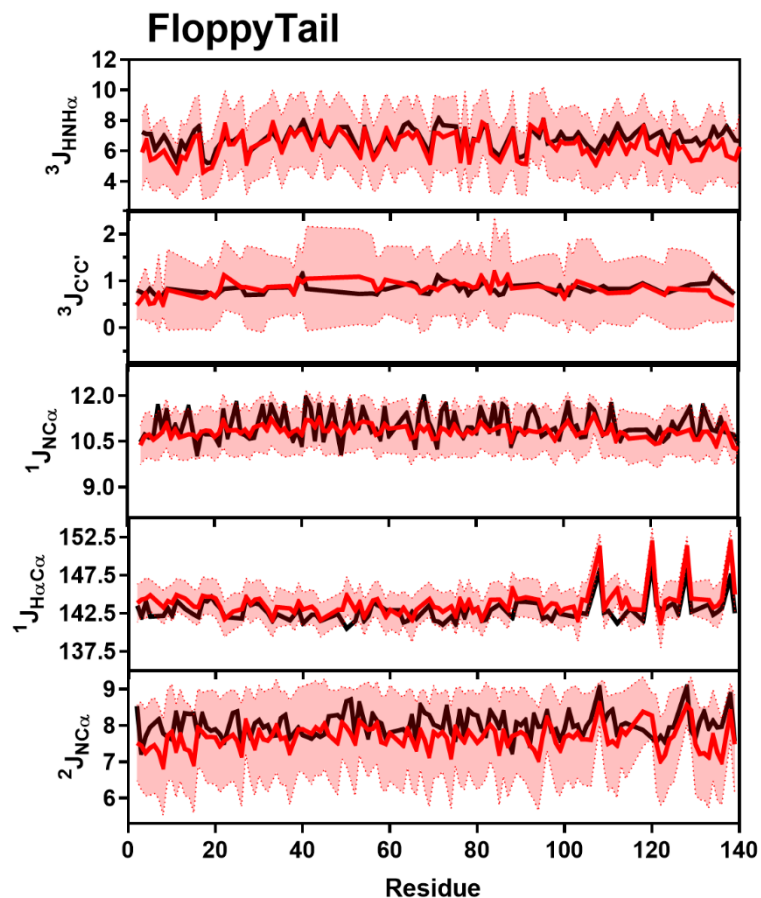


Figure 6.88: Comparison of Simulated FloppyTail and Experimental J-Coupling Data. Simulated Chemical Shift values from the FloppyTail ensemble (red) overlaid on experimental data (black) of $^3J_{\text{HNH}\alpha}$ (Top), $^3J_{\text{HNH}\alpha}$ (Upper Middle), $^1J_{\text{H}\alpha\text{C}\alpha}$ (Middle), $^1J_{\text{NC}\alpha}$ (Lower Middle) and $^2J_{\text{NC}\alpha}$ (Bottom) from Mantsyzov *et al.* and Lee *et al.*^{247, 252}

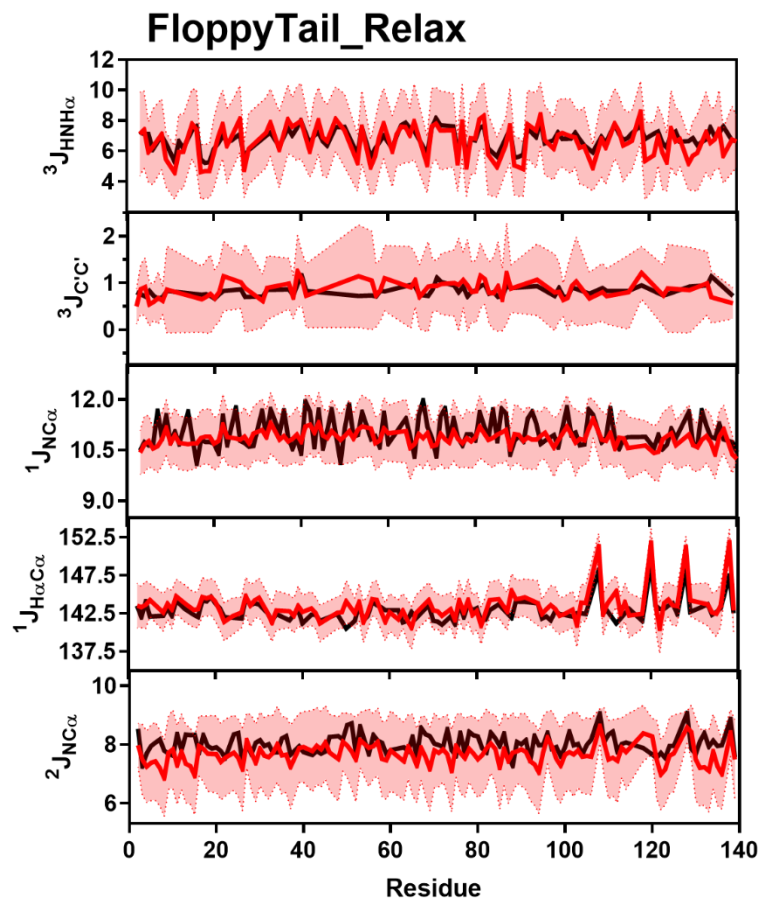


Figure 6.89: Comparison of Simulated FloppyTail_Relax and Experimental J-Coupling Data. Simulated Chemical Shift values from the FloppyTail_Relax ensemble (red) overlaid on experimental data (black) of $^3J_{\text{HNH}\alpha}$ (Top), $^3J_{\text{HNH}\alpha}$ (Upper Middle), $^1J_{\text{H}\alpha\text{C}\alpha}$ (Middle), $^1J_{\text{NC}\alpha}$ (Lower Middle) and $^2J_{\text{NC}\alpha}$ (Bottom) from Mantsyzov *et al.* and Lee *et. al.*^{247, 252}

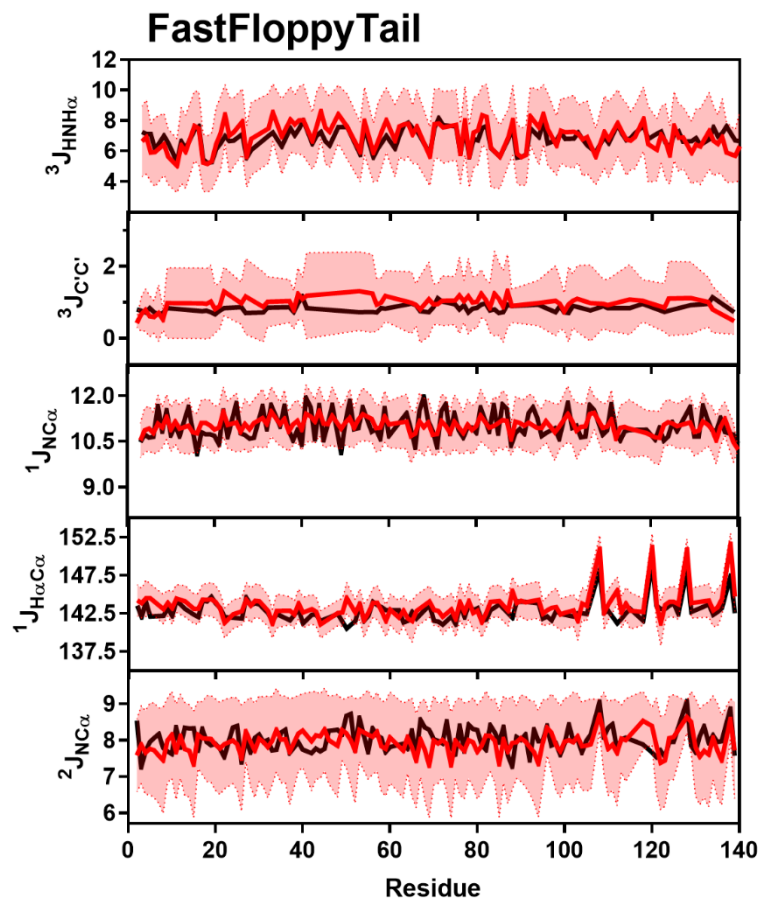


Figure 6.90: Comparison of Simulated FastFloppyTail and Experimental J-Coupling Data. Simulated Chemical Shift values from the FastFloppyTail ensemble (red) overlaid on experimental data (black) of $^3J_{\text{HNH}\alpha}$ (Top), $^3J_{\text{HNH}\alpha}$ (Upper Middle), $^1J_{\text{H}\alpha\text{C}\alpha}$ (Middle), $^1J_{\text{NC}\alpha}$ (Lower Middle) and $^2J_{\text{NC}\alpha}$ (Bottom) from Mantsyzov *et al.* and Lee *et. al.*^{247, 252}

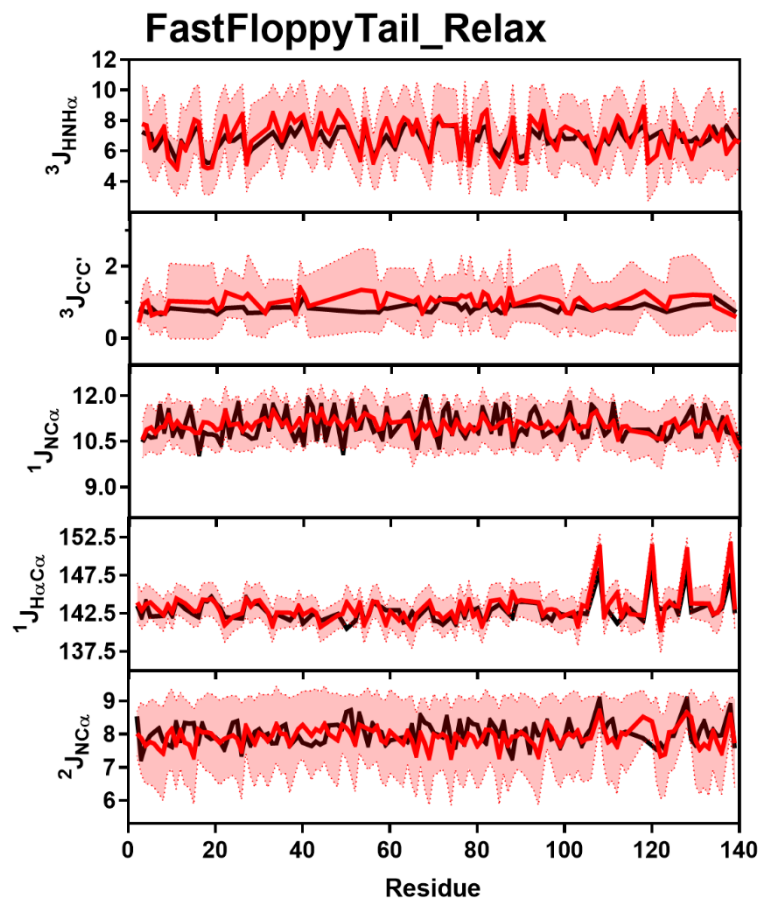


Figure 6.91: Comparison of Simulated FastFloppyTail_Relax and Experimental J-Coupling Data. Simulated Chemical Shift values from the FastFloppyTail_Relax ensemble (red) overlaid on experimental data (black) of $^3J_{\text{HNH}\alpha}$ (Top), $^3J_{\text{HNH}\alpha}$ (Upper Middle), $^1J_{\text{H}\alpha\text{C}\alpha}$ (Middle), $^1J_{\text{NC}\alpha}$ (Lower Middle) and $^2J_{\text{NC}\alpha}$ (Bottom) from Mantsyzov *et al.* and Lee *et. al.*^{247, 252}

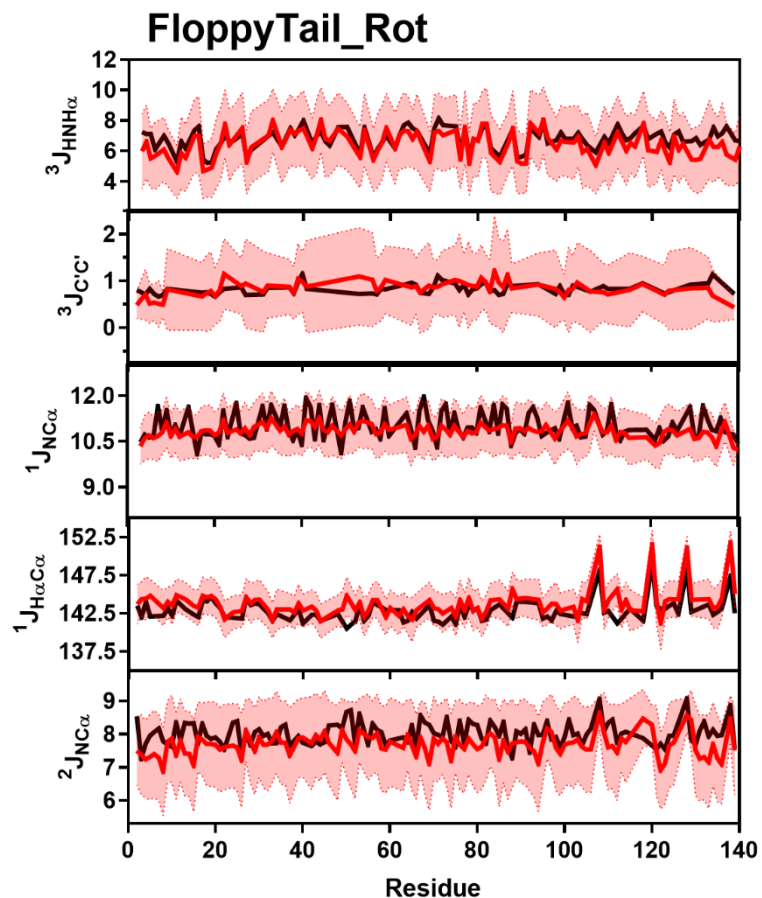


Figure 6.92: Comparison of Simulated FloppyTail_Rot and Experimental J-Coupling Data. Simulated Chemical Shift values from the FloppyTail_Rot ensemble (red) overlaid on experimental data (black) of $^3J_{\text{HNH}\alpha}$ (Top), $^3J_{\text{HNH}\alpha}$ (Upper Middle), $^1J_{\text{H}\alpha\text{C}\alpha}$ (Middle), $^1J_{\text{NC}\alpha}$ (Lower Middle) and $^2J_{\text{NC}\alpha}$ (Bottom) from Mantsyzov *et al.* and Lee *et. al.*^{247, 252}

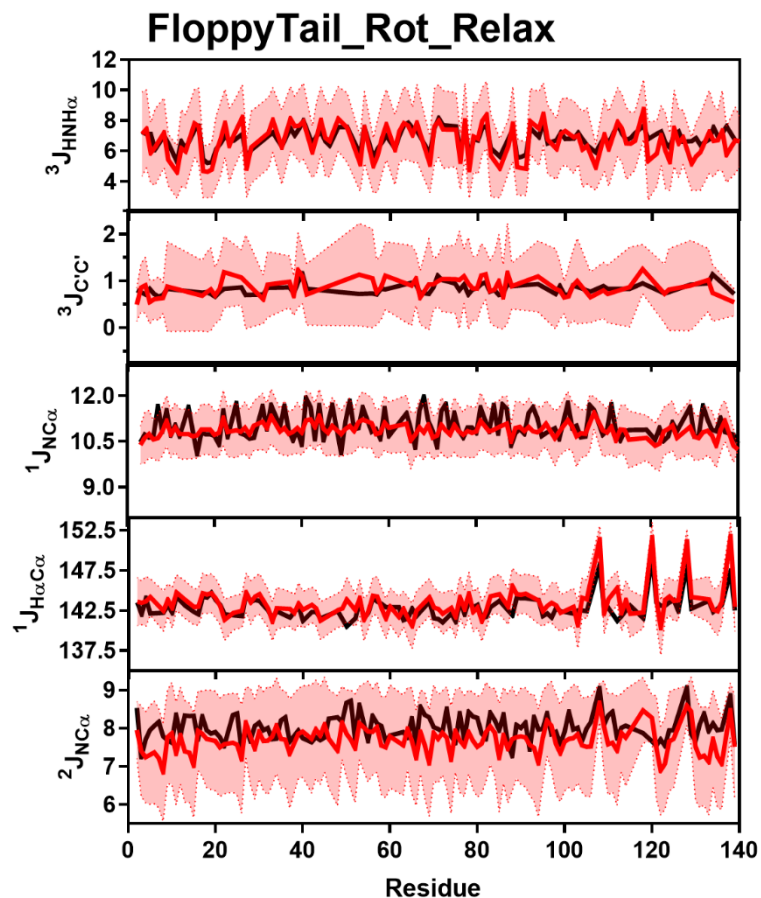


Figure 6.93: Comparison of Simulated FloppyTail_Rot_Relax and Experimental J-Coupling Data. Simulated Chemical Shift values from the FloppyTail_Rot_Relax ensemble (red) overlaid on experimental data (black) of $^3J_{\text{HNH}\alpha}$ (Top), $^3J_{\text{HNH}\alpha}$ (Upper Middle), $^1J_{\text{H}\alpha\text{C}\alpha}$ (Middle), $^1J_{\text{NC}\alpha}$ (Lower Middle) and $^2J_{\text{NC}\alpha}$ (Bottom) from Mantsyzov *et al.* and Lee *et. al.*^{247, 252}

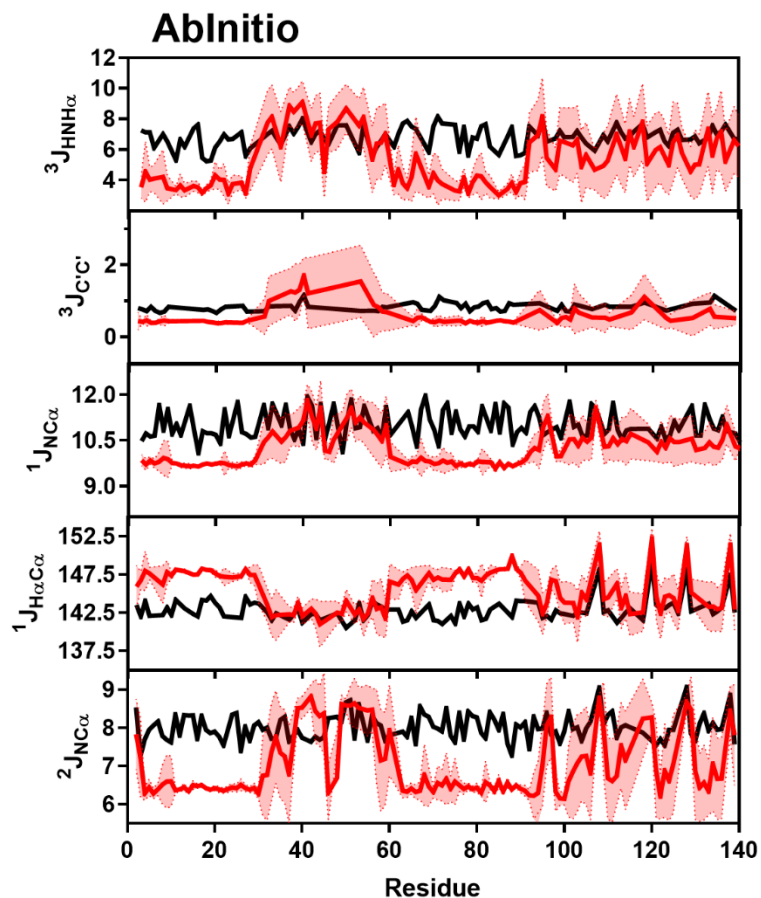


Figure 6.94: Comparison of Simulated AbInitio and Experimental J-Coupling Data. Simulated Chemical Shift values from the AbInitio ensemble (red) overlaid on experimental data (black) of ${}^3J_{\text{HNH}\alpha}$ (Top), ${}^3J_{\text{HNH}\alpha}$ (Upper Middle), ${}^1J_{\text{H}\alpha\text{C}\alpha}$ (Middle), ${}^1J_{\text{NC}\alpha}$ (Lower Middle) and ${}^2J_{\text{NC}\alpha}$ (Bottom) from Mantsyzov *et al.* and Lee *et. al.*^{247, 252}

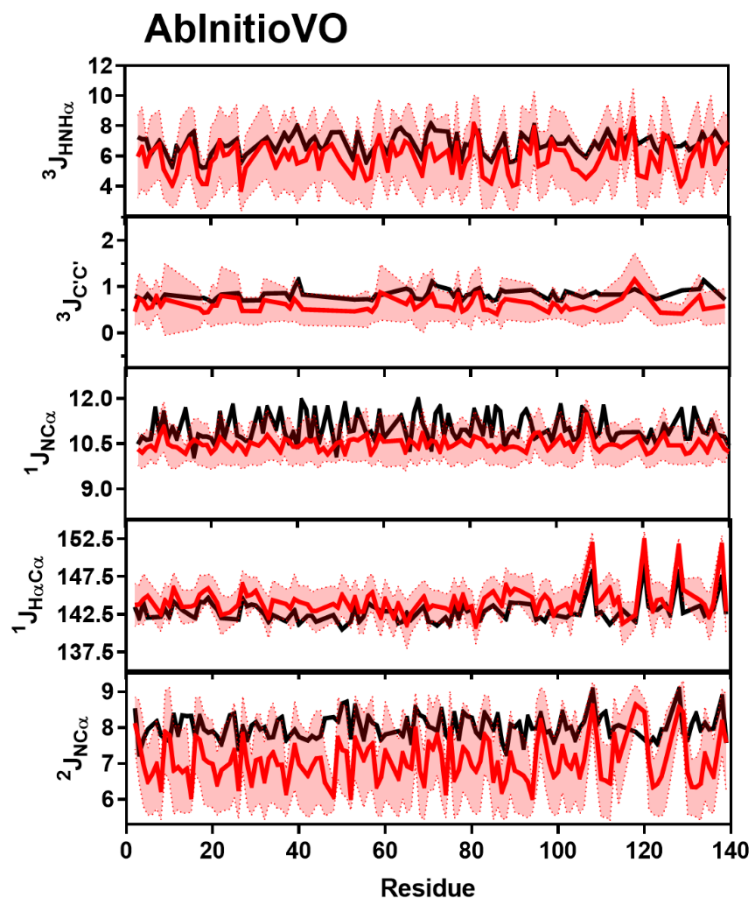


Figure 6.95: Comparison of Simulated AbInitioVO and Experimental J-Coupling Data. Simulated Chemical Shift values from the AbInitioVO ensemble (red) overlaid on experimental data (black) of $^3J_{\text{HNH}\alpha}$ (Top), $^3J_{\text{HNH}\alpha}$ (Upper Middle), $^1J_{\text{H}\alpha\text{C}\alpha}$ (Middle), $^1J_{\text{NC}\alpha}$ (Lower Middle) and $^2J_{\text{NC}\alpha}$ (Bottom) from Mantsyzov *et al.* and Lee *et al.*^{247, 252}

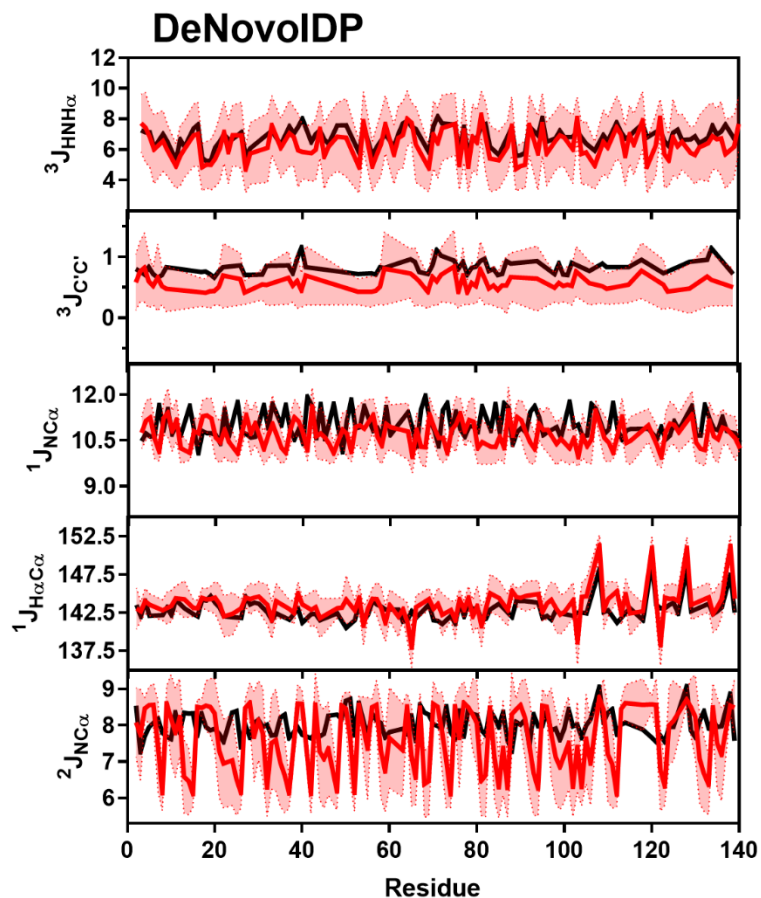


Figure 6.96: Comparison of Simulated DeNovoIDP and Experimental J-Coupling Data. Simulated Chemical Shift values from the DeNovoIDP ensemble (red) overlaid on experimental data (black) of $^3J_{\text{HNH}\alpha}$ (Top), $^3J_{\text{HNH}\alpha}$ (Upper Middle), $^1J_{\text{H}\alpha\text{C}\alpha}$ (Middle), $^1J_{\text{NC}\alpha}$ (Lower Middle) and $^2J_{\text{NC}\alpha}$ (Bottom) from Mantsyzov *et al.* and Lee *et al.*^{247, 252}

Comparison of AbInitio and AbInitioVO Simulations to Experimental Data

Table 6.5: Comparison of AbInitio and AbInitioVO Outputs to PDB Structures

PDB ID	# of Res.	Ordered Region	Fold Type	Min RMSD		<Rg>		Exp Rg
				AbInitio	AbInitio VO	AbInitio	AbInitio VO	
Ordered								
1B3A	67	-	α/β	2.68	3.31	12.0	11.8	-
1BK2	57	-	β	2.33	2.16	11.4	11.1	-
1BQ9	54	-	β	2.64	3.91	11.0	10.7	-
1ENH	54	-	α	0.77	1.18	10.7	10.5	-
1HZ6	64	-	α/β	2.16	1.99	12.4	12.1	-
1PGX	70	-	α	3.71	3.54	13.0	13.0	-
1R69	63	-	α	0.99	1.14	10.7	10.5	-
1SHF	59	-	β	3.30	1.52	11.8	11.2	-
1UBI	76	-	α/β	2.72	2.59	12.5	12.3	-
5CRO	61	-	α/β	4.35	4.56	11.3	11.0	-
Partially-Ordered								
1D7Q	144	27-115	α/β	6.38	6.41	17.4	21.8	-
1EJF	160	1-110	β	6.97	7.73	18.4	26.5	-
1FOX	76	1-76	α/β	1.10	1.22	12.7	12.1	-
1GHC	225	41-114	α/β	3.04	2.33	21.6	47.6	-
1LWM	93	26-93	α	1.43	3.61	15.2	17.5	-
1VZS	76	1-49	α	2.02	1.98	13.3	12.8	-
2L42	100	14-90	α/β	6.70	5.32	14.3	14.0	-
2LSU	110	1-89	α	1.06	1.64	14.9	13.9	-
Disordered								
ASYN	140	-	-	-	-	17.4	28.4	33.0 / 26.6
DSH3	59	-	-	-	-	12.3	11.5	10.3 / 18.3
NTAL	125	-	-	-	-	15.9	27.0	27.5
PAAA	142	-	-	-	-	18.4	19.6	22.4
SIC1	92	-	-	-	-	14.0	30.6	32.1
TAUK	130	-	-	-	-	16.0	15.9	-

* Experimental values for the radius of gyration were determined for ASYN from SAXS²²⁸ and NMR²⁰⁵ experiments, for DSH3 from NMR experiments capturing the Rg of the folded²⁵⁸ and unfolded²⁵⁹ states, for NTAL from NMR experiments²⁶⁰, for PAAA from a simulated ensemble generated from SAXS and NMR experiments²⁶¹, for SIC1 from simulated ensembles from SAXS and NMR experiments²⁶² and was not reported for TAUK.¹²⁴

Comparison of Radii of Gyration:

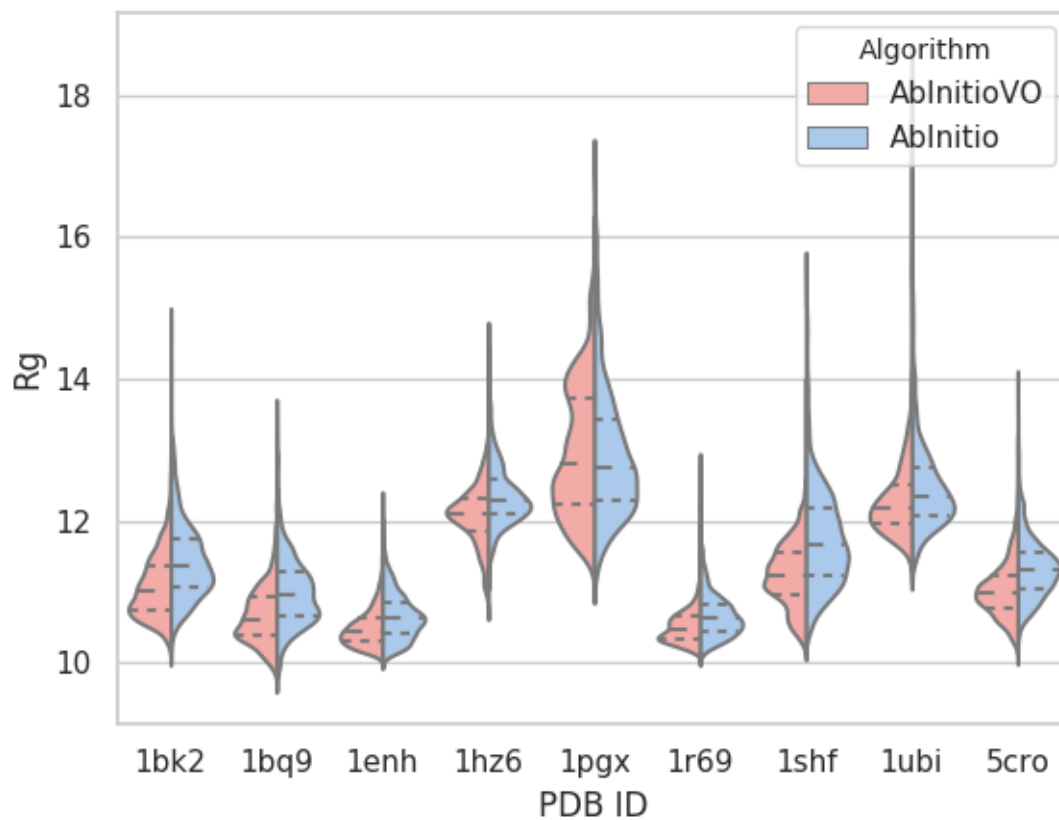


Figure 6.97: Radii of Gyration of Ordered Proteins.

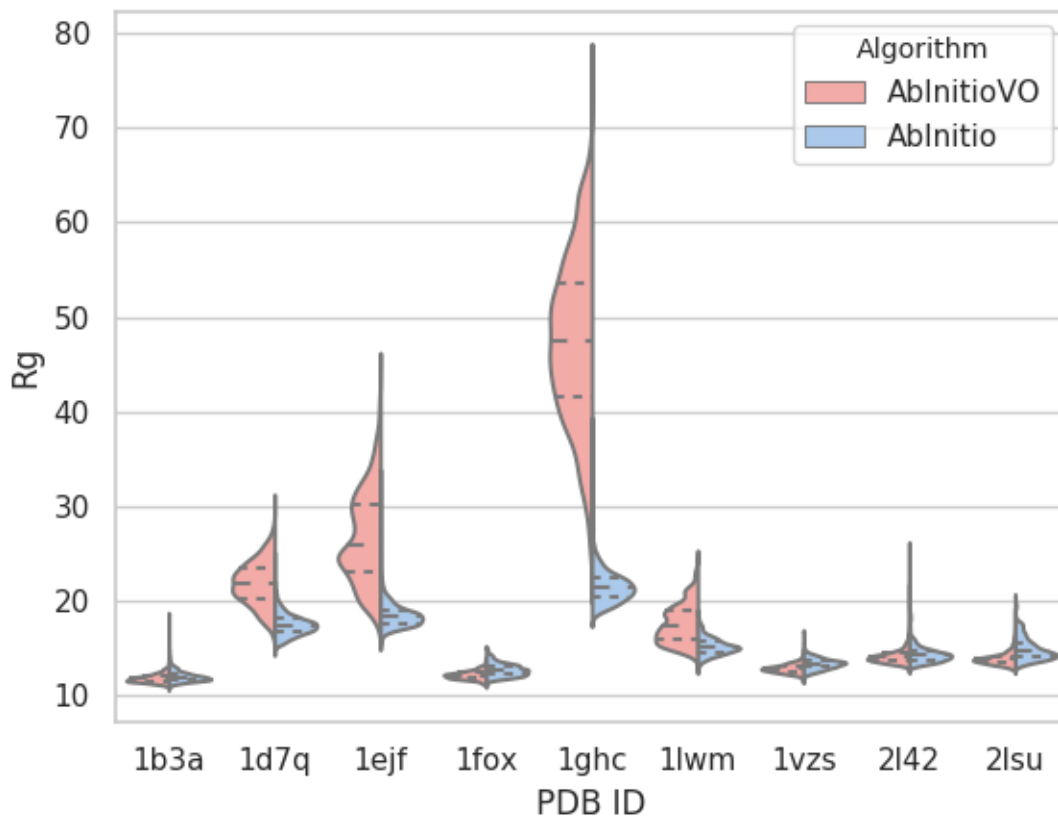


Figure 6.98: Radii of Gyration of Partially-Ordered Proteins.

