

Examination of Amyloid Structures *in vitro* and *ex vivo*
with 2D IR Spectroscopy

By
Tianqi Zhang

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY
(CHEMISTRY)

at the
UNIVERSITY OF WISCONSIN - MADISON
2016

Date of final oral examination: November 22, 2016 at 1:00 PM

This dissertation is approved by the following members of the Final Oral Committee:

Martin T. Zanni, Professor, Chemistry

Qiang Cui, Professor, Chemistry

Eric R. Strieter, Assistant Professor, Chemistry

James C. Weisshaar, Professor, Chemistry

Randall H. Goldsmith, Assistant Professor, Chemistry

Abstract

Two-dimensional infrared (2D IR) spectroscopy is a powerful tool for studying biomolecules, especially amyloid forming proteins. The goal of my research is to understand the formation of cataracts at the molecular level for both individual proteins and in lens tissues. The enhanced spectral resolution and unique cross-peak features of 2D IR provide markers to probe and measure the protein structures that are difficult to examine otherwise. The spectroscopy is enhanced with biochemical methods that are also an integral part of this research. During my time in the Zanni group, I studied isotope labeled proteins as well as lens tissues. The aggregation and amyloid formation of segmentally isotope-labeled γ D-crystallin was examined when treated with acid, heat, or UV-irradiation. To investigate if all lens crystallin proteins forms amyloid structures in these conditions, I initiated the research on a different crystallin protein, α B-crystallin. α B-crystallin forms molecular chaperon that might bind to misfolded or unfolded proteins. The results of my study shows that α B-crystallin readily forms amyloid fibers when treated with acid or heat but only partially unfolds when irradiated with UV light, a known risk factor for age-related cataracts. Thus, the hypothesis of α B-crystallin protecting the lens from forming cataracts sounds plausible. With this hypothesis in mind, I studied the interaction between uniformly ^{13}C labeled γ D-crystallin and unlabeled α B-crystallin with 2D IR spectroscopy. With isotope labels the two kinds of proteins can be spectrally resolved although they are in the same solution. This is one of the first attempts to study protein-protein interaction with 2D IR spectroscopy. All of these above studies of crystallin proteins *in vitro* build a foundation for investigation of a more complicated system – the *ex vivo* lens tissues. I examined the ocular lens tissues to study crystallin protein structure directly in the lens. This study validated previous research of crystallin proteins *in vitro*

with the first reported 2D IR measurement of tissue samples *in situ*. We observed amyloid structure in UV-irradiated lens tissues with 2D IR spectroscopy and verified the observation with other biophysical methods. Our results lead us to conclude that age-related cataracts are an amyloid disease. We hypothesize that the cataract onset is delayed by α -crystallin proteins in the lens. These findings from our study of cataracts with 2D IR spectroscopy establishes that 2D IR spectroscopy is a promising tool for tackling complicated biological questions.

Table of Contents

ABSTRACT	I
TABLE OF CONTENTS	III
LIST OF FIGURES	VII
LIST OF TABLES	IX
ACKNOWLEDGEMENTS	X
1 CHAPTER 1	1
INTRODUCTION	1
1.1 Lens and cataracts	1
1.1.1 <i>Sight</i>	1
1.1.2 <i>Cataracts and cataract surgeries</i>	2
1.1.3 <i>Cataracts is a protein aggregation disease</i>	4
1.1.4 <i>Crystallin proteins</i>	4
1.2 2D IR spectroscopy	6
1.2.1 <i>Linear IR versus 2D IR</i>	7
1.2.2 <i>2D IR spectrum</i>	8
1.2.3 <i>Coupling and cross peaks</i>	8
1.3 Study cataract formation with 2D IR spectroscopy	11
1.3.1 <i>γD-crystallin</i>	11
1.3.2 <i>αB-crystallin</i>	12
1.3.3 <i>Lens tissues</i>	13
1.4 2D IR application in other peptides	13
1.4.1 <i>Human amylin</i>	14
1.4.2 <i>Transmembrane peptides</i>	15
1.5 Summary and outline	15
1.6 References	18
2 CHAPTER 2	24
EXPERIMENTAL METHODS	24
2.1 Introduction.....	24
2.2 Reagents and Buffers	24
2.3 Biochemistry protocols	29
2.3.1 <i>PCR and DNA gel</i>	29
2.3.2 <i>Sequencing</i>	33
2.3.3 <i>Gene and plasmid construction</i>	34
2.3.4 <i>Making Electrocompetent cells and heat-shock-competent cells</i>	36
2.3.5 <i>Electroporation</i>	38
2.3.6 <i>SDS-PAGE</i>	40
2.3.7 <i>Point mutation using a commercial kit</i>	47
2.3.8 <i>Addition of long DNA pieces into a plasmid</i>	52
2.3.9 <i>Thioflavin T (ThT) fluorescence plate reader</i>	58
2.3.10 <i>ThT fluorescence of tissue samples</i>	66
2.3.11 <i>Columns</i>	66
2.4 Expression and purification of peptide and protein	78

2.4.1	<i>Expression and purification of γD-crystallin</i>	78
2.4.2	<i>Expression and purification of γD-crystallin CTD and NTD</i>	82
2.4.3	<i>Native chemical ligation</i>	84
2.4.4	<i>Expression and purification of αB-crystallin</i>	87
2.4.5	<i>Expression and purification of hIAPP and its variations using Chitin binding domain</i> ..	90
2.4.6	<i>Expression and purification of hIAPP and its variations using Histag</i>	94
2.5	Manual synthesis of isotope labeled peptides using Fmoc chemistry	99
2.5.1	<i>Manual SPPS without microwave</i>	100
2.5.2	<i>Manual SPPS with microwave</i>	102
2.6	Sample storage and preparation for 2D IR spectroscopy measurements	103
2.6.1	<i>Peptides and proteins in solution</i>	103
2.6.2	<i>Preparation and measurement of transmembrane peptides</i>	105
2.7	Sample preparation for TEM samples	107
2.8	References	108
3	CHAPTER 3.....	109
	AMYLOID FORMATION IN HUMAN γD-CRYSTALLIN INDUCED BY UV-B PHOTODAMAGE.....	109
3.1	Abstract	109
3.2	Introduction	110
3.3	Material and Methods	112
3.3.1	<i>Preparation of native and denatured γD-crystallin samples</i>	112
3.3.2	<i>Sample preparation for mass spectrometry</i>	113
3.3.3	<i>Trypsin digestion of γD-crystallin</i>	113
3.3.4	<i>Sample preparation for transmission electron microscopy</i>	113
3.3.5	<i>Thioflavin T assay of UV-irradiated γD-crystallin</i>	114
3.3.6	<i>Sample preparation and data collection for 2D IR spectroscopy</i>	114
3.4	Results	114
3.4.1	<i>UV-vis spectra of γD-crystallin before and after UV-B irradiation</i>	116
3.4.2	<i>UV-B irradiation cleavages the backbone of γD-crystallin</i>	116
3.4.3	<i>Mass spectrometry of UV-B irradiated γD-crystallin identifies semitryptic fragments</i> ..	118
3.4.4	<i>Structural analysis of UV-B irradiated γD-crystallin</i>	123
3.5	Discussion	130
3.6	Conclusion	138
3.7	References	139
4	CHAPTER 4.....	146
	AN ALTERNATIVE STRUCTURAL ISOFORM IN AMYLOID-LIKE AGGREGATES FORMED FROM THERMALLY DENATURED HUMAN γD-CRYSTALLIN	146
4.1	Abstract	146
4.2	Introduction	147
4.3	Material and Methods	150
4.4	Results and Discussion	151
4.5	Conclusions	162
4.6	Acknowledgements	165
4.7	References	166
5	CHAPTER 5.....	172
	AMYLOID IDENTIFIED IN UV-INDUCED CATARACTS OF PORCINE LENSES USING 2D IR SPECTROSCOPY	172

5.1	Abstract	172
5.2	Introduction	173
5.3	Material and Methods	175
5.4	Results	177
5.4.1	<i>Signature of amyloid formation in 2D IR spectra of in vitro proteins.....</i>	<i>177</i>
5.4.2	<i>2D IR spectroscopy identifies highly-ordered β-sheets in αB-crystallin amyloid structures that are not easily classified with TEM.</i>	<i>181</i>
5.4.3	<i>TEM images of porcine lens tissue treated with acid and UV-irradiation.....</i>	<i>182</i>
5.4.4	<i>2D IR spectra reveal amyloid fibers in sliced porcine lens tissues upon acid denaturation and UV-irradiation.</i>	<i>184</i>
5.4.5	<i>2D IR is more sensitive to amyloid formation than FTIR spectroscopy.....</i>	<i>188</i>
5.4.6	<i>Thioflavin T (ThT) fluorescence essay of lens tissue is consistent with the formation of amyloid fiber after UV-irradiation.....</i>	<i>188</i>
5.5	Discussion	189
5.6	Conclusion	197
5.7	Acknowledgements	197
5.8	References	198
6	CHAPTER 6.....	202
	GXXXG-MEDIATED PARALLEL AND ANTIPARALLEL DIMERIZATION OF TRANSMEMBRANE HELICES AND ITS INHIBITION BY CHOLESTEROL: SINGLE-MOLECULE FRET AND 2D IR STUDIES	202
6.1	Abstract	202
6.2	Introduction	203
6.3	Result and Discussion	203
6.4	Conclusion	210
6.5	References	210
7	CHAPTER 7.....	213
	CONCLUSIONS AND FUTURE WORK.....	213
7.1	Introduction	213
7.2	α B-crystallin as a molecular chaperone	214
7.2.1	<i>Chaperone activity of αB-crystallin towards γD-crystallin.....</i>	<i>214</i>
7.2.2	<i>Inhibition of hIAPP fiber elongation by αB-crystallin chaperone</i>	<i>219</i>
7.2.3	<i>αB-crystallin chaperone dynamics.....</i>	<i>221</i>
7.3	Fragments of hIAPP	222
7.3.1	<i>FGAIL fragment.....</i>	<i>222</i>
7.3.2	<i>CFGAILSS fragment</i>	<i>223</i>
7.3.3	<i>NNFGAILSS fragment.....</i>	<i>223</i>
7.4	Small peptide A6K.....	225
7.5	Apolipoprotein A1	226
7.6	References	228
A1.	APPENDIX 1.....	230
	SUPPORTING INFORMATION FOR CHAPTER 3: AMYLOID FIBER FORMATION IN HUMAN γD-CRYSTALLIN INDUCED BY UV-B PHOTODAMAGE.....	230
A2.	APPENDIX 2.....	235

SUPPORTING INFORMATION FOR CHAPTER 5: AMYLOID IDENTIFIED IN UV-INDUCED CATARACTS OF PORCINE LENSES USING 2D IR SPECTROSCOPY	235
A2.1 Supporting Material and Methods	235
A2.1.1 Purification of recombinant γ D-Crystallin.....	235
A2.1.2 Purification of recombinant α B-Crystallin	235
A2.1.3 Slices of porcine lenses	236
A2.1.4 2D IR spectra of protein solutions	236
A2.1.5 2D IR spectra of lens slices.....	236
A2.1.6 FTIR spectra of lens slices	237
A2.1.7 UV-irradiation of protein solution and lens tissues	237
A2.1.8 Acid treatment of protein solution and lens tissues.....	237
A2.1.9 ThT fluorescence essay of lens slices	237
A2.1.10 TEM of lens tissues and protein solutions.....	238
A2.2 Calculation of UV dosage from UV lamp in lab and its equivalent to the average sun exposure per person.	238
A2.3 Estimation of percent amyloid β -sheet in acid and UV treated lens tissues.	239
A2.4 Supporting Figures and legends.....	241
A2.5 Supporting Reference.....	244
A3. APPENDIX 3.....	246
SUPPORTING INFORMATION FOR CHAPTER 6: GXXXG-MEDIATED PARALLEL AND ANTIPARALLEL DIMERIZATION OF TRANSMEMBRANE HELICES AND ITS INHIBITION BY CHOLESTEROL: SINGLE-MOLECULE FRET AND 2D IR STUDIES	246
A3.1 Supporting Material and Methods	246
A3.1.1 Peptide Synthesis.....	246
A3.1.2 Single-Molecule FRET.....	248
A3.1.3 FTIR.....	249
A3.1.4 2D IR.....	250
A3.1.5 2D IR Simulation.....	251
A3.2 Supporting Discussion	251
A3.3 Supporting Figures.....	251
A3.4 Supporting References	256
A4. APPENDIX 4.....	257
PLASMIDS, GENES AND PEPTIDES	257
A4.1 Plasmids and genes	258
A4.1.1 γ D-crystallin.....	258
A4.1.2 α B-crystallin.....	264
A4.1.3 hIAPP.....	266
A4.2 Peptides.....	273
A5. APPENDIX 5.....	274
LIST OF PUBLICATIONS.....	274