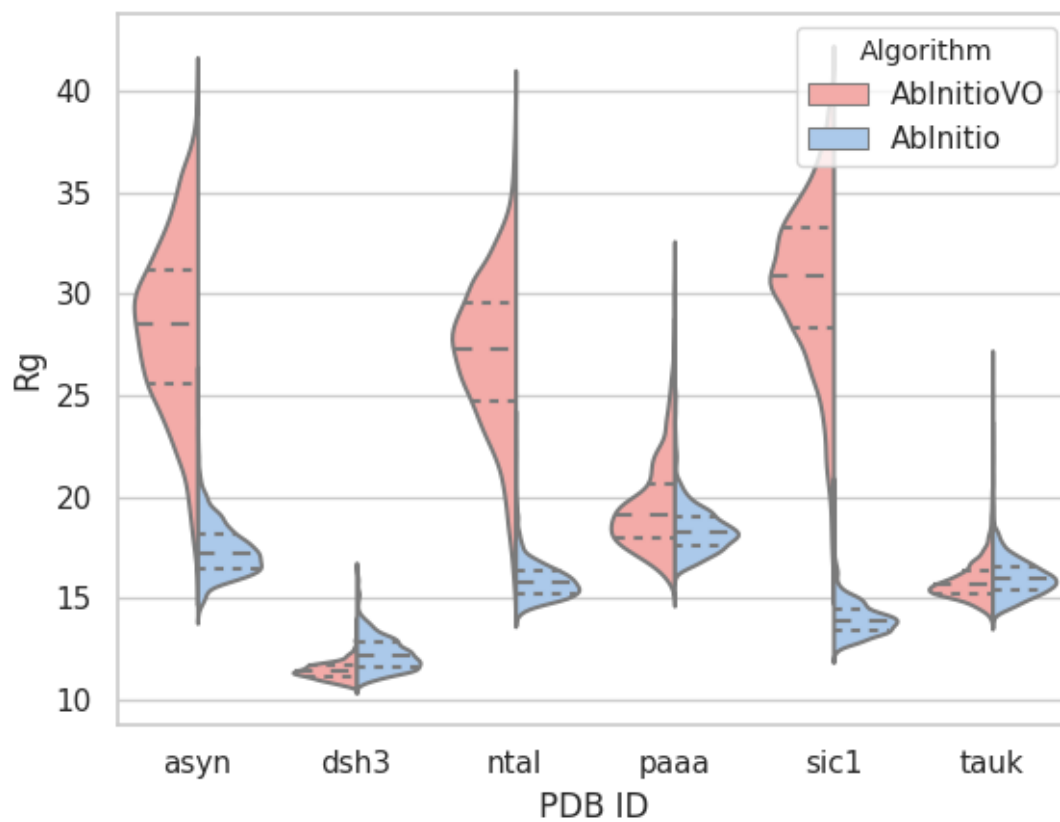


Figure 6.99: Radii of Gyration of Disordered Proteins.

Comparison of Folding Funnels

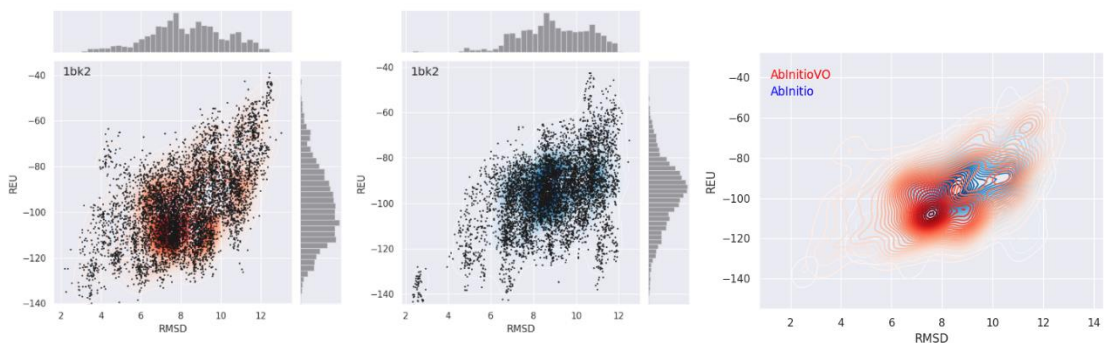


Figure 6.100: Folding Funnel Comparison for 1bk2. Folding funnel from AbInitioVO (left) and AbInitio (middle) simulations plotting C α RMSD versus Rosetta Energy Units (REU) from each structure compared to the PDB structure. For each folding funnel, histograms of the computed RMSD (top) and REU (right) are shown. KDE plot (right) showing overlay of AbInitiovo (red) and AbInitio (blue) folding funnels.

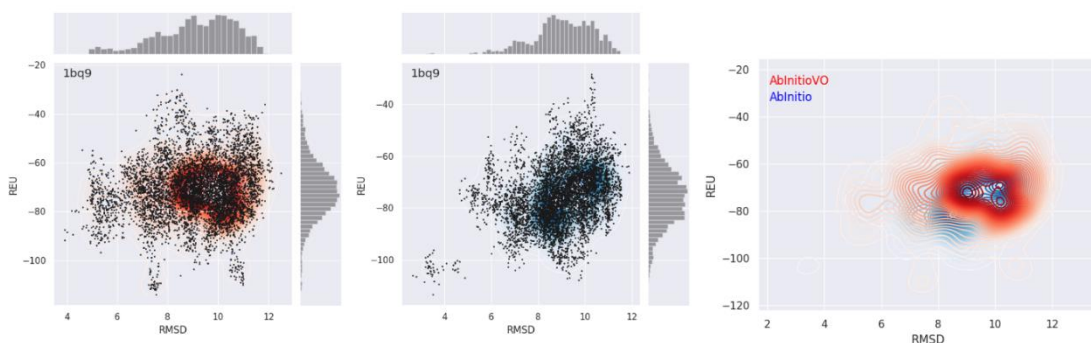


Figure 6.101: Folding Funnel Comparison for 1bq9. Folding funnel from AbInitioVO (left) and AbInitio (middle) simulations plotting C α RMSD versus Rosetta Energy Units (REU) from each structure compared to the PDB structure. For each folding funnel, histograms of the computed RMSD (top) and REU (right) are shown. KDE plot (right) showing overlay of AbInitiovo (red) and AbInitio (blue) folding funnels.

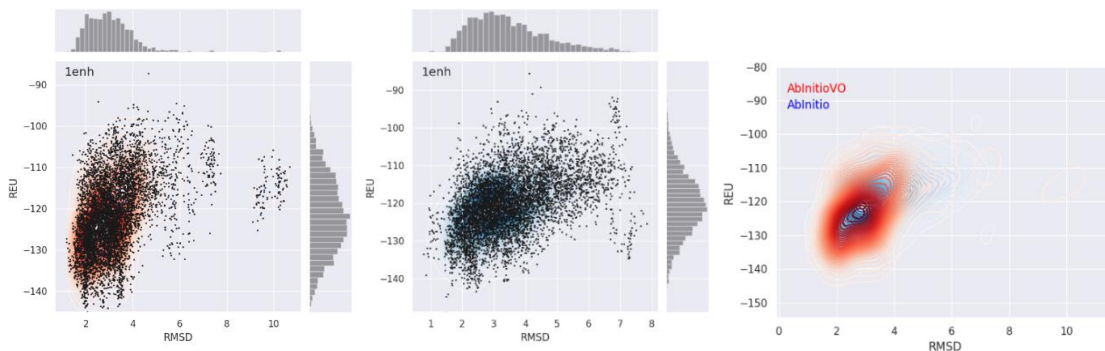


Figure 6.102: Folding Funnel Comparison for 1enh. Folding funnel from AbInitioVO (left) and AbInitio (middle) simulations plotting C α RMSD versus Rosetta Energy Units (REU) from each structure compared to the PDB structure. For each folding funnel, histograms of the computed RMSD (top) and REU (right) are shown. KDE plot (right) showing overlay of AbInitiovo (red) and AbInitio (blue) folding funnels.

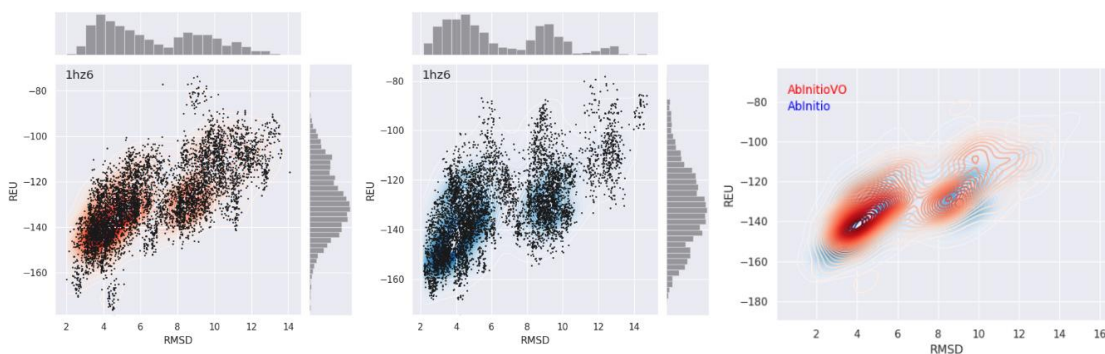


Figure 6.103: Folding Funnel Comparison for 1hz6. Folding funnel from AbInitioVO (left) and AbInitio (middle) simulations plotting C α RMSD versus Rosetta Energy Units (REU) from each structure compared to the PDB structure. For each folding funnel, histograms of the computed RMSD (top) and REU (right) are shown. KDE plot (right) showing overlay of AbInitiovo (red) and AbInitio (blue) folding funnels.

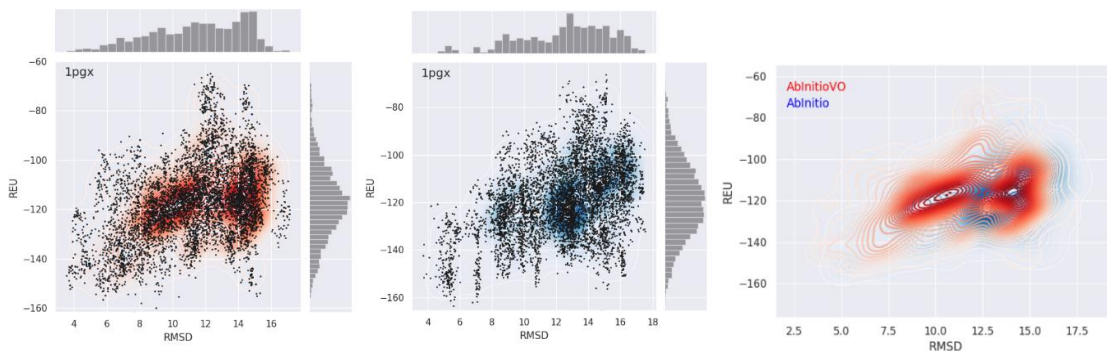


Figure 6.104: Folding Funnel Comparison for 1pgx. Folding funnel from AbInitioVO (left) and AbInitio (middle) simulations plotting $C\alpha$ RMSD versus Rosetta Energy Units (REU) from each structure compared to the PDB structure. For each folding funnel, histograms of the computed RMSD (top) and REU (right) are shown. KDE plot (right) showing overlay of AbInitiovo (red) and AbInitio (blue) folding funnels.

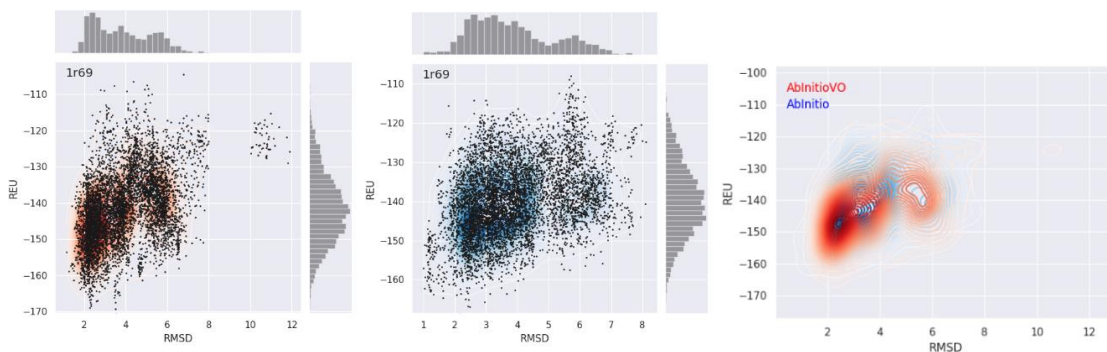


Figure 6.105: Folding Funnel Comparison for 1r69. Folding funnel from AbInitioVO (left) and AbInitio (middle) simulations plotting $C\alpha$ RMSD versus Rosetta Energy Units (REU) from each structure compared to the PDB structure. For each folding funnel, histograms of the computed RMSD (top) and REU (right) are shown. KDE plot (right) showing overlay of AbInitiovo (red) and AbInitio (blue) folding funnels.

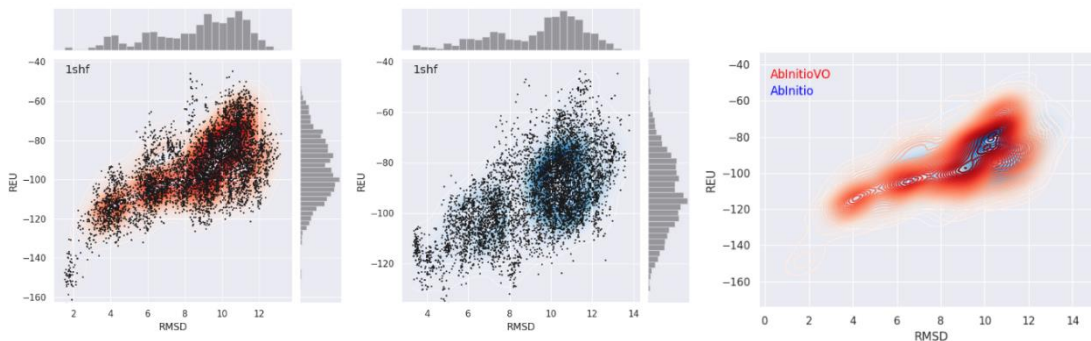


Figure 6.106: Folding Funnel Comparison for 1shf. Folding funnel from AbInitioVO (left) and AbInitio (middle) simulations plotting C α RMSD versus Rosetta Energy Units (REU) from each structure compared to the PDB structure. For each folding funnel, histograms of the computed RMSD (top) and REU (right) are shown. KDE plot (right) showing overlay of AbInitioVO (red) and AbInitio (blue) folding funnels.

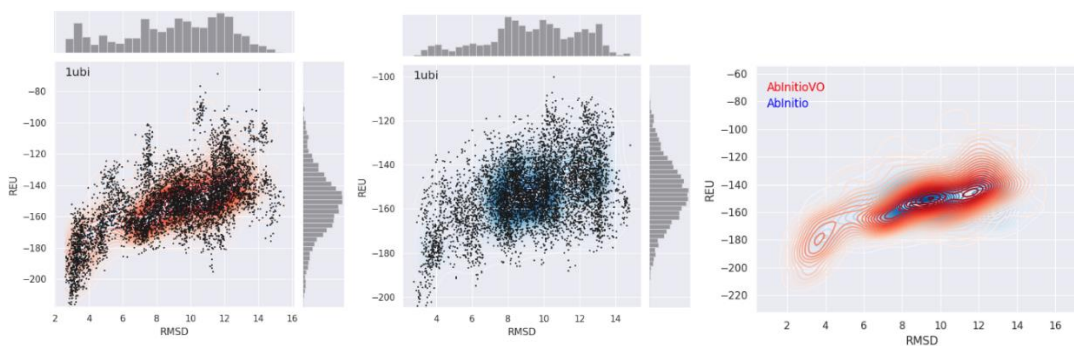


Figure 6.107: Folding Funnel Comparison for 1ubi. Folding funnel from AbInitioVO (left) and AbInitio (middle) simulations plotting C α RMSD versus Rosetta Energy Units (REU) from each structure compared to the PDB structure. For each folding funnel, histograms of the computed RMSD (top) and REU (right) are shown. KDE plot (right) showing overlay of AbInitioVO (red) and AbInitio (blue) folding funnels.

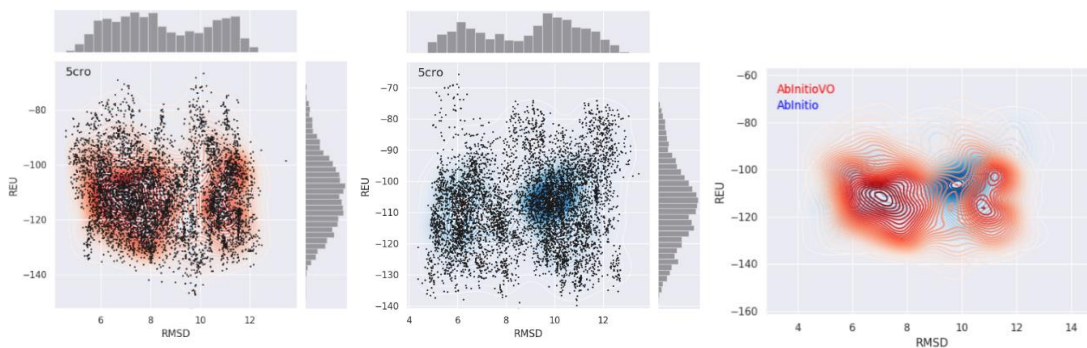


Figure 6.108: Folding Funnel Comparison for 5cro. Folding funnel from AbInitioVO (left) and AbInitio (middle) simulations plotting C α RMSD versus Rosetta Energy Units (REU) from each structure compared to the PDB structure. For each folding funnel, histograms of the computed RMSD (top) and REU (right) are shown. KDE plot (right) showing overlay of AbInitiovo (red) and AbInitio (blue) folding funnels.

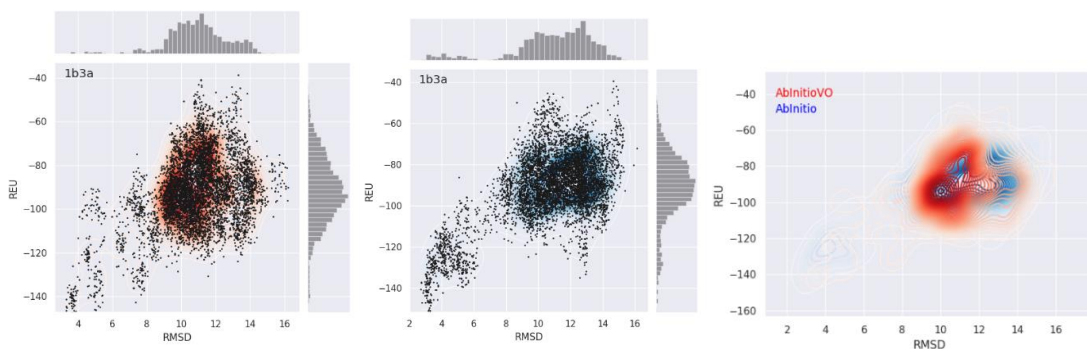


Figure 6.109: Folding Funnel Comparison for 1b3a. Folding funnel from AbInitioVO (left) and AbInitio (middle) simulations plotting C α RMSD versus Rosetta Energy Units (REU) from each structure compared to the folded domain of the PDB structure. For each folding funnel, histograms of the computed RMSD (top) and REU (right) are shown. KDE plot (right) showing overlay of AbInitiovo (red) and AbInitio (blue) folding funnels.

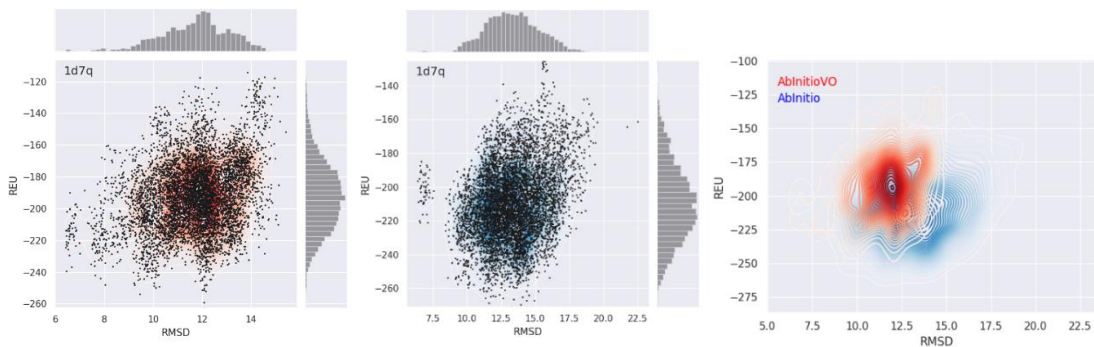


Figure 6.110: Folding Funnel Comparison for 1d7q. Folding funnel from AbInitioVO (left) and AbInitio (middle) simulations plotting C α RMSD versus Rosetta Energy Units (REU) from each structure compared to the folded domain of the PDB structure. For each folding funnel, histograms of the computed RMSD (top) and REU (right) are shown. KDE plot (right) showing overlay of AbInitioVO (red) and AbInitio (blue) folding funnels.

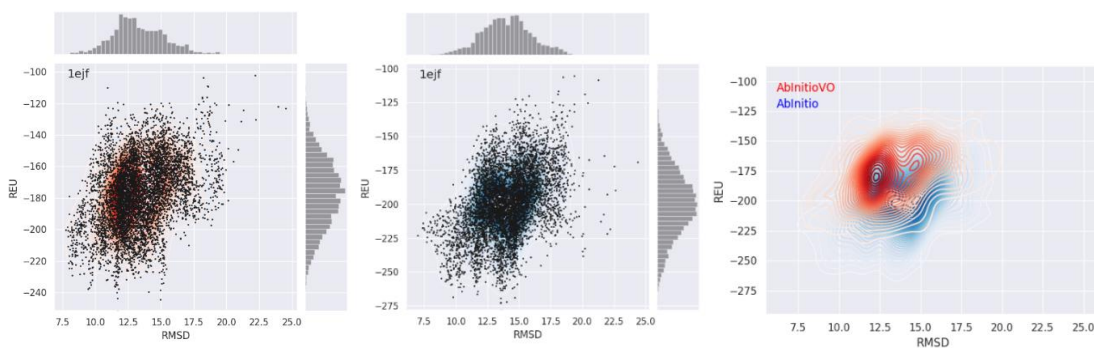


Figure 6.111: Folding Funnel Comparison for 1ejf. Folding funnel from AbInitioVO (left) and AbInitio (middle) simulations plotting C α RMSD versus Rosetta Energy Units (REU) from each structure compared to the folded domain of the PDB structure. For each folding funnel, histograms of the computed RMSD (top) and REU (right) are shown. KDE plot (right) showing overlay of AbInitioVO (red) and AbInitio (blue) folding funnels.

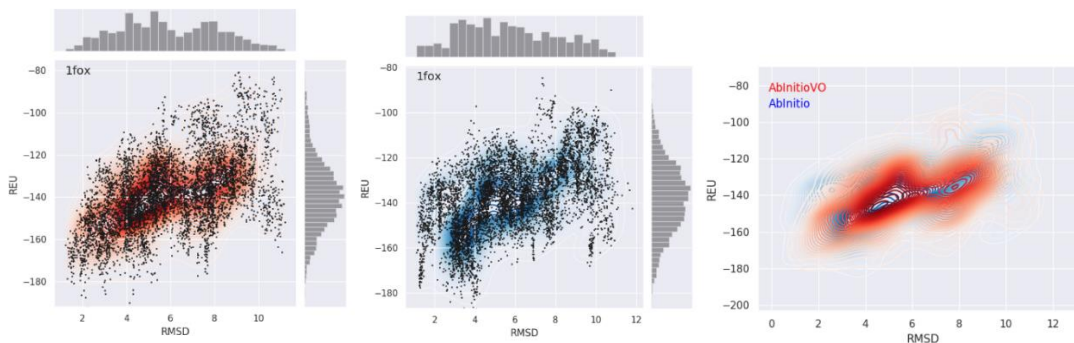


Figure 6.112: Folding Funnel Comparison for 1fox. Folding funnel from AbInitioVO (left) and AbInitio (middle) simulations plotting C α RMSD versus Rosetta Energy Units (REU) from each structure compared to the folded domain of the PDB structure. For each folding funnel, histograms of the computed RMSD (top) and REU (right) are shown. KDE plot (right) showing overlay of AbInitiovo (red) and AbInitio (blue) folding funnels.

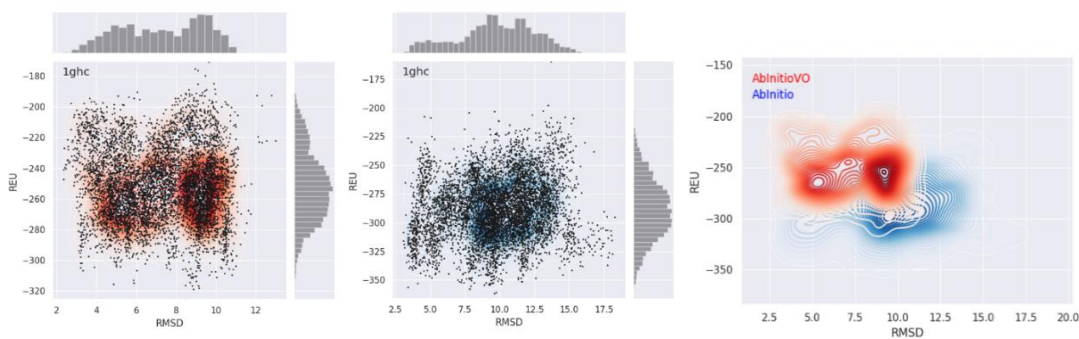


Figure 6.113: Folding Funnel Comparison for 1ghc. Folding funnel from AbInitioVO (left) and AbInitio (middle) simulations plotting C α RMSD versus Rosetta Energy Units (REU) from each structure compared to the folded domain of the PDB structure. For each folding funnel, histograms of the computed RMSD (top) and REU (right) are shown. KDE plot (right) showing overlay of AbInitiovo (red) and AbInitio (blue) folding funnels.

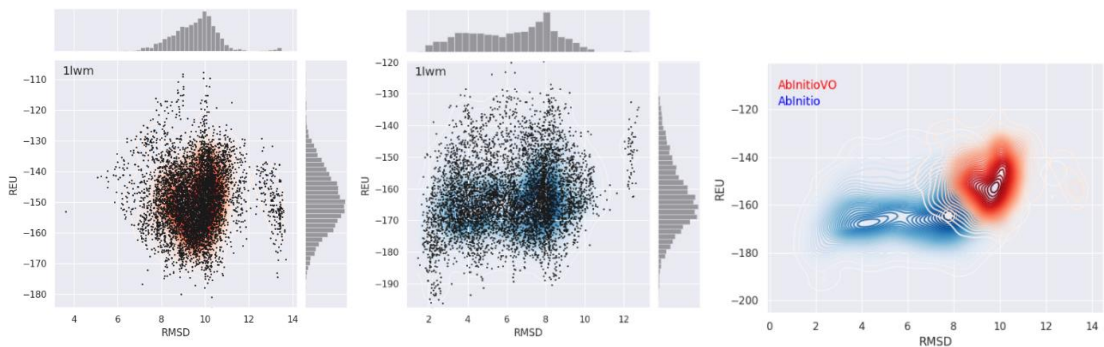


Figure 6.114: Folding Funnel Comparison for 1lwm. Folding funnel from AbInitioVO (left) and AbInitio (middle) simulations plotting C α RMSD versus Rosetta Energy Units (REU) from each structure compared to the folded domain of the PDB structure. For each folding funnel, histograms of the computed RMSD (top) and REU (right) are shown. KDE plot (right) showing overlay of AbInitioVO (red) and AbInitio (blue) folding funnels.

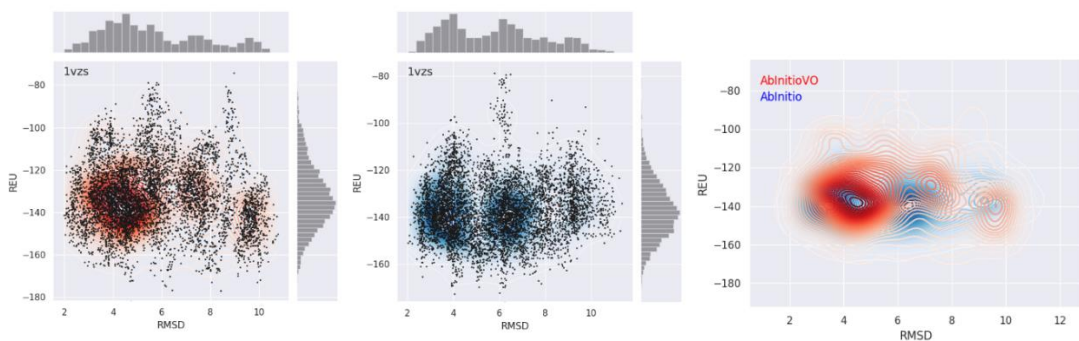


Figure 6.115: Folding Funnel Comparison for 1vzs. Folding funnel from AbInitioVO (left) and AbInitio (middle) simulations plotting C α RMSD versus Rosetta Energy Units (REU) from each structure compared to the folded domain of the PDB structure. For each folding funnel, histograms of the computed RMSD (top) and REU (right) are shown. KDE plot (right) showing overlay of AbInitioVO (red) and AbInitio (blue) folding funnels.

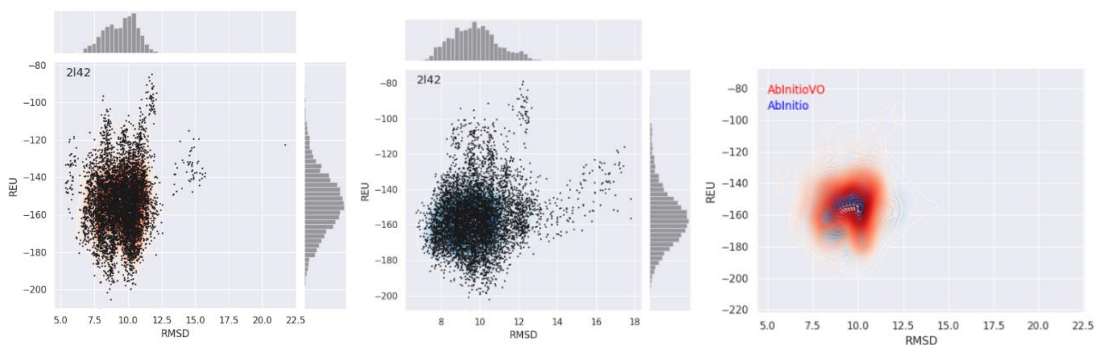


Figure 6.116: Folding Funnel Comparison for 2I42. Folding funnel from AbInitioVO (left) and AbInitio (middle) simulations plotting C α RMSD versus Rosetta Energy Units (REU) from each structure compared to the folded domain of the PDB structure. For each folding funnel, histograms of the computed RMSD (top) and REU (right) are shown. KDE plot (right) showing overlay of AbInitioVO (red) and AbInitio (blue) folding funnels.

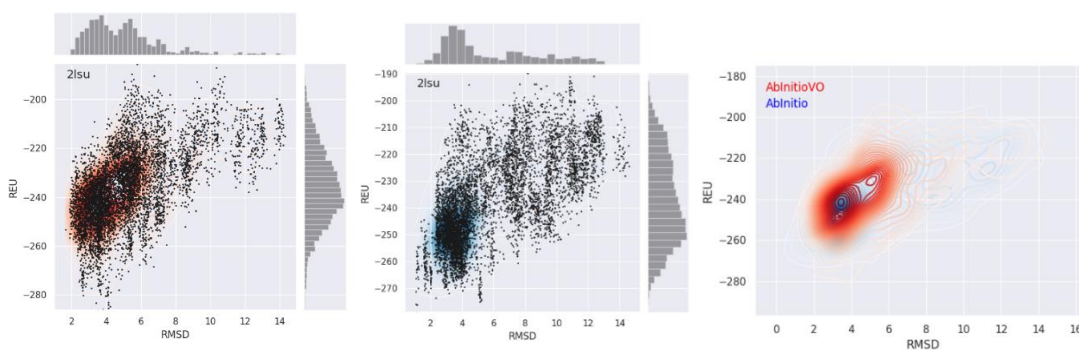


Figure 6.117: Folding Funnel Comparison for 2Isu. Folding funnel from AbInitioVO (left) and AbInitio (middle) simulations plotting C α RMSD versus Rosetta Energy Units (REU) from each structure compared to the folded domain of the PDB structure. For each folding funnel, histograms of the computed RMSD (top) and REU (right) are shown. KDE plot (right) showing overlay of AbInitioVO (red) and AbInitio (blue) folding funnels.

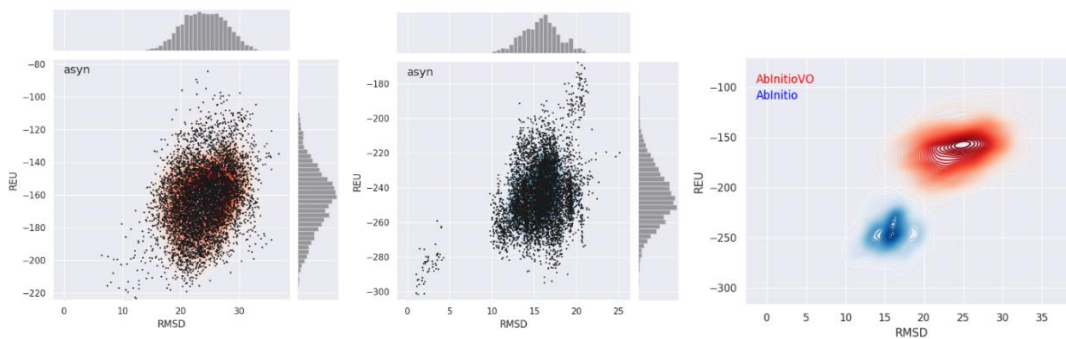


Figure 6.118: Folding Funnel Comparison for asyn. Folding funnel from AbInitioVO (left) and AbInitio (middle) simulations plotting C α RMSD versus Rosetta Energy Units (REU) from each structure compared to the lowest energy structure. For each folding funnel, histograms of the computed RMSD (top) and REU (right) are shown. KDE plot (right) showing overlay of AbInitioVO (red) and AbInitio (blue) folding funnels.

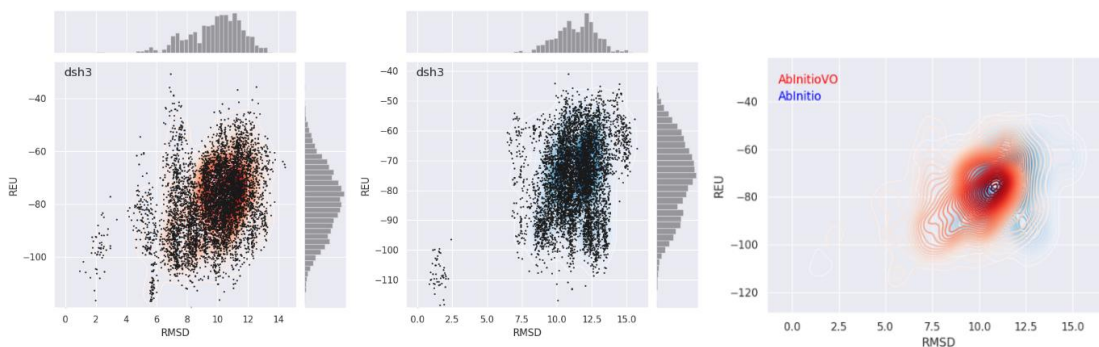


Figure 6.119: Folding Funnel Comparison for dsh3. Folding funnel from AbInitioVO (left) and AbInitio (middle) simulations plotting C α RMSD versus Rosetta Energy Units (REU) from each structure compared to the lowest energy structure. For each folding funnel, histograms of the computed RMSD (top) and REU (right) are shown. KDE plot (right) showing overlay of AbInitioVO (red) and AbInitio (blue) folding funnels.

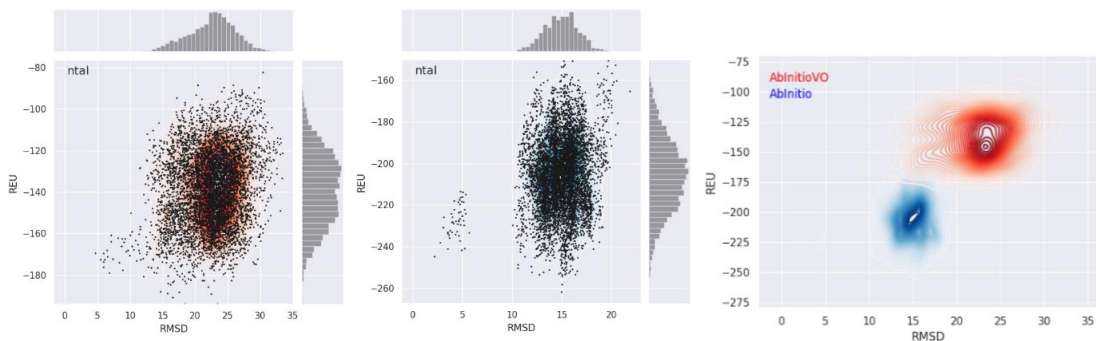


Figure 6.120: Folding Funnel Comparison for ntal. Folding funnel from AbInitioVO (left) and AbInitio (middle) simulations plotting C α RMSD versus Rosetta Energy Units (REU) from each structure compared to the lowest energy structure. For each folding funnel, histograms of the computed RMSD (top) and REU (right) are shown. KDE plot (right) showing overlay of AbInitioVO (red) and AbInitio (blue) folding funnels.

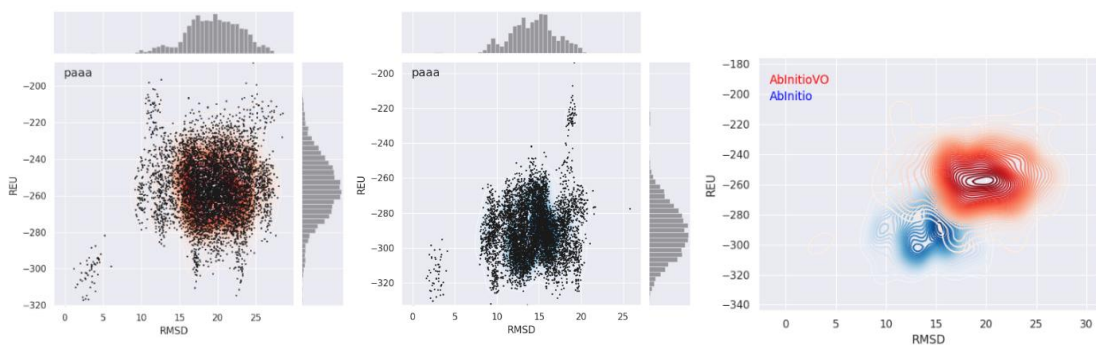


Figure 6.121: Folding Funnel Comparison for paaa. Folding funnel from AbInitioVO (left) and AbInitio (middle) simulations plotting C α RMSD versus Rosetta Energy Units (REU) from each structure compared to the lowest energy structure. For each folding funnel, histograms of the computed RMSD (top) and REU (right) are shown. KDE plot (right) showing overlay of AbInitioVO (red) and AbInitio (blue) folding funnels.

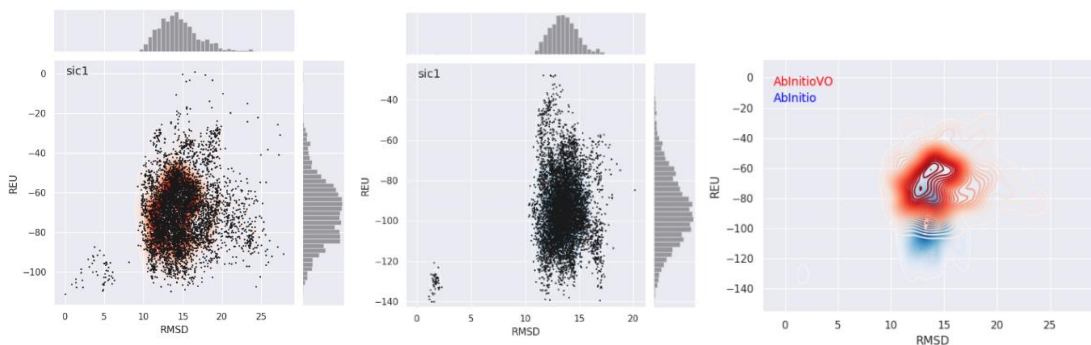


Figure 6.122: Folding Funnel Comparison for sic1. Folding funnel from AbInitioVO (left) and AbInitio (middle) simulations plotting C α RMSD versus Rosetta Energy Units (REU) from each structure compared to the lowest energy structure. For each folding funnel, histograms of the computed RMSD (top) and REU (right) are shown. KDE plot (right) showing overlay of AbInitioVO (red) and AbInitio (blue) folding funnels.

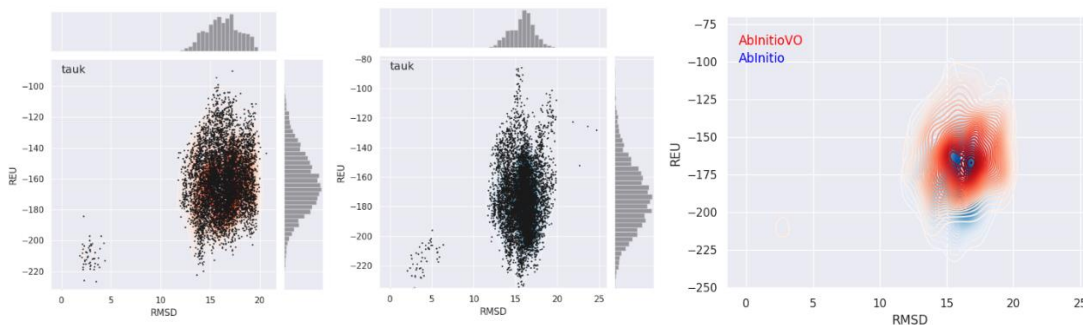


Figure 6.123: Folding Funnel Comparison for tauk. Folding funnel from AbInitioVO (left) and AbInitio (middle) simulations plotting C α RMSD versus Rosetta Energy Units (REU) from each structure compared to the lowest energy structure. For each folding funnel, histograms of the computed RMSD (top) and REU (right) are shown. KDE plot (right) showing overlay of AbInitioVO (red) and AbInitio (blue) folding funnels.

Comparisons of Folded Domain Structures of Ordered and Partially-Ordered Proteins

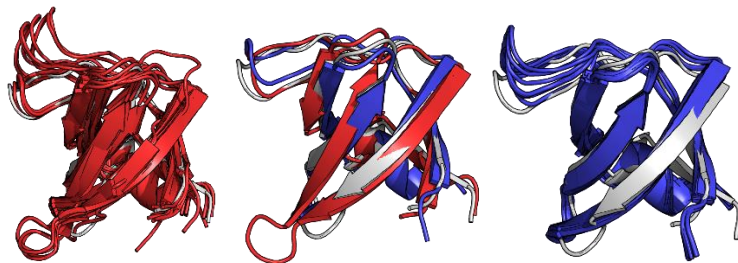


Figure 6.124: 1bk2 Lowest RMSD Structures. 10 lowest RMSD structures from AbInitioVO (left) and AbInitio (right) along with the overlay (center) of PDB 1bk2 (white) and the single lowest RMSD structure from AbInitioVO (red) and AbInitio (blue)

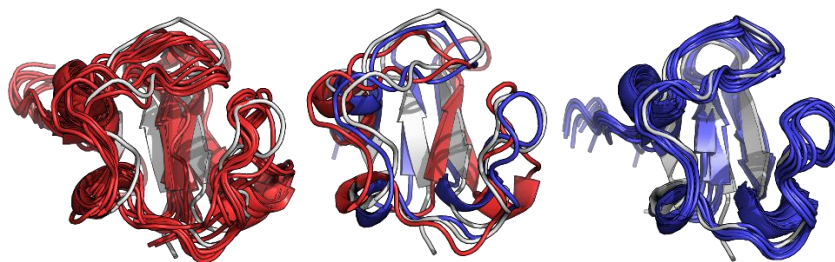


Figure 6.125: 1bq9 Lowest RMSD Structures. 10 lowest RMSD structures from AbInitioVO (left) and AbInitio (right) along with the overlay (center) of PDB 1bq9 (white) and the single lowest RMSD structure from AbInitioVO (red) and AbInitio (blue)

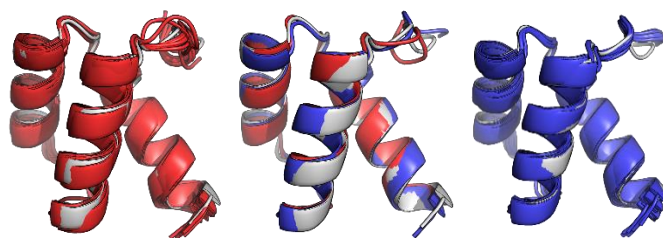


Figure 6.126: 1enh Lowest RMSD Structures. 10 lowest RMSD structures from AbInitioVO (left) and AbInitio (right) along with the overlay (center) of PDB 1enh

(white) and the single lowest RMSD structure from AbInItioVO (red) and AbInItio (blue)

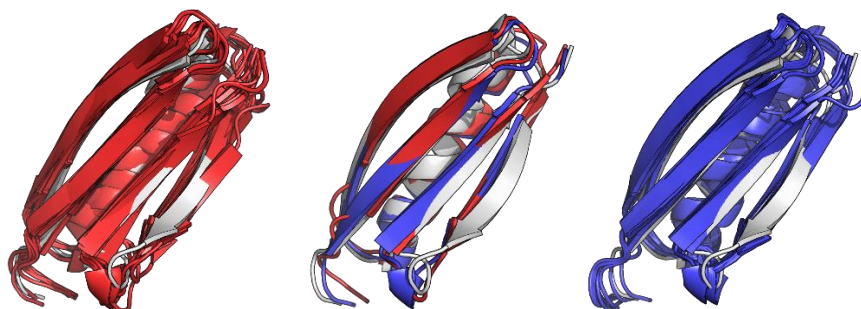


Figure 6.127: 1hz6 Lowest RMSD Structures. 10 lowest RMSD structures from AbInItioVO (left) and AbInItio (right) along with the overlay (center) of PDB 1hz6 (white) and the single lowest RMSD structure from AbInItioVO (red) and AbInItio (blue)

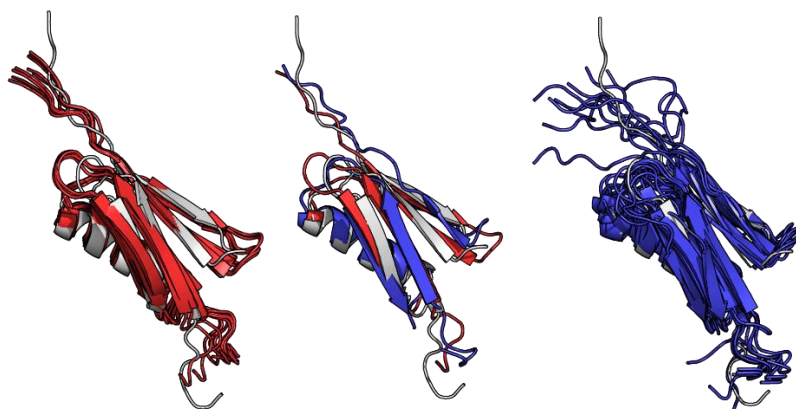


Figure 6.128: 1pgx Lowest RMSD Structures. 10 lowest RMSD structures from AbInItioVO (left) and AbInItio (right) along with the overlay (center) of PDB 1pgx (white) and the single lowest RMSD structure from AbInItioVO (red) and AbInItio (blue)

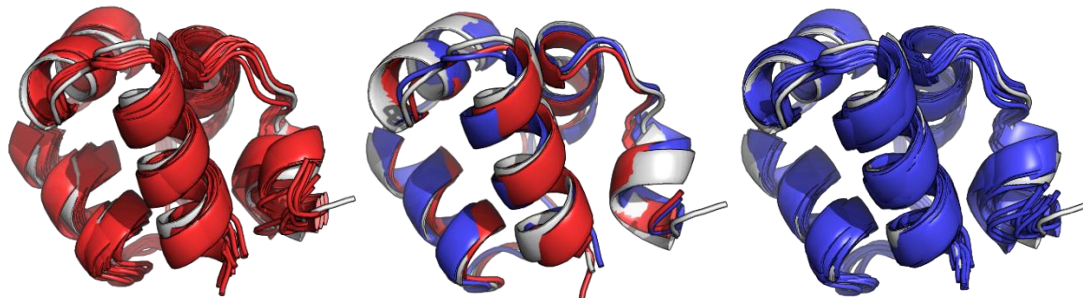


Figure 6.129: 1r69 Lowest RMSD Structures. 10 lowest RMSD structures from AbInitioVO (left) and AbInitio (right) along with the overlay (center) of PDB 1r69 (white) and the single lowest RMSD structure from AbInitioVO (red) and AbInitio (blue)

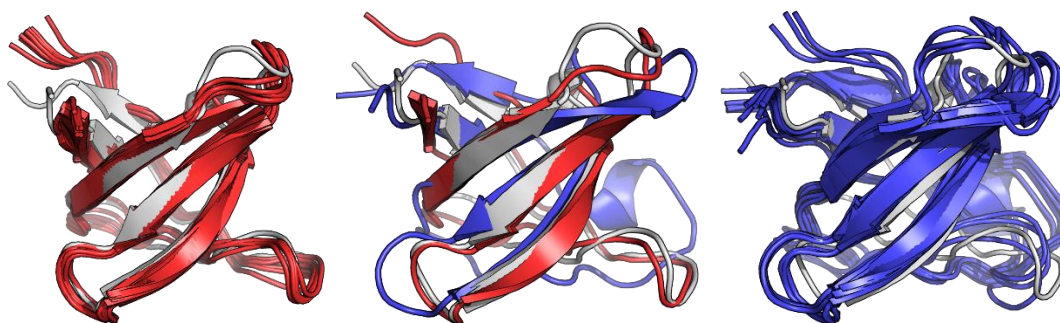


Figure 6.130: 1shf Lowest RMSD Structures. 10 lowest RMSD structures from AbInitioVO (left) and AbInitio (right) along with the overlay (center) of PDB 1shf (white) and the single lowest RMSD structure from AbInitioVO (red) and AbInitio (blue)

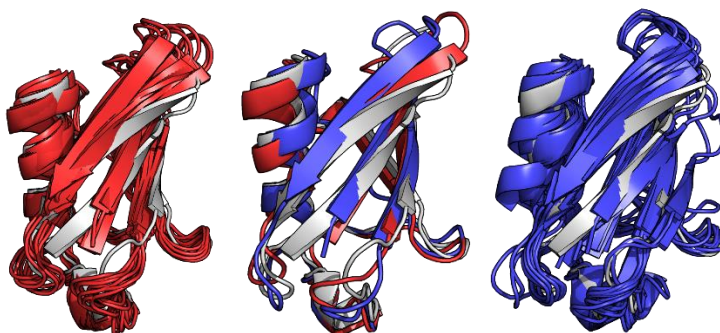


Figure 6.131: 1ubi Lowest RMSD Structures. 10 lowest RMSD structures from AbInitioVO (left) and AbInitio (right) along with the overlay (center) of PDB 1ubi (white) and the single lowest RMSD structure from AbInitioVO (red) and AbInitio (blue)

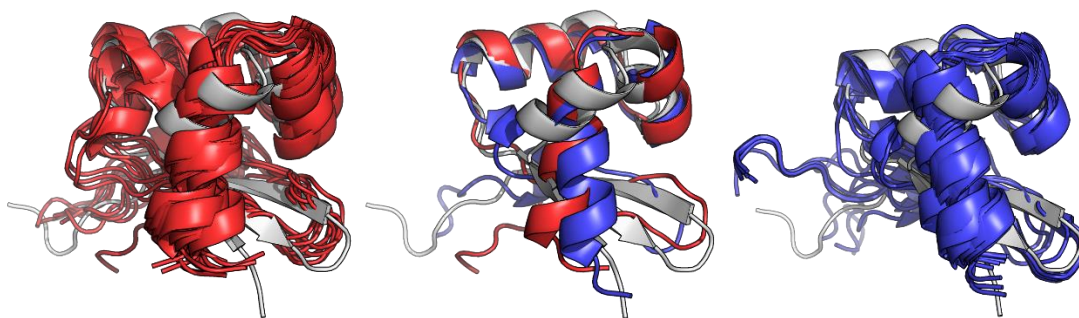


Figure 6.132: 5cro Lowest RMSD Structures. 10 lowest RMSD structures from AbInitioVO (left) and AbInitio (right) along with the overlay (center) of PDB 5cro (white) and the single lowest RMSD structure from AbInitioVO (red) and AbInitio (blue)

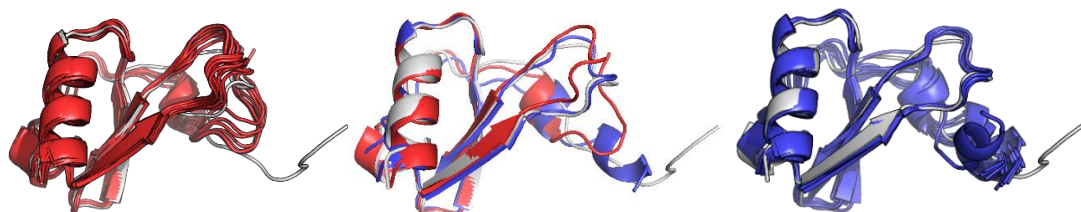


Figure 6.133: 1b3a Lowest RMSD Structures. 10 lowest RMSD structures from AbInitioVO (left) and AbInitio (right) along with the overlay (center) of PDB 1b3a (white) and the single lowest RMSD structure from AbInitioVO (red) and AbInitio (blue)

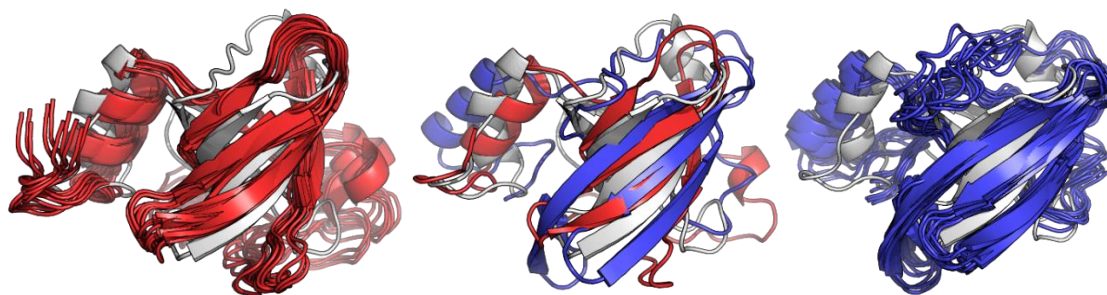


Figure 6.134: 1d7q Lowest RMSD Structures. 10 lowest RMSD structures from AbInitioVO (left) and AbInitio (right) along with the overlay (center) of PDB 1d7q (white) and the single lowest RMSD structure from AbInitioVO (red) and AbInitio (blue)

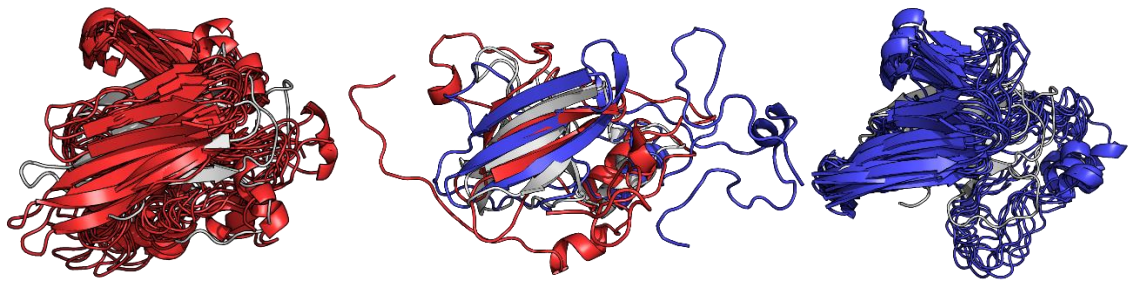


Figure 6.135: 1ejf Lowest RMSD Structures. 10 lowest RMSD structures from AbInitioVO (left) and AbInitio (right) along with the overlay (center) of PDB 1ejf (white) and the single lowest RMSD structure from AbInitioVO (red) and AbInitio (blue)

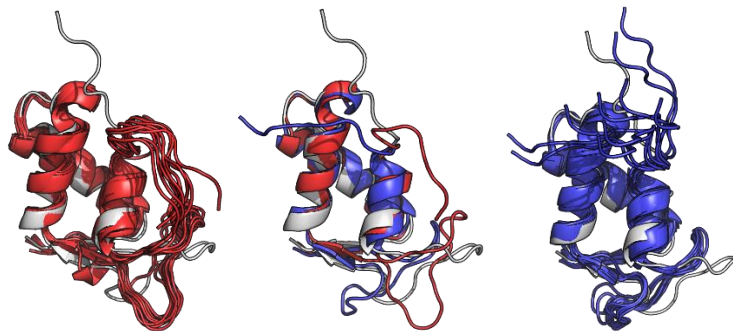


Figure 6.136: 1fox Lowest RMSD Structures. 10 lowest RMSD structures from AbInitioVO (left) and AbInitio (right) along with the overlay (center) of PDB 1fox (white) and the single lowest RMSD structure from AbInitioVO (red) and AbInitio (blue)

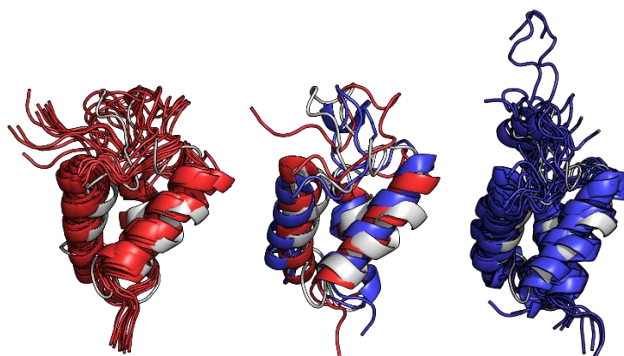


Figure 6.137: 1ghc Lowest RMSD Structures. 10 lowest RMSD structures from AbInitioVO (left) and AbInitio (right) along with the overlay (center) of PDB 1ghc

(white) and the single lowest RMSD structure from AbInitioVO (red) and AbInitio (blue)

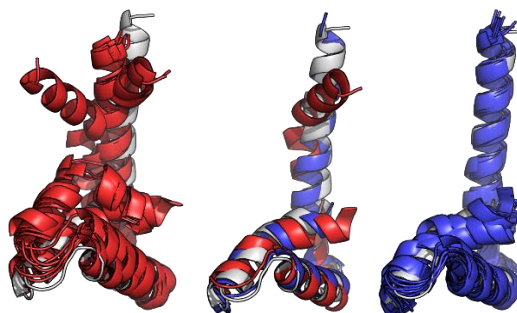


Figure 6.138: 1lwm Lowest RMSD Structures. 10 lowest RMSD structures from AbInitioVO (left) and AbInitio (right) along with the overlay (center) of PDB 1lwm (white) and the single lowest RMSD structure from AbInitioVO (red) and AbInitio (blue)

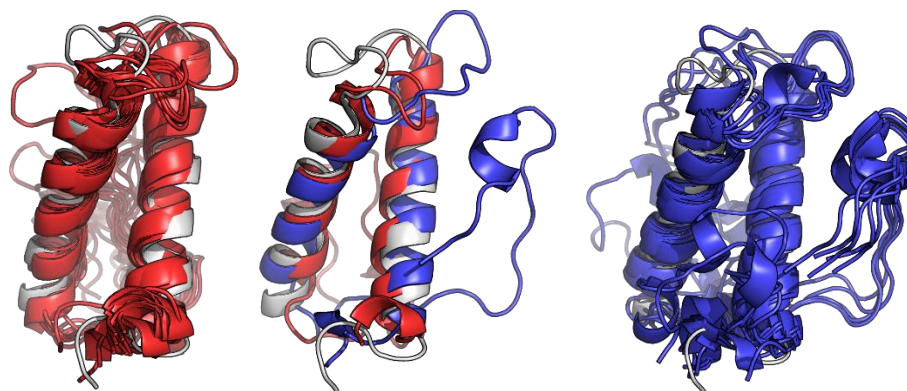


Figure 6.139: 1vzs Lowest RMSD Structures. 10 lowest RMSD structures from AbInitioVO (left) and AbInitio (right) along with the overlay (center) of PDB 1vzs (white) and the single lowest RMSD structure from AbInitioVO (red) and AbInitio (blue)

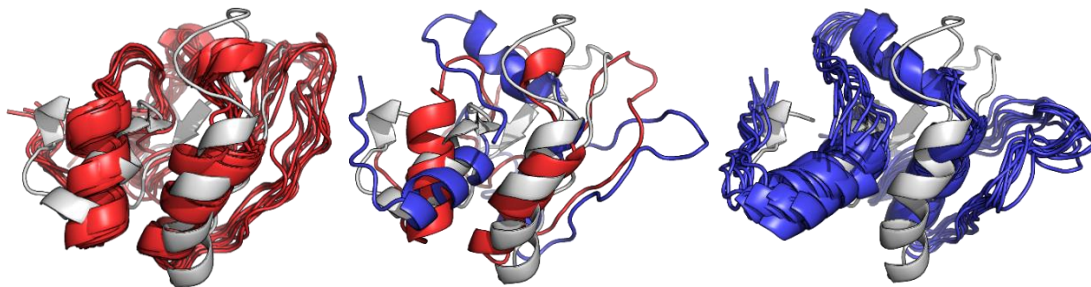


Figure 6.140: 2I42 Lowest RMSD Structures. 10 lowest RMSD structures from AbInitioVO (left) and AbInitio (right) along with the overlay (center) of PDB 2I42 (white) and the single lowest RMSD structure from AbInitioVO (red) and AbInitio (blue)

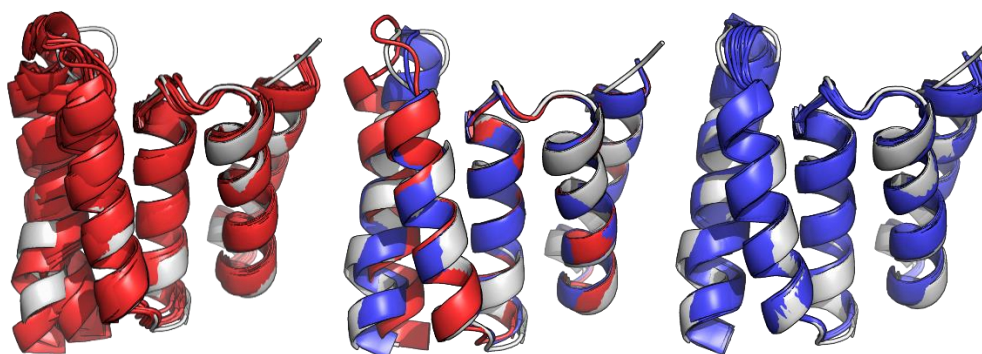


Figure 6.141: 2Isu Lowest RMSD Structures. 10 lowest RMSD structures from AbInitioVO (left) and AbInitio (right) along with the overlay (center) of PDB 2Isu (white) and the single lowest RMSD structure from AbInitioVO (red) and AbInitio (blue)

Structural Comparisons of Disordered Domains of Partially-Ordered and Disordered Proteins

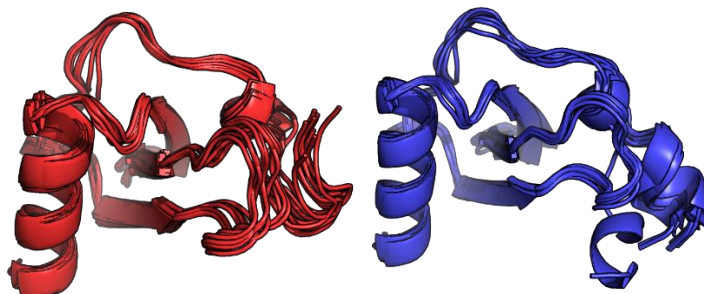


Figure 6.142: 1b3a Lowest RMSD Full Structures. 10 lowest RMSD structures from AbInitioVO (left) and AbInitio (right)

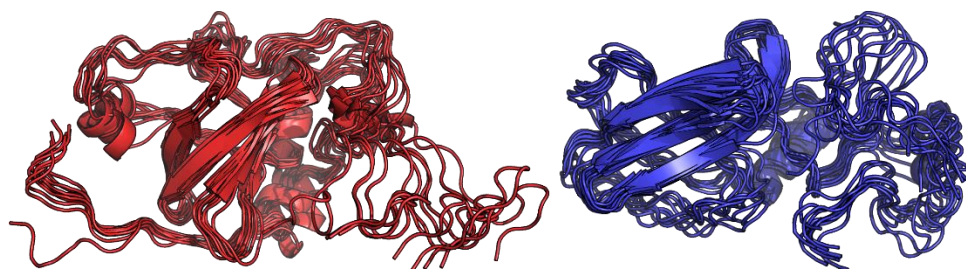


Figure 6.143: 1d7q Lowest RMSD Full Structures. 10 lowest RMSD structures from AbInitioVO (left) and AbInitio (right)

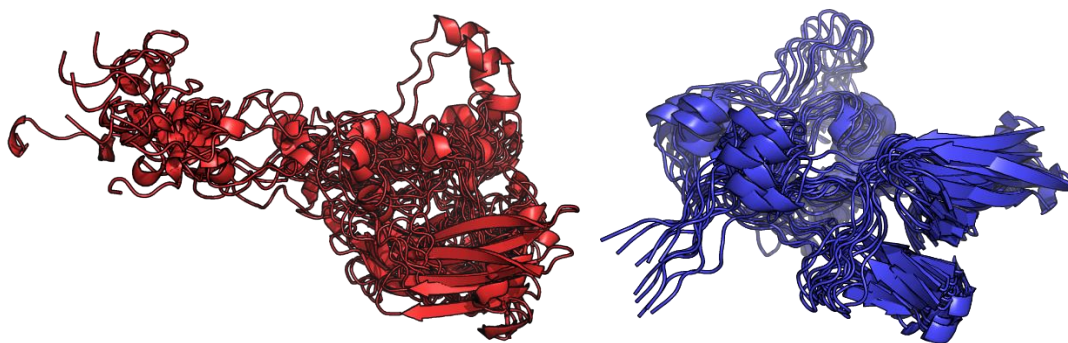


Figure 6.144: 1ejf Lowest RMSD Full Structures. 10 lowest RMSD structures from AbInitioVO (left) and AbInitio (right)

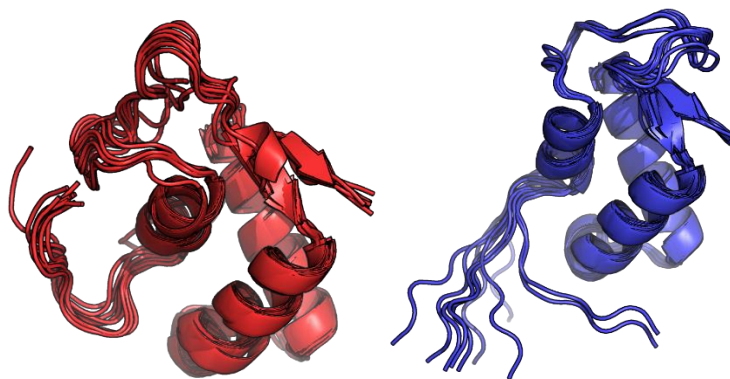


Figure 6.145: 1fox Lowest RMSD Full Structures. 10 lowest RMSD structures from AbInitioVO (left) and AbInitio (right)

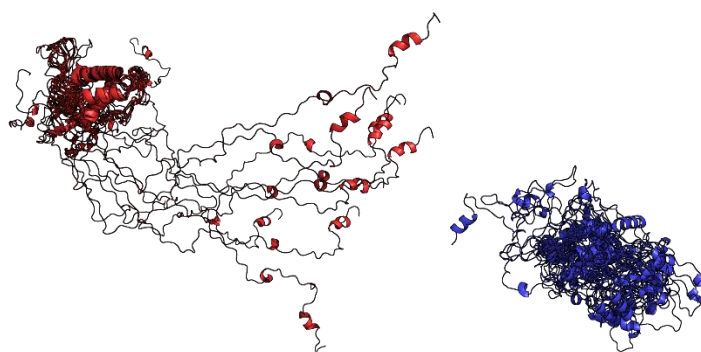


Figure 6.146: 1ghc Lowest RMSD Full Structures. 10 lowest RMSD structures from AbInitioVO (left) and AbInitio (right)

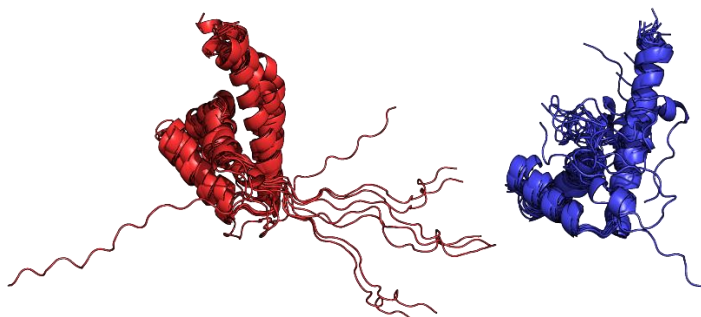


Figure 6.147: 1lwm Lowest RMSD Full Structures. 10 lowest RMSD structures from AbInitioVO (left) and AbInitio (right)