List of Figures

Figure 1.1 Carton illustration of 2D IR spectroscopy.....	10
Figure 1.2 2D IR spectra of human amylin solution.....	10
Figure 2.1 Setup of DNA agarose gel.....	32
Figure 2.2 Electroporator.....	38
Figure 2.3 Parts of gel casting setup.....	42
Figure 2.4 Picture illustration of gel casting setup.....	43
Figure 2.5 DNA agarose gel illuminated with UV lamp.....	57
Figure 2.6 Screen print illustration of ThT plate reader program.....	59
Figure 2.7 Screen print illustration of plate reader parameters.....	61
Figure 2.8 Setup of Assays and Groups.....	63
Figure 2.9 An example of ThT plate reader result.....	64
Figure 2.10 Kinetics plots of UV-irradiation induced γ D-crystallin amyloid fiber formation.....	65
Figure 2.11 Size Exclusion Chromatograph of γ D-crystallin.....	79
Figure 2.12 HPLC retention time trace of γ D-crystallin.....	80
Figure 2.13 MALDI spectrum of γ D-crystallin.....	81
Figure 2.14 Native chemical ligation schematic for γ D-crystallin.....	86
Figure 2.15 Size Exclusion Chromatograph of α B-crystallin.....	88
Figure 2.16 MALDI spectrum of α B-crystallin.....	89
Figure 2.17 HPLC trace of hIAPP L12P V32P purification.....	92
Figure 2.18 MALDI spectrum of ^{13}C - ^{15}N labeled L12P V32P hIAPP.....	93
Figure 2.19 MALDI spectrum of ^{13}C - ^{15}N labeled intein.....	97
Figure 2.20 MALDI spectrum of ^{13}C - ^{15}N labeled L12P V32P hIAPP with Histag.....	98
Figure 2.21 Manual peptide synthesis setup.....	101
Figure 3.1 UV-B photodamage products of human γ D-Crystallin.....	115
Figure 3.2 Representative CID spectra of tryptic and semitryptic peptides.....	119
Figure 3.3 Figure illustration of oxidative cleavage sites on γ D-Crystallin.....	122
Figure 3.4 Aggregates of human γ D-Crystallin.....	123
Figure 3.5 2D IR spectra of γ D-Crystallin variants.....	125
Figure 4.1 Characterization of thermally induced γ D-crystallin aggregates.....	152
Figure 4.2 2D IR spectra of γ D-crystallin amyloid aggregates.....	154
Figure 4.3 Isotope dilution of γ D-crystallin aggregates.....	155
Figure 4.4 2D IR spectra of segmentally labeled γ D-crystallin aggregates.....	158
Figure 4.5 Structural models for the thermally induced aggregation of γ D-crystallin.....	161
Figure 5.1 Cartoon illustration of amide I modes of (amyloid) β -sheets.....	174
Figure 5.2 2D IR spectra of γ D- and α B-crystallin proteins in deuterated buffer.....	178
Figure 5.3 2D IR spectra and TEM images of aggregated α B-crystallin.....	182
Figure 5.4 TEM images of porcine lens tissues upon acid and UV treatment.....	184
Figure 5.5 2D IR spectra of porcine lens tissues upon acid and UV treatment.....	185
Figure 5.6 FTIR and 2D IR diagonal slices of porcine lens tissues.....	187
Figure 5.7 ThT fluorescence histogram.....	189
Figure 6.1 Cartoon illustration of dimer formation.....	204
Figure 6.2 GXXXG-mediated association of transmembrane helices (25°C).....	206
Figure 6.3 Angle orientation and peakshift of isotope labeled residues.....	207
Figure 6.4 Kinetics of the monomer-dimer transition of GXXXG helices in POPC.....	209
Figure 7.1 2D IR spectra of aggregated γ D-crystallin and native α B-crystallin.....	216
Figure 7.2 2D IR spectra of aggregated γ D-crystallin, native α B-crystallin, and their subtraction result.....	218
Figure 7.3 Kinetics of hIAPP aggregation with and without α B-crystallin.....	220

Figure 7.4 Waiting time 2D IR spectra of aggregated hIAPP and α B-crystallin.....	221
Figure 7.5 2D IR spectrum of amidated FGAIL fragment in D ₂ O buffer.....	222
Figure 7.6 2D IR spectrum of FGAIL fragment in D ₂ O buffer	222
Figure 7.7 2D IR spectrum of CFGAILSS fragment in D ₂ O buffer	223
Figure 7.8 2D IR spectrum of NNFGAILSS fragment in D ₂ O buffer	224
Figure 7.9 Kinetics of NNFGAILSS aggregation.....	224
Figure 7.10 2D IR spectra of A ₆ K in aqueous buffers.	225
Figure 7.11 Crystal structure of Apolipoprotein A1 tetramer (PDB: 1AVA).....	226
Figure 7.12 2D IR spectrum of native ApoA-1	227
Figure A2. 1 2D IR spectra of UV-irradiated α B-crystallin proteins in D ₂ O buffer.....	241
Figure A2. 2 Kinetics of amyloid formation in acid treated lens tissues.	242
Figure A2. 3 FWHM of α B-crystallin protein 2D IR diagonal slices upon UV-irradiation.	242
Figure A2. 4 Fitting of FTIR and 2D IR diagonal slice of acid treated lens tissue.....	243
Figure A2. 5 Pump slices of 2D IR spectra.....	244
Figure A3. 1 Sequences and HPLC chromatograms of synthetic crosslinked transmembrane dimers. ...	252
Figure A3. 2 Fluorescence quenching of dehydroergosterol.	253
Figure A3. 3 IR spectra of the isotope-edited GXXXG helix.....	254
Figure A3. 4 Predicted dimer structure.....	255
Figure A3. 5 Self-quenching of NBD-labeled helices.	255

List of Tables

Table 2.1 Table of PCR cycles.....	30
Table 2.2 Table of % agarose and corresponding DNA fragment sizes.	31
Table 2.3 Recipe for SDS-PAGE gel.....	44
Table 2.4 PCR program for plasmid amplification, mutagenesis.	51
Table 2.5 PCR program for plasmid amplification.....	56
Table 2.6 Number output of a ThT plate reader result.....	64
Table 2.7 Microwave setting for Fmoc chemistry solid phase synthesis.....	103
Table 3.1 Representative IR frequencies of β -sheet amide I modes.	127
Table 6.1 Energetic contribution of GXXXG motif at 25°C	208
Table A1. 1.....	231
Table A1. 2.....	232
Table A1. 3.....	233
Table A1. 4.....	233
Table A1. 5.....	234
Table A1. 6.....	234

Acknowledgements

I want to take this chance to thank all my teachers, professors and mentors from elementary school to graduate school. I have been so lucky to be your student. You always put students before yourself. I would not be who I am today without your selfless efforts to make me a better person.

I want to thank Qiang Cui and Eric Strieter for serving on my TBO and dissertation committees. Your kind feedback and advice supported me throughout graduate school. Your questions and suggestions during my TBO really helped me and Marty plan future directions for this project. I am grateful to Randy Goldsmith and Jim Weisshaar for serving on my dissertation committee. Randy has been a great supporter and friend over the years. I have received endless help from Jim's research group throughout the last 5 years.

I owe Professor Martin Zanni my gratitude for taking me into his research group 5 years ago. I had never seen an ultrafast laser before touring the labs with him. He showed me a whole new world of chemistry that is fascinating and exciting. Marty has been an amazing advisor to me. He is always passionate and encouraging, guiding me through difficult times during my research. I wish to be as caring and patient a person as he is one day. Also thanks to Marty, I got to know the Zanni group. The Zanni group is a wonderful academic family that I am so proud to be a part of. During my 5 years of graduate school, I had the privilege to work and interact with these excellent scientists and fun people. I will miss every one of those happy, sad, exciting, frustrating, sunny, and rainy days I spent with this family.

The postdocs of the Zanni group are magical and superman-like. Sean Moran taught me almost everything I know about biochemistry. He is an extremely talented and passionate biochemist with tons of brilliant ideas all the time. It is not at all an exaggeration to say that Sean

is an encyclopedia of biochemical knowledge. I thank Sean especially for mentoring me and being extremely patient and kind to me. Chris Middleton knows everything about lasers, from small details to building complex systems. He is also magic at programming! Maksim Grechko is like a special agent who is good at everything. He is always calm but also has a great sense of humor. I had so many great times working with him on protein expression and purification. I want to thank Arnaldo Serrano especially for opening the door to the world of lasers and optics to me. I learned so much from him as we spent days aligning the laser table together. He always encouraged me to be independent but also ensured that I will never be left in the dark if anything goes wrong. Ayanjeet Ghosh is awesome to work with and dance with. His comments are always satirically funny and to the point. Andy Jones is a super cool scientist who is also a talented athlete and outdoorsman. I want to thank him for saving a group of people on our bike trip, which would have ended poorly if he hadn't been there. Michal Maj is one of a kind. He is an extremely passionate scientist and always full of energy, the perfect kind of lab partner. I greatly enjoyed working with him on the amylin project. He inspired and encouraged me to think out of the box. Justin Lomont is always patient with my stupidest questions and comments. He is always positive, kind and available. I cannot articulate how much I appreciate the postdocs in the Zanni group. They have brought so much into the group and taught me so many different things.

Lauren Buchanan was my host during the graduate school visiting weekend and later became one of my mentors. Thank you Lauren for showing me the best of Madison and how to be a successful scientist. Ann Woys was my mentor during a 3 week rotation before join the Zanni group. Her enthusiasm for science and work ethic made all the hard work in graduate school seem easy and super interesting. I admire her adventurous spirit for life and her inexhaustible energy. Emily was also one of my mentors on lasers and peptide synthesis. She is a sweetheart and an

inspiring person who always brings delicious deserts to cheer us up. We had lots of good memories during conferences. Jennifer Laaser taught me so much during graduate school, not just lasers but also a variety of facts about life. I admire her for her ability to be extremely efficient in just about everything she does. I feel extraordinarily lucky to be surrounded by so many wonderful women, and I look up to all of them. They inspire and motivate me to be a better person and show me how wonderful people can be.

I admire Sudipta Mukherjee for working on some very difficult and complicated problems. He is always nice and willing to help no matter how stressful life can be. David Skoff gave me the warmest welcome when I joined the group. He always brightens up the gloomy days. I thank Ha Dong Gyun for being here in the lab during those nights I had to work late. I wish him the best luck in life. Randy Mehlenbacher is one of the smartest and nicest people I have ever known. He has been an important support for me throughout the years. I cherish all the good memories we had together. It was my honor to fight together with Huong Kratochvil during our last year of graduate school. I admire her passion for science and her never-procrastinating personality. Tracey Reitz has been a great friend and support inside and outside the lab. I thank her for all the generous help and love she gave me. I will miss her a lot but I know she might just show up at my door one day to bring more joy to my life.

I respect Jiajung Ho for his dedication to science and nerdiness. However, he is the coolest nerd I have ever known. Tom McDonough, you are the legend. It has been so much fun to run, to play badminton and to do pushups with you. Thank you for always being there. I admire Kacie Rich for her amazing social skills. You are also the best workout partner because you are super encouraging and inspiring. Nick Kearns, you are a beast Annie. Josh Ostrander, it becomes a habit of mine to wait for your morning invitation to get coffee. The awesome coffee trips make every

day lovely. I had great times with Jack O'Connor inside or outside of the laboratory. Your smile is like the sunshine. Best luck with pursuing your dreams! Jessi Flach, thank you for being an awesome office mate and gym mate, and putting up with me. I want to thank Kaarin Evens for the helpful discussions about organic chemistry. I also admire her for being a super cool camping geek. Megan Petti, I am so glad you joined the Zanni group. I am sure you are going to enjoy the next 5 years with this great family.

Ariel Alperstein, it was an honor and a pleasure working with you for the past two years. You are a great scientist who is conscientious and detail oriented. I appreciate your contribution and help to the cataract project and your comments on this dissertation. I also want to thank you for taking over my responsibilities and taking care of all the biochemistry legacies. I am looking forward to reading about the great works you are doing. During my time in the Zanni group, I also had the pleasure to work with some amazing undergraduate students. Igor Luzhansky, Elliot Eklund, Noah Richter and Nicholas Yan, thank you for teaching me how to be a mentor and for putting up with me. I am extremely proud of you and I wish you the best luck.

My time in graduate school was accompanied by many excellent collaborators across the world. I had a great time working with Dr. Yoshiaki Yano. I respect his hardworking ethics and his devotion to science. I always think about the time we went to the Madison farmer's market at 6:30 AM after working all night on the laser. Dr. Daniel Raleigh and his group have been so helpful when it comes to peptides. I thank them for their generous sharing of knowledge and resources. It has been very nice knowing and working with Gary Chan and Dr. Giorgio Cavigliolo. We had a fruitful collaboration even though we did not know each other for very long. I want especially thank Ellen Weiss and the Public Affairs Committee of the Biophysical Society. Thank you for

giving me the opportunity to serve on this committee and to talk to policy makers. I am honored to be part of this important committee.

The Department of Chemistry at UW-Madison is a very lovely place because of the wonderful people. I want to thank all the professors and students who gave me tremendous help and great advice. This dissertation would not be possible without you. The Weisshaar group, the Strieter group, the Blackwell group, the Burstyn group, the Cavagnero group, and the Goldsmith group have been so generous with their instruments, time, chemicals, knowledge, friendship, and kindness. I want to thank Sonisilpa Mohapatra and Mainak Mustafi for their help on the tissue fluorescence assay. I also thank Matt Stolt for his help with microscopy and Ambar Rana for sharing knowledge and materials. Many thanks to the Keck center teaching assistants Michelle Boursier and Michael Welsh for training and help with the instruments. I want to thank Kyle Czech, Paul Hebert, and Jonathan Handali for their help on various subjects.

Additionally, this department has so many great staff who go above and beyond in their jobs and make everything easier for the graduate students. I thank April Sonnentag, Kristi Heming, Cheri Stephens, and Susan Martin for their assistance throughout my time in graduate school. I want to tell Martha Vestling that I would not be able to do the characterization of the proteins and peptides without you. I love talking to you about my projects. I am grateful for having such an awesome mass spectrometry facility in the department. I thank Tracy Dryer for his super cool glass-blowing class and his generous help to make my crazy ideas come true. Bruce Goldade is always there when I need a poster or something printed. Michael Bradley knows everything about shipping and he is my Santa Claus all year around. Desiree Bates and Arrietta Clauss are putting in so much effort to make graduate school and looking for jobs friendlier to the students. I thank Arrietta personally for introducing the Intel recruiters to our department that benefited many of us.

I also thank Arrietta for her support and encouragement when I decide to start the consulting club for Ph.D students. Jeff Nielsen, thank you for always being there to make sure things get done when we needed repairs in the lab. I feel safe working in the building at any time because we have you as the building manager. I want to thank Chad Skemp for taking care of my orders and requests. Craig Trewartha and Bill Goebel fixed anything I brought to the electronic shop and always had the parts in stock when I needed them. I thank Steve Myers and James Mullarkey in the Machine Shop for fixing our chillers and pumps countless times, and Kendall Schneider for helping us with fixing all kinds of instruments. I thank Jim Zernicke for an always well maintained stock room that has basically anything I need.

The University of Wisconsin-Madison is a unique place because there are no boundaries between the departments. I received so much help from people outside the chemistry department. Randy Massey has been helping me with TEM since I was a first year student. Thank you for being so patient with me and my messy samples. Karla Knobel gave us tremendous help with the cryostat, which made our tissue works possible. Leandro Teixeira taught me the basics of eye structure and the dissection of eyeball. His support was very important to me when the lens project was started. Travis Tangen and the rest of WARF outreach program encouraged and helped me to present my research to the public. The WI Idea STEM Fellows program has been very meaningful to me.