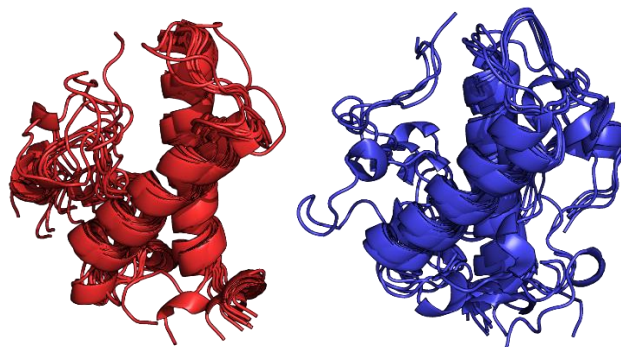


Figure 6.148: 1vzs Lowest RMSD Full Structures. 10 lowest RMSD structures from AbInitioVO (left) and AbInitio (right)

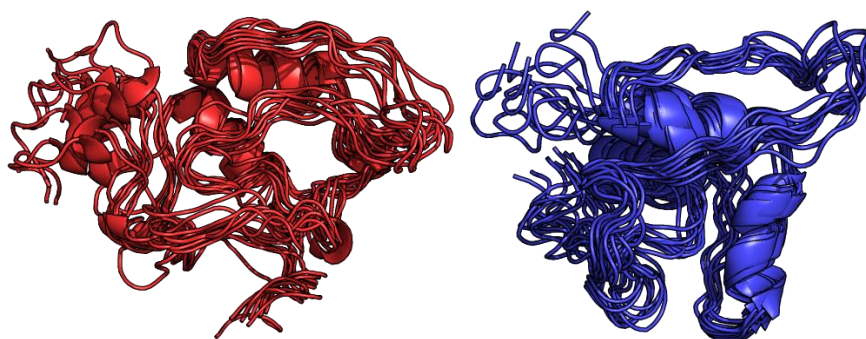


Figure 6.149: 2l42 Lowest RMSD Full Structures. 10 lowest RMSD structures from AbInitioVO (left) and AbInitio (right)

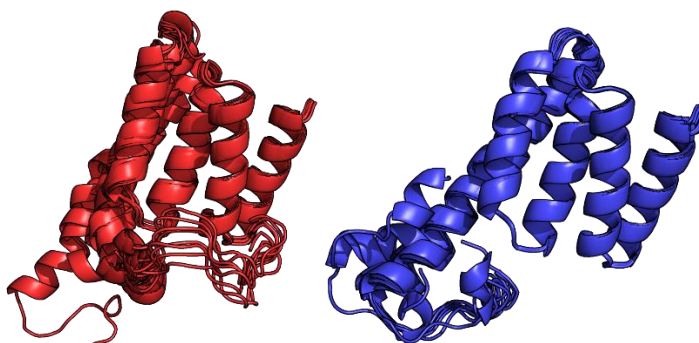


Figure 6.150: 2lsu Lowest RMSD Full Structures. 10 lowest RMSD structures from AbInitioVO (left) and AbInitio (right)

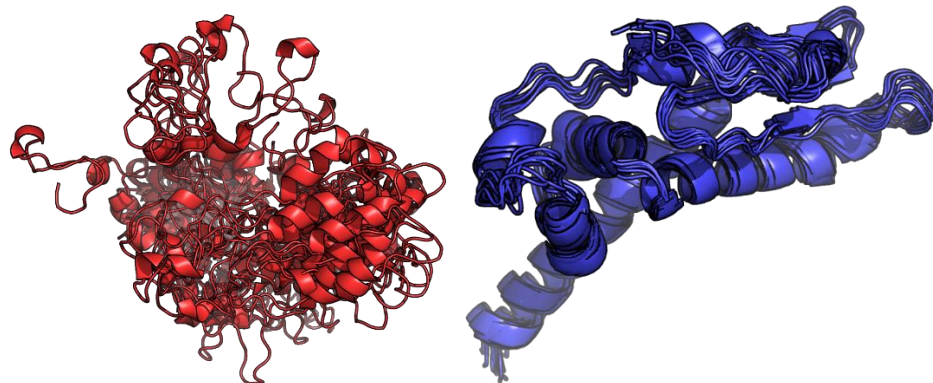


Figure 6.151: Asyn Lowest RMSD Full Structures. 10 lowest RMSD structures from AbInioVO (left) and AbInio (right)

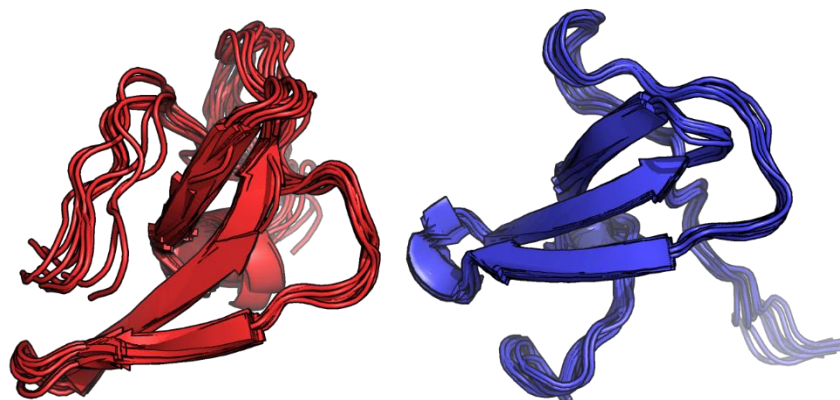


Figure 6.152: Dsh3 Lowest RMSD Full Structures. 10 lowest RMSD structures from AbInioVO (left) and AbInio (right)

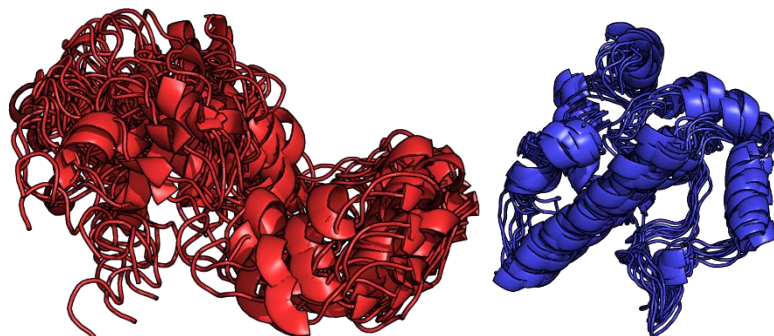


Figure 6.153: Ntal Lowest RMSD Full Structures. 10 lowest RMSD structures from AbInioVO (left) and AbInio (right)

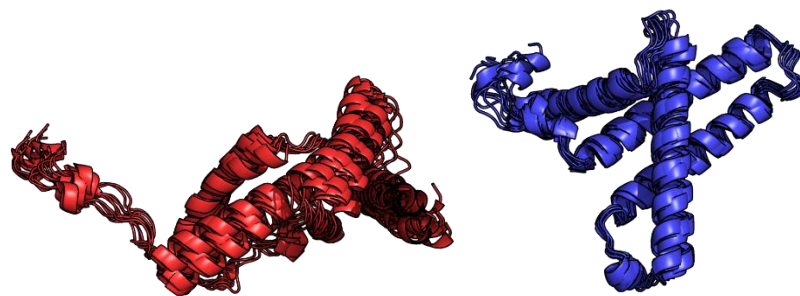


Figure 6.154: Paa Lowest RMSD Full Structures. 10 lowest RMSD structures from AbInitioVO (left) and AbInitio (right)

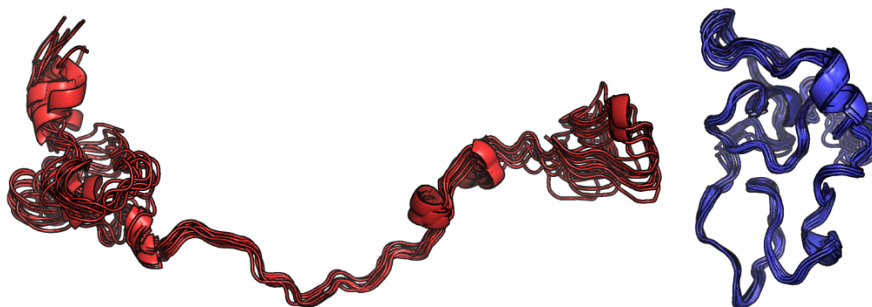


Figure 6.155: Sic1 Lowest RMSD Full Structures. 10 lowest RMSD structures from AbInitioVO (left) and AbInitio (right)

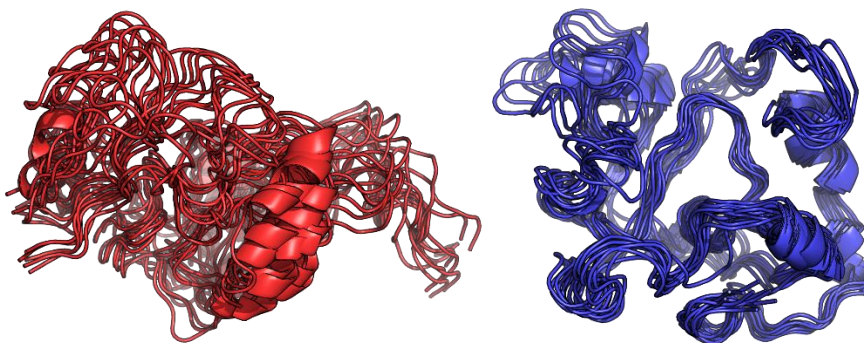


Figure 6.156: Tauk Lowest RMSD Full Structures. 10 lowest RMSD structures from AbInitioVO (left) and AbInitio (right)

**CHAPTER 7: HIGH-THROUGHPUT *IN SILICO* AND *IN VITRO* SCREENING
FOR COMPOUNDS THAT SELECTIVELY BIND TO ALPHA-SYNUCLEIN
FIBRILS.**

§ 7.1 Introduction

α -Synuclein (α S) is a 140 amino acid, intrinsically disordered protein which is abundantly expressed at the presynaptic termini of central nervous systems neurons.^{19, 263} When bound to membranes, α S takes on a partially helical structure and is involved in various physiological activities.^{23, 264} Although the function of α S is poorly understood, its localization, along with knockout and overexpression studies, suggests that alongside synapsin, VAMP2 and others, α S plays a significant role in maintaining synaptic vesicle reserve pools, neurotransmitter release and synapse function and plasticity.^{19, 27-28} Conversely, the potential pathological role of α S is well documented, where neuronal inclusions comprised principally of fibrillar α S termed Lewy bodies (LBs) and Lewy neurites (LNs) have long served as post-mortem hallmarks of Parkinson's Disease (PD).^{19, 22} Furthermore, similar aggregates can be observed in dementia with Lewy bodies (DLB), while a second form of aggregate has been identified in multiple system atrophy (MSA).³⁴ Compared to LBs and LNs which often appear in the substantia nigra, glia cell inclusions (GCIs) associated with MSA are found in oligodendrocytes of white matter tracts.³⁴ However, the inability to track fibril formation and localization in living patients has hindered the development of robust correlations between PD progression and fibril burden and prevents α S fibrils from serving as useful clinical markers.

To date, diagnosis of PD has primarily relied on the presentation of clinical symptoms, chiefly motor deficits. Although these symptoms are effective in tracking progression at later stages of disease, they are not observed until a

substantial degree of neuronal loss has occurred.³⁰ Moreover, these symptoms are not exclusive to PD, but are observed for other Parkinsonian syndromes.³³ Therefore, since differences in the presence and localization of fibrillar α S have already been established, methods for tracking deposits in patients could clarify diagnosis. Over the past decade, a breakthrough in the clinical evaluation of Alzheimer's disease (AD) was enabled in vivo imaging of amyloid β ($a\beta$) plaques with positron emission tomography (PET), a molecular imaging technique.²⁶⁵⁻²⁷¹ The development of $a\beta$ and tau specific PET probes has allowed researchers to determine that the formation of $a\beta$ aggregates precedes disease onset while tau-based neurofibrillary tangles (NFTs) occur later in disease progression.²⁷² These promising results from the imaging studies of AD patients have generated interest in the development of PET radiotracers to image α S and improve the diagnosis of PD.

Despite these advancements, the development of imaging probes with the requisite specificity for use in PD research and diagnosis has been challenging. In addition to LBs and LNs, PD patients also present neuronal aggregates comprised of $a\beta$ and tau.²⁷³ Although the Petersson and Mach laboratories as well as others have successfully identified compounds with moderate binding affinity for employment as PET imaging agents, no compound to date has displayed high enough affinity and specificity for use as a radioligand.²⁷⁴⁻²⁷⁵ Previously, following publication of the first solid-state NMR structure of α S fibrils by Rienstra and colleagues, we demonstrated that through the combined use of computational

docking, competition radioligand binding assays and photocrosslinking mass spectrometry we were able to posit binding sites for several previously developed compounds.^{47, 276}

Here, we explore the utility of another computational approach, exemplar-based *in silico* screening, in an effort to develop a molecule that potently and specifically binds to α S fibrils.²⁷⁷ Through the application of this method we have identified a molecular scaffold and confirmed through structure-activity relationship (SAR) studies that members of this isoxazole containing compound series have nanomolar affinity for α S fibrils. Moreover, several of these compounds have moderate specificity for α S over $\alpha\beta$ fibrils. Lastly, we demonstrate the potential of this molecule as a PET probe by imaging fibrillar α S deposits in mouse brain tissue using a radiolabeled analog of the identified molecule.

§ 7.2 Methods

In-Silico Exemplar-based Screening in Align-It.

The Rosetta Modeling Suite was used to map exemplars at each residue of the 2N0A PDB by running⁴⁷:

```
ROSETTA/make_exemplar.linuxgccrelease -database  
ROSETTA/database -in:file:s 2N0A.pdb -pocket_grid_size 12 -pocket_static_grid  
-pocket_filter_by_exemplar -pocket_surface_dist 1 -central_relax_pdb_num XX
```

Where residues for the previously identified sites 2, 3/13 and 9 corresponded to residues 156, 163 and 198.²⁷⁶ In order to allow Align-It to recognize the exemplar outputs from Rosetta, the hydrogen bond donating and

accepting parameters in Align-It were altered. Prior to compiling Align-It, the following lines were added at line 42 to the file hDonFuncCalc.cpp in the src directory:

```
if (a->GetAtomicNum() == 4)
{
    PharmacophorePoint p;

    p.func = HDON;

    p.point.x = a->x();
    p.point.y = a->y();
    p.point.z = a->z();

    p.hasNormal = false;

    p.alpha = funcSigma[HDON];

    pharmacophore->push_back(p);
}
```

Additionally, the following lines were added to line 42 to the file hAccFuncCalc.app in src directory:

```
if (atom->GetAtomicNum() == 10)
{
    if(_hAccCalcAccSurf(atom) < 0.02)
    {
        continue;
    }
}
```

```

    }
    PharmacophorePoint p;
    p.func = HACC;
    p.point.x = atom->x();
    p.point.y = atom->y();
    p.point.z = atom->z();
    p.hasNormal = false;
    p.alpha = funcSigma[HACC];
    pharmacophore->push_back(p);
}

```

The ZINC15, lead-like, commercially available compound database, consisting of ~ 10 million molecules was used for the initial screen against the three sites.²⁷⁸ Molecular alignments of target molecules to each selected exemplar were performed using Align-It which reduces each molecule/exemplar to a set of pharmacophores and reports a Tanimoto Coefficient for each alignment which captures both agreements in molecular features and their alignment in three-dimensional space.²⁷⁹ The top 30 compounds as quantified by the Tanimoto Coefficient from each search were retained and a subset of compounds from each search was selected by hand for experimentation. Since the compounds identified for site 3/13 were either too small or too similar to compounds previously explored,²⁷⁶ select compounds from the site 2 and site 9 screens were used in

subsequent experimental screens and the full set of compounds can be found in Figures 7.5 and 7.7.

Preparation of α S Monomer Fibrils.

Recombinant expression and purification of wild-type α S protein was performed as previously described.^{78, 276} Fibrils were prepared in a manner similar to previously described, where 100 μ M α S monomer was incubated in 20 mM tris 100 mM NaCl, pH 7.4 at 37 °C and shaken at 1300 rpm for 3 days.^{78, 276}

Preparation of $a\beta$ fibrils.

$a\beta$ fibrils were prepared as described previously.²⁸⁰ Briefly, monomer $a\beta$ (1 mg, Bachem) was dissolved in hexafluoroisopropanol (HFIP) at a concentration of 2 mg/mL and incubated for 1 hour at 37 °C until the peptide completely dissolved. Then HFIP was evaporated under air. The peptide powder was dissolved again in HFIP (2 mg/mL), aliquoted and left to dry overnight under vacuum. Aliquots were stored in a freezer at -20 °C.

To prepare fibrils, the HFIP-treated peptide aliquot was dissolved in 10 mM NaOH (10 μ L) solution, then the sample was diluted with 90 μ L 10 mM phosphate buffer, pH 7.4. Concentration of the peptide solution was confirmed by measuring the absorbance at 214 nm with NanoDrop 2000c spectrophotometer. Extinction coefficient (76848 $M^{-1}cm^{-1}$) was calculated using literature values.²⁸¹ Next, the solution was agitated by a continuous slow rotation at room temperature for 3 days and fibril formation was confirmed by TEM.

Aggregation and Disaggregation Fluorescence Polarization Assays.

Wild-type and fluorescently labeled α -synuclein was produced as previously described.¹¹⁹ Aggregations were performed at a total monomer concentration of 100 μ M with 1 % labeled α S in 20 mM tris 100 NaCl pH 7.4 in the presence of 10 and 100 μ M compound and with equivalent volume of DMSO as a control. At each timepoint an aliquot of the fibrilization reaction was removed and diluted 10-fold in buffer. Fluorescence polarization was measured in a half-area well plate at a total volume of 50 μ L on a Tecan F200 plate reader. Aggregations were performed in triplicate.

Disaggregation experiments were performed on labeled fibril samples prepared as described above in the absence of compound. A final concentration of 10 μ M fibrils was added to each well for each disaggregation experiment and each compound was added just prior to measurement to the final concentrations detailed in Figure 7.9 and total volumes of 50 μ L. Samples were shaken at 149 rpm at room temperature with polarization measurements taken every 150 second.

Photocrosslinking Assay.

Aggregations prepared as described above were dosed with equimolar photocrosslinking compound in DMSO and shaken overnight at 37 °C shaking at 250 rpm. A control was also prepared where the same volume of DMSO used to add compound was added and incubated as previously described. Following incubation samples were irradiated under a TLC lamp (365 nm) for 1 hour. Fibrillar samples were disaggregated by adding 20 mM SDS and boiled for 20 mins. SDS

and excess compound were removed through chloroform methanol precipitation.²⁸² Samples were digested with trypsin at 37 °C and the resultant peptides were analyzed by MALDI mass spectrometry to identify crosslinking site.

Screening compound library.

Compounds 1-17 were purchased from vendors that were listed on the ZINC15 compound library. Compounds were screened for α S binding and after identifying the lead compound, 6, the core structure of compound 6 and its pyrazole or oxadiazole derivatives were used for similarity search on the websites of Mcule, Inc., ChemDiv, Inc., MolPort, and Enamine, Ltd. The similarity threshold was set to higher than 0.8, then compound 18-56 were chosen and ordered from the four companies mentioned above.

In order to screen the purchased compounds for α S binding, 100 nM of each compounds were incubated for 1 hour at 37 °C with 100 nM ASyn fibrils and [³H]-Tg-190b (6 nM) or [³H]-BF2846 (3 nM) in 50 mM Tris-HCl, pH 7.4. Total binding was measured in the absence of competitor and nonspecific binding was determined in reactions containing cold Tg-190b (1 μ M) or BF2846 (0.5 μ M). After incubation, bound and free radioligand were separated by vacuum filtration through Whatman GF/C filters (Brandel) in a 24-sample harvester system (Brandel), followed by washing with buffer containing 10 mM Tris-HCl (pH 7.4) and 150 mM NaCl. Filters containing the bound ligand were mixed with 3 mL of scintillation cocktail (MicroScint-20, PerkinElmer Informatics, Inc.) and counted after 12 hours of incubation on a MicroBeta System (PerkinElmer Informatics, Inc.).

All data points were performed in triplicates. Percentage of bound radioligand relative to total binding was plotted and data was analyzed by One-Way ANOVA, comparing the mean of each data set to the mean of total binding.

Competition binding assay.

α S fibrils (100 nM for site 2 and 50 nM for site 9) or α β fibrils (100 nM) were mixed with site 2 ligand [3 H]-Tg-190b (6 nM) or site 9 ligand [3 H]-BF2846 (3 nM) and varying concentration of cold compounds. Compounds were diluted in 50 mM Tris-HCl (pH 7.4) and mixed with fibrils and radioligand in a total volume of 150 μ L. Total binding was measured in the absence of competitor and nonspecific binding was determined in reactions containing cold Tg-190b (1 μ M) or BF2846 (0.5 μ M). In a duplicate set of binding reaction, fibrils were replaced with equal volume of buffer to measure the amount of radioligand binding to the filter paper. Reactions were incubated at 37 $^{\circ}$ C for 1 hour. After incubation bound and free radioligand were separated by vacuum filtration through Whatman GF/C filters (Brandel) in a 24-sample harvester system (Brandel), followed by washing with buffer containing 10 mM Tris-HCl (pH 7.4) and 150 mM NaCl. Filters containing the bound ligand were mixed with 3 mL of scintillation cocktail (MicroScint-20, PerkinElmer Informatics, Inc.) and counted after 12 hours of incubation on a MicroBeta System (PerkinElmer Informatics, Inc.). All data points were performed in triplicates. K_i values were calculated by fitting the data to the equation below by nonlinear regression, using GraphPad Prism software:

$$\log EC_{50} = \log(10^{\log K_i} * (1 + [\text{radioligand}]/K_d)) \quad (\text{Eq. 7.1})$$

$$Y = \text{Bottom} + (\text{Top}-\text{Bottom}) / (1+ 10^{(X-\log EC_{50})}) \quad (\text{Eq. 7.2})$$

Where logEC50 is the log of the concentration of competitor that results in binding half-way between Bottom and Top; logKi is the log of the molar equilibrium dissociation constant of unlabeled ligand; [radioligand] is the concentration of hot ligand in nM; Kd is the equilibrium dissociation constant of the hot ligand in nM; Top and Bottom are plateaus in the units of Y axis.

Saturation binding assay.

α S (50 nM) or $\alpha\beta$ (100 nM) fibrils were incubated for 1 hour at 37 °C with increasing concentrations of [¹²⁵I]**61** in 50 mM Tris-HCl, pH 7.4, in a total volume of 150 μ L. Nonspecific binding was determined in a duplicate set of binding reactions containing 2 μ M cold **52**. To measure the amount of radioligand binding to the filter paper, fibrils were replaced with equal volume of buffer in a duplicate set of binding reaction. After incubation, bound and free radioligand were separated by vacuum filtration through Whatman GF/C filters (Brandel) in a 24-sample harvester system (Brandel), followed by washing with buffer containing 10 mM Tris-HCl (pH 7.4) and 150 mM NaCl. Filters containing the bound ligand were counted immediately on 2470 WIZARD Automatic Gamma Counter (Perkin Elmer). All data points were performed in triplicate. The equilibrium dissociation constant (K_d) and the maximal number of binding sites (B_{max}) were determined by fitting the data to the equation $Y = B_{\text{max}} * X / (K_d + X)$, using GraphPad Prism software.

Animals.

A53T (B6C3-Tg(Prnp-SNCA*A53T)83Vle/J) and B6C3F1/J mice were obtained from The Jackson Laboratory. All animal studies were performed under protocols approved by the University of Pennsylvania Institutional Animal Care and Use Committee. Animals were euthanized by cervical dislocation under isoflurane anesthesia at 17 months of age and the brain was extracted for autoradiography and microscopy.

In vitro autoradiography.

Blocks of mouse brain tissue were frozen in optimal cutting temperature compound (OCT, Tissue-Tek, Sakura Finetek, USA). The frozen tissue was sliced into 10 μ M thick sections in a Leica CM1950 cryostat and mounted onto Apex Superior Adhesive slides (Leica). Frozen sections of both A53T and B6C3F1/J mouse brain tissue were thawed at RT for 20 min, then washed with 40% ethanol in PBS for 5 min. Next, sections were incubated (1 hour at RT) with 40% ethanol in PBS containing either [¹²⁵I]61 (6 nM) alone or [¹²⁵I]61 (6 nM) with 20 μ M cold Tg-190b. After incubation, sections were washed in ice-cold 40% ethanol in PBS (2x30 sec), followed by a wash in ice-cold DI water (1 min). Sections were dried in a stream of air and then exposed to a phosphor screen (Fujifilm) and the screen was imaged on a Typhoon FLA 7000 phosphor imager (GE Healthcare Europe). Raw autoradiography images were imported to MATLAB R2017b (MathWorks Inc., Natick, MA) to extract individual image. Each autoradiography image was

manually registered to corresponding staining image by using PMOD image analysis software (version 3.7; PMOD Technologies Ltd, Zurich, Switzerland).

Immunofluorescence.

Frozen brain tissue sections, adjacent to the ones used for autoradiography, were thawed at room temperature for 20 min. Sections were fixed with 4% paraformaldehyde in PBS, washed with PBS three-times, then permeabilized with 0.1% Triton X-100 in PBS. Sections were blocked with 10% normal goat serum (Fisher Scientific) at room temperature for 1 hour, then with goat F(ab) anti-mouse IgG H&L (1:1000 in 1% normal goat serum in PBS with 0.2% Tween-20; ab6668) for 1 hour at RT. After blocking, sections were incubated with primary Anti- α S (phosphor S129) antibody (81A; 1:1000 in 1% normal goat serum in PBST) overnight at 4 °C. After three washes with PBST, the tissue was incubated with a secondary antibody labeled with Alexa Fluor 488 (1:500 in 1% normal goat serum in PBST) for 1 hour at room temperature. Tissue was washed with PBST twice, then with PBS once and coverslipped. The fluorescent images were acquired by a Zeiss Axio Imager M2 microscope.

§ 7.3 Results

Exemplar-Based In-Silico Screen

Through methods developed by Karanicolas and colleagues, using the Rosetta Modeling Suite, exemplars can be facilely generated for any protein of interest given an input structure.²⁸³ An exemplar is a pseudoligand designed to be

a perfect molecular complement to a surface exposed pocket on a protein of interest (Fig 7.1b).^{277, 283} Following selection of an anchor residue, the protein is cast onto a three-dimensional grid, and grid points that correspond to the protein pocket are “chemotyped” by the adjacent functional features on the protein. Based on whether the protein surface presents hydrogen bond donating or accepting moieties or a hydrophobic patch, grid points are assigned characteristics that are complementary to the protein surface. Subsequently, molecular alignment of compounds from a database allows for rapid *in silico* screening to identify molecular architectures that satisfy the chemical features (i.e. hydrogen bonding/accepting) captured by the exemplar (Fig 7.1b).²⁷⁷

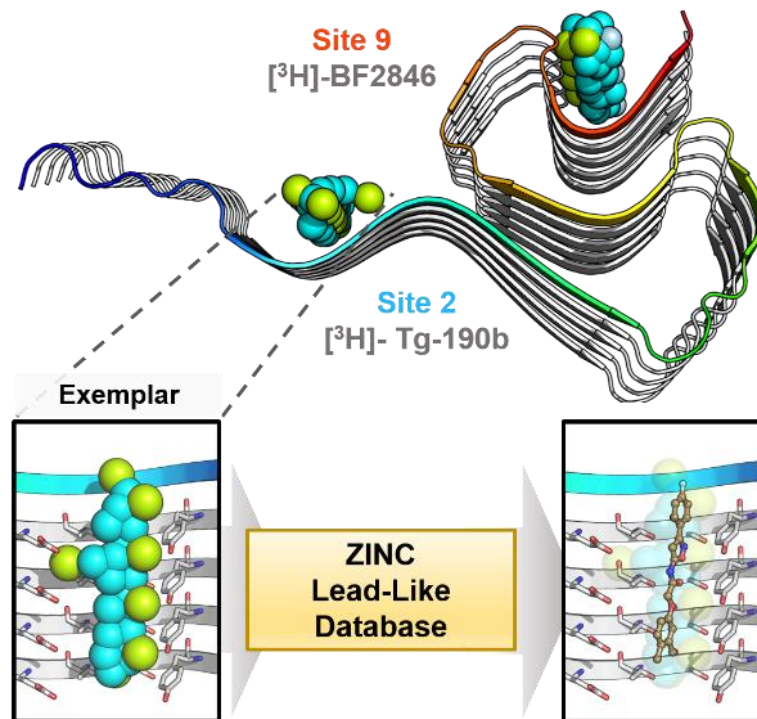


Figure 7.1: Exemplar-based in-silico screening. (a) PDB 2N0A with exemplars of previously described sites 2 (Y39-S42-T44) and 9 (G86-F94-K96) shown as spheres representing hydrophobic (cyan), hydrogen bond donating (yellow) and accepting (light blue) pharmacophores. Compounds used in site-specific competition binding experiments are listed with for each site in gray. (b) Workflow for identifying small molecules.²⁷⁶ A zoom in on the site 2 exemplar is shown on the left with compound **6** docked in the conformation identified from the ZINC database by Align-It.²⁷⁸⁻²⁷⁹

In order to assess the potential efficacy of using an exemplar-based approach, we targeted sites that had been identified through our prior efforts as the binding sites for a myriad of previously described α S fibril radioligands (Fig 7.1a).²⁷⁶ Our previous work highlights two compounds, [3 H]Tg-190b and [3 H]BF2846 (Fig 7.2), which selectively bind to site 2 and 9 respectively and can be utilized for *in vitro* competition binding assays to screen compounds identified *in silico*.

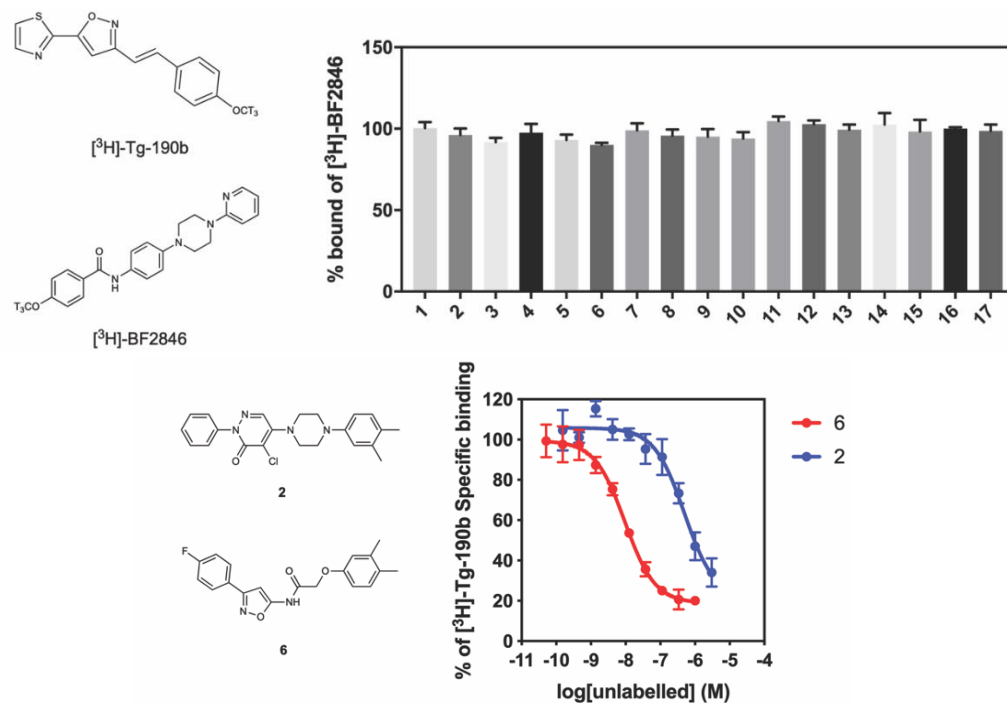


Figure 7.2: Radioligand Structure and Initial Binding Assay. (Top Left) Molecular structure of radioligands used in the screening and competition binding assays. (Top Left) Radioligand displacement assay testing binding affinity of each compounds against a site 9 radioligand, $[^3\text{H}]\text{BF2846}$. (Bottom Left) Molecular structures of the two top hits (2 and 6) from the initial screen. (Bottom Right) Competition binding curves in αS fibrils (100 nM) for compounds 2 and 6. αS fibrils were incubated with $[^3\text{H}]\text{Tg-190b}$ (6 nM) and increasing concentrations of competitors (2 and 6). Compounds 2 and 6 have K_i values of 85.6 nM and 1.18 nM, respectively. Data points represent mean \pm s.d. (n=3).

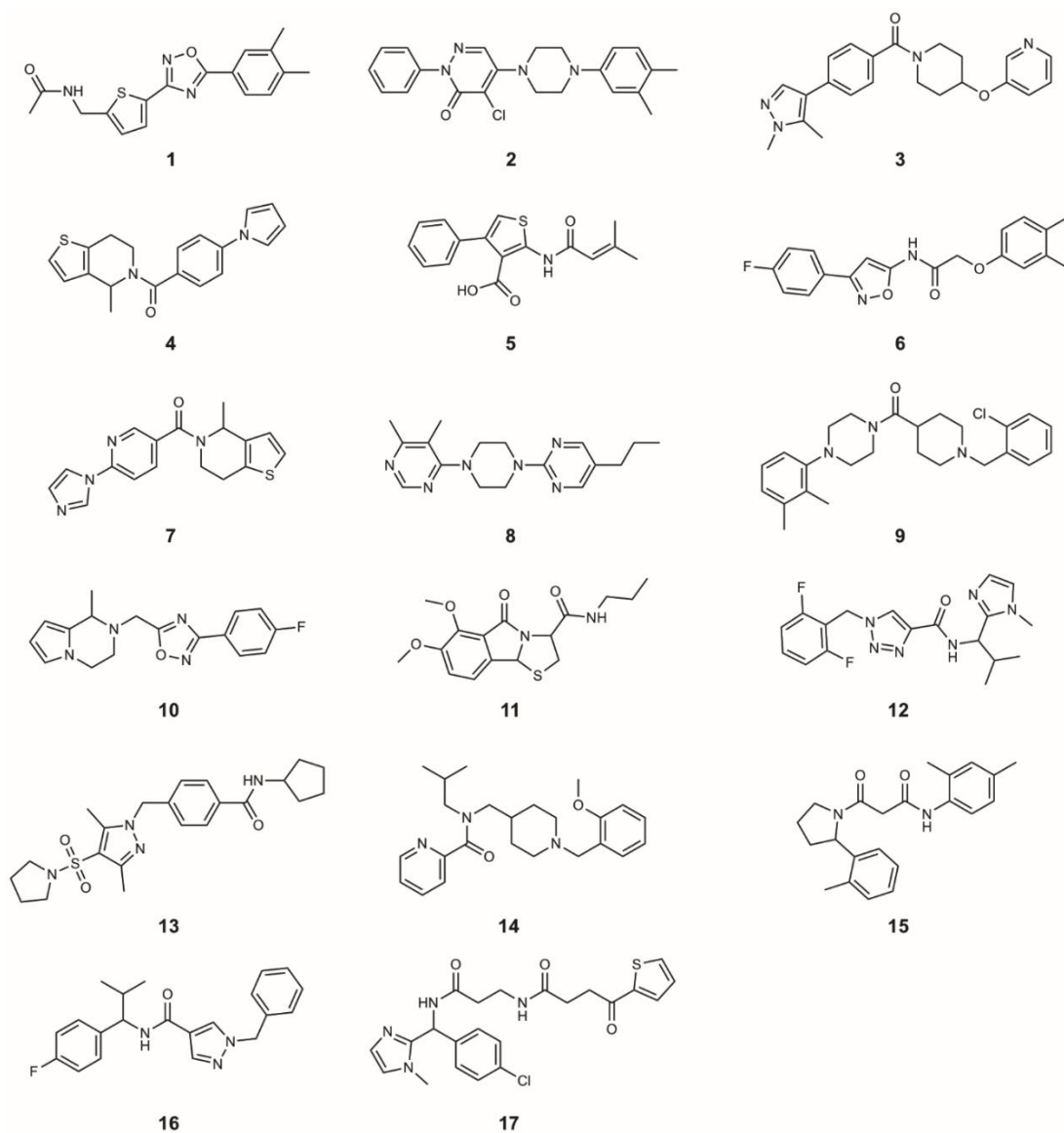


Figure 7.3. Molecular structures of 17 compounds from the Exemplar screening. Compound 6 showed high affinity for Asyn fibrils *in vitro* and its core structure was used for similarity search.

Therefore, we selected these two sites as our initial targets. Using the 2N0A PDB structure deposited by Rienstra and colleagues, pocket templates were generated using residues 44 and 86 of the central strand of the fibril as anchor

points for sites 2 and 9 respectively. The exemplar for each site was screened using Align-It against ~ 10 million commercially available, lead-like molecules from the ZINC15 database. The top 50 molecules which displayed the best overlap with each exemplar pseudoligand were retained. From the top 50 molecules identified in the initial screen at sites 2 and 9 (Figs 7.4-5 and 7.6-7), 17 compounds were purchased and were employed in high throughput screening to determine their relative affinities for α S fibrils. Of the set of molecules selected, 2 molecules were able to displace the site 2 specific radioligand [3 H]Tg-190b (Fig 7.2a). Competitive binding experiments revealed that compounds **2** and **6** displayed inhibitor constants (K_i) of 85.6 nM and 1.18 nM, respectively, against a site 2 specific radioligand (Fig 7.2b).²⁸⁴

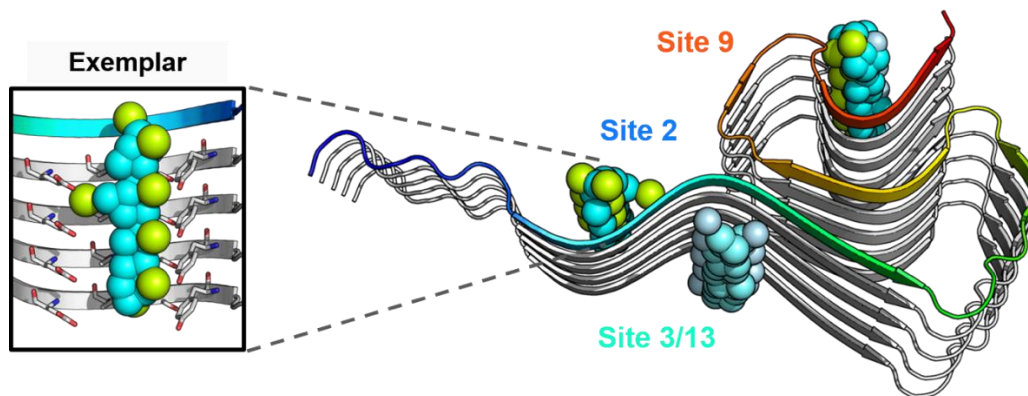


Figure 7.4: Site 2 Exemplar from PDB 2N0A fibril.

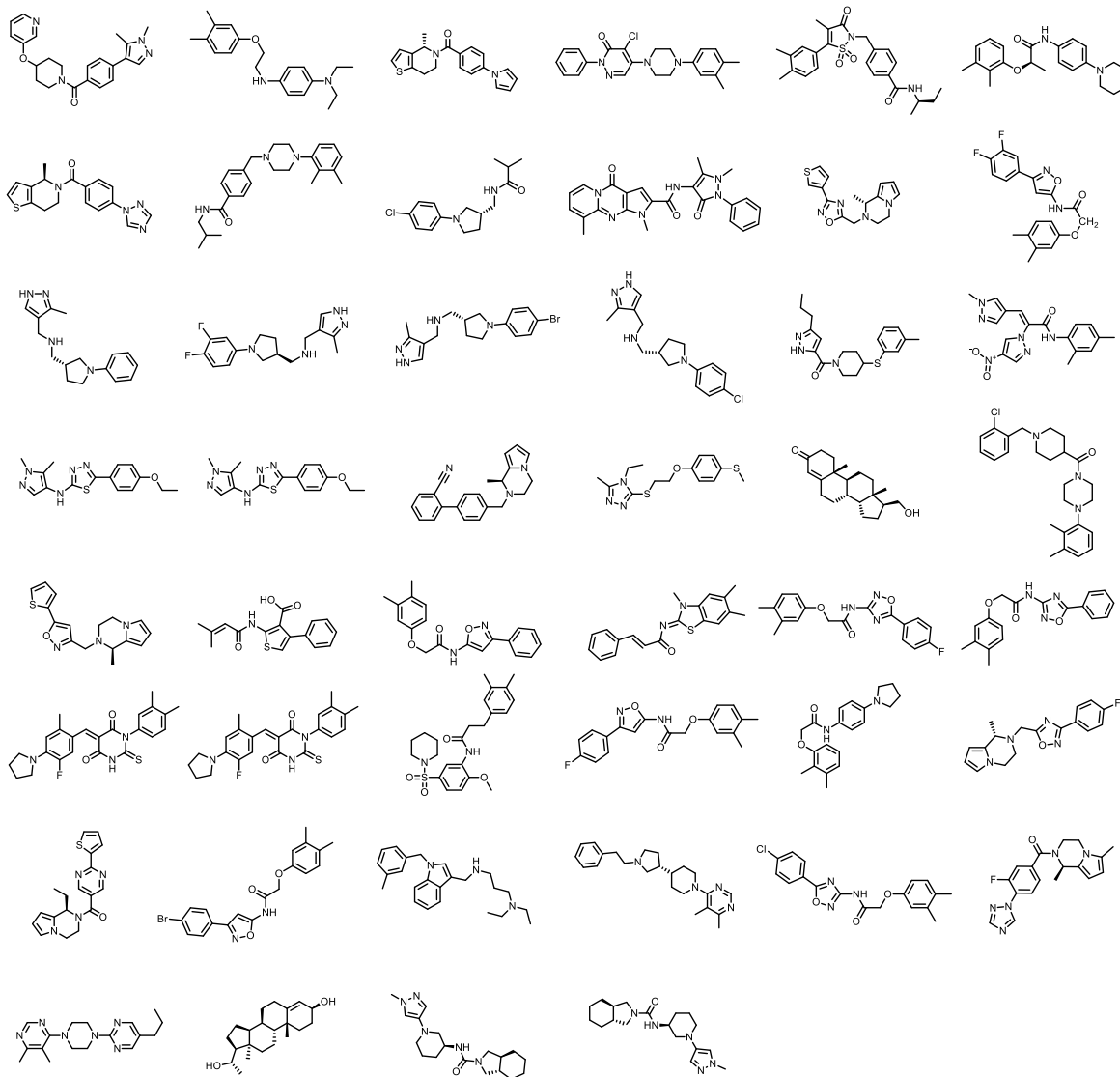


Figure 7.5: Top Hits from ZINC15 database for Site 2 Exemplar.

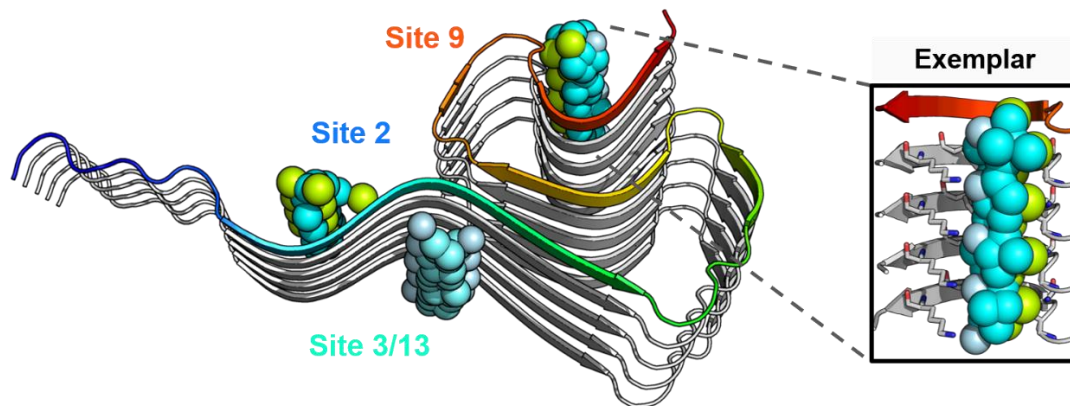


Figure 7.6: Site 9 Exemplar from PDB 2N0A fibril.

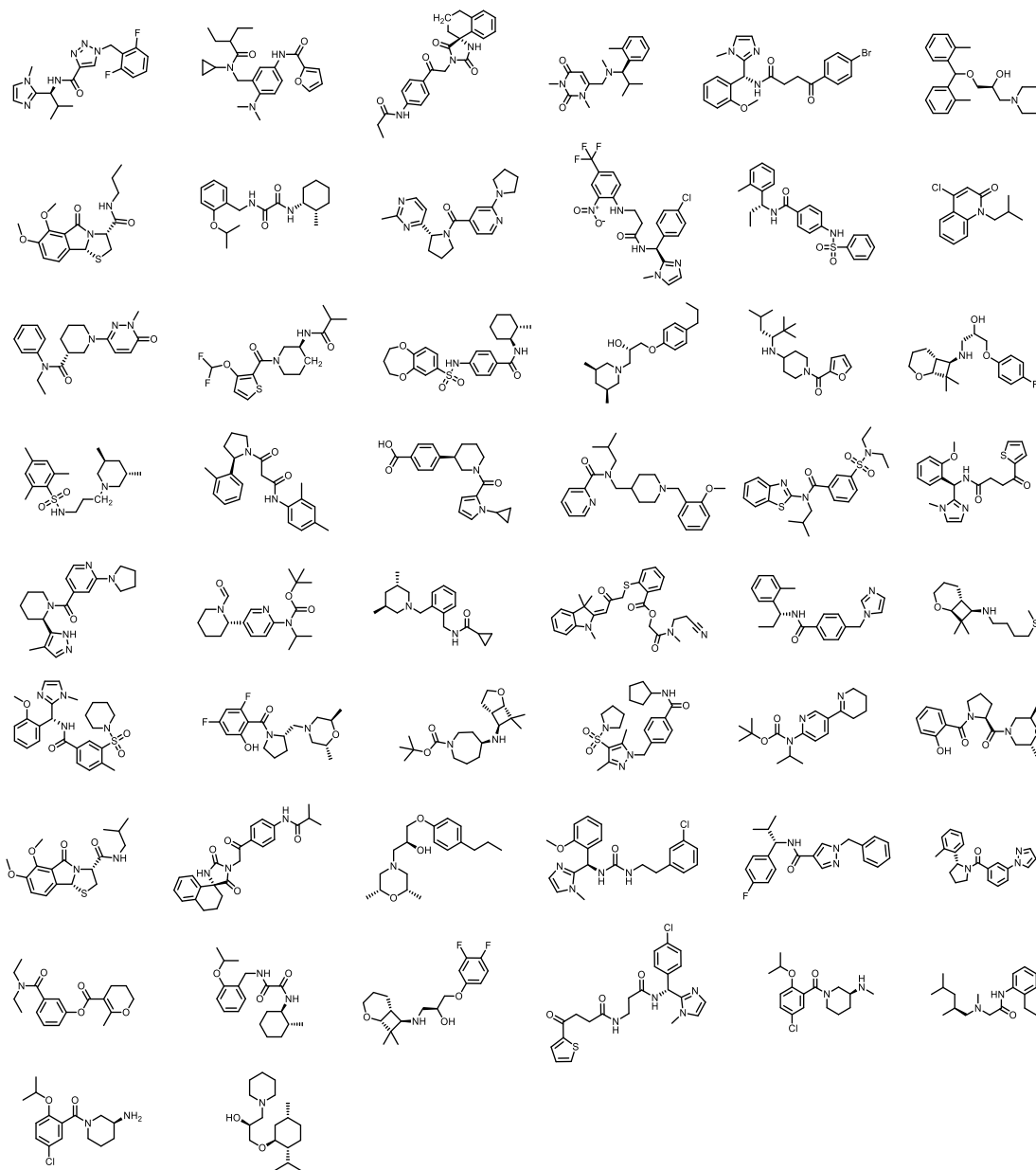


Figure 7.7: Top Hits from ZINC15 database for Site 9 Exemplar.

Lead Compound Binds to Site 2 and Does Not Affect α S Aggregation

Following the identification of two potential lead compounds, we decided to further investigate the more potent binder to confirm that the compound 1) binds

to the target site on α S fibrils and that 2) the compound is non-perturbing and does not affect the aggregation of α S. We posit that a suitable imaging compound should not perturb the aggregation process, nor should it remodel or disaggregate fibrils. Therefore, a derivative of compound **6** was synthesized by Marshall Lougee with a photocrosslinking group to confirm the binding location (Fig. 7.8). Following incubation of compound 62 with α S fibrils and irradiation with 365 nm light, the sample was analyzed by full-protein and trypsin digest MALDI mass spectrometry (Fig. 7.8). Analysis of the resultant data revealed a clear single mass shift, suggesting that most of the compound binds at a single site. Furthermore, analysis of the digested sample confirmed that crosslinking could be observed at the target site, site 2, initially used in the *in silico* screen (Fig. 7.8).

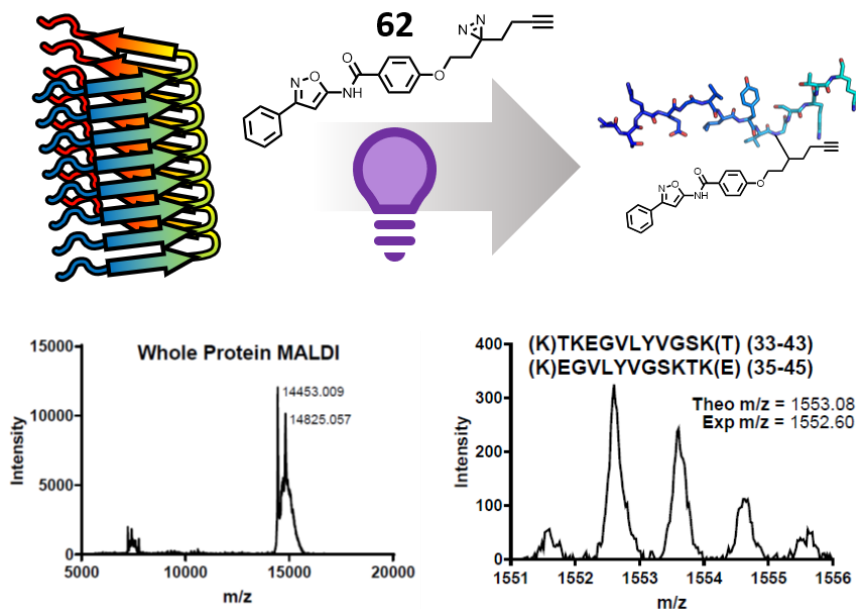


Figure 7.8: Photocrosslinking of radioligand analog to α S fibrils. Top: Experimental scheme illustrating that following incubation of 62, a photocrosslinkable analog of 6 synthesized by Marshall Lougee, with α S fibrils that irradiation with UV light and fibril digest results in identification of the compound binding site. Bottom: MALDI mass spectrometry data before (left) and after (right) digestion of α S with trypsin following photocrosslinking with a BJ-1-094 analog.

To confirm that the probe does not affect the aggregation state or propensity of α S, we generated fluorescently labeled α S for use in fluorescence polarization (FP) assays performed by Marshall Lougee. By attaching fluorescein-maleimide to a Y136C mutant α S construct FP can be used to monitor fibril aggregation and stability as well as the compound's effect on these processes, as previously described.^{78, 119} The compound did not show any significant effect on the rate of aggregation or on the stability of α S fibrils (Fig 7.9).

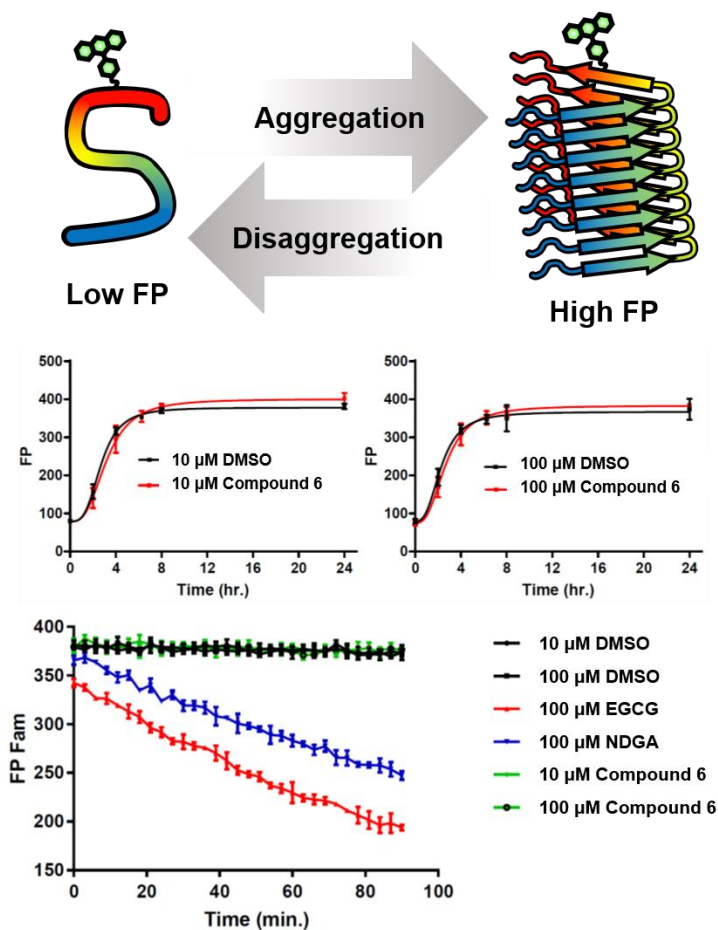


Figure 7.9: Impact of identified molecule on α S fibril aggregation and stability. Top: Experimental scheme illustrating that incorporation of fluorescently labeled monomer into fibrils results in a significant increase in fluorescent polarization allowing aggregation to disaggregation to be tracked. Middle: Aggregation of α S in the presence of 1% labeled monomer along with 10 and 100 μ M **6** compound. Bottom: Assessment of potential remodeling of 1% labeled α S fibrils by **6** compared to EGCG and NDGA as a positive control and DMSO as a negative control.

SAR Screen to Improve Affinity/Specificity

In order to improve the affinity and selectivity of compound **6**, Zsofia Lengyel-Zhand conducted an SAR screen by culling structurally similar compounds from the Mcule library via similarity search.²⁸⁵ In total 39 molecules, summarized in Figure 7.10 and Figure 7.11, assayed the impact of various

substituents on the two terminal aryl rings as well as the importance of the central isoxazole by replacement with a pyrazole or oxadiazole. High throughput screening of the selected compounds revealed that the majority of them effectively displace the site 2 radioligand (Fig. 7.10b), however only 3 of the compounds showed binding to site 9 (Fig. 7.10c). One of these compounds features the replacement of the central isoxazole with a pyrazole (compound **55**) and interestingly was the only non-isoxazole compound to show any affinity for either site of α S fibrils. Moreover, all binding competent compounds with a bromine substitution at the *para* position of the leftmost ring (as drawn in Fig 3a) display a high affinity for both site 2 and site 9 (compounds **28** and **31**). Furthermore, we find that ortho substitutions on the rightmost ring (compounds **18**, **22**, **27**, **30**, **38**, **47**, **49** and **54**) hinder binding while meta substitutions (compounds **20**, **28**, **40**, **41** and **52**) are tolerated. That same rightmost aryl system is also largely intolerant to the addition of halogens (compounds **23**, **25**, **32**, **35**, **37**, **44**, **50**) but accommodates more electron donating groups such as methyl (compounds **28**, **36** and **40**) and methoxy (**19**, **21**, **24**, **31**, **34** and **39**) groups. Lastly, we find that branching points such as the insertion of a methyl group at the methyl ether adjacent to the central amide eliminates binding (compound **23** and **35**).

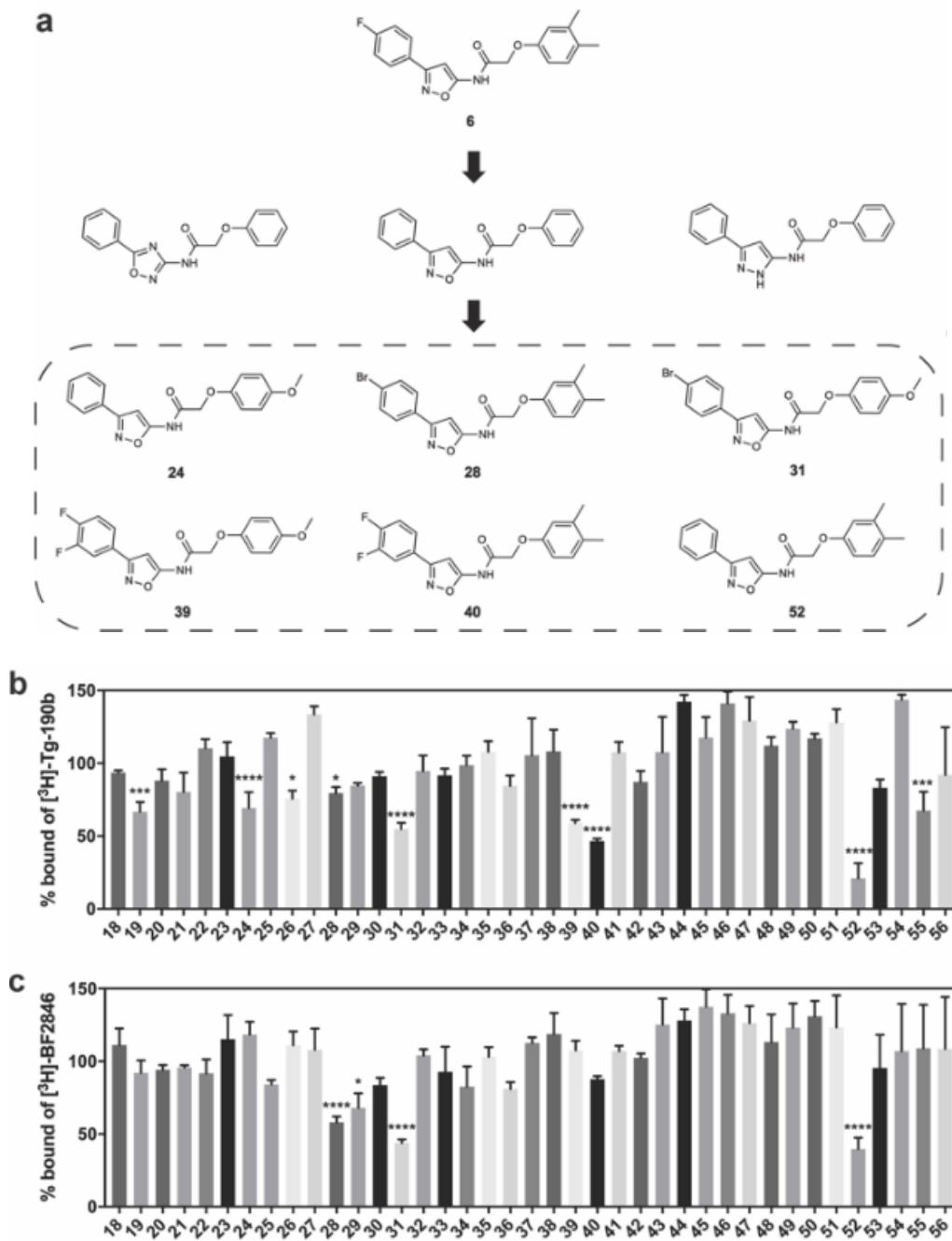


Figure 7.10: SAR screen based on exemplar compound (6). (a) Example set of compounds identified through similarity search. The core structure of **6** and its pyrazole and oxadiazole derivatives were used for similarity search, with similarity threshold set to 0.8 or higher. (b,c) Radioligand displacement assay testing binding affinity of each compounds against a site 2 (b) or site 9 (c) radioligand. Data points represent mean \pm s.d. (n=3).

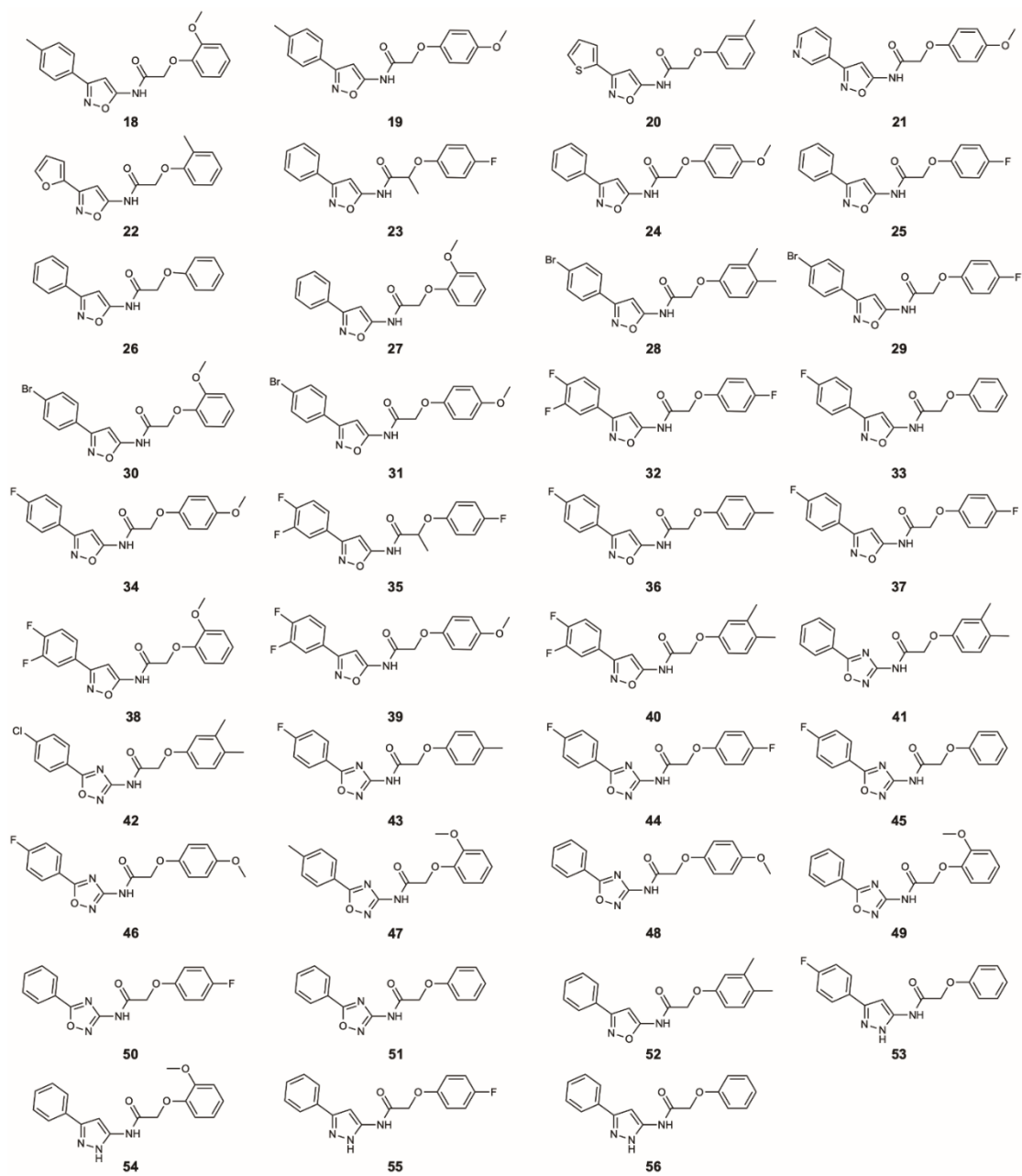


Figure 7.11. Molecular structure of the 39 compounds from the similarity search.

Ligand	α S Fibrils K_i (nM)		$a\beta$ Fibrils
	[3 H]Tg-190b	[3 H]BF2846	[3 H]BF2846
6	1.18 (0.90-1.56)	32.7 (20.7-51.5)	7.59 (3.14-18.7)
24	47.9 (22.3-101)	139 (87.1-268)	352 (206-609)
28	1.97 (1.51-2.58)	8.57 (3.77-19.5)	4.13 (2.10-8.54)
31	34.4 (13.6-84.6)	7.53 (5.21-10.9)	16.8 (11.0-25.5)
39	2.83 (1.55-5.14)	>1000	19.3 (7.07-53.9)
40	5.63 (2.05-15.5)	>1000	15.7 (4.97-48.3)
52	8.39 (4.94-14.2)	>1000	11.2 (3.69-34.2)

Table 7.1: Comparison of K_i values in α S and $a\beta$ (using $a\beta$ 42) fibrils. Values were determined by competition binding assay with [3 H]Tg-190b or [3 H]BF2846. 95% confidence intervals for K_i values are shown in parentheses (n=3).

High throughput screening of the SAR library was analyzed with One-Way ANOVA and compounds showing the highest affinity for α S fibrils ($P < 0.0001$) were selected for in-depth characterization. We measured their affinity for both site 2 and site 9 in α S and also in $a\beta$ fibrils (Table 7.1). Site 9 radioligand [3 H]BF2846, has similar affinity for $a\beta$ fibrils and α S fibrils, allowing it to also be employed in $a\beta$ fibril competition assay. Gratifyingly, we observed that the most potent α S fibril binders all displayed a preference for α S over $a\beta$ fibrils with selectivity ranging from 2- to 7-fold.

Synthesis and characterization of radioligand

Despite elucidating several important features of the base-scaffold in the SAR study, the initial compound (**6**) identified in the in-silico screen remained the most potent binder, thus we set out to synthesize and further characterize its iodinated derivative, compound **61**, as well as its radiolabeled isotopolog [¹²⁵I]**61**. Synthesis of the [¹²⁵I]**61** compound was performed by Bienneke Janssen as detailed in Figure 7.12.

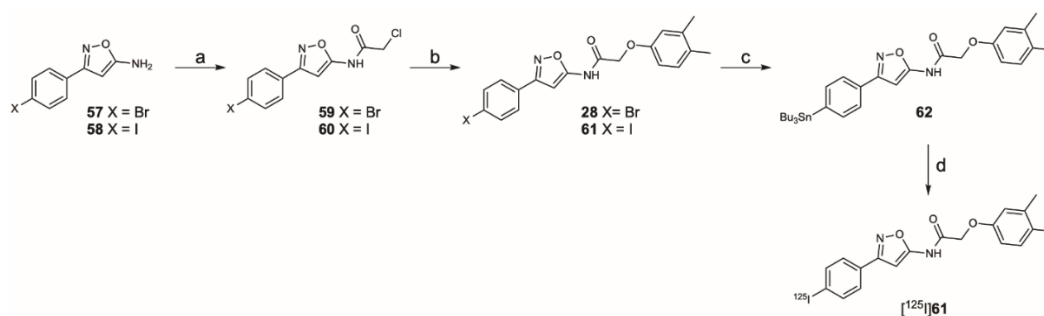


Figure 7.12: Synthesis route towards **61** and [¹²⁵I]**61**. Reagents and conditions: (a) 2-chloroacetyl chloride, TEA, CH₂Cl₂, 0 °C – rt, 20 h (**59** 67%; **60** 65%); (b) 3,4-dimethylphenol, Cs₂CO₃, MeCN, 60 °C, 14-20 h, (**28** 36%; **61** 42%); (c) (SnBu₃)₂, Pd(PPh₃)₄, toluene, 110 °C, 3 h (55%); (d) [¹²⁵I]NaI, H₂O₂, AcOH, MeOH, 60 min, 57% radiochemical yield, radiochemical purity >99%, molar activity of 81 GBq·μmol⁻¹.