Madison felt like a home to me the first time I met her. Now she is a home filled with friends. Thank you JJ for being a great friend. Thank you, the bike ride crew for the awesome and confidence-building trips. Thank you Jessi, Julia and Morgan for all the gym classes together. Thank you Heejun and Sudheer for all the grocery shopping parties. I will miss all these good times.

Friends are always friends no matter where they are. Dawn, thank you for letting me be part of such a beautiful wedding even though we have not seen each other for 16 years. Huaming, I hope you will never change and I will always love you like a brother. Thank you for being there for me every time when I am confused or frustrated. Cynthia and Danny, thank you so much for your support and help during the job search. I miss you lots. Maria, it is such an honor to practice cases with you. I wish you all the best.

I have the best families. My parents-in-law never hesitated for a second to take me as a part of the family. Thank you for giving me such a sweet family in the US. Thank you so much Mom and Dad. Caroline, you are the best sister-in-law. You have the ability to make everything better. Bill is awesome, although sometimes I wonder if he is an alien because he runs 100-mile races. Aunt Dale, thank you for being a great aunt who always cheers for us. Uncle Eric and Bruce, thank you for visiting us and please visit us more often. I want to thank my parents for their support and love throughout the years. They have been my closest friends and my best allies. They encourage to do anything I want but they might sue me for not going home often enough.

I am very lucky that I married my best friend. It is so perfect because he loves cooking and I love eating. Kevin Heylman, thank you for everything. You changed the track of my life and made me a better person. I cherish our past, am happy in our present, and dream for our future.

1 Chapter 1

Introduction

1.1 Lens and cataracts

1.1.1 Sight

There are five senses: sight, taste, hearing, touch and smell. These senses help us build a perception of the world around us. Among them, sight is often considered to be the most important because it allows us to appreciate the beauty that nature offers and navigate daily life. Humans have examined our sense of vision since Plato's time and many famous philosophers had their own theories of how sight works.¹ We now know the mechanics of vision is enabled through an elegant and delicate mechanism of tissues. However, that mechanism can fail on macroscopic to microscopic levels.

Vision is achieved in the eye by capturing and focusing the light onto the retina.² The mammalian eye works like a camera with two optical lenses, each with distinct functions. The first lens is the cornea, which accounts for around 80% of the focusing power of the eye.^{2,3} Despite its high focusing power, the cornea cannot change its focal point based on the distance of the object. Thus, with solely the cornea we would only be able to see things at a single distance from the eye. The pupil is like the aperture of a camera and controls the amount of light that reaches the next lens, the ocular lens. The ocular lens has muscles attached on both sides of it that pulls it to change its shape given the distance of the object. Thus, the eye is like an autofocus camera. The retina is the film where the focused images are captured.^{2,4} It sends the images as electrical impulses to the brain.⁵ To enable vision, all of the above parts have to be functional and work together.

Macroscopic damages, such as wounds and injuries, to any component of the eye can cause impaired vision or even blindness. However, the most common cause of blindness, cataracts, is not mechanical failure. Rather, it is microscopic damage that occurs on the molecular level. It is estimated that 39 million people are blind in the world, of which 56% are caused by cataracts.⁶ On a different scale, cataracts affect over 22 million people alone in the America and costs the country \$3.6 billion per year for treatments.⁷ The focus of this dissertation is to understand the formation of cataracts, this microscopic cause of blindness, on the molecular level.

1.1.2 Cataracts and cataract surgeries

Cataracts occur in the ocular lens. As discussed above, the ocular lens focuses light onto the retina. To do so, it has to be transparent as well as maintain the correct refractive index. These properties are achieved by relinquishing most of the organelles and creating a high protein concentration in the lens cells.⁸ Because the organelles are degraded during cell differentiation, the lens proteins are not regenerated, and so the proteins in the lens must have longer lifetimes than most other proteins in the body that are regenerated.⁹ Nevertheless, the lens proteins are not invulnerable. Environmental stresses can damage the lens proteins through oxidation, phosphorylation, cleavage, and other types of posttranslational modification.¹⁰ When lens proteins are damaged, they can form aggregates that accumulates as a white patch in the lens and blocks light from reaching the retina, thereby forming cataracts.¹⁰

The word cataract comes from the Latin word “cataracta” meaning waterfall, possibly referring to how the white deposit in the eye lens matches the whiteness from turbulence of air bubbles mixed in water.¹¹ This name reveals the extensive presence of cataracts in human history. In fact, cataract disease might be one of the most ancient diseases that has accompanied the history of human development. The treatment of cataracts also started early in human history. The first

documented cataract surgery was performed as astonishingly early as around 5th century BC.¹² This type of cataract surgery was called “couching” and was a rather barbarian approach, in which a matured and hardened cataract lens is knocked with a hammer so that it is detached from the muscles that are holding it. The lens then falls into the cavity of the eye ball. Vision can be restored but is blurry due to a 20% loss in focusing power.¹³ “Discission” is another surgical approach in which the cataract lens is cut and exposed to aqueous content that absorbs the cataract material. This approach leads to high risks of complications like glaucoma.¹² The technology of cataract surgery advanced quickly during the 1900s. In 1967, Dr. Charles Kelman revolutionized cataract surgery by using ultrasonic waves to break up cataractous material before it was extracted.¹⁴ This practice, usually referred to as *Phacoemulsification*, does not destroy the lens capsule and the incision can be much smaller. Additionally, unmaturing cataracts can also be operated on with this technique so that patients do not have to wait for cataracts to fully mature in their eyes before surgery.¹³ Phacoemulsification is a milestone of modern cataract surgery and makes cataract surgery commonplace.

In the present days, cataract surgery is one of the most common surgical procedures performed.¹⁵ In 2015, 20 million cataract surgeries were done worldwide.¹⁵ These advanced and widespread surgery techniques may let us believe that cataracts should no longer be a concern in our modern time. However, despite advanced technologies and the number of surgeries performed annually, the number of people who are affected or blind from cataracts continues to increase. This phenomenon is actually a sign of something positive – an extended life expectancy. Thanks to the development of modern medical science and vaccination, the average human life expectancy in 2010 was 67 years, more than doubled that of 1900’s 31 year expectancy.¹⁶ Increase in lifespan increases the prevalence of age-associated diseases. Along with type II diabetes and Alzheimer’s

disease, cataracts are categorized as age-associated diseases. With the upward trend in cataract diseases and the demand of cataract surgeries, there has been concern that cataract surgery, the only effective treatment for cataracts, cannot catch up with the increasing demand,^{17, 18} not to mention concerns about the costly medical expenses. Thus, finding an alternative cure or prevention method for cataracts is one important and urgent task nowadays. Since cataracts are caused by microscopic changes, a rudimentary understanding of cataract formation on the molecular level may provide insights into finding a nonsurgical treatment.

1.1.3 Cataracts is a protein aggregation disease

Originally, cataracts were thought to be deposits of calcium or another white compound in the eye.¹³ The connection between cataracts and protein aggregation was not made until the 1920s with the efforts of W. S. Duke-Elder,^{19, 20} when it became apparent that the white deposits in cataract lenses are nothing but proteins. Thereafter, the development and availability of spectroscopy and analytical instruments have brought tremendous advances to the study of cataracts. The major molecular compositions of lenses are identified.^{21, 22} The native structures of several crystallin proteins are determined by X-ray crystallography.²³ With the discovery of recombinant DNA and protein purification, each lens protein can be expressed and studied *in vitro* in isolation with other proteins. Meanwhile, several crystallin proteins are extracted from the water-insoluble part of the cataractous lenses.^{24, 25} The study of cataracts is now moving from macroscopic properties of tissue failure to the molecular level understanding of protein failure.

1.1.4 Crystallin proteins

On the molecular level, the ocular lens is made of 60% water and 38% proteins,²⁶ 90% of which are crystallin proteins.^{21, 27} These proteins are very stable compared to other kinds of proteins in the human body due to their composition and tight packing.²⁸ Their lifetime can be over

50 years while the other proteins in the human body have an average lifetime of 24 hours.²⁹ There are three families of crystallin proteins: α , β and γ -crystallins, all of which have been identified in the aggregates of cataract lenses.³⁰ While all three families of crystallin proteins function as structural proteins, α -crystallins can act as molecular chaperones that bind to other unfolded or misfolded proteins.³¹ There are two kinds of α -crystallins, α A and α B-crystallin. α A-crystallin is only found in the lens while α B-crystallin is found throughout the body, such as muscle tissues.³² Its expression level is sometimes linked to several types of cancers.^{33, 34} Although the chaperone activity of α -crystallin is largely unknown, the research of such molecular chaperones has very promising applications. It has been shown that α -crystallins can inhibit the aggregation of A β amyloid fibers (Alzheimer's disease) and some other peptides.³⁵⁻³⁷ Understanding the mechanism of α -crystallin chaperone function can provide insights into making protein and peptide therapies for protein aggregation diseases.

β and γ -crystallin proteins are mainly structural proteins with slight differences in their native conformations.³⁸ These crystallin proteins not only maintain the correct refractive index in lens but also absorb most of the UVB-irradiation hitting the eye so that the retina is shielded from the damage.² However, the energy from the absorbed radiation must go somewhere. It has been shown that γ D-crystallin may funnel the radiation energy through 4 tryptophan residues located on both of its domains into thermal energy to protect the proteins from UV-irradiation.^{39, 40} Meanwhile, the irradiation can damage γ D-crystallin, resulting in oxidized side chains or cleaved proteins that are prone to aggregation.^{39, 41, 42} *In vitro* experiments on UV-irradiated γ D-crystallin identify cleavage product kynurenine, a yellow compound, as a derivative of tryptophan due to oxidation and cleavage.^{43, 44} This compound is likely to be responsible for the pale yellow coloring often seen in cataractous lenses.

During the past half century, many breakthroughs have been made by the studies on crystallin proteins. Several specific posttranslational modifications were identified in the proteins purified from aged and cataractous lenses.^{25, 45-47} The aggregation pathways of various lens proteins have been examined in different conditions.^{27, 28, 48-55} Along with that, several risk factors of cataract formation have been recognized and investigated.^{10, 56} On one hand, most of the previous studies on the molecular level uses *in vitro* proteins. The aggregation mechanisms of *in vitro* proteins may or may not be the same as *in vivo* proteins due to the very different protein concentration and protein composition. Crystallin proteins have a much higher concentration in the lens than what is used in most of the *in vitro* studies.⁵⁷ In the lens, all three families of crystallin protein are present while usually solution with single kind of proteins is examined in *in vitro* studies. However, evidence that is needed to validate these *in vitro* results is lacking. On the other hand, lens tissues themselves are difficult to measure because they have high protein concentration.^{58, 59} Most of the tools mentioned above cannot be used for lens tissues without any further processing, such as purification or fixation. Some tools can be used to measure lens tissues directly, but it is difficult to extract clear and useful information from the results. Therefore, the exact mechanism of cataract formation is still largely unknown despite the combined efforts of all the methods and technologies mentioned above.

1.2 2D IR spectroscopy

In this dissertation I will introduce 2D IR spectroscopy as a new approach to study cataract formation through both *in vitro* crystallin proteins (Chapter 3 and 4) and *ex vivo* lens tissues (Chapter 5). IR spectroscopy probes the vibrational transitions in molecules. Some of these vibrational transitions, like the amide I mode (carbonyl bond on the protein backbone), are especially sensitive to the local environment and structure, and thus are good gauges for measuring

protein secondary structures.^{60, 61} Therefore, IR spectroscopy is a great tool to study protein conformational changes, such as folding, unfolding and aggregation. Largely based on absorption spectroscopy, IR spectroscopy can measure a wide range of samples including solids, liquids and gases. In biophysics, IR spectroscopy is often used to measure both *in vitro* protein solutions and *ex vivo* tissue samples without fixation or further purification.⁶² 2D IR is a nonlinear IR spectroscopy technique that measures the same transitions as standard IR spectroscopy but provides much more information about molecular structures and couplings. There are many articles^{60, 61, 63-65} and books⁶⁶⁻⁶⁸ that explain the concepts and principles of 2D IR spectroscopy. In this section, I focus on introducing some key features of 2D IR spectroscopy that are needed for application to biological samples.

1.2.1 Linear IR versus 2D IR

In linear IR spectroscopy, the intensity of light transmitted through a sample is measured to examine the amount of energy absorbed. Unlike linear IR spectroscopy, 2D IR is a nonlinear spectroscopy which uses three pulses to create the signal. The pulse sequence is shown in Figure 1.1A. A sample interacts with two pump pulses and a probe pulse some time later. For this reason, the signal intensity of linear IR scales to the 2nd power of the transition dipole, but scales to the 4th power in 2D IR. This higher scaling on the transition dipole makes 2D IR extremely sensitive to vibrational transitions with large transition dipoles, such as highly ordered β -sheets in amyloid fibers.⁶⁹ Because the residues in amyloid β -sheets are strongly coupled, a very large transition dipoles is formed along the fiber axis, making even a small amount of amyloid easy to detect with 2D IR spectroscopy.⁷⁰⁻⁷² Figure 1.1B shows the overlay of a traditional FTIR spectrum and a diagonal slice of 2D IR spectrum of the same sample. This sample of acid treated lens contains amyloid structures that only appears as a small shoulder around 1620 cm^{-1} in FTIR. The peak at

the same frequency is much more prominent in 2D IR. More examples of enhanced spectra resolution in 2D IR spectra are discussed in Chapter 5.

1.2.2 2D IR spectrum

2D IR spectroscopy can generate more than a diagonal slice. A standard output of 2D IR is a two-dimensional frequency-frequency correlation spectrum that provides much more information about molecular dynamics and coupling than a linear IR spectra. As we mentioned above, molecules interact with light 3 times in 2D IR spectroscopy. Assume the vibrational state of all the molecules is $v=0$ before excitation. The first two pump pulses create a population at $v=1$. When the sample is probed with a third pulse, two events can happen. Some molecules are excited to the $v=2$ level (excited state absorption) while the rest of the molecules go back to the ground state $v=0$ through stimulated emission (Figure 1.1C). As color coded in Figure 1.1 C and D, stimulated emission results in the blue peak on the diagonal (Figure 1.1D) and the excitation corresponds to the red peak at a lower frequency. Thus, 2D IR peaks appear in pairs that are opposite in sign. The frequency difference between the blue peak (fundamental peak) and the red peak (the overtone) is a result of anharmonicity of the Morse potential as shown in Figure 1.1C.