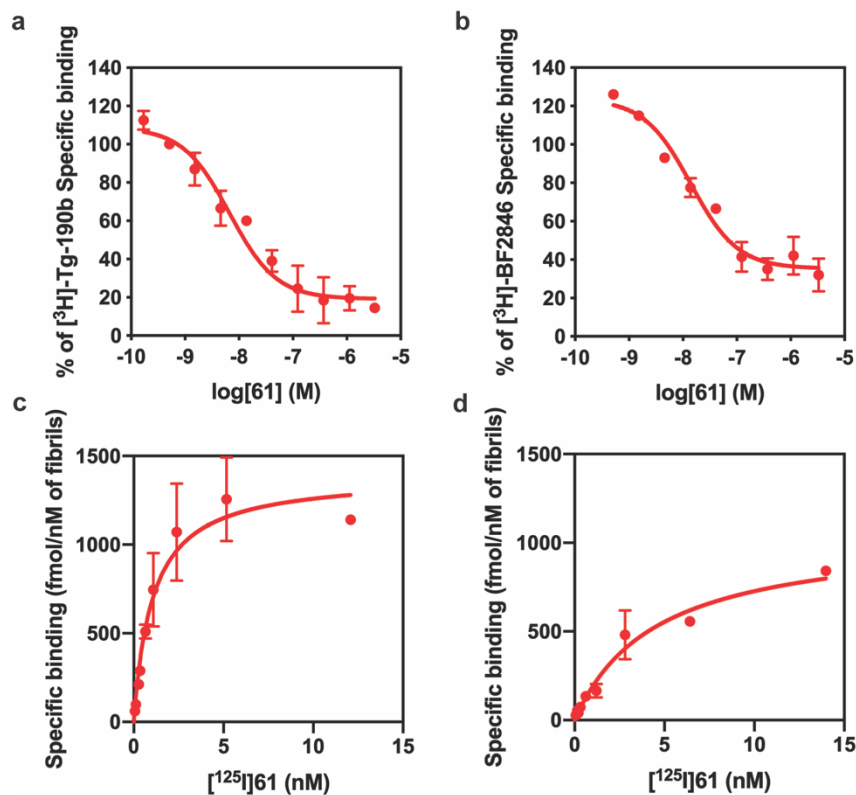


Figure 7.13: Characterization of the lead compound. (a,b) Competition binding curves in α S fibrils with [3 H]Tg-190b (a) and in α β 42 fibrils with [3 H]BF2846 (b). Compound **61** has a K_i of 1.84 nM and 4.66 nM in α S and α β 42 fibrils, respectively. (c,d) Saturation binding curves for [125 I]**61**. K_d and B_{max} values were obtained for α S fibrils (c, $K_d = 1.06$ nM and $B_{max} = 1392$ fmol/nM of fibrils) and α β 42 fibrils (d, $K_d = 4.56$ nM and $B_{max} = 1058$ fmol/nM of fibrils). Data points represent mean \pm s.d. (n=3).

The potential efficacy of this radioligand as an imaging probe was tested by Zsofia Lengyel-Zhand by *in vitro* autoradiography studies. Images were obtained by incubating sagittal brain sections from 17-month old PD mouse model (A53T) and C6C3F1/J control with [125 I]**61** and subsequent exposure to storage phosphor screens. The autoradiograms are shown in Fig 7.14b. and clearly demonstrate that there is increased signal in A53T mouse brain compared with the control across the entire brain section. Adjacent brain sections were also stained with PS129 anti-

α S antibody (Fig 7.14) and colocalization between the autoradiogram and fluorescence images was observed in the medulla, pons and midbrain regions containing the substantia nigra.

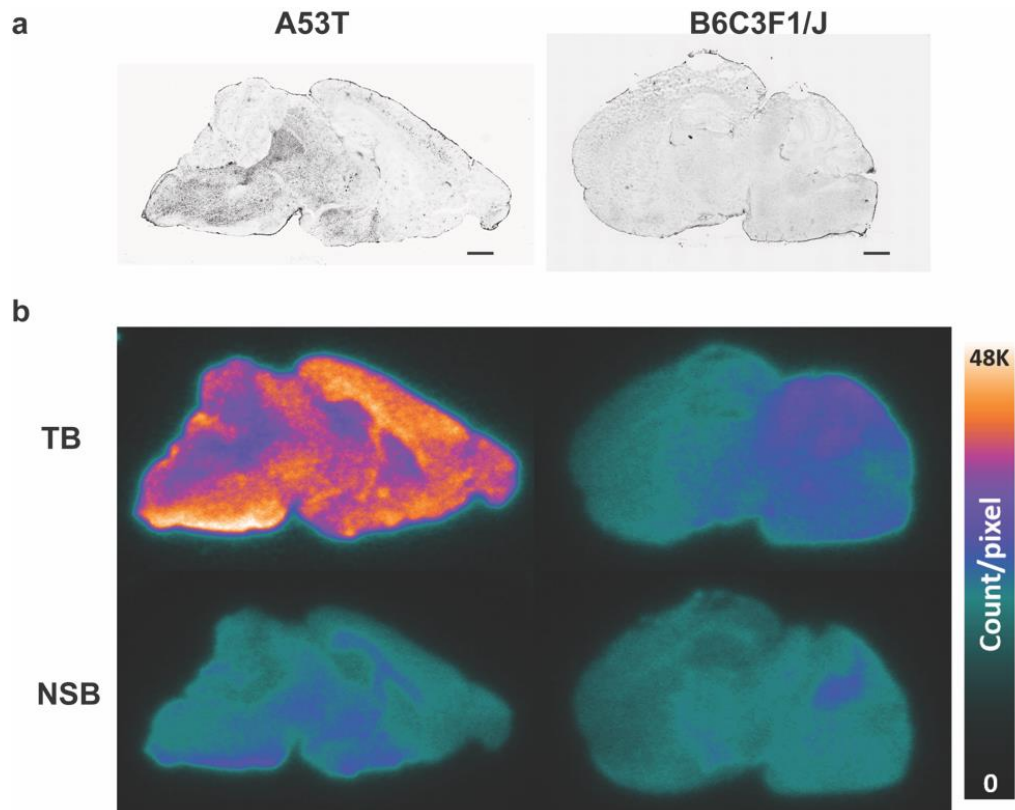


Figure 7.14: In vitro autoradiography on mouse brain tissue sections to assess [125 I]61 binding. (a) Immunofluorescence staining of A53T and normal (B6C3F1/J) mouse brain sections with PS129 anti- α S antibody (scale bar: 1000 μ m). (b) Autoradiograms showing the binding of [125 I]61 in A53T and normal mouse brain sections. Upper two sections are total binding (TB) and lower two sections are non-specific binding (NSB), defined using non-labeled Tg-190b (20 μ M).

§ 7.4 Discussion

The development of a PET tracer to image Parkinson's disease is a high priority in the field of radiopharmaceutical research, since a radioligand that binds to α S could greatly improve the clinical diagnosis of PD. Despite the progress in recent years, a radiotracer for imaging α S aggregates with high affinity and selectivity has yet to be developed. In this work, we utilized a combination of exemplar-based in silico screening and radioligand binding studies to identify several compounds that bind to α S fibrils with nanomolar affinity and moderate selectivity over $a\beta$. Our approach is unique and represents the first use of this method in radioligand development; we envision further exploring this method towards development of a PET tracer targeting α S.

Although we were able to successfully identify a radioligand to image α S in vitro, there are several potential avenues for improving the compound identification process by modifying the library composition. The current library contained single rotamers from the ZINC15 commercially available, lead-like library.²⁷⁸ After identifying similar compounds through the SAR screen which afforded a reduced base scaffold, and our ultimate radioligand, we looked to see if this scaffold was contained in our searched database and why this compound series may have been missed in the initial screen. We verified that this reduced scaffold was indeed present in the original library and that the single rotamer of this series of compounds contained in the database did not align favorably with the input exemplar. Therefore, we envision that the incorporation of more rotamers alone would likely improve our hit rate and may intrinsically lead to new scaffolds.

CHAPTER 8: FUTURE DIRECTIONS AND OUTLOOK

The efforts described in this thesis work have afforded novel computational and experimental methods for studying disordered and aggregation prone proteins. Early chapters highlight the developments of minimally perturbing fluorescence probes, in the form of unnatural amino acids and small scaffold protein modifications and showcase their employment in protein-based FRET experiments. Furthermore, this work describes an initial exploration of methods for predicting which sites are optimal for incorporation of a given unnatural amino acid probe, laying the groundwork for a predictive algorithm. The latter chapters focus on using FRET-based experimental methods, alongside experimentally restrained computational modeling, to gain a better understanding of the disordered ensemble of α -synuclein. This allowed for the development of new PyRosetta-based algorithms which, in tandem, allow for the accurate prediction of ordered and disordered proteins from sequence. Lastly, the work herein culminates in the identification of a small molecule probe, using exemplar modeling, which can selectively bind to α -synuclein fibrils and can be leveraged as a positron emission tomography imaging probe. Overall, this work has demonstrated the usefulness of the approaches and tools developed herein, providing new platforms for further experimentation on α -synuclein and other disordered proteins.

§ 8.1 Identification of Novel Compounds with Improved Therapeutic Potential via FRET-based Screening

This work has demonstrated that FRET-based fluorescence assays, alongside experimentally restrained molecular modeling, were able to capture

differences in the structural ensembles of α S in the presence and absence of 2 M TMAO. Although this model system may not feature a high degree of physiological relevance, it provides an important benchmark for the conditions under which ensemble FRET can be used. Indeed, this method has already been employed to study the effects of small molecules such as epigallocatechin gallate (EGCG), nordihydroguaiaretic acid (NDGA) and NDGA derivatives on the structure of α S in a plate-reader based assay.²⁸⁶ Overall, these efforts demonstrate that high-throughput ensemble FRET-based screens may be useful in identifying new therapeutic candidates for Parkinson's Disease.

Since fibrils have long served as the pathological hallmark of PD several groups have attempted to identify compounds that disaggregate α S fibrils.^{19, 287} Fibrils are notoriously stable, and incomplete dissociating of fibrillar architectures generally leads to shedding of highly-mobile oligomers, which are still highly cytotoxic.²⁸⁸ Moreover, the identification of poly-ubiquitinylation and other degradation markers on fibrillar α S suggests that endogenous cell machinery may not be sufficient to clear these bulky, insoluble species.²⁸⁹⁻²⁹¹ Therefore, in contrast to clearing fibrils, which appears to be quite difficult, small molecules which hinder fibrils' ability to recruit monomer may be a more efficacious route to the development of a novel therapeutic.^{286-287, 292-293} This approach could be leveraged from the fibril end, where a small molecule which targets fibrils prevents fibril ends from templating naïve monomer, or from the monomer end, where a small molecule stabilizes the soluble form of α S or lowers the available pools by pushing monomer towards degradation.^{287, 292-293} The former approach appears more

reasonable in light of the myriad of small molecules that have been identified that target fibrillar architectures and is discussed later in this chapter.²⁷⁶ However, the latter approaches present a clear barrier in identifying small molecules that specifically target the disordered conformation of α S, as described earlier.

The approach of targeting a protein's disordered state, though still in its infancy, is not novel to this thesis.²⁹⁴⁻²⁹⁷ Indeed, others have successfully identified molecules which are capable of binding to disordered proteins using NMR-based experimental and MD-based computational approaches.^{14, 297-298} Furthermore, the work of Kelley and coworkers with transthyretin supports the notion that a protein's soluble form might be stabilized relative to fibril formation.²⁹⁹ Therefore, the next major application of the developed ensemble-based FRET approach is to perform high-throughput screens with the goal of identifying molecules that bind to, and impact the structural ensemble, of α S. Recently, the National Cancer Institute at the National Institute of Health has begun to provide natural products libraries to labs free of charge.³⁰⁰ These libraries contain 150,000 pre-fractionated marine, microbial and plant extracts and are provided in 384-well plates. Therefore, future work will be focused on developing the ensemble FRET-based approach, described herein, into a robust high-throughput strategy aimed at identifying compounds that modulates the conformational ensemble of α S.

§ 8.2 Continued Exemplar-based Screening

In addition to these clear experimental next steps, further computational efforts towards the identification of PD probes are underway. The final chapter of

this thesis work outlined a general workflow for identifying novel molecular scaffolds that are capable of selectively binding to α S fibrils. Although the compound identified was sufficient for vetting the validity of this approach, the selectivity and solubility of the molecule could stand to improve. Therefore, a second *in silico* search is currently being prepared, making use of the improvements noted in that chapter, which will hopefully lead to an improved hit. Additionally, although the recently identified compound boasts improved specificity for α S fibrils over other protein fibrils, there is also an ongoing effort to further enhance compounds selectivity. To solve this problem, exemplars will once again be employed, and a library of exemplars has been crafted from tau and $\alpha\beta$ fibril structures deposited in the PDB. After a set of potential binders has been identified using a target exemplar, compounds will be counter screened against this off-target exemplar library. This method, which has been previously demonstrated by Karanicolas and coworkers, allows for the rapid removal of non-specific compounds from a set of potential binders.³⁰¹ Furthermore, efforts are ongoing for developing a second tier *in silico* screening approach to cull false positions from potential hits identified in the initial screen. The basis of this approach to date has employed modeling in PyRosetta, along with the development of a custom score function, to predict experimental binding data. The concept of custom score functions is discussed later in this chapter. The identification, testing and development of new compounds is the subject of a recently approved U19 grant from the NIH which involves the Petersson Lab, along with many others. Overall, the exemplar-based methods employed in the latter portion of this thesis will be

employed to generate new lead compounds for α S fibrils from the multitude of cryo-EM structures recently deposited in the Protein Data Bank and can be extended to other fibrillar proteins for identifying targeting compounds.^{50, 302-303}

The final future focus on this topic is on the combination of disordered ensemble prediction with exemplar-based modeling to identify compounds that bind disordered proteins. Unlike the previous focus on targeting fibrils, the lack of static pockets in the disordered state makes pocket identification and thus exemplar selection difficult. However, the previously described FastFloppyTail method for generating accurate models of disordered proteins at unparalleled speeds provides a potential avenue to overcoming this sampling issue. By crafting composite exemplars from similar protein regions across many structures, one may be able to capture the necessary three-dimensional architecture and chemical interactions sufficient for identifying binders. Moreover, the efforts of Nath and colleagues as well as others suggests that this may be a viable strategy.^{298, 304} Lastly, the recent excitement surrounding proteolysis targeting chimera (PROTAC) further increases the potential utility for small molecules which are capable of selectively targeting a disordered protein.³⁰⁵ Ultimately this approach could open the class of disordered proteins to new therapeutic approaches.

§ 8.3 Novel Approaches in Rosetta Targeting Experimental Optimization

Finally, an ongoing computational effort is focused on developing custom score functions. In the Rosetta Modeling Suite, score functions are utilized alongside the Metropolis criterion to perform Monte Carlo sampling of protein

structures.²⁴⁰ Here, custom score functions allow a sequence or molecule to be scored, post-sampling, to predict experimental measurables. Although these are comprised of the same energy terms as the canonical Rosetta score functions, which have been optimized to serve as a surrogate for overall energetics, the reweighting of these terms is optimized to predict some experiment of interest. This requires a significant experimental dataset whose fundamental experimental measurable can be traced to a structural phenomenon. Interestingly, this is not an approach previously employed within the Rosetta community which has largely focused on optimizing the core scoring performance of Rosetta (i.e., matching structural or thermodynamic data).¹⁹⁴ A preliminary example of how this may be employed is demonstrated in this thesis work in the chapter entitled “Systematic Evaluation of Soluble Protein Expression Using a Fluorescent Unnatural Amino Acid Reveals No Reliable Predictions of Tolerability.” Following an initial demonstration that the energetic differences between the mutant and native proteins as assessed by the canonical Rosetta energy indeed was insufficient for the prediction of soluble protein fractions, it was demonstrated that individual components were able to demonstrate some correlative capacity. Therefore, it is likely that recombination of these terms can afford a score function that is capable of predicting the soluble fraction of a given mutant provided the initial dataset utilized for training is sufficient. To this end, further experimentation is underway with a more extensive library of proteins and unnatural amino acid mutants.

Although the concept of custom score functions was conceived with the goal of predicting tolerability of unnatural amino acid incorporation, there are many

opportunities to extend this approach to predict other phenomena of interest. Exemplar-based modeling provides an efficient platform for identifying potential lead molecules but, as previously demonstrates, delivers several false-positives. Therefore, a custom score function is currently in development which utilizes binding data from the SAR study present herein to generate a second-tier screening algorithm. This approach docks each compound to the input fibril structure and uses the resulting structure as the basis for custom score evaluation. Preliminary efforts, not presented in this work, have shown significant promise in the capabilities of this method. Lastly, this approach is being extended to other systems of interest in the Petersson lab which are focused on design. Custom score functions from peptide proteolysis and logP data are providing algorithms for designing stabilized peptides while directed evolution data will be used to improve *in silico* directed evolution of aminoacyl tRNA synthetase.

Overall, the methods and tools developed herein have been and continue to be employed towards challenging problems for disordered and aggregation prone proteins and beyond. This work has focused on developing generalized approaches and employing these and other methods to elucidate the structural complexities of disordered proteins. Moreover, this work has extended previously developed methods to the class of aggregation prone proteins and demonstrated the efficacy of these approaches. Lastly, these developments are continuing to be improved and are being leveraged for new protein systems of interest.

BIBLIOGRAPHY

1. Voet, D.; Voet, J. G., *Biochemistry*. 2nd ed.; J. Wiley & Sons: New York, 1995; p xvii, 1361 p.
2. Oldfield, C. J.; Dunker, A. K., Intrinsically Disordered Proteins and Intrinsically Disordered Protein Regions. *Annual Review of Biochemistry* **2014**, *83* (1), 553-584.
3. van der Lee, R.; Buljan, M.; Lang, B.; Weatheritt, R. J.; Daughdrill, G. W.; Dunker, A. K.; Fuxreiter, M.; Gough, J.; Gsponer, J.; Jones, D. T.; Kim, P. M.; Kriwacki, R. W.; Oldfield, C. J.; Pappu, R. V.; Tompa, P.; Uversky, V. N.; Wright, P. E.; Babu, M. M., Classification of Intrinsically Disordered Regions and Proteins. *Chemical Reviews* **2014**, *114* (13), 6589-6631.
4. Dunker, A. K.; Brown, C. J.; Lawson, J. D.; Iakoucheva, L. M.; Obradovic, Z., Intrinsic disorder and protein function. *Biochemistry* **2002**, *41* (21), 6573-6582.
5. Uversky, V. N.; Oldfield, C. J.; Dunker, A. K., Intrinsically Disordered Proteins in Human Diseases: Introducing the D2 Concept. *Annual Review of Biophysics* **2008**, *37* (1), 215-246.
6. Tompa, P., Intrinsically unstructured proteins. *Trends in Biochemical Sciences* **2002**, *27* (10), 527-533.
7. Darling, A. L.; Uversky, V. N., Intrinsic Disorder and Posttranslational Modifications: The Darker Side of the Biological Dark Matter. *Frontiers in Genetics* **2018**, *9* (158).
8. Wang, J.; Choi, J.-M.; Holehouse, A. S.; Lee, H. O.; Zhang, X.; Jahnel, M.; Maharana, S.; Lemaitre, R.; Pozniakovsky, A.; Drechsel, D.; Poser, I.; Pappu, R. V.; Alberti, S.; Hyman, A. A., A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell* **2018**, *174* (3), 688-699.e16.
9. Chong, S.; Dugast-Darzacq, C.; Liu, Z.; Dong, P.; Dailey, G. M.; Cattoglio, C.; Heckert, A.; Banala, S.; Lavis, L.; Darzacq, X.; Tjian, R., Imaging dynamic and selective low-complexity domain interactions that control gene transcription. *Science* **2018**, *361* (6400), eaar2555.
10. McSwiggen, D. T.; Hansen, A. S.; Teves, S. S.; Marie-Nelly, H.; Hao, Y.; Heckert, A. B.; Umemoto, K. K.; Dugast-Darzacq, C.; Tjian, R.; Darzacq, X., Evidence for DNA-mediated nuclear compartmentalization distinct from phase separation. *eLife* **2019**, *8*, e47098.
11. Borgia, A.; Borgia, M. B.; Bugge, K.; Kissling, V. M.; Heidarsson, P. O.; Fernandes, C. B.; Sottini, A.; Soranno, A.; Buholzer, K. J.; Nettels, D.; Kragelund, B. B.; Best, R. B.; Schuler, B., Extreme disorder in an ultrahigh-affinity protein complex. *Nature* **2018**, *555* (7694), 61-+.
12. Berlow, R. B.; Dyson, H. J.; Wright, P. E., Expanding the Paradigm: Intrinsically Disordered Proteins and Allosteric Regulation. *Journal of Molecular Biology* **2018**, *430* (16), 2309-2320.
13. Wright, P. E.; Dyson, H. J., Intrinsically disordered proteins in cellular signalling and regulation. *Nature reviews. Molecular cell biology* **2015**, *16* (1), 18-29.

14. Tsafou, K.; Tiwari, P. B.; Forman-Kay, J. D.; Metallo, S. J.; Toretzky, J. A., Targeting Intrinsically Disordered Transcription Factors: Changing the Paradigm. *Journal of Molecular Biology* **2018**, *430* (16), 2321-2341.
15. Scheckel, C.; Aguzzi, A., Prions, prionoids and protein misfolding disorders. *Nature Reviews Genetics* **2018**, *19* (7), 405-418.
16. Chiti, F.; Dobson, C. M., Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress Over the Last Decade. *Annual Review of Biochemistry* **2017**, *86* (1), 27-68.
17. Akter, R.; Cao, P.; Noor, H.; Ridgway, Z.; Tu, L.-H.; Wang, H.; Wong, A. G.; Zhang, X.; Abedini, A.; Schmidt, A. M.; Raleigh, D. P., Islet Amyloid Polypeptide: Structure, Function, and Pathophysiology. *Journal of Diabetes Research* **2016**, *2016*, 18.
18. Kelly, J. W., The alternative conformations of amyloidogenic proteins and their multi-step assembly pathways. *Curr. Opin. Struct. Biol.* **1998**, *8* (1), 101-106.
19. Lashuel, H. A.; Overk, C. R.; Oueslati, A.; Masliah, E., The many faces of α -synuclein: from structure and toxicity to therapeutic target. *Nature Reviews Neuroscience* **2013**, *14* (1), 38-48.
20. Auluck, P. K.; Caraveo, G.; Lindquist, S., α -Synuclein: membrane interactions and toxicity in Parkinson's disease. *Annual Review of Cell and Developmental Biology* **2010**, *26*, 211-33.
21. Lavedan, C., The Synuclein Family. *Genome Research* **1998**, *8* (9), 871-880.
22. Spillantini, M. G.; Crowther, R. A.; Jakes, R.; Hasegawa, M.; Goedert, M., α -synuclein in filamentous inclusions of Lewy bodies from Parkinson's disease and dementia with Lewy bodies. *Proceedings of the National Academy of Sciences of the United States of America* **1998**, *95* (11), 6469-6473.
23. Ulmer, T. S.; Bax, A.; Cole, N. B.; Nussbaum, R. L., Structure and Dynamics of Micelle-bound Human α -Synuclein. *Journal of Biological Chemistry* **2005**, *280* (10), 9595-9603.
24. Trexler, A. J.; Rhoades, E., α -Synuclein Binds Large Unilamellar Vesicles as an Extended Helix. *Biochemistry* **2009**, *48* (11), 2304-2306.
25. Eliezer, D.; Kutluay, E.; Bussell, R.; Browne, G., Conformational properties of α -synuclein in its free and lipid-associated states. *Journal of Molecular Biology* **2001**, *307* (4), 1061-1073.
26. Han, J.; Pluhackova, K.; Böckmann, R. A., The Multifaceted Role of SNARE Proteins in Membrane Fusion. *Front Physiol* **2017**, *8*, 5-5.
27. Atias, M.; Tevet, Y.; Sun, J.; Stavsky, A.; Tal, S.; Kahn, J.; Roy, S.; Gitler, D., Synapsins regulate α -synuclein functions. *Proceedings of the National Academy of Sciences* **2019**, *116* (23), 11116-11118.
28. Sun, J.; Wang, L.; Bao, H.; Premi, S.; Das, U.; Chapman, E. R.; Roy, S., Functional cooperation of α -synuclein and VAMP2 in synaptic vesicle recycling. *Proceedings of the National Academy of Sciences* **2019**, *116* (23), 11113-11115.
29. DeWitt, D. C.; Rhoades, E., α -Synuclein Can Inhibit SNARE-Mediated Vesicle Fusion through Direct Interactions with Lipid Bilayers. *Biochemistry* **2013**, *52* (14), 2385-2387.

30. Cheng, H.-C.; Ulane, C. M.; Burke, R. E., Clinical progression in Parkinson disease and the neurobiology of axons. *Ann Neurol* **2010**, *67* (6), 715-725.
31. Recasens, A.; Dehay, B., Alpha-synuclein spreading in Parkinson's disease. *Front Neuroanat* **2014**, *8*, 159-159.
32. Vargas, J. Y.; Grudina, C.; Zurzolo, C., The prion-like spreading of alpha-synuclein: From in vitro to in vivo models of Parkinson's disease. *Ageing research reviews* **2019**, *50*, 89-101.
33. Kotzbauer, P. T.; Tu, Z.; Mach, R. H., Current status of the development of PET radiotracers for imaging alpha synuclein aggregates in Lewy bodies and Lewy neurites. *Clinical and Translational Imaging* **2017**, *5* (1), 3-14.
34. McCormack, A.; Chegeni, N.; Chegini, F.; Colella, A.; Power, J.; Keating, D.; Chataway, T., Purification of alpha-synuclein containing inclusions from human post mortem brain tissue. *Journal of Neuroscience Methods* **2016**, *266*, 141-150.
35. Zhang, G.; Xia, Y.; Wan, F.; Ma, K.; Guo, X.; Kou, L.; Yin, S.; Han, C.; Liu, L.; Huang, J.; Xiong, N.; Wang, T., New Perspectives on Roles of Alpha-Synuclein in Parkinson's Disease. *Front Aging Neurosci* **2018**, *10*, 370-370.
36. Thakur, P.; Breger, L. S.; Lundblad, M.; Wan, O. W.; Mattsson, B.; Luk, K. C.; Lee, V. M. Y.; Trojanowski, J. Q.; Björklund, A., Modeling Parkinson's disease pathology by combination of fibril seeds and α -synuclein overexpression in the rat brain. *Proceedings of the National Academy of Sciences* **2017**, *114* (39), E8284-E8293.
37. Volpicelli-Daley, L. A.; Luk, K. C.; Patel, T. P.; Tanik, S. A.; Riddle, D. M.; Stieber, A.; Meaney, D. F.; Trojanowski, J. Q.; Lee, V. M., Exogenous alpha-synuclein fibrils induce Lewy body pathology leading to synaptic dysfunction and neuron death. *Neuron* **2011**, *72* (1), 57-71.
38. Luk, K. C.; Kehm, V.; Carroll, J.; Zhang, B.; O'Brien, P.; Trojanowski, J. Q.; Lee, V. M., Pathological alpha-synuclein transmission initiates Parkinson-like neurodegeneration in nontransgenic mice. *Science* **2012**, *338* (6109), 949-53.
39. Chartier-Harlin, M. C.; Kachergus, J.; Roumier, C.; Mouroux, V.; Douay, X.; Lincoln, S.; Levecque, C.; Larvor, L.; Andrieux, J.; Hulihan, M.; Waucquier, N.; Defebvre, L.; Amouyel, P.; Farrer, M.; Destee, A., alpha-synuclein locus duplication as a cause of familial Parkinson's disease. *Lancet* **2004**, *364* (9440), 1167-1169.
40. Singleton, A. B.; Farrer, M.; Johnson, J.; Singleton, A.; Hague, S.; Kachergus, J.; Hulihan, M.; Peuralinna, T.; Dutra, A.; Nussbaum, R.; Lincoln, S.; Crawley, A.; Hanson, M.; Maraganore, D.; Adler, C.; Cookson, M. R.; Muentert, M.; Baptista, M.; Miller, D.; Blancato, J.; Hardy, J.; Gwinn-Hardy, K., alpha-synuclein locus triplication causes Parkinson's disease. *Science* **2003**, *302* (5646), 841-841.
41. Polymeropoulos, M. H.; Lavedan, C.; Leroy, E.; Ide, S. E.; Dehejia, A.; Dutra, A.; Pike, B.; Root, H.; Rubenstein, J.; Boyer, R.; Stenroos, E. S.; Chandrasekharappa, S.; Athanassiadou, A.; Papapetropoulos, T.; Johnson, W. G.; Lazzarini, A. M.; Duvoisin, R. C.; Di Iorio, G.; Golbe, L. I.; Nussbaum, R. L., Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. *Science* **1997**, *276* (5321), 2045-7.

42. Kruger, R.; Kuhn, W.; Muller, T.; Woitalla, D.; Graeber, M.; Kosel, S.; Przuntek, H.; Epplen, J. T.; Schols, L.; Riess, O., Ala30Pro mutation in the gene encoding alpha-synuclein in Parkinson's disease. *Nature genetics* **1998**, *18* (2), 106-8.
43. Zarranz, J. J.; Alegre, J.; Gomez-Esteban, J. C.; Lezcano, E.; Ros, R.; Ampuero, I.; Vidal, L.; Hoenicka, J.; Rodriguez, O.; Atares, B.; Llorens, V.; Gomez Tortosa, E.; del Ser, T.; Munoz, D. G.; de Yebenes, J. G., The new mutation, E46K, of alpha-synuclein causes Parkinson and Lewy body dementia. *Ann Neurol* **2004**, *55* (2), 164-73.
44. Proukakis, C.; Houlden, H.; Schapira, A. H., Somatic alpha-synuclein mutations in Parkinson's disease: hypothesis and preliminary data. *Movement disorders : official journal of the Movement Disorder Society* **2013**, *28* (6), 705-12.
45. Lesage, S.; Anheim, M.; Letournel, F.; Bousset, L.; Honore, A.; Rozas, N.; Pieri, L.; Madiona, K.; Durr, A.; Melki, R.; Verny, C.; Brice, A., G51D alpha-synuclein mutation causes a novel parkinsonian-pyramidal syndrome. *Ann Neurol* **2013**, *73* (4), 459-71.
46. Fujioka, S.; Ogaki, K.; Tacik, P. M.; Uitti, R. J.; Ross, O. A.; Wszolek, Z. K., Update on novel familial forms of Parkinson's disease and multiple system atrophy. *Parkinsonism & related disorders* **2014**, *20 Suppl 1*, S29-34.
47. Tuttle, M. D.; Comellas, G.; Nieuwkoop, A. J.; Covell, D. J.; Berthold, D. A.; Kloepper, K. D.; Courtney, J. M.; Kim, J. K.; Barclay, A. M.; Kendall, A.; Wan, W.; Stubbs, G.; Schwieters, C. D.; Lee, V. M. Y.; George, J. M.; Rienstra, C. M., Solid-state NMR structure of a pathogenic fibril of full-length human α -synuclein. *Nat Struct Mol Biol* **2016**, *23* (5), 409-415.
48. Li, B.; Ge, P.; Murray, K. A.; Sheth, P.; Zhang, M.; Nair, G.; Sawaya, M. R.; Shin, W. S.; Boyer, D. R.; Ye, S.; Eisenberg, D. S.; Zhou, Z. H.; Jiang, L., Cryo-EM of full-length α -synuclein reveals fibril polymorphs with a common structural kernel. *Nature Communications* **2018**, *9* (1), 3609.
49. Li, Y.; Zhao, C.; Luo, F.; Liu, Z.; Gui, X.; Luo, Z.; Zhang, X.; Li, D.; Liu, C.; Li, X., Amyloid fibril structure of α -synuclein determined by cryo-electron microscopy. *Cell Research* **2018**, *28* (9), 897-903.
50. Guerrero-Ferreira, R.; Taylor, N. M. I.; Mona, D.; Ringler, P.; Lauer, M. E.; Riek, R.; Britschgi, M.; Stahlberg, H., Cryo-EM structure of alpha-synuclein fibrils. *eLife* **2018**, *7*, e36402.
51. Ferrie, J. J.; Ieda, N.; Haney, C. M.; Walters, C. R.; Sungwienwong, I.; Yoon, J.; Petersson, E. J., Multicolor protein FRET with tryptophan, selective coumarin-cysteine labeling, and genetic acridonylalanine encoding. *Chemical Communications* **2017**, *53* (80), 11072-11075.
52. Sungwienwong, I.; Ferrie, J. J.; Jun, J. V.; Liu, C.; Barrett, T. M.; Hostetler, Z. M.; Ieda, N.; Hendricks, A.; Muthusamy, A. K.; Kohli, R. M.; Chenoweth, D. M.; Petersson, G. A.; Petersson, E. J., Improving the fluorescent probe acridonylalanine through a combination of theory and experiment. *Journal of Physical Organic Chemistry* **2018**, *31* (8), e3813.
53. Ferrie, J. J.; Haney, C. M.; Yoon, J.; Pan, B.; Lin, Y.-C.; Fakhraai, Z.; Rhoades, E.; Nath, A.; Petersson, E. J., Using a FRET Library with Multiple Probe

- Pairs To Drive Monte Carlo Simulations of α -Synuclein. *Biophysical Journal* **2018**, *114* (1), 53-64.
54. Bonomi, M.; Pellarin, R.; Vendruscolo, M., Simultaneous Determination of Protein Structure and Dynamics Using Cryo-Electron Microscopy. *Biophysical Journal* **2018**, *114* (7), 1604-1613.
55. Luchinat, E.; Banci, L., In-cell NMR: a topical review. *IUCrJ* **2017**, *4* (Pt 2), 108-118.
56. Graether, S. P., Troubleshooting Guide to Expressing Intrinsically Disordered Proteins for Use in NMR Experiments. *Frontiers in Molecular Biosciences* **2019**, *5* (118).
57. Lakowicz, J. R., *Principles of Fluorescence Spectroscopy*. Springer US: 2013.
58. Förster, T., Zwischenmolekulare Energiewanderung und Fluoreszenz. *Annalen der Physik* **1948**, *437* (1-2), 55-75.
59. LeBlanc, S. J.; Kulkarni, P.; Weninger, K. R., Single Molecule FRET: A Powerful Tool to Study Intrinsically Disordered Proteins. *Biomolecules* **2018**, *8* (4), 140.
60. Schuler, B.; Soranno, A.; Hofmann, H.; Nettels, D., Single-Molecule FRET Spectroscopy and the Polymer Physics of Unfolded and Intrinsically Disordered Proteins. *Annual Review of Biophysics* **2016**, *45* (1), 207-231.
61. Konrat, R., NMR contributions to structural dynamics studies of intrinsically disordered proteins. *J Magn Reson* **2014**, *241* (100), 74-85.
62. Luchinat, E.; Banci, L., A Unique Tool for Cellular Structural Biology: In-cell NMR. *Journal of Biological Chemistry* **2016**, *291* (8), 3776-3784.
63. Haas, E., Ensemble FRET methods in studies of intrinsically disordered proteins. *Methods in molecular biology (Clifton, N.J.)* **2012**, *895*, 467-98.
64. Zheng, W.; Borgia, A.; Buholzer, K.; Grishaev, A.; Schuler, B.; Best, R. B., Probing the Action of Chemical Denaturant on an Intrinsically Disordered Protein by Simulation and Experiment. *Journal of the American Chemical Society* **2016**, *138* (36), 11702-11713.
65. Zheng, W.; Zerze, G. H.; Borgia, A.; Mittal, J.; Schuler, B.; Best, R. B., Inferring properties of disordered chains from FRET transfer efficiencies. *J Chem Phys* **2018**, *148* (12), 123329-123329.
66. Riback, J. A.; Bowman, M. A.; Zmyslowski, A. M.; Knoverek, C. R.; Jumper, J. M.; Hinshaw, J. R.; Kaye, E. B.; Freed, K. F.; Clark, P. L.; Sosnick, T. R., Innovative scattering analysis shows that hydrophobic disordered proteins are expanded in water. *Science* **2017**, *358* (6360), 238-241.
67. Ferreon, A. C. M.; Gambin, Y.; Lemke, E. A.; Deniz, A. A., Interplay of alpha-synuclein binding and conformational switching probed by single-molecule fluorescence. *Proceedings of the National Academy of Sciences of the United States of America* **2009**, *106* (14), 5645-5650.
68. Nath, A.; Sammalkorpi, M.; DeWitt, David C.; Trexler, Adam J.; Elbaum-Garfinkle, S.; O'Hern, Corey S.; Rhoades, E., The Conformational Ensembles of α -Synuclein and Tau: Combining Single-Molecule FRET and Simulations. *Biophysical Journal* **2012**, *103* (9), 1940-1949.