1.2.3 Coupling and cross peaks

Note that on Figure 1.1D there is more than one pair of peaks on the diagonal. This means there are two vibrational modes in the spectral range that we are probing. Two vibrational modes can be coupled, which means the frequency of one vibration is correlated to that of the other one. The most straightforward indication of coupling is the appearance of pairs of peaks off-diagonal, the cross peaks (as shown in Figure 1.1D). For the cross peaks below the diagonal in Figure 1.1D, the fundamental peak (blue) has the same pump frequency as the lower transition (ω_1) on diagonal and the same probe frequency as the higher transition (ω_2) on diagonal. We can think of it as some

vibrational transitions that are excited at the lower pump frequency ω_1 (lower energy level) gain energy from the environment. Thus, when the molecules are probed, the transitions appear at a higher frequency ω_2 (higher energy level).

Cross peaks are an important feature in 2D IR spectroscopy because they are a clear indication of coupling, but that is not all. In some cases, cross peaks show us the transitions that are too weak to see directly on the diagonal. The intensity of the cross peak depends on that of both of the coupled vibrational transitions. Thus, even though one of the transitions is very weak, as long as the other transition is strong, we can still see cross peaks. Figure 1.2B shows a typical spectrum of amyloid fibers with two pairs of cross peaks off-diagonal (red box). Even though only one pair of diagonal peaks are intense enough to be observed, the appearance and position of the cross peaks indicate a weaker transition around 1690cm^{-1} due to the cross peaks. This feature of 2D IR turns out to be crucial in the study of amyloid structure formation in lens tissues (see Chapter 5). Note that in Figure 1.2A, we do not see cross peaks in the spectrum when the peptides are mostly random coils. Thus, cross peaks can be used as an indicator for amyloid formation in addition to the appearance of the diagonal amyloid peak $\sim 1620\text{ cm}^{-1}$ (Figure 1.2B).

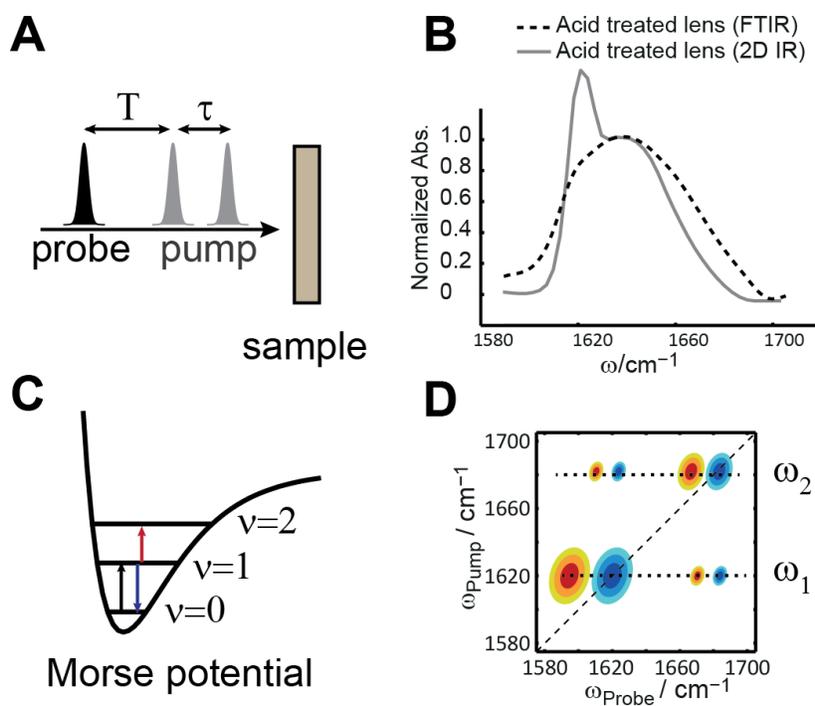


Figure 1.1 Cartoon illustration of 2D IR spectroscopy.

(A) 2D IR pulse sequence. (B) Overlay of a FTIR spectrum and a 2D IR diagonal slice. (C) Morse potential. (D) An illustration of a typical 2D IR spectrum with cross peaks.

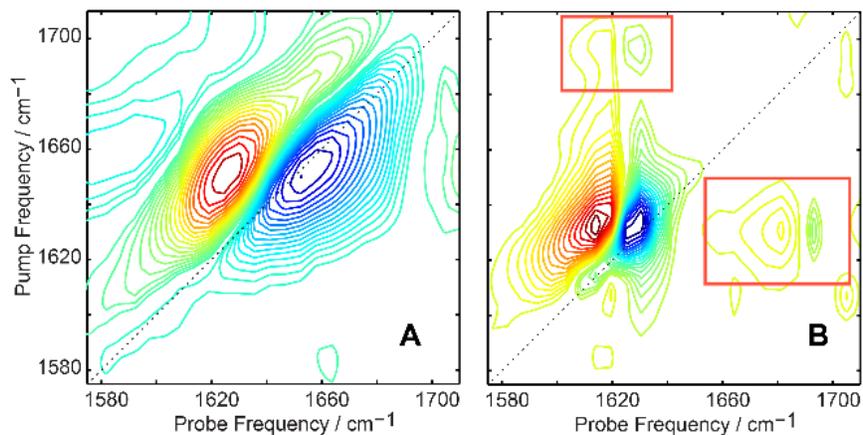


Figure 1.2 2D IR spectra of human amylin solution.

(A) Amylin in random coil state. (B) Amylin aggregated into amyloid fibers.

1.3 Study cataract formation with 2D IR spectroscopy

1.3.1 γ D-crystallin

With the features discussed above, 2D IR has been successfully used in the study of crystallin proteins (Chapter 3-5) and lens tissues (Chapter 5).⁵³ γ D-crystallin is a lens crystallin protein with two symmetric domains. Both of the isolated domains of γ D-crystallin as well as the full length protein form amyloid fiber when treated with acid.⁹ Previous studies by the Zanni group on segmentally isotope-labeled γ D-crystallin show that in the full length proteins, only the C-terminal domain forms amyloid fibers when the whole protein is treated with acid.^{53, 72} The N-terminal domain unfolds and does not participate in amyloid formation.⁷² This result was unexpected given the previous studies and knowledge about this protein. When unlabeled full length γ D-crystallin was studied, domain-specific information was not available.⁹ It was observed that the full length protein also forms amyloid fibers with higher fiber content than either isolated domains. Although both the isolated C-terminal and the N-terminal domain forms amyloid fibers when treated with acid, the N-terminal is more prone to amyloid formation than the C-terminal domain.⁹ It was concluded that the full length γ D-crystallin forms amyloid fibers with its N-terminal domain more prone to aggregation. This conclusion was later proved wrong by our 2D IR experiments on segmentally isotope-labeled γ D-crystallin.^{9, 53}

During my time in the Zanni group, I have been involved in the study of γ D-crystallin aggregation in other denaturing conditions, such as UV-irradiation (Chapter 3) and heat denaturation (Chapter 4). It is found that γ D-crystallin forms amyloid fibers in both of these conditions but with very different morphologies and mechanisms than in acid. UV-irradiation induces amyloid formation at the C-terminal domain of γ D-crystallin while unfolds the N-terminal domain. However, along with the fiber formation, the protein is also cleaved at specific sites. With

heat denaturation, we found that both domains of γ D-crystallin are incorporated into the final fiber structure possibly through cross-linking of the C-terminal and the N-terminal domains. These studies on γ D-crystallin help us to form a baseline to appreciate the complexity of crystallin protein aggregations.

1.3.2 α B-crystallin

As discussed above, α B-crystallin proteins function as molecular chaperones that can potentially bind to unfolded or misfolded proteins to prevent further aggregation. However, like the other crystallin proteins, this particular one can also form amyloid fibers in different conditions. The aggregation of α B-crystallin is examined with acid treatment, heat denaturation and UV-irradiation (Chapter 5). When treated with acid or heat, α B-crystallin readily forms amyloid fibers. These treatments are what we consider as extreme treatments that would not occur naturally unless there is an accident. UV-irradiation, however, is a major risk factor for cataract formation¹⁰ and is what an eye is exposed to regularly all the time when we are outside. When irradiated with UV light, α B-crystallin partially unfolds instead of forming amyloid structures like γ D-crystallin does. This observation is in line with a theory that, in the ocular lens, α B-crystallin protects other proteins from aggregation or forming long amyloid structures when exposed to UV-irradiation.⁷³ When UV-irradiated, α B-crystallin unfolds slightly and resists forming amyloid structures, which cannot function as chaperones. The unfolding is shown to enhance the chaperone activity of α B-crystallin.⁷⁴ Perhaps α B-crystallin also provides protection by preventing the other crystallin proteins from forming amyloid fibers in the lens. When the amount of misfolded crystallin in the lens exceeds the binding capacity of α B-crystallin, cataracts form. This hypothesis might explain why α -crystallin is evolutionarily so abundant in the lens (40%). However, this theory is formed

from *in vitro* experiments. More physiological evidence is needed to test and validate it. Our studies are aiming to provide this test in lens tissues.

1.3.3 Lens tissues

One of the major directions of my research has been the development of techniques and applications of 2D IR that expands from simple *in vitro* proteins to complex *ex vivo* tissues. The study of lens tissues is the first application of 2D IR spectroscopy to tissues (Chapter 5). By comparing 2D IR measurement of *in vitro* lens proteins to *ex vivo* lens tissues, the previous *in vitro* studies are validated. Meanwhile, the known mechanisms of *in vitro* lens protein aggregation help explaining the features from the *ex vivo* samples. The results of this study show that both acid treatment and UV-irradiation can cause crystallin proteins in the lens to aggregate into amyloid structures. The amyloid structures identified in UV-induced cataracts are the first evidence that age-related cataracts may be an amyloid disease. An interesting observation is that the morphology and the amount of amyloid structures in *in vitro* γ D-crystallin samples and in *ex vivo* lens tissues are rather different. This difference may be explained by the chaperone activity of α -crystallins.

Because 2D IR can be used to study both *in vitro* samples and *ex vivo* tissues without fixation and purification, it will be an important and suitable technique for similar tissue studies in the future. Using 2D IR to study proteins as well as tissue samples provides a link between microscopic protein structure and tissue failure.

1.4 2D IR application in other peptides

The study of cataract formation shows that 2D IR spectroscopy provides a unique perspective and offers insights into the research that is otherwise not available using other technique. In the rest of this chapter, I briefly discuss some other peptide systems I worked with using isotope-edited 2D IR spectroscopy, namely human amylin and transmembrane helices.

1.4.1 Human amylin

Peptides are smaller than proteins but it does not mean they are any less important. Many hormones we are familiar with are small peptides, such as insulin and glucagon that manipulate blood sugar levels, and amylin that controls appetite.⁷⁵ Besides their biological functions, some peptides form amyloid aggregates that are implicated in various diseases. Amyloid- β is a peptide that is found aggregated in the brains of Alzheimer patients.⁷⁶ Human amylin, co-secreted with insulin, is found aggregated in patients with type II diabetes.⁷⁷ While soluble in their native state, these small peptides can form amyloid fibers in the body and *in vitro*. These amyloid fibers are resistant to protein degradation while their final structure or intermediates can be cytotoxic.^{78,79}

In the Zanni group, we study the aggregation of human amylin using 2D IR spectroscopy in combination with electron microscopy (EM) and Thioflavin T (ThT) fluorescence. I contributed to the amylin projects as a biochemist in several ways. First, I modified existing protocols for peptide synthesis. Using Fmoc chemistry and solid phase synthesis, we can synthesize amylin with single residue isotope labeling.⁸⁰ However, the usage of isotope labeled material can be expensive. I introduced manual addition of reagents in the synthesis of human amylin. The new protocol can reduce the usage of expensive isotope labeled material by 50%. Second, I developed robust protocol for expressing human amylin and mutants in *E. coli*. A fully isotope-labeled peptide sometimes is desired in 2D IR and NMR studies. Although synthesis of amylin is easy with solid phase synthesis, the synthesis of a fully labeled peptide is not realistic due to the cost. In this case, expression using recombinant DNA can be the alternative way to obtain fully ¹³C-¹⁵N isotope labeled amylin. The protocols described in this dissertation can produce amylin peptides with correct modifications at high yield (details in Chapter 2). Third, I developed and tested protocol for Thioflavin T (ThT) fluorescence assays. Amyloid structures can be detected through binding

of fluorescent dyes like Thioflavin T (ThT).⁸¹ This method is generally accepted as a standard for measuring the presence of amyloid fibers. I successfully developed protocols and tested experimental conditions for ThT fluorescence assays used in the Zanni group (Chapter 2).

1.4.2 Transmembrane peptides

We collaborated with Dr. Yoshiaki Yano from Kyoto University to perform 2D IR spectroscopy on model α -helical transmembrane peptide AALALAA-AGLALGA-AALALAA. The GXXXG motif in the transmembrane helices encourages transient dimerization of the peptides.^{82, 83} We are interested in the interhelical packing and the orientations of this peptide in lipids. With isotope labeled residues, 2D IR measures the coupling between the labeled residues that reveals the interface of the transient dimers. Single-molecule FRET and polarized-attenuated total internal reflection (PATR) FTIR reveal the angle and orientation of the peptides dimers. This unique combination of techniques allows us to make a precise prediction of the transient species. This research is a good example of how isotope-edited 2D IR contributes to solving a complicated question (details in Chapter 6).

1.5 Summary and outline

2D IR is a powerful tool to study many biological systems, especially with biochemically designed samples. In this dissertation I provide detailed protocols for preparing isotope-labeled peptide and protein samples for 2D IR spectroscopy (**Chapter 2**). Some of these biochemical protocols are tweaked to fit our need in the lab but they can be generalized to other peptide and protein systems. I also include other common biochemical methods that can be used to complement and help validate 2D IR experiments, such as Thioflavin T fluorescence and electron microscopy.

The major component of my research is the study of cataracts. In **Chapter 3**, I talk about the experiment on UV-irradiated segmentally isotope-labeled γ D-crystallin that Dr. Sean Moran

and I had collaborated on. This experiment was one of the first to observe amyloid formation as well as protein site-specific cleavage in UV-irradiated γ D-crystallin proteins. In UV-irradiated γ D-crystallin, the C-terminal domain forms amyloid while the N-terminal domain unfolds, which is similar to in acid denatured γ D-crystallin. We also studied the aggregation of segmentally isotope-labeled γ D-crystallin with heat denaturation (**Chapter 4**). Heat denaturation is another method commonly used to induce protein aggregation. With 2D IR spectroscopy, we learn that when γ D-crystallin is denatured by heat, not only the C-terminal but also the N-terminal domain is incorporated in the final amyloid aggregate. By comparing the experimental and computational results of isotope dilution experiments, we concluded that the incorporation of N-terminal domain in heat induced γ D-crystallin amyloid structure is due to covalent bond formation between the C-terminal and the N-terminal domains. **Chapter 5** shows the comparison between the amyloid formation of γ D-crystallin and α B-crystallin, as well as measurement of tissue samples that I completed with Ariel Alperstein. We have been interested in the chaperone activity of α B-crystallin and its function in cataract formation. What we find interesting is that α B-crystallin does not form amyloid fibers when irradiated with UV light but only partially unfolds. Meanwhile, the unfolding is shown to enhance the chaperone activity of α B-crystallin.^{74, 84} Our hypothesis is that α B-crystallin partially unfolds and binds to misfolded γ D-crystallin to prevent the elongation of amyloid structures as the lens is exposed to UV-irradiation. With this hypothesis we study amyloid formation in lens tissues that are irradiated with UV light. From 2D IR and TEM measurements we conclude that amyloid structure can be induced in the lens tissues with UV-irradiation but the amount of amyloid structure is small compared to the total lens protein content. There appears to be a plateau where the increment of amyloid stops, which can be explained by the binding of α B-crystallin chaperone to misfolded γ D-crystallin to prevent fiber elongation.

During my time in the Zanni group, I collaborated with many great scientists on different projects. **Chapter 6** describes the research I worked in collaboration with Dr. Yoshiaki Yano on transmembrane helices. According to single molecule FRET, the α -helices with GXXXG motif forms antiparallel dimers in lipid bilayers. We use 2D IR spectroscopy to measure the coupling between isotope-labeled residues in these dimers to determine the angle between the two α -helices. These two techniques together help us understand the dynamics of helix-helix interaction in lipid bilayers and its dependence on the lipid compositions.

During my time in the Zanni group, I have used 2D IR spectroscopy for the studies of a variety of samples. Not all of them are described in the previous chapters. **Chapter 7** includes some different projects that are still ongoing, as well as the conclusion of this dissertation. By showing the preliminary data for these projects, I hope to provide the framework and a booster for my colleagues who will study these systems in the future.

Throughout this dissertation, I hope to demonstrate that 2D IR spectroscopy can be applied to a wide range of studies and provide unique insights into them. As showed from our studies on the crystallin proteins and lens tissues, 2D IR spectroscopy is extremely sensitive to the presence of amyloid structure that may not be detected by traditional methods. I look forward to see 2D IR spectroscopy being used to study and detect amyloid structure in other biological systems.

1.6 References

- (1) Findlen, P., and Bence, R. A history of the eye.
- (2) Hejtmancik, J. F., and Shiels, A. (2015) Overview of the Lens.
- (3) Goldstein, E. B., and Brockmole, J. (2016) *Sensation and perception*, Cengage Learning.
- (4) Bloemendal, H., Jong, W. d., Jaenicke, R., Lubsen, N. H., Slingsby, C., and Tardieu, A. (2004) Ageing and vision: structure, stability and function of lens crystallins, *Progress in Biophysics & Molecular Biology* **86**, 407–485.
- (5) Rodieck, R. W. (1973) The vertebrate retina: principles of structure and function.
- (6) Organization, W. H. (2010) Visual Impairment and Blindness 2010, World Health Organization.
- (7) Steinberg, E. P., Javitt, J. C., Sharkey, P. D., and et al. (1993) The content and cost of cataract surgery, *Archives of Ophthalmology* **111**, 1041-1049.
- (8) Hejtmancik, J. F., Riazuddin, S. A., McGreal, R., Liu, W., Cvekl, A., and Shiels, A. (2015) Lens Biology and Biochemistry.
- (9) Papanikolopoulou, K., Mills-Henry, I., Thol, S. L., Wang, Y., Gross, A. A. R., Kirschner, D. A., Decatur, S. M., and King, J. (2008) Formation of amyloid fibrils in vitro by human γ D-crystallin and its isolated domains, *Molecular Vision* **14**, 81-89.
- (10) West, S. K., and Valmadrid, C. T. (1995) Epidemiology of Risk Factors for Age-Related Cataract, *Survey of Ophthalmology* **39**, 323-334.
- (11) (1999) World Wide Words: Cataract.
- (12) (2008) Cataract Surgery in Antiquity, American Academy of Ophthalmology.
- (13) Mandal, A. (2014) Cataract History.
- (14) Kelman, C. D. (1967) Phaco-emulsification and aspiration: a new technique of cataract removal: a preliminary report, *American journal of ophthalmology* **64**, 23-35.
- (15) Lindstrom, R. (2015) Thoughts on Cataract Surgery: 2015, Review of Ophthalmology.
- (16) Prentice, T. (2006) Health, history and hard choices: Funding dilemmas in a fast in a fast-changing world, (Organization, W. H., Ed.).
- (17) Kessel, L. (2011) Can we meet the future demands for cataract surgery?, *Acta Ophthalmol* **89**, e289-290.
- (18) Erie, J. C. (2014) Rising cataract surgery rates: demand and supply, *Ophthalmology* **121**, 2-4.
- (19) Duke-Elder, W. S. (1927) An Ophthalmic Radiation Lamp: With a Note on Its Diagnostic and Therapeutic Applications, *The British journal of ophthalmology* **11**, 67-72.

- (20) Duke-Elder, W. S. (1928) Ultra-Violet Light in the Treatment of Ophthalmological Diseases II.- Local Phototherapy, *The British journal of ophthalmology* **12**, 353-373.
- (21) Spector, A. (1965) The soluble proteins of the lens, *Investigative Ophthalmology* **4**, 579-591.
- (22) Mathies, R., and Yu, N.-T. (1978) Raman Spectroscopy with Intensified Vidicon Detectors: A Study of Intact Bovine Lens Proteins, *JOURNAL OF RAMAN SPECTROSCOPY* **7**, 349-352.
- (23) Slingsby, C., Norledge, B., Simpson, A., Bateman, O. A., Wright, G., Driessen, H. P. C., Lindley, P. F., Moss, D. S., and Bax, B. (1997) X-ray diffraction and structure of crystallins, *Progress in retinal and eye research* **16**, 3-29.
- (24) Maraini, G., and Mangili, R. (1973) Characteristics of alpha-crystallin in human senile cataract, *Experimental eye research* **16**, 123-129.
- (25) Truscott, R. J. W., and Augusteyn, R. C. (1977) Oxidative Changes in Human Lens Proteins during Senile Nuclear Cataract Formation, *Biochimica et biophysica acta* **492**, 43-52.
- (26) Krishna, C. M., Uppuluri, S., Riesz, P., Zigler, J. S., and Balasubramanian, D. (1991) A Study of the Photodynamic Efficiencies of Some Eye Lens Constituents, *Photochemistry and photobiology* **54**, 51-58.
- (27) Ecroyd, H., and Carver, J. A. (2009) Crystallin proteins and amyloid fibrils, *Cellular and molecular life sciences : CMLS* **66**, 62-81.
- (28) Moreau, K. L., and King, J. A. (2012) Protein misfolding and aggregation in cataract disease and prospects for prevention, *Trends in molecular medicine* **18**, 273-282.
- (29) Toyama, B. H., Savas, J. N., Park, S. K., Harris, M. S., Ingolia, N. T., Yates, J. R., 3rd, and Hetzer, M. W. (2013) Identification of long-lived proteins reveals exceptional stability of essential cellular structures, *Cell* **154**, 971-982.
- (30) Zhu, J., Shao, J., Yao, Y., Chu, Z. D., Yu, Q. Q., Zhao, W., Lin, Q., and Zhang, Z. Y. (2013) Differential proteomics analysis of proteins from human diabetic and age-related cataractous lenses, *Journal of research in medical sciences : the official journal of Isfahan University of Medical Sciences* **18**, 984-989.
- (31) Horwitz, J. (1992) Alpha-Crystallin Can Function as a Molecular Chaperone, *Proceedings of the National Academy of Sciences of the United States of America* **89**, 10449-10453.
- (32) Bennardini, F., Wrzosek, A., and Chiesi, M. (1992) Alpha B-crystallin in cardiac tissue. Association with actin and desmin filaments, *Circulation Research* **71**, 288-294.
- (33) Chin, D., Boyle, G. M., Williams, R. M., Ferguson, K., Pandeya, N., Pedley, J., Campbell, C. M., Theile, D. R., Parsons, P. G., and Coman, W. B. (2005) Alpha B-crystallin, a new independent marker for poor prognosis in head and neck cancer, *The Laryngoscope* **115**, 1239-1242.
- (34) Cherneva, R., Petrov, D., Georgiev, O., Slavova, Y., Toncheva, D., Stamenova, M., and Trifonova, N. (2010) Expression profile of the small heat-shock protein alpha-B-crystallin in operated-on non-small-cell lung cancer patients: clinical implication, *European journal of cardio-*

- thoracic surgery : official journal of the European Association for Cardio-thoracic Surgery* **37**, 44-50.
- (35) Raman, B., Ban, T., Sakai, M., Pasta, S. Y., Ramakrishna, T., Naiki, H., Goto, Y., and Rao Ch, M. (2005) AlphaB-crystallin, a small heat-shock protein, prevents the amyloid fibril growth of an amyloid beta-peptide and beta2-microglobulin, *The Biochemical journal* **392**, 573-581.
- (36) Tanaka, N., Tanaka, R., Tokuhara, M., Kunugi, S., Lee, Y. F., and Hamada, D. (2008) Amyloid fibril formation and chaperone-like activity of peptides from alphaA-crystallin, *Biochemistry* **47**, 2961-2967.
- (37) Hochberg, G. K., Ecroyd, H., Liu, C., Cox, D., Cascio, D., Sawaya, M. R., Collier, M. P., Stroud, J., Carver, J. A., Baldwin, A. J., Robinson, C. V., Eisenberg, D. S., Benesch, J. L., and Laganowsky, A. (2014) The structured core domain of alphaB-crystallin can prevent amyloid fibrillation and associated toxicity, *Proceedings of the National Academy of Sciences of the United States of America* **111**, E1562-1570.
- (38) Serebryany, E., and King, J. A. (2014) The betagamma-crystallins: native state stability and pathways to aggregation, *Progress in biophysics and molecular biology* **115**, 32-41.
- (39) Schafheimer, N., and King, J. (2013) Tryptophan cluster protects human gammaD-crystallin from ultraviolet radiation-induced photoaggregation in vitro, *Photochemistry and photobiology* **89**, 1106-1115.
- (40) Chen, J., Toptygin, D., Brand, L., and King, J. (2008) Mechanism of the efficient tryptophan fluorescence quenching in human gammaD-crystallin studied by time-resolved fluorescence, *Biochemistry* **47**, 10705-10721.
- (41) Moran, S. D., Zhang, T. O., Decatur, S. M., and Zanni, M. T. (2013) Amyloid fiber formation in human gammaD-Crystallin induced by UV-B photodamage, *Biochemistry* **52**, 6169-6181.
- (42) Xia, Z., Yang, Z., Huynh, T., King, J. A., and Zhou, R. (2013) UV-radiation induced disruption of dry-cavities in human gammaD-crystallin results in decreased stability and faster unfolding, *Scientific reports* **3**, 1560.
- (43) Dyer, J. M., Bringans, S. D., and Bryson, W. G. (2006) Characterisation of photo-oxidation products within photoyellowed wool proteins: tryptophan and tyrosine derived chromophores, *Photochemical & photobiological sciences : Official journal of the European Photochemistry Association and the European Society for Photobiology* **5**, 698-706.
- (44) Mizdrak, J., Hains, P. G., Truscott, R. J., Jamie, J. F., and Davies, M. J. (2008) Tryptophan-derived ultraviolet filter compounds covalently bound to lens proteins are photosensitizers of oxidative damage, *Free radical biology & medicine* **44**, 1108-1119.
- (45) Fujii, N., Ishibashi, Y., Satoh, K., Fujino, M., and Harada, K. (1994) Simultaneous Racemization and Isomerization at Specific Aspartic-Acid Residues in Alpha-B-Crystallin from the Aged Human Lens, *Bba-Protein Struct M* **1204**, 157-163.
- (46) Fris, M., Tessem, M. B., Saether, O., and Midelfart, A. (2006) Biochemical changes in selenite cataract model measured by high-resolution MAS H NMR spectroscopy, *Acta ophthalmologica Scandinavica* **84**, 684-692.