69. Roy, R.; Hohng, S.; Ha, T., A practical guide to single-molecule FRET. *Nat Methods* **2008**, *5* (6), 507-516.
70. Riback, J. A.; Bowman, M. A.; Zmyslowski, A. M.; Plaxco, K. W.; Clark, P. L.; Sosnick, T. R., Commonly used FRET fluorophores promote collapse of an otherwise disordered protein. *Proceedings of the National Academy of Sciences* **2019**, *116* (18), 8889-8894.
71. Afitska, K.; Fucikova, A.; Shvadchak, V. V.; Yushchenko, D. A., Modification of C Terminus Provides New Insights into the Mechanism of α -Synuclein Aggregation. *Biophysical journal* **2017**, *113* (10), 2182-2191.
72. Bajar, B. T.; Wang, E. S.; Zhang, S.; Lin, M. Z.; Chu, J., A Guide to Fluorescent Protein FRET Pairs. *Sensors (Basel)* **2016**, *16* (9), 1488.
73. Xia, T.; Li, N.; Fang, X., Single-molecule fluorescence imaging in living cells. *Annual review of physical chemistry* **2013**, *64*, 459-80.
74. Cole, N. B., Site-specific protein labeling with SNAP-tags. *Current protocols in protein science* **2013**, *73*, 30.1.1-30.1.
75. England, C. G.; Luo, H.; Cai, W., HaloTag technology: a versatile platform for biomedical applications. *Bioconjugate chemistry* **2015**, *26* (6), 975-986.
76. Jun, J. V.; Petersson, E. J.; Chenoweth, D. M., Rational Design and Facile Synthesis of a Highly Tunable Quinoline-Based Fluorescent Small-Molecule Scaffold for Live Cell Imaging. *Journal of the American Chemical Society* **2018**, *140* (30), 9486-9493.
77. Speight, L. C.; Muthusamy, A. K.; Goldberg, J. M.; Warner, J. B.; Wissner, R. F.; Willi, T. S.; Woodman, B. F.; Mehl, R. A.; Petersson, E. J., Efficient Synthesis and In Vivo Incorporation of Acridon-2-ylalanine, a Fluorescent Amino Acid for Lifetime and Förster Resonance Energy Transfer/Luminescence Resonance Energy Transfer Studies. *Journal of the American Chemical Society* **2013**, *135* (50), 18806-18814.
78. Haney, C. M.; Wissner, R. F.; Warner, J. B.; Wang, Y. J.; Ferrie, J. J.; Covell, D. J.; Karpowicz, R. J.; Lee, V. M.-Y.; Petersson, E. J., Comparison of strategies for non-perturbing labeling of α -synuclein to study amyloidogenesis. *Organic & biomolecular chemistry* **2016**, *14* (5), 1584-1592.
79. Huang, Y.; Ferrie, J. J.; Chen, X.; Zhang, Y.; Szantai-Kis, D. M.; Chenoweth, D. M.; Petersson, E. J., Electronic interactions of $i, i + 1$ dithioamides: increased fluorescence quenching and evidence for n -to- π^* interactions. *Chemical Communications* **2016**, *52* (50), 7798-7801.
80. Petersson, E. J.; Goldberg, J. M.; Wissner, R. F., On the use of thioamides as fluorescence quenching probes for tracking protein folding and stability. *Physical Chemistry Chemical Physics* **2014**, *16* (15), 6827-6837.
81. Walters, C. R.; Ferrie, J. J.; Petersson, E. J., Dithioamide substitutions in proteins: effects on thermostability, peptide binding, and fluorescence quenching in calmodulin. *Chemical Communications* **2018**, *54* (14), 1766-1769.
82. Noren, C. J.; Anthony-Cahill, S. J.; Griffith, M. C.; Schultz, P. G., A general method for site-specific incorporation of unnatural amino acids into proteins. *Science* **1989**, *244* (4901), 182-8.

83. England, P. M., Unnatural Amino Acid Mutagenesis: A Precise Tool for Probing Protein Structure and Function. *Biochemistry* **2004**, *43*(37), 11623-11629.
84. Heckler, T. G.; Chang, L. H.; Zama, Y.; Naka, T.; Chorghade, M. S.; Hecht, S. M., T4 RNA ligase mediated preparation of novel "chemically misacylated" tRNAPhe. *Biochemistry* **1984**, *23* (7), 1468-1473.
85. Bain, J. D.; Diala, E. S.; Glabe, C. G.; Dix, T. A.; Chamberlin, A. R., Biosynthetic site-specific incorporation of a non-natural amino acid into a polypeptide. *Journal of the American Chemical Society* **1989**, *111* (20), 8013-8014.
86. Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G., Expanding the genetic code of Escherichia coli. *Science* **2001**, *292* (5516), 498-500.
87. Wang, L.; Xie, J.; Schultz, P. G., EXPANDING THE GENETIC CODE. *Annual Review of Biophysics and Biomolecular Structure* **2006**, *35* (1), 225-249.
88. Packer, M. S.; Liu, D. R., Methods for the directed evolution of proteins. *Nature Reviews Genetics* **2015**, *16*, 379.
89. Cooley, R. B.; Karplus, P. A.; Mehl, R. A., Gleaning unexpected fruits from hard-won synthetases: probing principles of permissivity in non-canonical amino acid-tRNA synthetases. *ChemBiochem* **2014**, *15* (12), 1810-1819.
90. Smolskaya, S.; Andreev, Y. A., Site-Specific Incorporation of Unnatural Amino Acids into Escherichia coli Recombinant Protein: Methodology Development and Recent Achievement. *Biomolecules* **2019**, *9* (7), 255.
91. Anderson, J. C.; Wu, N.; Santoro, S. W.; Lakshman, V.; King, D. S.; Schultz, P. G., An expanded genetic code with a functional quadruplet codon. *Proceedings of the National Academy of Sciences of the United States of America* **2004**, *101* (20), 7566-7571.
92. Short, G. F., 3rd; Golovine, S. Y.; Hecht, S. M., Effects of release factor 1 on in vitro protein translation and the elaboration of proteins containing unnatural amino acids. *Biochemistry* **1999**, *38* (27), 8808-19.
93. Lajoie, M. J.; Rovner, A. J.; Goodman, D. B.; Aerni, H. R.; Haimovich, A. D.; Kuznetsov, G.; Mercer, J. A.; Wang, H. H.; Carr, P. A.; Mosberg, J. A.; Rohland, N.; Schultz, P. G.; Jacobson, J. M.; Rinehart, J.; Church, G. M.; Isaacs, F. J., Genomically recoded organisms expand biological functions. *Science* **2013**, *342* (6156), 357-60.
94. Johnson, D. B. F.; Xu, J.; Shen, Z.; Takimoto, J. K.; Schultz, M. D.; Schmitz, R. J.; Xiang, Z.; Ecker, J. R.; Briggs, S. P.; Wang, L., RF1 knockout allows ribosomal incorporation of unnatural amino acids at multiple sites. *Nat Chem Biol* **2011**, *7* (11), 779-786.
95. Batjargal, S.; Walters, C. R.; Petersson, E. J., Inteins as Traceless Purification Tags for Unnatural Amino Acid Proteins. *Journal of the American Chemical Society* **2015**, *137* (5), 1734-1737.
96. Shah, N. H.; Muir, T. W., Inteins: Nature's Gift to Protein Chemists. *Chem Sci* **2014**, *5* (1), 446-461.
97. Piana, S.; Donchev, A. G.; Robustelli, P.; Shaw, D. E., Water Dispersion Interactions Strongly Influence Simulated Structural Properties of Disordered Protein States. *The Journal of Physical Chemistry B* **2015**, *119* (16), 5113-5123.

98. Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmuller, H.; MacKerell, A. D., Jr., CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat Methods* **2017**, *14* (1), 71-73.
99. Robustelli, P.; Piana, S.; Shaw, D. E., Developing a molecular dynamics force field for both folded and disordered protein states. *Proceedings of the National Academy of Sciences* **2018**.
100. Song, D.; Luo, R.; Chen, H.-F., The IDP-Specific Force Field ff14IDPSFF Improves the Conformer Sampling of Intrinsically Disordered Proteins. *Journal of Chemical Information and Modeling* **2017**, *57* (5), 1166-1178.
101. Best, R. B., Computational and theoretical advances in studies of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* **2017**, *42*, 147-154.
102. Shaw, D. E.; Deneroff, M. M.; Dror, R. O.; Kuskin, J. S.; Larson, R. H.; Salmon, J. K.; Young, C.; Batson, B.; Bowers, K. J.; Chao, J. C.; Eastwood, M. P.; Gagliardo, J.; Grossman, J. P.; Ho, C. R.; Ierardi, D. J.; Istv; #225; Kolossv, n.; #225; ry; Klepeis, J. L.; Layman, T.; McLeavey, C.; Moraes, M. A.; Mueller, R.; Priest, E. C.; Shan, Y.; Spengler, J.; Theobald, M.; Towles, B.; Wang, S. C., Anton, a special-purpose machine for molecular dynamics simulation. *Commun. ACM* **2008**, *51* (7), 91-97.
103. Shaw, D. E.; Grossman, J. P.; Bank, J. A.; Batson, B.; Butts, J. A.; Chao, J. C.; Deneroff, M. M.; Dror, R. O.; Even, A.; Fenton, C. H.; Forte, A.; Gagliardo, J.; Gill, G.; Greskamp, B.; Ho, C. R.; Ierardi, D. J.; Iserovich, L.; Kuskin, J. S.; Larson, R. H.; Layman, T.; Lee, L.-S.; Lerer, A. K.; Li, C.; Killebrew, D.; Mackenzie, K. M.; Mok, S. Y.-H.; Moraes, M. A.; Mueller, R.; Nociolo, L. J.; Peticolas, J. L.; Quan, T.; Ramot, D.; Salmon, J. K.; Scarpazza, D. P.; Schafer, U. B.; Siddique, N.; Snyder, C. W.; Spengler, J.; Tang, P. T. P.; Theobald, M.; Toma, H.; Towles, B.; Vitale, B.; Wang, S. C.; Young, C., Anton 2: raising the bar for performance and programmability in a special-purpose molecular dynamics supercomputer. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, IEEE Press: New Orleans, Louisiana, 2014; pp 41-53.
104. Dedmon, M. M.; Lindorff-Larsen, K.; Christodoulou, J.; Vendruscolo, M.; Dobson, C. M., Mapping Long-Range Interactions in α -Synuclein using Spin-Label NMR and Ensemble Molecular Dynamics Simulations. *Journal of the American Chemical Society* **2005**, *127* (2), 476-477.
105. Fisher, C. K.; Huang, A.; Stultz, C. M., Modeling intrinsically disordered proteins with bayesian statistics. *J Am Chem Soc* **2010**, *132* (42), 14919-27.
106. Krzeminski, M.; Marsh, J. A.; Neale, C.; Choy, W.-Y.; Forman-Kay, J. D., Characterization of disordered proteins with ENSEMBLE. *Bioinformatics* **2012**, *29* (3), 398-399.
107. Marsh, J. A.; Forman-Kay, J. D., Ensemble modeling of protein disordered states: Experimental restraint contributions and validation. *Proteins-Structure Function and Bioinformatics* **2012**, *80* (2), 556-572.
108. Salmon, L.; Nodet, G.; Ozenne, V.; Yin, G.; Jensen, M. R.; Zweckstetter, M.; Blackledge, M., NMR characterization of long-range order in intrinsically disordered proteins. *J Am Chem Soc* **2010**, *132* (24), 8407-18.

109. Wu, H.; Wolynes, P. G.; Papoian, G. A., AWSEM-IDP: A Coarse-Grained Force Field for Intrinsically Disordered Proteins. *The Journal of Physical Chemistry B* **2018**, *122* (49), 11115-11125.
110. Das, P.; Matysiak, S.; Mittal, J., Looking at the Disordered Proteins through the Computational Microscope. *ACS Cent Sci* **2018**, *4* (5), 534-542.
111. Mittal, A.; Lyle, N.; Harmon, T. S.; Pappu, R. V., Hamiltonian Switch Metropolis Monte Carlo Simulations for Improved Conformational Sampling of Intrinsically Disordered Regions Tethered to Ordered Domains of Proteins. *Journal of Chemical Theory and Computation* **2014**, *10* (8), 3550-3562.
112. Koehler Leman, J.; Weitzner, B. D.; Lewis, S. M.; Consortium, R.; Bonneau, R., Macromolecular Modeling and Design in Rosetta: New Methods and Frameworks. *Preprints* **2019**, 2019040263
113. Kaufmann, K. W.; Lemmon, G. H.; Deluca, S. L.; Sheehan, J. H.; Meiler, J., Practically useful: what the Rosetta protein modeling suite can do for you. *Biochemistry* **2010**, *49* (14), 2987-98.
114. Chaudhury, S.; Lyskov, S.; Gray, J. J., PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta. *Bioinformatics* **2010**, *26* (5), 689-691.
115. Haney, C. M.; Wissner, R. F.; Petersson, E. J., Multiply labeling proteins for studies of folding and stability. *Current Opinion in Chemical Biology* **2015**, *28*, 123-130.
116. Giepmans, B. N. G.; Adams, S. R.; Ellisman, M. H.; Tsien, R. Y., The Fluorescent Toolbox for Assessing Protein Location and Function. *Science* **2006**, *312* (5771), 217-224.
117. Lang, K.; Chin, J. W., Cellular Incorporation of Unnatural Amino Acids and Bioorthogonal Labeling of Proteins. *Chemical Reviews* **2014**, *114* (9), 4764-4806.
118. Sungwienwong, I.; Hostetler, Z. M.; Blizzard, R. J.; Porter, J. J.; Driggers, C. M.; Mbengi, L. Z.; Villegas, J. A.; Speight, L. C.; Saven, J. G.; Perona, J. J.; Kohli, R. M.; Mehl, R. A.; Petersson, E. J., Improving target amino acid selectivity in a permissive aminoacyl tRNA synthetase through counter-selection. *Organic & Biomolecular Chemistry* **2017**, *15* (17), 3603-3610.
119. Haney, C. M.; Cleveland, C. L.; Wissner, R. F.; Owei, L.; Robustelli, J.; Daniels, M. J.; Canyurt, M.; Rodriguez, P.; Ischiropoulos, H.; Baumgart, T.; Petersson, E. J., Site-Specific Fluorescence Polarization for Studying the Disaggregation of α -Synuclein Fibrils by Small Molecules. *Biochemistry* **2017**, *56* (5), 683-691.
120. Contessa, G. M.; Orsale, M.; Melino, S.; Torre, V.; Paci, M.; Desideri, A.; Cicero, D. O., Structure of calmodulin complexed with an olfactory CNG channel fragment and role of the central linker: Residual dipolar coupling to evaluate calmodulin binding modes outside the kinase family. *Journal of Biomolecular NMR* **2005**, *31* (3), 185-199.
121. Kainosho, M.; Torizawa, T.; Iwashita, Y.; Terauchi, T.; Mei Ono, A.; Güntert, P., Optimal isotope labelling for NMR protein structure determinations. *Nature* **2006**, *440* (7080), 52-57.

122. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 09*, Gaussian, Inc.: Pittsburgh PA, 2009.
123. Szymanska, A.; Wegner, K.; Lankiewicz, L., Synthesis of N-[(tert-butoxy)carbonyl]-3-(9,10-dihydro-9-oxoacridin-2-yl)-L-alanine, a new fluorescent amino acid derivative. *Helv. Chim. Acta* **2003**, *86* (10), 3326-3331.
124. Ozenne, V.; Schneider, R.; Yao, M.; Huang, J.-r.; Salmon, L.; Zweckstetter, M.; Jensen, M. R.; Blackledge, M., Mapping the Potential Energy Landscape of Intrinsically Disordered Proteins at Amino Acid Resolution. *Journal of the American Chemical Society* **2012**, *134* (36), 15138-15148.
125. Varadi, M.; Kosol, S.; Lebrun, P.; Valentini, E.; Blackledge, M.; Dunker, A. K.; Felli, I. C.; Forman-Kay, J. D.; Kriwacki, R. W.; Pierattelli, R.; Sussman, J.; Svergun, D. I.; Uversky, V. N.; Vendruscolo, M.; Wishart, D.; Wright, P. E.; Tompa, P., pE-DB: a database of structural ensembles of intrinsically disordered and of unfolded proteins. *Nucleic Acids Research* **2014**, *42* (D1), D326-D335.
126. Allison, J. R.; Varnai, P.; Dobson, C. M.; Vendruscolo, M., Determination of the Free Energy Landscape of α -Synuclein Using Spin Label Nuclear Magnetic Resonance Measurements. *Journal of the American Chemical Society* **2009**, *131* (51), 18314-18326.
127. Tong, H.; Zhao, J.; Li, X.; Zhang, Y.; Ma, S.; Lou, K.; Wang, W., Orchestration of dual cyclization processes and dual quenching mechanisms for enhanced selectivity and drastic fluorescence turn-on detection of cysteine. *Chemical Communications* **2017**, *53* (25), 3583-3586.
128. Chin, D.; Means, A. R., Calmodulin: a prototypical calcium sensor. *Trends Cell Biol* **2000**, *10* (8), 322-328.
129. Vetter, S. W.; Leclerc, E., Novel aspects of calmodulin target recognition and activation. *Eur J Biochem* **2003**, *270* (3), 404-414.
130. Spillantini, M. G.; Schmidt, M. L.; Lee, V. M. Y.; Trojanowski, J. Q.; Jakes, R.; Goedert, M., alpha-synuclein in Lewy bodies. *Nature* **1997**, *388* (6645), 839-840.
131. Wang, A.; Bolen, D. W., A Naturally Occurring Protective System in Urea-Rich Cells: Mechanism of Osmolyte Protection of Proteins against Urea Denaturation. *Biochemistry* **1997**, *36* (30), 9101-9108.

132. Canchi, D. R.; García, A. E., Cosolvent Effects on Protein Stability. *Annual review of physical chemistry* **2013**, *64* (1), 273-293.
133. Uversky, V. N.; Li, J.; Fink, A. L., Trimethylamine-N-oxide-induced folding of α -synuclein. *FEBS Letters* **2001**, *509* (1), 31-35.
134. Moosa, M. M.; Ferreon, A. C. M.; Deniz, A. A., Forced Folding of a Disordered Protein Accesses an Alternative Folding Landscape. *ChemPhysChem* **2015**, *16* (1), 90-94.
135. Ferreon, A. C. M.; Moosa, M. M.; Gambin, Y.; Deniz, A. A., Counteracting chemical chaperone effects on the single-molecule α -synuclein structural landscape. *Proceedings of the National Academy of Sciences* **2012**, *109* (44), 17826-17831.
136. Wissner, R. F.; Batjargal, S.; Fadzen, C. M.; Petersson, E. J., Labeling Proteins with Fluorophore/Thioamide Förster Resonant Energy Transfer Pairs by Combining Unnatural Amino Acid Mutagenesis and Native Chemical Ligation. *Journal of the American Chemical Society* **2013**, *135* (17), 6529-6540.
137. Anthis, N. J.; Doucleff, M.; Clore, G. M., Transient, Sparsely Populated Compact States of Apo and Calcium-Loaded Calmodulin Probed by Paramagnetic Relaxation Enhancement: Interplay of Conformational Selection and Induced Fit. *Journal of the American Chemical Society* **2011**, *133* (46), 18966-18974.
138. Liu, M. Y.; Chen, T. Y.; Ahamed, B.; Li, J.; Yau, K. W., CALCIUM-CALMODULIN MODULATION OF THE OLFACTORY CYCLIC NUCLEOTIDE-GATED CATION CHANNEL. *Science* **1994**, *266* (5189), 1348-1354.
139. Wu, G.; Berka, V.; Tsai, A.-L., Binding kinetics of calmodulin with target peptides of three nitric oxide synthase isozymes. *Journal of Inorganic Biochemistry* **2011**, *105* (9), 1226-1237.
140. Brown, S. E.; Martin, S. R.; Bayley, P. M., Kinetic Control of the Dissociation Pathway of Calmodulin-Peptide Complexes. *Journal of Biological Chemistry* **1997**, *272* (6), 3389-3397.
141. Kovacs, E.; Tóth, J.; Vértessy, B. G.; Liliom, K., Dissociation of Calmodulin-Target Peptide Complexes by the Lipid Mediator Sphingosylphosphorylcholine: IMPLICATIONS IN CALCIUM SIGNALING. *Journal of Biological Chemistry* **2010**, *285* (3), 1799-1808.
142. Lee, N. K.; Kapanidis, A. N.; Koh, H. R.; Korlann, Y.; Ho, S. O.; Kim, Y.; Gassman, N.; Kim, S. K.; Weiss, S., Three-color alternating-laser excitation of single molecules: Monitoring multiple interactions and distances. *Biophysical Journal* **2007**, *92* (1), 303-312.
143. Lee, J.; Lee, S.; Ragunathan, K.; Joo, C.; Ha, T.; Hohng, S., Single-Molecule Four-Color FRET. *Angewandte Chemie-International Edition* **2010**, *49* (51), 9922-9925.
144. Sun, Y. S.; Wallrabe, H.; Booker, C. F.; Day, R. N.; Periasamy, A., Three-Color Spectral FRET Microscopy Localizes Three Interacting Proteins in Living Cells. *Biophysical Journal* **2010**, *99* (4), 1274-1283.
145. Voss, S.; Zhao, L.; Chen, X.; Gerhard, F.; Wu, Y. W., Generation of an intramolecular three-color fluorescence resonance energy transfer probe by site-specific protein labeling. *Journal of Peptide Science* **2014**, *20* (2), 115-120.

146. Rogers, J. M. G.; Lippert, L. G.; Gai, F., Non-natural amino acid fluorophores for one- and two-step fluorescence resonance energy transfer applications. *Analytical Biochemistry* **2010**, *399* (2), 182-189.
147. Brown, M. P.; Royer, C., Fluorescence spectroscopy as a tool to investigate protein interactions. *Curr. Opin. Biotechnol.* **1997**, *8* (1), 45-49.
148. Eftink, M. R.; Shastry, M. C. R., Fluorescence methods for studying kinetics of protein-folding reactions. *Fluorescence Spectroscopy* **1997**, *278*, 258-286.
149. Beechem, J. M.; Brand, L., TIME-RESOLVED FLUORESCENCE OF PROTEINS. *Annual Review of Biochemistry* **1985**, *54*, 43-71.
150. Eftink, M. R.; Ghiron, C. A., FLUORESCENCE QUENCHING STUDIES WITH PROTEINS. *Analytical Biochemistry* **1981**, *114* (2), 199-227.
151. Van, D. M. B. W.; Coker, G., III; Chen, S. S. Y., *Resonance Energy Transfer: Theory and Data*. VCH: 1994; p approx. 192 pp.
152. Forster, T., Transfer mechanisms of electronic excitation. *Discuss. Faraday Soc.* **1959**, No. 27, 7-17.
153. Wu, P.; Brand, L., Resonance energy transfer: methods and applications. *Anal. Biochem.* **1994**, *218* (1), 1-13.
154. Goldberg, J. M.; Batjargal, S.; Chen, B. S.; Petersson, E. J., Thioamide Quenching of Fluorescent Probes Through Photoinduced Electron Transfer: Mechanistic Studies and Applications. *Journal of the American Chemical Society* **2013**.
155. Speight, L. C.; Samanta, M.; Petersson, E. J., Minimalist Approaches to Protein Labelling: Getting the Most Fluorescent Bang for Your Steric Buck. *Australian Journal of Chemistry* **2014**, *67* (5), 686-700.
156. Hamada, H.; Kameshima, N.; Szymanska, A.; Wegner, K.; Lankiewicz, L.; Shinohara, H.; Taki, M.; Sisido, M., Position-specific incorporation of a highly photodurable and blue-laser excitable fluorescent amino acid into proteins for fluorescence sensing. *Bioorganic & Medicinal Chemistry* **2005**, *13* (10), 3379-3384.
157. Taki, M.; Yamazaki, Y.; Suzuki, Y.; Sisido, M., Introduction of a Highly Photodurable and Common-laser Excitable Fluorescent Amino Acid into a Peptide as a FRET Acceptor for Protease Cleavage Detection. *Chemistry Letters* **2010**, *39* (8), 818-819.
158. Hoshaka, T.; Kajihara, D.; Ashizuka, Y.; Murakami, H.; Sisido, M., Efficient incorporation of nonnatural amino acids with large aromatic groups into streptavidin in in vitro protein synthesizing systems. *Journal of the American Chemical Society* **1999**, *121* (1), 34-40.
159. Hill, E. K.; de Mello, A. J.; Birrell, H.; Charlwood, J.; Camilleri, P., Steady state and time-resolved fluorescence of 2-aminoacridone sugar derivatives. *Journal of the Chemical Society-Perkin Transactions 2* **1998**, (11), 2337-2341.
160. Lozanov, V.; Ivanov, I. P.; Benkova, B.; Mitev, V., Peptide substrate for caspase-3 with 2-aminoacridone as reporting group. *Amino Acids* **2009**, *36* (3), 581-586.

161. Silva, V. L. M.; Silva, A. M. S.; Cavaleiro, J. A. S., New Synthesis of 2,3-Diarylacridin-9(10H)-ones and (E)-2-Phenyl-4-styrylfuro 3,2-c quinolines. *Synlett* **2010**, (17), 2565-2570.
162. Huang, C.; Yan, S. J.; Li, Y. M.; Huang, R.; Lin, J., Synthesis of polyhalo acridones as pH-sensitive fluorescence probes. *Bioorganic & Medicinal Chemistry Letters* **2010**, 20 (15), 4665-4669.
163. Schulman, S. G.; Sturgeon, R. J., VARIATIONS OF FLUORESCENCE EFFICIENCIES OF 9-(10H)-ACRIDONE AND ITS 4-METHOXY DERIVATIVE WITH PH EXCITED-STATE PROTON-TRANSFER IN VERY WEAK ACIDS AND BASES. *Analytica Chimica Acta* **1977**, 93 (1), 239-247.
164. Smith, J. A.; West, R. M.; Allen, M., Acridones and quinacridones: Novel fluorophores for fluorescence lifetime studies. *Journal of Fluorescence* **2004**, 14 (2), 151-171.
165. Zanker, V.; Cmiel, E., Photochemische Reduktionen von Nitro-9-acridonen in H-haltigen, O2-freien Lösungsmitteln. *Justus Liebigs Annalen der Chemie* **1975**, 1975 (9), 1576-1581.
166. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; J. A. Montgomery, J.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J., Gaussian 16, Revision A.03. Gaussian, Inc.: Wallingford CT, 2016.
167. Foresman, J. B.; Frisch, M. J., *Exploring Chemistry with Electronic Structure Methods*. 3 ed.; Gaussian, Inc.: Wallingford, CT, 2015.
168. Austin, A.; Petersson, G. A.; Frisch, M. J.; Dobek, F. J.; Scalmani, G.; Throssell, K., A Density Functional with Spherical Atom Dispersion Terms. *Journal of Chemical Theory and Computation* **2012**, 8 (12), 4989-5007.
169. Young, T. S.; Schultz, P. G., Beyond the canonical 20 amino acids: expanding the genetic lexicon. *The Journal of biological chemistry* **2010**, 285 (15), 11039-44.
170. Neumann-Staubitz, P.; Neumann, H., The use of unnatural amino acids to study and engineer protein function. *Curr Opin Struct Biol* **2016**, 38, 119-28.
171. Xiao, H.; Schultz, P. G., At the Interface of Chemical and Biological Synthesis: An Expanded Genetic Code. *Cold Spring Harb Perspect Biol* **2016**, 8 (9), a023945.

172. Chatterjee, A.; Sun, S. B.; Furman, J. L.; Xiao, H.; Schultz, P. G., A versatile platform for single- and multiple-unnatural amino acid mutagenesis in *Escherichia coli*. *Biochemistry* **2013**, *52* (10), 1828-37.
173. Goerke, A. R.; Swartz, J. R., High-level cell-free synthesis yields of proteins containing site-specific non-natural amino acids. *Biotechnology and bioengineering* **2009**, *102* (2), 400-16.
174. Albayrak, C.; Swartz, J. R., Cell-free co-production of an orthogonal transfer RNA activates efficient site-specific non-natural amino acid incorporation. *Nucleic Acids Res* **2013**, *41* (11), 5949-63.
175. Hammill, J. T.; Miyake-Stoner, S.; Hazen, J. L.; Jackson, J. C.; Mehl, R. A., Preparation of site-specifically labeled fluorinated proteins for ¹⁹F-NMR structural characterization. *Nature protocols* **2007**, *2* (10), 2601-7.
176. Hino, N.; Hayashi, A.; Sakamoto, K.; Yokoyama, S., Site-specific incorporation of non-natural amino acids into proteins in mammalian cells with an expanded genetic code. *Nature protocols* **2006**, *1* (6), 2957-62.
177. Zhang, B.; Xu, H.; Chen, J.; Zheng, Y.; Wu, Y.; Si, L.; Wu, L.; Zhang, C.; Xia, G.; Zhang, L.; Zhou, D., Development of next generation of therapeutic IFN- α 2b via genetic code expansion. *Acta biomaterialia* **2015**, *19*, 100-11.
178. Zheng, Y.; Yu, F.; Wu, Y.; Si, L.; Xu, H.; Zhang, C.; Xia, Q.; Xiao, S.; Wang, Q.; He, Q.; Chen, P.; Wang, J.; Taira, K.; Zhang, L.; Zhou, D., Broadening the versatility of lentiviral vectors as a tool in nucleic acid research via genetic code expansion. *Nucleic Acids Res* **2015**, *43* (11), e73.
179. Lau, K. F.; Dill, K. A., Theory for protein mutability and biogenesis. *Proceedings of the National Academy of Sciences of the United States of America* **1990**, *87* (2), 638-642.
180. Markiewicz, P.; Kleina, L. G.; Cruz, C.; Ehret, S.; Miller, J. H., Genetic studies of the lac repressor. XIV. Analysis of 4000 altered *Escherichia coli* lac repressors reveals essential and non-essential residues, as well as "spacers" which do not require a specific sequence. *J Mol Biol* **1994**, *240* (5), 421-33.
181. Lim, W. A.; Sauer, R. T., Alternative packing arrangements in the hydrophobic core of lambda repressor. *Nature* **1989**, *339* (6219), 31-6.
182. Campbell-Valois, F.-X.; Tarassov, K.; Michnick, S. W., Massive sequence perturbation of a small protein. *Proceedings of the National Academy of Sciences of the United States of America* **2005**, *102* (42), 14988-14993.
183. Romero, P. A.; Tran, T. M.; Abate, A. R., Dissecting enzyme function with microfluidic-based deep mutational scanning. *Proc Natl Acad Sci U S A* **2015**, *112* (23), 7159-64.
184. Luo, J.; Uprety, R.; Naro, Y.; Chou, C.; Nguyen, D. P.; Chin, J. W.; Deiters, A., Genetically encoded optochemical probes for simultaneous fluorescence reporting and light activation of protein function with two-photon excitation. *J Am Chem Soc* **2014**, *136* (44), 15551-8.
185. Reddington, S. C.; Baldwin, A. J.; Thompson, R.; Brancale, A.; Tippmann, E. M.; Jones, D. D., Directed evolution of GFP with non-natural amino acids identifies residues for augmenting and photoswitching fluorescence. *Chem Sci* **2015**, *6* (2), 1159-1166.

186. Arpino, J. A. J.; Baldwin, A. J.; McGarrity, A. R.; Tippmann, E. M.; Jones, D. D., In-Frame Amber Stop Codon Replacement Mutagenesis for the Directed Evolution of Proteins Containing Non-Canonical Amino Acids: Identification of Residues Open to Bio-Orthogonal Modification. *PLOS ONE* **2015**, *10* (5), e0127504.
187. Drew, K.; Renfrew, P. D.; Craven, T. W.; Butterfoss, G. L.; Chou, F. C.; Lyskov, S.; Bullock, B. N.; Watkins, A.; Labonte, J. W.; Pacella, M.; Kilambi, K. P.; Leaver-Fay, A.; Kuhlman, B.; Gray, J. J.; Bradley, P.; Kirshenbaum, K.; Arora, P. S.; Das, R.; Bonneau, R., Adding diverse noncanonical backbones to rosetta: enabling peptidomimetic design. *PLoS One* **2013**, *8* (7), e67051.
188. Huang, P.-S.; Ban, Y.-E. A.; Richter, F.; Andre, I.; Vernon, R.; Schief, W. R.; Baker, D., RosettaRemodel: A Generalized Framework for Flexible Backbone Protein Design. *PLOS ONE* **2011**, *6* (8), e24109.
189. Luo, Y.; Pfuetzner, R. A.; Mosimann, S.; Paetzel, M.; Frey, E. A.; Cherney, M.; Kim, B.; Little, J. W.; Strynadka, N. C., Crystal structure of LexA: a conformational switch for regulation of self-cleavage. *Cell* **2001**, *106* (5), 585-94.
190. Conway, P.; Tyka, M. D.; DiMaio, F.; Konerding, D. E.; Baker, D., Relaxation of backbone bond geometry improves protein energy landscape modeling. *Protein science : a publication of the Protein Society* **2014**, *23* (1), 47-55.
191. Chen, Z.; Yang, H.; Pavletich, N. P., Mechanism of homologous recombination from the RecA-ssDNA/dsDNA structures. *Nature* **2008**, *453* (7194), 489-4.
192. Smith, C. A.; Kortemme, T., Backrub-Like Backbone Simulation Recapitulates Natural Protein Conformational Variability and Improves Mutant Side-Chain Prediction. *Journal of Molecular Biology* **2008**, *380* (4), 742-756.
193. Kellogg, E. H.; Leaver-Fay, A.; Baker, D., Role of conformational sampling in computing mutation-induced changes in protein structure and stability. *Proteins* **2011**, *79* (3), 830-8.
194. Alford, R. F.; Leaver-Fay, A.; Jeliaskov, J. R.; O'Meara, M. J.; DiMaio, F. P.; Park, H.; Shapovalov, M. V.; Renfrew, P. D.; Mulligan, V. K.; Kappel, K.; Labonte, J. W.; Pacella, M. S.; Bonneau, R.; Bradley, P.; Dunbrack, R. L.; Das, R.; Baker, D.; Kuhlman, B.; Kortemme, T.; Gray, J. J., The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *Journal of Chemical Theory and Computation* **2017**, *13* (6), 3031-3048.
195. Bendor, Jacob T.; Logan, Todd P.; Edwards, Robert H., The Function of Alpha-Synuclein. *Neuron* **79** (6), 1044-1066.
196. Jao, C. C.; Hegde, B. G.; Chen, J.; Haworth, I. S.; Langen, R., Structure of membrane-bound alpha-synuclein from site-directed spin labeling and computational refinement. *Proceedings of the National Academy of Sciences* **2008**, *105* (50), 19666-71.
197. Uversky, V. N., Intrinsically Disordered Proteins and Their Environment: Effects of Strong Denaturants, Temperature, pH, Counter Ions, Membranes, Binding Partners, Osmolytes, and Macromolecular Crowding. *The Protein Journal* **2009**, *28* (7), 305-325.