- (47) GARNER, M. H., DEBDUTTA Roy, ROSENFELD, L., GARNER, W. H., and SPECTOR, A. (1981) Biochemical evidence for membrane disintegration in human cataracts, *Proc. NatL Acad. Sci. USA* **78**, 1892-1895.
- (48) Kosinski-Collins, M. S., and King, J. (2003) In vitro unfolding, refolding, and polymerization of human gammaD crystallin, a protein involved in cataract formation, *Protein science : a publication of the Protein Society* **12**, 480-490.
- (49) Acosta-Sampson, L., and King, J. (2010) Partially folded aggregation intermediates of human gammaD-, gammaC-, and gammaS-crystallin are recognized and bound by human alphaB-crystallin chaperone, *Journal of molecular biology* **401**, 134-152.
- (50) Simon, S., Michiel, M., Skouri-Panet, F., Lechaire, J. P., Vicart, P., and Tardieu, A. (2007) Residue R120 is essential for the quaternary structure and functional integrity of human alphaB-crystallin, *Biochemistry* **46**, 9605-9614.
- (51) Raju, M., Mooney, B. P., Thakkar, K. M., Giblin, F. J., Schey, K. L., and Sharma, K. K. (2015) Role of alphaA-crystallin-derived alphaA66-80 peptide in guinea pig lens crystallin aggregation and insolubilization, *Experimental eye research* **132**, 151-160.
- (52) Lee, S., Mahler, B., Toward, J., Jones, B., Wyatt, K., Dong, L., Wistow, G., and Wu, Z. (2010) A single destabilizing mutation (F9S) promotes concerted unfolding of an entire globular domain in gammaS-crystallin, *Journal of molecular biology* **399**, 320-330.
- (53) Moran, S. D., Decatur, S. M., and Zanni, M. T. (2012) Structural and sequence analysis of the human gammaD-crystallin amyloid fibril core using 2D IR spectroscopy, segmental ¹³C labeling, and mass spectrometry, *Journal of the American Chemical Society* **134**, 18410-18416.
- (54) Sugiyama, M., Fujii, N., Morimoto, Y., Kurabayashi, S., Vigild, M. E., Nakagawa, T., Sato, T., Itoh, K., Mori, K., and Fukunaga, T. (2008) Structural evolution of human recombinant alpha B-crystallin under UV irradiation, *Biomacromolecules* **9**, 431-434.
- (55) Schafheimer, N., Wang, Z., Schey, K., and King, J. (2014) Tyrosine/cysteine cluster sensitizing human gammaD-crystallin to ultraviolet radiation-induced photoaggregation in vitro, *Biochemistry* **53**, 979-990.
- (56) Cucinotta, F. A., Manuel, F. K., Jones, J., Iszard, G., Murrey, J., Djojonegro, B., and Wear, M. (2001) Space radiation and cataracts in astronauts, *Radiation research* **156**, 460-466.
- (57) Vendra, V. P., Khan, I., Chandani, S., Muniyandi, A., and Balasubramanian, D. (2015) Gamma crystallins of the human eye lens, *Biochimica et biophysica acta*.
- (58) Marsili, S., Salganik, R. I., Albright, C. D., Freel, C. D., Johnsen, S., Peiffer, R. L., and Costello, M. J. (2004) Cataract formation in a strain of rats selected for high oxidative stress, *Experimental eye research* **79**, 595-612.
- (59) Costello, M. J., Burette, A., Weber, M., Metlapally, S., Gilliland, K. O., Fowler, W. C., Mohamed, A., and Johnsen, S. (2012) Electron tomography of fiber cell cytoplasm and dense cores of multilamellar bodies from human age-related nuclear cataracts, *Experimental eye research* **101**, 72-81.

- (60) Hahn, S., Kim, S. S., Lee, C., and Cho, M. (2005) Characteristic two-dimensional IR spectroscopic features of antiparallel and parallel beta-sheet polypeptides: simulation studies, *The Journal of chemical physics* **123**, 084905.
- (61) Baiz, C. R., Peng, C. S., Reppert, M. E., Jones, K. C., and Tokmakoff, A. (2012) Coherent two-dimensional infrared spectroscopy: Quantitative analysis of protein secondary structure in solution, *Analyst* **137**, 1793-1799.
- (62) Lin, S.-Y., Li, M.-J., and Cheng, W.-T. (2007) FT-IR and Raman vibrational microspectroscopies used for spectral biodiagnosis of human tissues, *Spectroscopy* **21**, 1-30.
- (63) Shim, S. H., and Zanni, M. T. (2009) How to turn your pump-probe instrument into a multidimensional spectrometer: 2D IR and Vis spectroscopies via pulse shaping, *Physical chemistry chemical physics : PCCP* **11**, 748-761.
- (64) Middleton, C. T., Woys, A. M., Mukherjee, S. S., and Zanni, M. T. (2010) Residue-specific structural kinetics of proteins through the union of isotope labeling, mid-IR pulse shaping, and coherent 2D IR spectroscopy, *Methods* **52**, 12-22.
- (65) Strasfeld, D. B., Ling, Y. L., Gupta, R., Raleigh, D. P., and Zanni, M. T. (2009) Strategies for extracting structural information from 2D IR spectroscopy of amyloid: application to islet amyloid polypeptide, *The journal of physical chemistry. B* **113**, 15679-15691.
- (66) Hamm, P., and Zanni, M. (2011) *Concepts and Methods of 2D Infrared Spectroscopy*, Cambridge University Press.
- (67) Cho, M. (2009) *Two-dimensional optical spectroscopy*, CRC press.
- (68) Mukamel, S. (1999) *Principles of nonlinear optical spectroscopy*, Oxford University Press on Demand.
- (69) Buchanan, L. E., Dunkelberger, E. B., Tran, H. Q., Cheng, P. N., Chiu, C. C., Cao, P., Raleigh, D. P., de Pablo, J. J., Nowick, J. S., and Zanni, M. T. (2013) Mechanism of IAPP amyloid fibril formation involves an intermediate with a transient beta-sheet, *Proceedings of the National Academy of Sciences of the United States of America* **110**, 19285-19290.
- (70) Nelson, R., Sawaya, M. R., Balbirnie, M., Madsen, A. O., Riek, C., Grothe, R., and Eisenberg, D. (2005) Structure of the cross-beta spine of amyloid-like fibrils, *Nature* **435**, 773-778.
- (71) Stroud, J. C., Liu, C., Teng, P. K., and Eisenberg, D. (2012) Toxic fibrillar oligomers of amyloid-beta have cross-beta structure, *Proceedings of the National Academy of Sciences of the United States of America* **109**, 7717-7722.
- (72) Moran, S. D., Woys, A. M., Buchanan, L. E., Bixby, E., Decatur, S. M., and Zanni, M. T. (2012) Two-dimensional IR spectroscopy and segmental ¹³C labeling reveals the domain structure of human γ D-crystallin amyloid fibrils, *PNAS*.
- (73) Mach, H., Trautman, P. A., Thomson, J. A., Lewis, R. V., and Middaugh, C. R. (1990) Inhibition of alpha-Crystallin Aggregation by gamma-Crystallin, *J. Biol. Chem.* **265**, 4844-4848.

- (74) Foit, L., George, J. S., Zhang, B. W., Brooks, C. L., 3rd, and Bardwell, J. C. (2013) Chaperone activation by unfolding, *Proceedings of the National Academy of Sciences of the United States of America* **110**, E1254-1262.
- (75) Khemtémourian, L., Killian, J. A., Hoppener, J. W., and Engel, M. F. (2008) Recent insights in islet amyloid polypeptide-induced membrane disruption and its role in beta-cell death in type 2 diabetes mellitus, *Experimental diabetes research* **2008**, 421287.
- (76) Brody, J. A., and Grant, M. D. (2001) Age-associated diseases and conditions: implications for decreasing late life morbidity, *Aging* **13**, 64-67.
- (77) Meng, F., Marek, P., Potter, K. J., Verchere, C. B., and Raleigh, D. P. (2008) Rifampicin does not prevent amyloid fibril formation by human islet amyloid polypeptide but does inhibit fibril thioflavin-T interactions: implications for mechanistic studies of beta-cell death, *Biochemistry* **47**, 6016-6024.
- (78) Tenidis, K., Waldner, M., and Bernhagen, J. (2000) Identification of a Penta- and Hexapeptide of Islet Amyloid Polypeptide (IAPP) with Amyloidogenic and Cytotoxic Properties, *J. Mol. Biol* **295**, 1055-1071.
- (79) Kirkitadze, M. D., Condrón, M. M., and Teplow, D. B. (2001) Identification and characterization of key kinetic intermediates in amyloid beta-protein fibrillogenesis, *Journal of molecular biology* **312**, 1103-1119.
- (80) Marek, P., Woys, A. M., Sutton, K., Zanni, M. T., and Raleigh, D. P. (2010) Efficient Microwave-Assisted Synthesis of Human Islet Amyloid Polypeptide Designed to Facilitate the Specific Incorporation of Labeled Amino Acids, *Organic letters* **12**, 4848-4851.
- (81) Sulatskaya, A. I., Kuznetsova, I. M., and Turoverov, K. K. (2011) Interaction of thioflavin T with amyloid fibrils: stoichiometry and affinity of dye binding, absorption spectra of bound dye, *The journal of physical chemistry. B* **115**, 11519-11524.
- (82) Russ, W. P., and Engelman, D. M. (2000) The GxxxG motif: a framework for transmembrane helix-helix association, *Journal of molecular biology* **296**, 911-919.
- (83) Teese, M. G., and Langosch, D. (2015) Role of GxxxG Motifs in Transmembrane Domain Interactions, *Biochemistry* **54**, 5125-5135.
- (84) Palmieri, V., Maulucci, G., Maiorana, A., Papi, M., and De Spirito, M. (2013) alpha-crystallin modulates its chaperone activity by varying the exposed surface, *Chembiochem : a European journal of chemical biology* **14**, 2362-2370.

2 Chapter 2

Experimental Methods*

2.1 Introduction

There is one aspect that is unique about the Zanni group and I truly appreciate – although the group is known for spectroscopy, we proudly design and produce most of the biological sample in house. During my time in the Zanni group, I had the privilege to work on both ultrafast laser systems and biochemical procedures. One idea that became clear to me is that the preparation of samples can sometimes determine what kind of information we will be able to extract from the signals. For example, for a small peptide or for a large protein, the isotope-labeling strategies and the sample design are very different. Many excellent theses have elaborated the techniques and tricks for the standard 2D IR setup. In this chapter, I want to focus on the biochemistry methods. I hope to not only describe the key biochemical protocols but also explain the rudimental principle behind the practices. Many of these protocols, although written for a particular protein or system, can be applied to others with appropriate modification.

2.2 Reagents and Buffers

All the reagents and buffers in the biochemistry protocols introduced in this chapter are listed below. Some of them do not have a standard name and can be mistaken for different

* Some material in this chapter was originally published as: Tianqi O. Zhang, Maksim Grechko, Sean D. Moran and Martin T. Zanni. “Isotope-Labeled Amyloids via Synthesis, Expression, and Chemical Ligation for Use in FTIR, 2D IR, and NMR Studies.” *Protein Amyloid Aggregation, Methods and Protocols*. 2016. Vol 1345, pp 21-41. DOI 10.1007/978-1-4939-2978-8-2

protocols. To avoid confusion, I added a letter name (or use its original name) for each reagent or buffer in bold. In the individual protocols, only the bolded letter names will be used.

- 1) Unlabeled E. coli protein expression in Lysogeny broth (LB)
 - a) **LB** (1L): Dissolve 10 g tryptone, 5 g yeast extract and 10 g NaCl in 1 L H₂O. Autoclave.
 - b) Agar plate LB (**ALB**) (20 plates): Dissolve 5 g tryptone, 2.5 g yeast extract, 5 g NaCl and 10 g agar in 0.5 L H₂O. Autoclave.
- 2) Isotope-labeled expression
 - a) Minimal Media buffer (**MMB**, 0.5 L): Dissolve 3 g of K₂HPO₄, 4.5 g of KH₂PO₄, 1.25 g of NH₄Cl (or (NH₄)₂SO₄. Use ¹⁵NH₄Cl if ¹⁵N expression is desired), and 0.25 g of citric acid monohydrate. Adjust pH to 6.6 using KOH. Autoclave.
 - b) Minimal Media stock solution (**MMSS**): 1 g of glucose (use ¹³C glucose if ¹³C label is desired), 35 mg of thiamine hydrochloride, 10 mL of 50g/L CaCl₂, 325 μL of vitamin solution, 500 μL of 100 mg/mL ampicillin, and 0.2 g celtone base powder (optional). Filter with 0.2 μM sterile filters.
 - c) Minimal Media (**MM**): in autoclaved MMB add all stock solution prepared above and 500 μL of autoclaved trace element solution.
 - d) **Vitamin solution**: Dissolve 0.4 g of each of the following: pantothenic acid (calcium salt), choline chloride, folic acid, nicotinamide, pyridoxal hydrochloride, and thiamine hydrochloride in 800 mL of H₂O. In the same solution, dissolve 0.8 g of myoinositol and 0.04 g of riboflavin. Adjust solution to pH 7.2 and filter with 0.2 μm sterile filter. Aliquot the solution to 0.35 mL in sterile Eppendorf tubes. Avoid light and place them in -80 °C freezer for long-term storage.

- e) **Stock MgSO₄ solution:** Weigh 10 g of anhydrous MgSO₄ and dissolve in 200 mL of H₂O. Autoclave.
 - f) **Trace element solution:** Dissolve 16.2 g of FeCl₃·6H₂O, 2.4 g of ZnSO₄·7H₂O, 4.2 g of CoCl₂·6H₂O, 4.2 g of Na₂MoO₄·2H₂O, 4.8 g of CuSO₄·5H₂O, 1.2 g of H₃BO₃, 3.0 g of MnSO₄, and 30 mL of 37 % HCl into 570 mL of H₂O. Autoclave.
- 3) Ni-NTA column:
- a) Ni-NTA resuspension buffer (**NiRB**): 50 mM Na₂HPO₄, 500 mM NaCl, 50 mM imidazole, pH 7.5
 - b) Ni-NTA elution buffer (**NiEB**): 50 mM Na₂HPO₄, 500 mM NaCl, 500 mM imidazole, pH 7.5
 - c) Ni-NTA stripping solution (**NiSS**): 100 mM EDTA, 50 mM Na₂HPO₄, 500 mM NaCl, pH 8.0
 - d) Ni-NTA reload solution (**NiRS**): 100 mM NiSO₄ (or solutions of other metal ions)
- 4) Ion Exchange (Q-sepharose) column:
- a) Buffer A (IECA): 20 mM Tris, pH 8.5 at 4 °C
 - b) Buffer B (IECB): 20 mM Tris and 500 mM NaCl, pH 8.5 at 4 °C
 - c) Buffer Clean (IECC): 20 mM Tris and 2.0 M NaCl, pH 6.0
 - d) Storage solution (IECS): 20% ethanol in H₂O
- 5) Reverse Phase column
- a) Buffer A (RPA): 0.046 % HCl in H₂O. Dissolve 5 mL of 37 % HCl in 4 L of H₂O in a fume hood. Mix well.
 - b) Buffer B (RPB): 80% (v/v) acetonitrile in H₂O with 0.046% HCl. In a fume hood, mix 3.2 L acetonitrile and 0.8 L H₂O, and then add in 5 mL of 37% HCl.

- 6) Ni-NTA column for hIAPP:
 - a) Ni-NTA hIAPP resuspension buffer (**NiRBhIAPP**): 50 mM Na₂HPO₄, 500 mM NaCl, 2 M urea, pH 7.5
 - b) Ni-NTA hIAPP elution buffer (**NiEBhIAPP**): 50 mM Na₂HPO₄, 500 mM NaCl, 500 mM imidazole, 2 M urea, pH 7.5
- 7) Chitin binding column for hIAPP
 - a) hIAPP Chitin binding buffer (**ChihIAPPB**): 20 mM HEPES, 0.1 mM EDTA, 50 mM NaCl, 2 M urea, pH 8.0
 - b) hIAPP Chitin cleavage buffer (**ChihIAPPC**): 100 mM DTT and 2 M ammonium bicarbonate in 50 mL **ChihIAPPB**, pH 8.0
 - c) Stripping (**ChiStrip**): 0.3 M NaOH in **ChihIAPPB**
- 8) Chitin binding column for segmentally labeling γ D-crystallin, N-terminal domain
 - a) Chitin binding buffer (**Chi γ DB**): 20 mM HEPES, 200 mM NaCl, pH 8.5
 - b) γ D-crystallin chitin cleavage buffer (**Chi γ DC**): Chi γ DB with 50 mM MESNA, pH 8.5
 - c) γ D-crystallin chitin Storage buffer (**Chi γ DS**): 5 mM Bis-Tris, pH 6.5, 250 mM NaCl
- 9) SDS-PAGE
 - a) Running buffer (**PAGERB**, 4L): In a 4 liter bottle, add 12.12 g of Tris base (powder), 57.6 g of glycine, and 4L of H₂O. Add 4 g of Sodium dodecyl sulfate (SDS) after the addition of H₂O.
 - b) Loading buffer (**PAGELB**, 5X): 0.025% (m/v) bromophenol blue, 50% (v/v) Glycerol (50%), 10% (m/v) SDS, 0.25 M Tris-Cl, pH 6.8

- c) Staining solution (**PAGESS**): Dissolve 0.25 g Coomassie Brilliant Blue in 400 mL methanol, add 100 mL acetic acid in the hood. Fill up to 1L with H₂O. Filter with filter paper.
 - d) Destaining solution (**PAGEDS**): 40% (v/v) Methanol, 10% (v/v) acetic acid in H₂O.
- 10) DNA gel
- a) **TAE stock solution (10X)**: Dissolve 48.4g Tris base, 20ml of 0.5M EDTA at pH 8.0, and 11.44 mL Glacial acetic acid in H₂O to make 1000 mL total solution.
- 11) 2D IR protein buffer
- a) 20 mM Tris in D₂O, pD 7.4
 - b) 20 mM phosphate, 100 mM NaCl in D₂O, pD 7.4
 - c) 20 mM phosphate, 100 mM NaCl in D₂O, pH 3.0
 - d) 1 mM phosphate in D₂O, pH 5.5
- 12) Factor Xa
- a) Factor Xa cleavage buffer (FXaCB): 10 mM Tris, 20 mM NaCl, pH 7.5
 - b) CaCl₂ stock solution: 1.0 M CaCl₂.
- 13) ThT stock solution
- a) ThT (1000X): 25 mM ThT in H₂O

2.3 Biochemistry protocols

2.3.1 PCR and DNA gel

PCR is short for polymerase chain reaction. The basic idea of PCR is to copy a piece of target DNA with a specific primer and terminator. The target DNA can be in a solution or simply a small amount of colony picked from an agar plate (colony PCR). The PCR product can be visualized and analyzed by agarose gel. Ethidium bromide is added to agarose gel to stain the DNA molecules in gel. This molecule is a potential mutagen. Handle it carefully in the hood. Liquid Ethidium bromide waste should be treated as chemical hazardous waste. As I am preparing this dissertation, agarose gel containing ethidium bromide can be placed in the trash if they are dried and placed in a sealed container. You should check the most updated regulation before disposing of ethidium bromide.

A typical PCR reaction can be setup as below:

- 1) Thaw everything on ice. Add the following elements in a PCR tube. Follow the order and leave the tube on ice.
 - a) 1 μ L of Primer 1 (forward)
 - b) 1 μ L of Primer 2 (reverse)
 - c) 1 μ L of dNTP
 - d) 0.5 μ L of target DNA (or a pipette tip pick of bacteria colony)
 - e) H₂O to bring the total volume to 50 μ L
 - f) 5 μ L of 10X reaction buffer
 - g) 0.5 μ L enzyme DNA polymerase
- 2) Mix the solution well and avoid bubbles.
- 3) Program the PCR machine (G-storm) as following:

# of Cycles	Temp (°C)	Time	Note
1	95	2 min	Initial melting of the template DNA
30	95	5 sec	Detaching polymerase and melting of the DNA
	Annealing temp	15 sec	Binding of polymerase
	72	15 sec (or 30 sec/1k bp)	Extension of primers. This temperature might differ depending on the polymerase.
1	72	5 min	Final extension of the primers. This temperature might differ depending on the polymerase.
1	4	Infinity	Storage

Table 2.1 Table of PCR cycles

The temperature and extension time are enzyme and template specific.

Annealing temp (usually 48-65 C) can be roughly calculated as:

$$T_{\text{Annealing}} = 2 * (\# \text{ of A+T}) + 4 * (\# \text{ of C+G}) - 5$$

- 4) While the PCR is running, set up DNA gel.
 - a) Clean the DNA gel casket (Figure 2.1A). Place the gel plate (Figure 2.1B) perpendicular to the walls so that a well is sealed by the rubber bands (Figure 2.1C). Insert the desired well divider. Two well dividers can be used in one gel to fit more samples but the gel will be half the length. Use thicker side of the well divider if there is plenty of sample (Figure 2.1D).
 - b) In a beaker, dilute 30 mL of 10X TAE buffer into 270 mL H₂O to make 1X TAE buffer.
 - c) In a 200 ml flask, add 50 mL of 1X TAE buffer made in previous step. Dissolve 1 g of agarose in the flask. Microwave this solution in microwave for 1.5 min (or until boils). 1g agarose /50mL solution makes 2% agarose gel. The percentage of the gel usually lies between 0.7% - 2%. The longer the DNA fragment is, the smaller percentage of agarose should be used. The table below provides a general guide for choosing the percentage of agarose gel.

Agarose (%)	DNA (kbp)
-------------	-----------

0.7%	5-8
1.0%	3-5
1.5%	1-3
2.0%	0.1-1

Table 2.2 Table of % agarose and corresponding DNA fragment sizes.

- d) With gloves, carefully bring the flask to the hood. Add 10 μL of ethidium bromide (10 mg/mL from Sigma) to the solution when it cools down to around 60 C. Mix the solution by gentle swirling. Avoid bubbles.
 - e) In the hood, pour the solution (all 50 mL) slowly from a corner. Let the agarose gel solidify in the hood for at least 30 min.
- 5) After agarose gel solidifies, remove the entire well (with the gel). Rotate the well 90 degrees and place it back in the casket. Pour the rest of the 1X solution (250 mL) into the casket. Remove the spacers by pulling them straight up.
 - 6) Mix 50 μL of PCR product with 12 μL 5X DNA loading buffer. Add the mixture slowly to a well by inserting the pipette tip half way inside the well. The well size depends on the spacer used. Observe from the side of the casket and make sure the samples do not overflow.
 - 7) Add 5 μL of DNA ladder in a well without samples. Choose the appropriate ladder range for different samples.
 - 8) Slide on the cover lid. Attach the electric cords into the side of the cover lid. The black cord should be on the side of the sample wells (Figure 2.1F). With one hand, plug in the other side of the electric cords into the power supply (Figure 2.1G). Turn on the power supply and set at constant (K) ampere mode. Set it at 100 V and 400 mA. The dyes should move to the bottom of the gel after around 60-80 min.

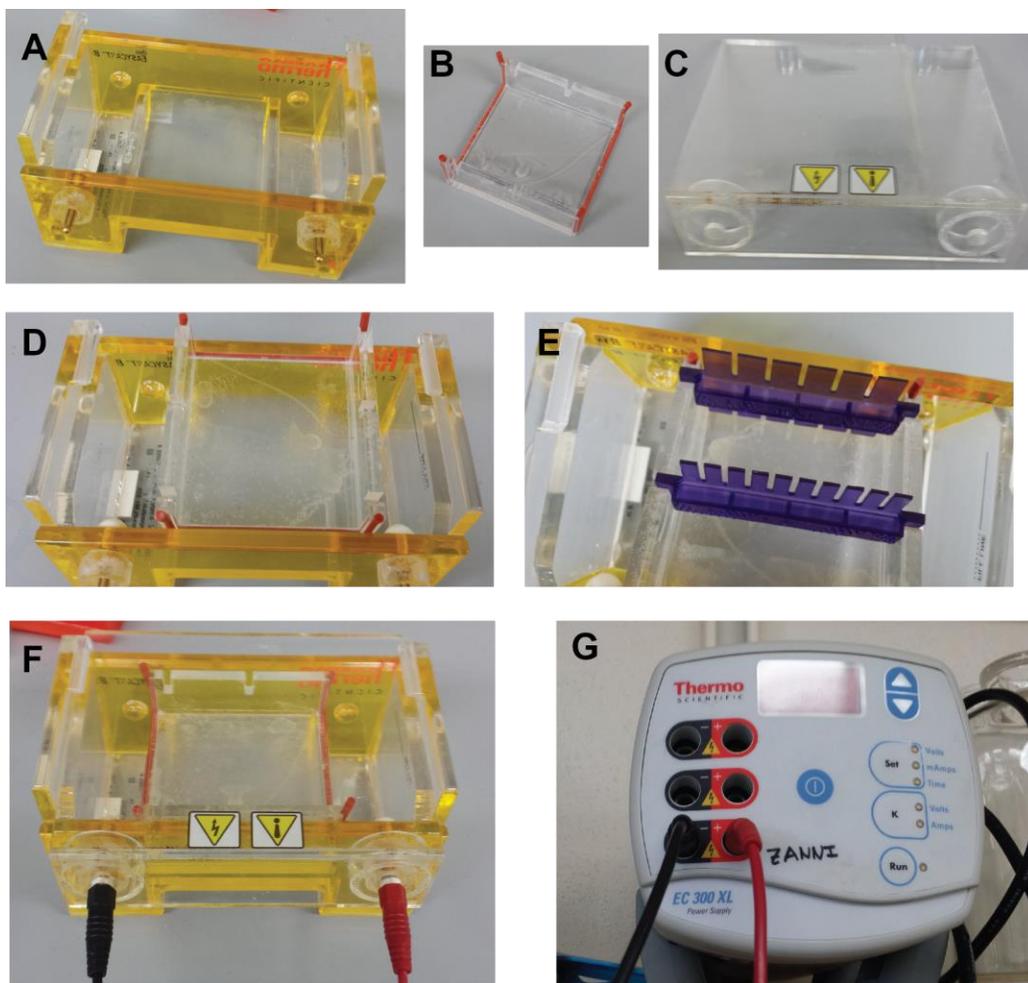


Figure 2.1 Setup of DNA agarose gel

The apparatus of DNA agarose gel includes: (A) gel casket, (B) gel plate, (C) casket lid. (D) To pour DNA gel, the gel plate needs to be placed in the casket perpendicularly as shown. (E) Two well dividers are available. Both can be used on the same gel with sacrificed gel length. (F) Place the lid on the casket before running the gel. Gel will run from the black electrode to the red electrode. (G) Power supply with electrodes of matching colors.

- 9) After electrophoresis, turn off the power supply with one hand. Remove the cover lid. With gloves, gently transfer the agarose gel onto the UV lamp viewer. Place the cover on top of the gel. UV light can cause damage to eyes and skin. Do not look directly onto the UV lamp.
- 10) The protein bands should be fluorescent with pink light when viewed under UV light. A picture can be taken at this point.

- 11) PCR product can be purified from gel (this is actually important to do when amplify a plasmid). If purification is desired, make sure the gel is not exposed to the UV light for too long since it can be damaged. While the UV light is on, quickly cut out the band with the correct number of base pairs. Gel extraction can be done then with spin columns in a kit.

2.3.2 Sequencing

Gene Sequencing allows us to get the exact sequence of a piece of DNA. After receiving a new plasmid or gene, it is a good practice to amplify them and obtain the sequencing results before moving to the next step. UW-Madison has a sequencing facility on campus. Visit <https://www.biotech.wisc.edu/> to get to the exact location. Before submitting a sequencing sample, you need to setup an account with the Biotech Center. Here I describe how to prepare a sample for sequencing.

- 1) Use strip tubes (8 in a row). Use the whole length strip tube even if you have less than 8 samples.
- 2) For each sample, mix the following to obtain a total 24 μ L solution
 - a) 400-500 ng plasmid or 40-50 ng PCR product
 - b) 10-20 pmol primer

Note that only one primer is needed for each sample. For a longer sequence, you can submit two samples with the same template, with either a forward primer or a reverse primer. Thus, if one sequencing cannot cover the whole length, the opposite side can make up for the missing pieces.

- 3) Wrap the samples in aluminum foil. Put a tape on it with your name, “Zanni group”, “number of samples”, “Discount sequencing striptube”, and the date.
- 4) Take the sample to biotech center and submit the order using their computer on site.

- 5) Your sequencing data should be uploaded within 1-3 days. Visit <https://www.biotech.wisc.edu/services/dnaseq> and “Access your data”. Login with your UW login. The data should be organized by dates under your name.
- 6) The data are ab1 files, which can be open using *FinchTV* and many other software programs.

2.3.3 Gene and plasmid construction

There are two common ways to make small peptides: solid phase synthesis (SPS) and *E. coli* expression. For a 0.1 mM scale synthesis, the SPS provides much higher yield than a 2 L expression in *E. coli*. The SPS also allows the synthesis of single isotope labeled peptides. However, when a uniformly labeled peptide is desired, for example ^{13}C - ^{15}N labeled peptide for NMR studies, expression in *E. coli* is more economically feasible.