198. Smith, W. W.; Schreck, C. F.; Hashem, N.; Soltani, S.; Nath, A.; Rhoades, E.; O'Hern, C. S., Molecular simulations of the fluctuating conformational dynamics of intrinsically disordered proteins. *Physical Review E* **2012**, *86*, 041910.
199. Jeschke, G., Ensemble models of proteins and protein domains based on distance distribution restraints. *Proteins: Structure, Function, and Bioinformatics* **2016**, *84* (4), 554-560.
200. Bertoncini, C. W.; Jung, Y.-S.; Fernandez, C. O.; Hoyer, W.; Griesinger, C.; Jovin, T. M.; Zweckstetter, M., Release of long-range tertiary interactions potentiates aggregation of natively unstructured α -synuclein. *Proceedings of the National Academy of Sciences of the United States of America* **2005**, *102* (5), 1430-1435.
201. Sung, Y. H.; Eliezer, D., Residual structure, backbone dynamics, and interactions within the synuclein family. *Journal of Molecular Biology* **2007**, *372* (3), 689-707.
202. Aznauryan, M.; Delgado, L.; Soranno, A.; Nettels, D.; Huang, J. R.; Labhardt, A. M.; Grzesiek, S.; Schuler, B., Comprehensive structural and dynamical view of an unfolded protein from the combination of single-molecule FRET, NMR, and SAXS. *Proceedings of the National Academy of Sciences of the United States of America* **2016**, *113* (37), E5389-E5398.
203. Ruzza, P.; Hussain, R.; Biondi, B.; Calderan, A.; Tessari, I.; Bubacco, L.; Siligardi, G., Effects of Trehalose on Thermodynamic Properties of Alpha-synuclein Revealed through Synchrotron Radiation Circular Dichroism. *Biomolecules* **2015**, *5* (2), 724.
204. McNulty, B. C.; Tripathy, A.; Young, G. B.; Charlton, L. M.; Orans, J.; Pielak, G. J., Temperature-induced reversible conformational change in the first 100 residues of alpha-synuclein. *Protein Science* **2006**, *15* (3), 602-608.
205. Morar, A. S.; Olteanu, A.; Young, G. B.; Pielak, G. J., Solvent-induced collapse of α -synuclein and acid-denatured cytochrome c. *Protein Science* **2001**, *10* (11), 2195-2199.
206. Bai, J.; Liu, M.; Pielak, G. J.; Li, C., Macromolecular and Small Molecular Crowding Have Similar Effects on α -Synuclein Structure. *ChemPhysChem* **2017**, *18* (1), 55-58.
207. Yancey, P. H., Organic osmolytes as compatible, metabolic and counteracting cytoprotectants in high osmolarity and other stresses. *Journal of Experimental Biology* **2005**, *208* (15), 2819-2830.
208. Knake, L.; Schwaab, G.; Kartaschew, K.; Havenith, M., Solvation Dynamics of Trimethylamine N-Oxide in Aqueous Solution Probed by Terahertz Spectroscopy. *The Journal of Physical Chemistry B* **2015**, *119* (43), 13842-13851.
209. Chang, Y.-C.; Oas, T. G., Osmolyte-Induced Folding of an Intrinsically Disordered Protein: Folding Mechanism in the Absence of Ligand. *Biochemistry* **2010**, *49* (25), 5086-5096.
210. Kokubo, H.; Hu, C. Y.; Pettitt, B. M., Peptide Conformational Preferences in Osmolyte Solutions: Transfer Free Energies of Decalanine. *Journal of the American Chemical Society* **2011**, *133* (6), 1849-1858.

211. Hu, C. Y.; Kokubo, H.; Lynch, G. C.; Bolen, D. W.; Pettitt, B. M., Backbone additivity in the transfer model of protein solvation. *Protein Science* **2010**, *19* (5), 1011-1022.
212. Holthauzen, L. M. F.; Rosgen, J.; Bolen, D. W., Hydrogen Bonding Progressively Strengthens upon Transfer of the Protein Urea-Denatured State to Water and Protecting Osmolytes. *Biochemistry* **2010**, *49* (6), 1310-1318.
213. Baskakov, I.; Bolen, D. W., Forcing thermodynamically unfolded proteins to fold. *Journal of Biological Chemistry* **1998**, *273* (9), 4831-4.
214. Larini, L.; Shea, J.-E., Double Resolution Model for Studying TMAO/Water Effective Interactions. *The Journal of Physical Chemistry B* **2013**, *117* (42), 13268-13277.
215. Doi, H.; Watanabe, Y.; Aida, M., Influence of Trimethylamine *N*-Oxide (TMAO) on the Three-dimensional Distribution and Alignment of Solvent Molecules in Aqueous Solution. *Chemistry Letters* **2014**, *43* (6), 865-867.
216. Sagle, L. B.; Cimat, K.; Litosh, V. A.; Liu, Y.; Flores, S. C.; Chen, X.; Yu, B.; Cremer, P. S., Methyl Groups of Trimethylamine *N*-Oxide Orient Away from Hydrophobic Interfaces. *Journal of the American Chemical Society* **2011**, *133* (46), 18707-18712.
217. Ma, J.; Pazos, I. M.; Gai, F., Microscopic insights into the protein-stabilizing effect of trimethylamine *N*-oxide (TMAO). *Proceedings of the National Academy of Sciences* **2014**, *111* (23), 8476-8481.
218. Cho, S. S.; Reddy, G.; Straub, J. E.; Thirumalai, D., Entropic Stabilization of Proteins by TMAO. *The Journal of Physical Chemistry B* **2011**, *115* (45), 13401-13407.
219. Rao, J. N.; Jao, C. C.; Hegde, B. G.; Langen, R.; Ulmer, T. S., A Combinatorial NMR and EPR Approach for Evaluating the Structural Ensemble of Partially Folded Proteins. *Journal of the American Chemical Society* **2010**, *132* (25), 8657-8668.
220. Batjargal, S.; Wang, Y. J.; Goldberg, J. M.; Wissner, R. F.; Petersson, E. J., Native Chemical Ligation of Thioamide-Containing Peptides: Development and Application to the Synthesis of Labeled α -Synuclein for Misfolding Studies. *Journal of the American Chemical Society* **2012**, *134* (22), 9172-9182.
221. Glasscock, J. M.; Zhu, Y.; Chowdhury, P.; Tang, J.; Gai, F., Using an amino acid fluorescence resonance energy transfer pair to probe protein unfolding: application to the villin headpiece subdomain and the LysM domain. *Biochemistry* **2008**, *47* (42), 11070-6.
222. Taskent-Sezgin, H.; Chung, J.; Patsalo, V.; Miyake-Stoner, S. J.; Miller, A. M.; Brewer, S. H.; Mehl, R. A.; Green, D. F.; Raleigh, D. P.; Carrico, I., Interpretation of *p*-Cyanophenylalanine Fluorescence in Proteins in Terms of Solvent Exposure and Contribution of Side-Chain Quenchers: A Combined Fluorescence, IR and Molecular Dynamics Study. *Biochemistry* **2009**, *48* (38), 9040-9046.
223. Gill, S. C.; von Hippel, P. H., Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem* **1989**, *182* (2), 319-26.

224. Horne, W. S.; Johnson, L. M.; Ketas, T. J.; Klasse, P. J.; Lu, M.; Moore, J. P.; Gellman, S. H., Structural and biological mimicry of protein surface recognition by α/β -peptide foldamers. *Proceedings of the National Academy of Sciences* **2009**, *106* (35), 14751-14756.
225. Kim, D. E.; Chivian, D.; Baker, D., Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Research* **2004**, *32* (Web Server issue), W526-W531.
226. Ortega, A.; Amoros, D.; Garcia de la Torre, J., Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models. *Biophys J* **2011**, *101* (4), 892-8.
227. Li, X.-H.; Culver, J. A.; Rhoades, E., Tau Binds to Multiple Tubulin Dimers with Helical Structure. *Journal of the American Chemical Society* **2015**, *137* (29), 9218-9221.
228. Araki, K.; Yagi, N.; Nakatani, R.; Sekiguchi, H.; So, M.; Yagi, H.; Ohta, N.; Nagai, Y.; Goto, Y.; Mochizuki, H., A small-angle X-ray scattering study of alpha-synuclein from human red blood cells. **2016**, *6*, 30473.
229. Song, J.; Gomes, G.-N.; Shi, T.; Gradinaru, C. C.; Chan, H. S., Conformational Heterogeneity and FRET Data Interpretation for Dimensions of Unfolded Proteins. *Biophysical Journal* **2017**, *113* (5), 1012-1024.
230. Kakish, J.; Allen, K. J. H.; Harkness, T. A.; Krol, E. S.; Lee, J. S., Novel Dimer Compounds That Bind α -Synuclein Can Rescue Cell Growth in a Yeast Model Overexpressing α -Synuclein. A Possible Prevention Strategy for Parkinson's Disease. *ACS Chemical Neuroscience* **2016**, *7* (12), 1671-1680.
231. Dikiy, I.; Fauvet, B.; Jovičić, A.; Mahul-Mellier, A.-L.; Desobry, C.; El-Turk, F.; Gitler, A. D.; Lashuel, H. A.; Eliezer, D., Semisynthetic and in Vitro Phosphorylation of Alpha-Synuclein at Y39 Promotes Functional Partly Helical Membrane-Bound States Resembling Those Induced by PD Mutations. *ACS Chemical Biology* **2016**, *11* (9), 2428-2437.
232. Lewis, Y. E.; Galesic, A.; Levine, P. M.; De Leon, C. A.; Lamiri, N.; Brennan, C. K.; Pratt, M. R., O-GlcNAcylation of α -Synuclein at Serine 87 Reduces Aggregation without Affecting Membrane Binding. *ACS Chemical Biology* **2017**, *12* (4), 1020-1027.
233. Del Mar, C.; Greenbaum, E. A.; Mayne, L.; Englander, S. W.; Woods, V. L., Structure and properties of α -synuclein and other amyloids determined at the amino acid level. *Proceedings of the National Academy of Sciences* **2005**, *102* (43), 15477-15482.
234. Uversky, V. N., Introduction to Intrinsically Disordered Proteins (IDPs). *Chemical Reviews* **2014**, *114* (13), 6557-6560.
235. Martin, E. W.; Holehouse, A. S.; Grace, C. R.; Hughes, A.; Pappu, R. V.; Mittag, T., Sequence Determinants of the Conformational Properties of an Intrinsically Disordered Protein Prior to and upon Multisite Phosphorylation. *Journal of the American Chemical Society* **2016**, *138* (47), 15323-15335.
236. Kleiger, G.; Saha, A.; Lewis, S.; Kuhlman, B.; Deshaies, R. J., Rapid E2-E3 assembly and disassembly enable processive ubiquitylation of cullin-RING ubiquitin ligase substrates. *Cell* **2009**, *139* (5), 957-68.

237. Simons, K. T.; Kooperberg, C.; Huang, E.; Baker, D., Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. *J Mol Biol* **1997**, *268* (1), 209-25.
238. Bradley, P.; Misura, K. M. S.; Baker, D., Toward High-Resolution de Novo Structure Prediction for Small Proteins. *Science* **2005**, *309* (5742), 1868-1871.
239. Raman, S.; Vernon, R.; Thompson, J.; Tyka, M.; Sadreyev, R.; Pei, J.; Kim, D.; Kellogg, E.; DiMaio, F.; Lange, O.; Kinch, L.; Sheffler, W.; Kim, B.-H.; Das, R.; Grishin, N. V.; Baker, D., Structure prediction for CASP8 with all-atom refinement using Rosetta. *Proteins* **2009**, *77 Suppl 9* (0 9), 89-99.
240. Rohl, C. A.; Strauss, C. E.; Misura, K. M.; Baker, D., Protein structure prediction using Rosetta. *Methods in enzymology* **2004**, *383*, 66-93.
241. Gront, D.; Kulp, D. W.; Vernon, R. M.; Strauss, C. E. M.; Baker, D., Generalized Fragment Picking in Rosetta: Design, Protocols and Applications. *PLOS ONE* **2011**, *6* (8), e23294.
242. Leman, J. K.; Ralf, M.; Mert, K.; Nils, W.; Jens, M., Simultaneous prediction of protein secondary structure and transmembrane spans. *Proteins: Structure, Function, and Bioinformatics* **2013**, *81* (7), 1127-1140.
243. Buchan, D. W. A.; Minnici, F.; Nugent, T. C. O.; Bryson, K.; Jones, D. T., Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Research* **2013**, *41* (Web Server issue), W349-W357.
244. Wang, S.; Li, W.; Liu, S.; Xu, J., RaptorX-Property: a web server for protein structure property prediction. *Nucleic Acids Research* **2016**, *44* (W1), W430-W435.
245. Shen, Y.; Bax, A., SPARTA+: a modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network. *J Biomol NMR* **2010**, *48* (1), 13-22.
246. Zweckstetter, M., NMR: prediction of molecular alignment from structure using the PALES software. *Nature protocols* **2008**, *3*, 679.
247. Mantsyzov, A. B.; Maltsev, A. S.; Ying, J.; Shen, Y.; Hummer, G.; Bax, A., A maximum entropy approach to the study of residue-specific backbone angle distributions in α -synuclein, an intrinsically disordered protein. *Protein Science* **2014**, *23* (9), 1275-1290.
248. Kabsch, W.; Sander, C., Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, *22* (12), 2577-637.
249. Vögeli, B.; Ying, J.; Grishaev, A.; Bax, A., Limits on Variations in Protein Backbone Dynamics from Precise Measurements of Scalar Couplings. *Journal of the American Chemical Society* **2007**, *129* (30), 9377-9385.
250. Vuister, G. W.; Delaglio, F.; Bax, A., The use of 1JC alpha H alpha coupling constants as a probe for protein backbone conformation. *J Biomol NMR* **1993**, *3* (1), 67-80.
251. Wirmer, J.; Schwalbe, H., Angular dependence of 1J(Ni,C α i) and 2J(Ni,C α (i - 1)) coupling constants measured in J-modulated HSQCs. *Journal of Biomolecular NMR* **2002**, *23* (1), 47-55.

252. Lee, J. H.; Li, F.; Grishaev, A.; Bax, A., Quantitative Residue-Specific Protein Backbone Torsion Angle Dynamics from Concerted Measurement of 3J Couplings. *Journal of the American Chemical Society* **2015**, *137* (4), 1432-1435.
253. Lee, J. C.; Gray, H. B.; Winkler, J. R., Tertiary Contact Formation in α -Synuclein Probed by Electron Transfer. *Journal of the American Chemical Society* **2005**, *127* (47), 16388-16389.
254. Grupi, A.; Haas, E., Segmental conformational disorder and dynamics in the intrinsically disordered protein alpha-synuclein and its chain length dependence. *J Mol Biol* **2011**, *405* (5), 1267-83.
255. Yang, S.; Julien, R.; Alexander, G.; Ad, B., Prediction of nearest neighbor effects on backbone torsion angles and NMR scalar coupling constants in disordered proteins. *Protein Science* **2018**, *27* (1), 146-158.
256. Hofmann, H.; Soranno, A.; Borgia, A.; Gast, K.; Nettels, D.; Schuler, B., Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with single-molecule spectroscopy. *Proceedings of the National Academy of Sciences* **2012**, *109* (40), 16155-16160.
257. Camilloni, C.; De Simone, A.; Vranken, W. F.; Vendruscolo, M., Determination of secondary structure populations in disordered states of proteins using nuclear magnetic resonance chemical shifts. *Biochemistry* **2012**, *51* (11), 2224-31.
258. Bezsonova, I.; Singer, A.; Choy, W.-Y.; Tollinger, M.; Forman-Kay, J. D., Structural Comparison of the Unstable drkN SH3 Domain and a Stable Mutant. *Biochemistry* **2005**, *44* (47), 15550-15560.
259. Marsh, J. A.; Forman-Kay, J. D., Structure and Disorder in an Unfolded State under Nondenaturing Conditions from Ensemble Models Consistent with a Large Number of Experimental Restraints. *Journal of Molecular Biology* **2009**, *391* (2), 359-374.
260. Jensen, M. R.; Communie, G.; Ribeiro, E. A.; Martinez, N.; Desfosses, A.; Salmon, L.; Mollica, L.; Gabel, F.; Jamin, M.; Longhi, S.; Ruigrok, R. W. H.; Blackledge, M., Intrinsic disorder in measles virus nucleocapsids. *Proceedings of the National Academy of Sciences* **2011**, *108* (24), 9839-9844.
261. Sterckx, Yann G. J.; Volkov, Alexander N.; Vranken, Wim F.; Kragelj, J.; Jensen, Malene R.; Buts, L.; Garcia-Pino, A.; Jové, T.; Van Melderen, L.; Blackledge, M.; van Nuland, Nico A. J.; Loris, R., Small-Angle X-Ray Scattering- and Nuclear Magnetic Resonance-Derived Conformational Ensemble of the Highly Flexible Antitoxin PaaA2. *Structure* **2014**, *22* (6), 854-865.
262. Mittag, T.; Marsh, J.; Grishaev, A.; Orlicky, S.; Lin, H.; Sicheri, F.; Tyers, M.; Forman-Kay, J. D., Structure/Function Implications in a Dynamic Complex of the Intrinsically Disordered Sic1 with the Cdc4 Subunit of an SCF Ubiquitin Ligase. *Structure* **2010**, *18* (4), 494-506.
263. Clayton, D. F.; George, J. M., Synucleins in synaptic plasticity and neurodegenerative disorders. *Journal of Neuroscience Research* **1999**, *58* (1), 120-129.
264. Dikiy, I.; Eliezer, D., Folding and misfolding of alpha-synuclein on membranes. *Biochim Biophys Acta* **2012**, *1818* (4), 1013-1018.

265. Klunk, W. E.; Engler, H.; Nordberg, A.; Wang, Y. M.; Blomqvist, G.; Holt, D. P.; Bergstrom, M.; Savitcheva, I.; Huang, G. F.; Estrada, S.; Ausen, B.; Debnath, M. L.; Barletta, J.; Price, J. C.; Sandell, J.; Lopresti, B. J.; Wall, A.; Koivisto, P.; Antoni, G.; Mathis, C. A.; Langstrom, B., Imaging brain amyloid in Alzheimer's disease with Pittsburgh Compound-B. *Ann Neurol* **2004**, *55* (3), 306-319.
266. Wong, D. F.; Rosenberg, P. B.; Zhou, Y.; Kumar, A.; Raymont, V.; Ravert, H. T.; Dannals, R. F.; Nandi, A.; Brasic, J. R.; Ye, W.; Hilton, J.; Lyketsos, C.; Kung, H. F.; Joshi, A. D.; Skovronsky, D. M.; Pontecorvo, M. J., In Vivo Imaging of Amyloid Deposition in Alzheimer Disease Using the Radioligand F-18-AV-45 (Flbetapir F 18). *Journal of Nuclear Medicine* **2010**, *51* (6), 913-920.
267. Rowe, C. C.; Ackerman, U.; Browne, W.; Mulligan, R.; Pike, K. L.; O'Keefe, G.; Tochon-Danguy, H.; Chan, G.; Berlangieri, S. U.; Jones, G.; Dickinson-Rowe, K. L.; Kung, H. P.; Zhang, W.; Kung, M. P.; Skovronsky, D.; Dyrks, T.; Hall, G.; Krause, S.; Friebe, M.; Lehman, L.; Lindemann, S.; Dinkelborg, L. M.; Masters, C. L.; Villemagne, V. L., Imaging of amyloid beta in Alzheimer's disease with F-18-BAY94-9172, a novel PET tracer: proof of mechanism. *Lancet Neurology* **2008**, *7* (2), 129-135.
268. Kudo, Y.; Okamura, N.; Furumoto, S.; Tashiro, M.; Furukawa, K.; Maruyama, M.; Itoh, M.; Iwata, R.; Yanai, K.; Arai, H., 2-(2-(2-dimethylaminothiazol-5-yl ethenyl)-6-(2-fluoro ethoxy)benzoxazole: A novel PET agent for in vivo detection of dense amyloid plaques in Alzheimer's disease patients. *Journal of Nuclear Medicine* **2007**, *48* (4), 553-561.
269. Chien, D. T.; Bahri, S.; Szardenings, A. K.; Walsh, J. C.; Mu, F.; Su, M.-Y.; Shankle, W. R.; Elizarov, A.; Kolb, H. C., Early Clinical PET Imaging Results with the Novel PHF-Tau Radioligand F-18 -T807. *Journal of Alzheimers Disease* **2013**, *34* (2), 457-468.
270. Lockhart, S. N.; Baker, S. L.; Okamura, N.; Furukawa, K.; Ishiki, A.; Furumoto, S.; Tashiro, M.; Yanai, K.; Arai, H.; Kudo, Y.; Harada, R.; Tomita, N.; Hiraoka, K.; Watanuki, S.; Jagust, W. J., Dynamic PET Measures of Tau Accumulation in Cognitively Normal Older Adults and Alzheimer's Disease Patients Measured Using 18F THK-5351. *Plos One* **2016**, *11* (6).
271. Maruyama, M.; Shimada, H.; Sahara, T.; Shinotoh, H.; Ji, B.; Maeda, J.; Zhang, M.-R.; Trojanowski, J. Q.; Lee, V. M. Y.; Ono, M.; Masamoto, K.; Takano, H.; Sahara, N.; Iwata, N.; Okamura, N.; Furumoto, S.; Kudo, Y.; Chang, Q.; Saido, T. C.; Takashima, A.; Lewis, J.; Jang, M.-K.; Aoki, I.; Ito, H.; Higuchi, M., Imaging of Tau Pathology in a Tauopathy Mouse Model and in Alzheimer Patients Compared to Normal Controls. *Neuron* **2013**, *79* (6), 1094-1108.
272. Jack, C. R., Jr.; Barrio, J. R.; Kepe, V., Cerebral amyloid PET imaging in Alzheimer's disease. *Acta Neuropathol* **2013**, *126* (5), 643-657.
273. Muntane, G.; Dalfo, E.; Martinez, A.; Ferrer, I., Phosphorylation of tau and alpha-synuclein in synaptic-enriched fractions of the frontal cortex in Alzheimer's disease, and in Parkinson's disease and related alpha-synucleinopathies. *Neuroscience* **2008**, *152* (4), 913-923.
274. Chu, W.; Zhou, D.; Gaba, V.; Liu, J.; Li, S.; Peng, X.; Xu, J.; Dhavale, D.; Bagchi, D. P.; d'Avignon, A.; Shakerdige, N. B.; Bacskai, B. J.; Tu, Z.; Kotzbauer,

- P. T.; Mach, R. H., Design, Synthesis, and Characterization of 3-(Benzylidene)indolin-2-one Derivatives as Ligands for alpha-Synuclein Fibrils. *Journal of Medicinal Chemistry* **2015**, *58* (15), 6002-6017.
275. Zhang, X.; Jin, H.; Padakanti, P. K.; Li, J.; Yang, H.; Fan, J.; Mach, R. H.; Kotzbauer, P.; Tu, Z., Radiosynthesis and in Vivo Evaluation of Two PET Radioligands for Imaging alpha-Synuclein. *Applied Sciences-Basel* **2014**, *4* (1), 66-78.
276. Hsieh, C.-J.; Ferrie, J. J.; Xu, K.; Lee, I.; Graham, T. J. A.; Tu, Z.; Yu, J.; Dhavale, D.; Kotzbauer, P.; Petersson, E. J.; Mach, R. H., Alpha Synuclein Fibrils Contain Multiple Binding Sites for Small Molecules. *ACS Chemical Neuroscience* **2018**, *9* (11), 2521-2527.
277. Johnson, D. K.; Karanicolas, J., Ultra-High-Throughput Structure-Based Virtual Screening for Small-Molecule Inhibitors of Protein-Protein Interactions. *J Chem Inf Model* **2016**, *56* (2), 399-411.
278. Sterling, T.; Irwin, J. J., ZINC 15--Ligand Discovery for Everyone. *Journal of chemical information and modeling* **2015**, *55* (11), 2324-2337.
279. Taminau, J.; Thijs, G.; De Winter, H., Pharao: pharmacophore alignment and optimization. *Journal of molecular graphics & modelling* **2008**, *27* (2), 161-9.
280. Xiao, Y.; Ma, B.; McElheny, D.; Parthasarathy, S.; Long, F.; Hoshi, M.; Nussinov, R.; Ishii, Y., A β (1–42) fibril structure illuminates self-recognition and replication of amyloid in Alzheimer's disease. *Nature structural & molecular biology* **2015**, *22* (6), 499.
281. Kuipers, B. J.; Gruppen, H., Prediction of molar extinction coefficients of proteins and peptides using UV absorption of the constituent amino acids at 214 nm to enable quantitative reverse phase high-performance liquid chromatography– mass spectrometry analysis. *Journal of agricultural and food chemistry* **2007**, *55* (14), 5445-5451.
282. Puchades, M.; Westman, A.; Blennow, K.; Davidsson, P., Removal of sodium dodecyl sulfate from protein samples prior to matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Communications in Mass Spectrometry* **1999**, *13* (5), 344-349.
283. Johnson, D. K.; Karanicolas, J., Druggable Protein Interaction Sites Are More Predisposed to Surface Pocket Formation than the Rest of the Protein Surface. *PLOS Computational Biology* **2013**, *9* (3), e1002951.
284. Hsieh, C.-J.; Xu, K.; Lee, I.; Graham, T. J. A.; Tu, Z.; Dhavale, D.; Kotzbauer, P.; Mach, R. H., Chalcones and Five-Membered Heterocyclic Isosteres Bind to Alpha Synuclein Fibrils in Vitro. *Acs Omega* **2018**, *3* (4), 4486-4493.
285. Kiss, R.; Sandor, M.; Szalai, F. A., <http://Mcule.com>: a public web service for drug discovery. *Journal of Cheminformatics* **2012**, *4* (1), P17.
286. Daniels, M. J.; Nourse, J. B.; Kim, H.; Sainati, V.; Schiavina, M.; Murralli, M. G.; Pan, B.; Ferrie, J. J.; Haney, C. M.; Moons, R.; Gould, N. S.; Natalello, A.; Grandori, R.; Sobott, F.; Petersson, E. J.; Rhoades, E.; Pierattelli, R.; Felli, I.; Uversky, V. N.; Caldwell, K. A.; Caldwell, G. A.; Krol, E. S.; Ischiropoulos, H., Cyclized NDGA modifies dynamic α -synuclein monomers preventing aggregation and toxicity. *Scientific Reports* **2019**, *9* (1), 2937.

287. Pujols, J.; Peña-Díaz, S.; Lázaro, D. F.; Peccati, F.; Pinheiro, F.; González, D.; Carija, A.; Navarro, S.; Conde-Giménez, M.; García, J.; Guardiola, S.; Giralt, E.; Salvatella, X.; Sancho, J.; Sodupe, M.; Outeiro, T. F.; Dalfó, E.; Ventura, S., Small molecule inhibits α -synuclein aggregation, disrupts amyloid fibrils, and prevents degeneration of dopaminergic neurons. *Proceedings of the National Academy of Sciences* **2018**, *115* (41), 10481-10486.
288. Alam, P.; Bousset, L.; Melki, R.; Otzen, D. E., α -synuclein oligomers and fibrils: a spectrum of species, a spectrum of toxicities. *Journal of Neurochemistry* **2019**, *150* (5), 522-534.
289. Rott, R.; Szargel, R.; Shani, V.; Bisharat, S.; Engelender, S., alpha-Synuclein ubiquitination and novel therapeutic targets for Parkinson's disease. *CNS & neurological disorders drug targets* **2014**, *13* (4), 630-7.
290. Rott, R.; Szargel, R.; Shani, V.; Hamza, H.; Savyon, M.; Abd Elghani, F.; Bandopadhyay, R.; Engelender, S., SUMOylation and ubiquitination reciprocally regulate α -synuclein degradation and pathological aggregation. *Proceedings of the National Academy of Sciences* **2017**, *114* (50), 13176-13181.
291. Tofaris, G. K.; Razaq, A.; Ghetti, B.; Lilley, K. S.; Spillantini, M. G., Ubiquitination of alpha-synuclein in Lewy bodies is a pathological event not associated with impairment of proteasome function. *The Journal of biological chemistry* **2003**, *278* (45), 44405-11.
292. Pujols, J.; Pena-Diaz, S.; Conde-Gimenez, M.; Pinheiro, F.; Navarro, S.; Sancho, J.; Ventura, S., High-Throughput Screening Methodology to Identify Alpha-Synuclein Aggregation Inhibitors. *International journal of molecular sciences* **2017**, *18* (3).
293. Kurnik, M.; Sahin, C.; Andersen, C. B.; Lorenzen, N.; Giehm, L.; Mohammad-Beigi, H.; Jessen, C. M.; Pedersen, J. S.; Christiansen, G.; Petersen, S. V.; Staal, R.; Krishnamurthy, G.; Pitts, K.; Reinhart, P. H.; Mulder, F. A. A.; Mente, S.; Hirst, W. D.; Otzen, D. E., Potent alpha-Synuclein Aggregation Inhibitors, Identified by High-Throughput Screening, Mainly Target the Monomeric State. *Cell chemical biology* **2018**, *25* (11), 1389-1402.e9.
294. Heller, G. T.; Aprile, F. A.; Vendruscolo, M., Methods of probing the interactions between small molecules and disordered proteins. *Cell Mol Life Sci* **2017**, *74* (17), 3225-3243.
295. Heller, G. T.; Bonomi, M.; Vendruscolo, M., Structural Ensemble Modulation upon Small-Molecule Binding to Disordered Proteins. *Journal of Molecular Biology* **2018**, *430* (16), 2288-2292.
296. Heller, G. T.; Sormanni, P.; Vendruscolo, M., Targeting disordered proteins with small molecules using entropy. *Trends in Biochemical Sciences* **2015**, *40* (9), 491-496.
297. Longhena, F.; Spano, P.; Bellucci, A., Targeting of Disordered Proteins by Small Molecules in Neurodegenerative Diseases. In *Targeting Trafficking in Drug Development*, Ulloa-Aguirre, A.; Tao, Y.-X., Eds. Springer International Publishing: Cham, 2018; pp 85-110.
298. Tóth, G.; Gardai, S. J.; Zago, W.; Bertocini, C. W.; Cremades, N.; Roy, S. L.; Tambe, M. A.; Rochet, J.-C.; Galvagnion, C.; Skibinski, G.; Finkbeiner, S.;

- Bova, M.; Regnstrom, K.; Chiou, S.-S.; Johnston, J.; Callaway, K.; Anderson, J. P.; Jobling, M. F.; Buell, A. K.; Yednock, T. A.; Knowles, T. P. J.; Vendruscolo, M.; Christodoulou, J.; Dobson, C. M.; Schenk, D.; McConlogue, L., Targeting the Intrinsically Disordered Structural Ensemble of α -Synuclein by Small Molecules as a Potential Therapeutic Strategy for Parkinson's Disease. *PLOS ONE* **2014**, *9* (2), e87133.
299. Bulawa, C. E.; Connelly, S.; Devit, M.; Wang, L.; Weigel, C.; Fleming, J. A.; Packman, J.; Powers, E. T.; Wiseman, R. L.; Foss, T. R.; Wilson, I. A.; Kelly, J. W.; Labaudinière, R., Tafamidis, a potent and selective transthyretin kinetic stabilizer that inhibits the amyloid cascade. *Proceedings of the National Academy of Sciences of the United States of America* **2012**, *109* (24), 9629-9634.
300. Thornburg, C. C.; Britt, J. R.; Evans, J. R.; Akee, R. K.; Whitt, J. A.; Trinh, S. K.; Harris, M. J.; Thompson, J. R.; Ewing, T. L.; Shipley, S. M.; Grothaus, P. G.; Newman, D. J.; Schneider, J. P.; Grkovic, T.; O'Keefe, B. R., NCI Program for Natural Product Discovery: A Publicly-Accessible Library of Natural Product Fractions for High-Throughput Screening. *ACS Chemical Biology* **2018**, *13* (9), 2484-2497.
301. Johnson, D. K.; Karanicolas, J., Selectivity by Small-Molecule Inhibitors of Protein Interactions Can Be Driven by Protein Surface Fluctuations. *PLOS Computational Biology* **2015**, *11* (2), e1004081.
302. Guerrero-Ferreira, R.; Taylor, N. M. I.; Arteni, A.-A.; Kumari, P.; Mona, D.; Ringler, P.; Britschgi, M.; Lauer, M. E.; Verasdock, J.; Riek, R.; Melki, R.; Meier, B. H.; Böckmann, A.; Bousset, L.; Stahlberg, H., Two new polymorphic structures of alpha-synuclein solved by cryo-electron microscopy. *bioRxiv* **2019**, 654582.
303. Vilar, M.; Chou, H.-T.; Luehrs, T.; Maji, S. K.; Riek-Loher, D.; Verel, R.; Manning, G.; Stahlberg, H.; Riek, R., The fold of alpha-synuclein fibrils. *Proceedings of the National Academy of Sciences of the United States of America* **2008**, *105* (25), 8637-8642.
304. Baggett, D. W.; Nath, A., The Rational Discovery of a Tau Aggregation Inhibitor. *Biochemistry* **2018**, *57* (42), 6099-6107.
305. Paiva, S. L.; Crews, C. M., Targeted protein degradation: elements of PROTAC design. *Curr Opin Chem Biol* **2019**, *50*, 111-119.