Here I describe the construction of hIAPP gene and the insertion into pTXB1 vector. hIAPP is a 37 residue long peptide, which means there are at least 111 base pairs in gene piece. pTXB1 is a vector from New England Biolabs. It utilizes intein mediated cleavage for protein purification and modification. The gene for desired peptide or protein can be purchased through commercial websites like Genscript but the turnaround time can be long. For short peptides like hIAPP, we can construct the gene using PCR in our lab to reduce the waiting time. For longer sequences, it's better to directly purchase the entire gene piece.

2.3.3.1 Step 1. Gene design

Each amino acid in *E. coli* is coded with several different combinations of DNA base pairs. When a mammalian peptide is expressed in *E. coli*, its gene needs to be modified according to *E. coli* codon usage. We also need to consider insertion of the gene into the vector using restriction enzymes (usually two different ones). For each vector, the restriction enzyme sites are listed in the vector maps. If specific features of a vector is needed, then the choices of restriction enzymes can

be very limited. For pTXB1, if we want to express a peptide that is fused with intein on its C-terminus, we have to use Sap1 and Nde1.

Sap1:

```
5' ... G C T C T T C N^N N N ...3'
3' ... C G A G A A G N N N N^...5'
```

Nde1:

```
5' ... C A^T A T G ...3'
3' ... G T A T^A C ...5'
```

In the above sites, the base pairs are recognition sequences and the ^ sign shows where the restriction enzyme cuts the sequence. The letter N means any base pair. For example, after cut by restriction enzyme Nde1, the two ends of the DNA will be as following:

Nde1 End 1:

```
5' ... C A
3' ... G T A T
```

Nde1 End 2:

```
T A T G ...3'
A C ...5'
```

The sequence of hIAPP is: KCNTATCATQRLANFLVHSSNCFGAILSSTNVGSNTY

The optimization can be done on websites like: <https://www.idtdna.com/>. The following DNA sequence is optimized based on *E. coli* codon usage.

```
AAGTGCAACACCGCCACCTGCGCCACACAACGCCTTGCCAACTTTCTGGTGCATAGCAGTAACA
ATTTTCGGTGCATTCTTTTCGTCCACTAATGTAGGCAGTAATACATAC
```

However, there are other factors to consider. First, we need to check for restriction enzyme sites within this sequence. For most restriction enzymes, the 5' to 3' sequence is identical to its reverse complement. There are some special ones like Sap1. It's a good practice to check for the restriction enzyme sites both in the original sequence and in the reverse complement.

2.3.4 Making Electrocompetent cells and heat-shock-competent cells

Competent cells can pick up plasmids from the solution when applied a voltage across or heat shocked. The majority of the process is the same for electrocompetent cells and heat-shock-competent cells. In the transformation, electrocompetent cells usually are more efficient, with shorter operation time. To make competent cells, especially cell lines without additional build-in antibiotics, it is important to keep everything sterile. The keys to make good competent cells are 1) follow the instruction as close as possible, 2) keep the cells cold (on ice or in the cold room) at all times, and 3) work as quickly as possible during centrifugation and resuspension steps.

2.3.4.1 Day 1

- 1) Make 0.5 L low salt LB solution
 - a) 5 g Tryptone
 - b) 2.5 g Yeast extract
 - c) 0.5 g NaCl
- 2) Make 1 L of 15% glycerol solution (for electrocompetent cells).
- 3) Autoclave the tips (1000 μ L and 100 μ L), 2 centrifuge tubes (500 mL), 1.7 mL microcentrifuge tubes (around 50 of them in a beaker) and the solutions listed above. *When autoclave centrifuge tubes, wrap the centrifuge tubes and the caps separately with aluminum foil.*
- 4) Prepare a 10 mL overnight culture from either a stock or a single colony from plates at the end of the day.

2.3.4.2 Day 2

- 5) Inoculate 0.5 L LB with 10 mL overnight culture and grow at 37 °C (220 rpm shaking) to 0.4-0.6 OD at 600 nm.

- 6) Add antibiotic (usually 30 $\mu\text{g}/\text{mL}$ Chloramphenicol for BL21, none for XL1)
- 7) When the cells are at the right OD, place the flask on ice for 20 minutes in the cold room.

Bring all the following to the cold room:

- a) centrifuge tubes
 - b) 10.50 rotor
 - c) Microcentrifuge tubes
 - d) 1000 mL and 100 mL pipettes
 - e) Pipette tips
 - f) A balance
 - g) Chem wipe
 - h) ethanol
- 8) Add cell culture to centrifuge tubes (250 mL in each). Balance before centrifuge. Centrifuge at 5000 x g for 10 minutes, 4 °C when everything has cooled.
 - 9) In the cold room, pour off supernatant, and tap the rim of the tube gently on clean Kimwipes to remove excess liquid. *You may lose some cells during this step, if they did not sediment completely. That's OK, the prep will be better if you just keep the cells that sedimented.*
 - 10) Resuspend cells in each tube with 250 mL ice cold, sterile 15% glycerol (v/v) in H₂O.
 - 11) Repeat steps 8)-10) two more times.
 - 12) After the last centrifuge, resuspend cells in 4 mL sterile 15% glycerol
 - 13) Aliquot 102 μL cells in sterile microcentrifuge tubes.
 - 14) *Outside of the cold room, freeze quickly on liquid nitrogen (15 sec) and store at -80 °C. You can put yourself in danger if you pour liquid nitrogen in the cold room!*

To make heat-shock-competent cells, replace 15% glycerol (v/v) in H₂O with 15% glycerol (v/v) in 0.1 M CaCl₂ solution.

2.3.5 Electroporation

There are two common used method to incorporate foreign plasmid into cells. Heat-shock and electroporation. Electroporation is a powerful and efficient tool that allows for the incorporation of plasmids into different cells for protein expression, imaging, or many other purposes. Different cell lines require different conditions and voltages for electroporation. Here I only describe the electroporation of plasmid into E. coli cell for protein overexpression. This protocol takes around 1.5 hours to complete.

- 1) Prepare autoclave LB at 37 °C or room temperature. Prepare three autoclaved agar plates per electroporation with the desired antibiotics. For better results, incubate the agar plates in 37 °C before use. Thaw plasmid on ice. The concentration of plasmid can be measured with the Nanodrop. Ideal plasmid concentration is 0.1-1 ng.
- 2) Thaw electrocompetent cells on ice. The production of electro- and heat-competent cells are described in the previous section. In our laboratory, the electrocompetent cells are in 100 μ L aliquots. Cool an electrocompetent cuvette on ice with the lid attached.



Figure 2.2 Electroporator

(A) Control Panel of eletroporator. (B) Cuvette holder that is moved from the electroporator.

- 3) Prepare the Eppendorf Electroporator 2510 (Figure 2.2A)

- a) Connect the power cord of the instrument.
- b) Set electroporator voltage to 1800 V (this voltage is specific to *E. coli*) and leave the instrument on 'Time constant'.
- c) Take 2 μL of plasmid and transfer into the electrocompetent cells. Mix well with a 20-200 μL pipette and avoid bubbles (bubbles can affect the efficiency of electroporation and lower your yield). Transfer the mixture to the electrocompetent cuvette with the cuvette tilted.
- d) From here, it is important to perform each step after another and make sure the cells are on ice all the time except when it is inside of the electroporator. Take a 1000 μL pipette and take up 900 μL autoclaved LB.
- e) Tap the cuvette on Kimwipes to get rid of air bubbles trapped inside. Dry the outside of the cuvette with Kimwipes. Insert the cuvette in the holder and slide them into the electroporator (Figure 2.2B). Double press 'Pulse' and then you should hear a beep and see a time constant showing on the screen.
- f) The time constant reflects the rate of voltage decaying to 1/3 the original set voltage. The ideal time constant of transformation differs by cell type and cuvette. For our purpose, the time constant is a measurement of how good the electrocompetent cells are. The ideal time constant is around 5.0.
- g) Take the cuvette holder out as soon as possible. Open lid on the cuvette and add the 900 μL LB into the cuvette right away. Gently mix several times.
- h) Pipette out the cell culture into a Falcon tube and incubate at 37 °C with shaking for 45-60 min.

- i) Make triangles with Pasteur pipettes and sterilize with ethanol and flame 3 times. In 3 agar plate pipette 2, 20 and 200 μL of cell culture, and 200, 180, 0 μL of LB, respectively. Thus, we will have a 100X, 10X and 1X dilution of the cell culture. Use the triangle Pasteur pipettes are spread out the cell culture at the middle of the plate evenly across the whole plate.
- j) Incubate these plate at 37 °C overnight without shaking. You should take the plates out after 15-20 hours, depending on the cell type. These plates can be saved in 4 °C fridge for up to a week.

2.3.6 SDS-PAGE

SDS-PAGE is a technique to separate and visualize protein content by its molecular weight. SDS-PAGE is similar to MALDI but it is less sensitive and accurate. However, SDS-PAGE can be used for quantitative measurements while MALDI cannot. PAGE is short for polyacrylamide gel electrophoresis, in which an electrical field is applied across the gel. In an electrical field, the movement speed of a molecule depends on its total net charge. However when a protein is folded, the three dimensional structure buries some of the side chains so that their charges are shielded. Thus for a folded protein, its movement through a gel depends on the fold, charges on the outer surface, amino acid composition, the electrical field and not so much on its molecular weight. In SDS-PAGE, this problem is solved by denaturing the proteins before running a gel.

The SDS-PAGE sample is usually denatured with a thiol molecule such as DTT or β -mercaptoethanol that reduces the disulfide bonds within the proteins. The proteins then lose their tertiary structure and become more linear molecules. SDS stands for sodium dodecyl sulfate. It is a surfactant that is commonly used as detergent. It is added in the gel as well as in the running buffer for SDS-PAGE. It helps proteins remain their linear form as well as coating the protein with

negative charges (because SDS binds uniformly to all side chains). Thus, the intrinsic charges on side chains (R) will not affect the movement of the protein. Because SDS binds uniformly to the whole protein, the number of total negative charges increases as the number of residue increases. Proteins in PAGE are then separated solely by their molecular weights. Linear and negative charged proteins move through the acrylamide gel matrix towards the positive anode. Proteins with different sizes have different frictions and interactions with the matrix. Smaller ones travels faster than larger proteins.

Polyacrylamide is formed through a polymerization reaction with acrylamide and bisacrylamide. Polyacrylamide is chemically inert so it will not react with proteins or small molecules. However, the monomers can be carcinogenic in the powder form so we work with acrylamide and bisacrylamide in a solution in the hood. If only acrylamide is used, we will get long chain polymers without pores. When acrylamide is polymerized with bisacrylamide, we can get gels with pores that allow proteins to travel through. The sizes of the pores in such gel can be controlled by the final concentration of the acrylamide and bisacrylamide polymer reaction. The polymerization reaction is catalyzed by free radicals, generated from $\text{NH}_4(\text{SO}_4)_2$ and TEMED in solution. $\text{NH}_4(\text{SO}_4)_2$ solution can degrade so it should be made fresh to ensure good radical generation and acrylamide polymerization.

Precasted SDS-PAGE gels can be purchased and used directly. However, these gels have a limited shelf life and are expensive. With the small amount of gels that we use in the group, it is more economically efficient to stack gel in house. We use a Hoefer SE260 setup that allows us to cast gels right before use. Precasted gels can be used on this setup too. It is recommended to learn the gel stacking skills from a senior student because the amount of details in this process. Here I provide a general description and protocol.

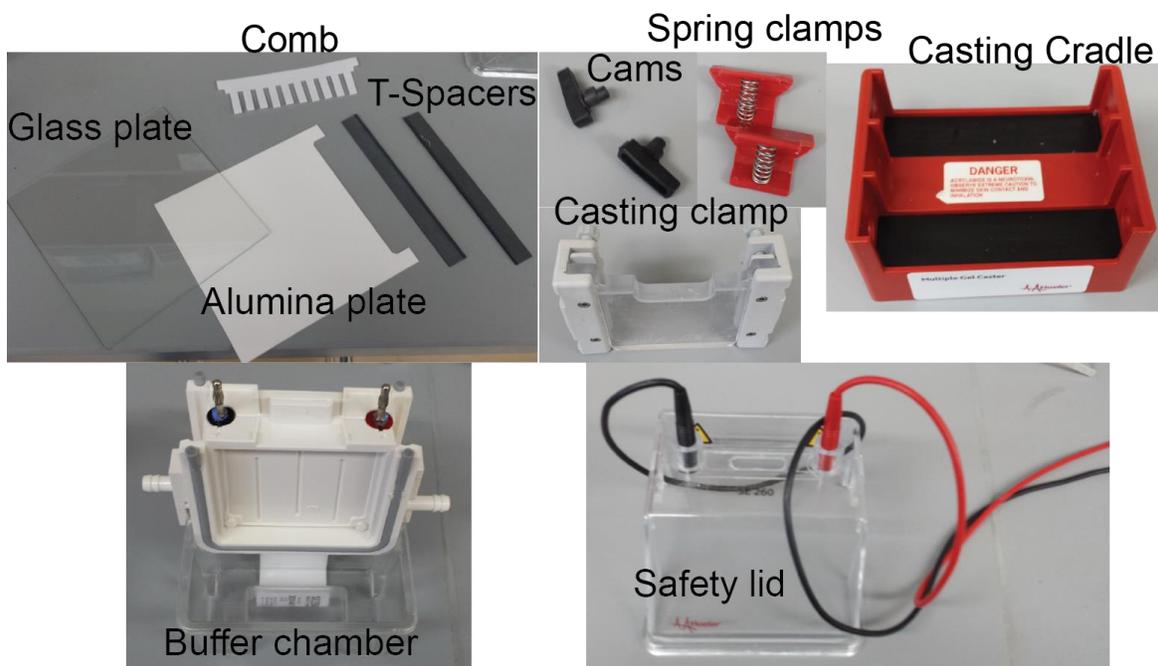


Figure 2.3 Parts of gel casting setup.

- 1) **Parts of the gel casting setup are shown in** Figure 2.3. Place two spacers in between of an alumina plate and a glass plate. Apply seal gel on the spacers to prevent leaking. Level the bottom of the two plates. Put this assembly in the casting cradle with 2 mm of the plates sticking out of the bottom of the cradle (Figure 2.4A). Tighten the screws evenly. If the screws are too tight, the glass cover can crack.

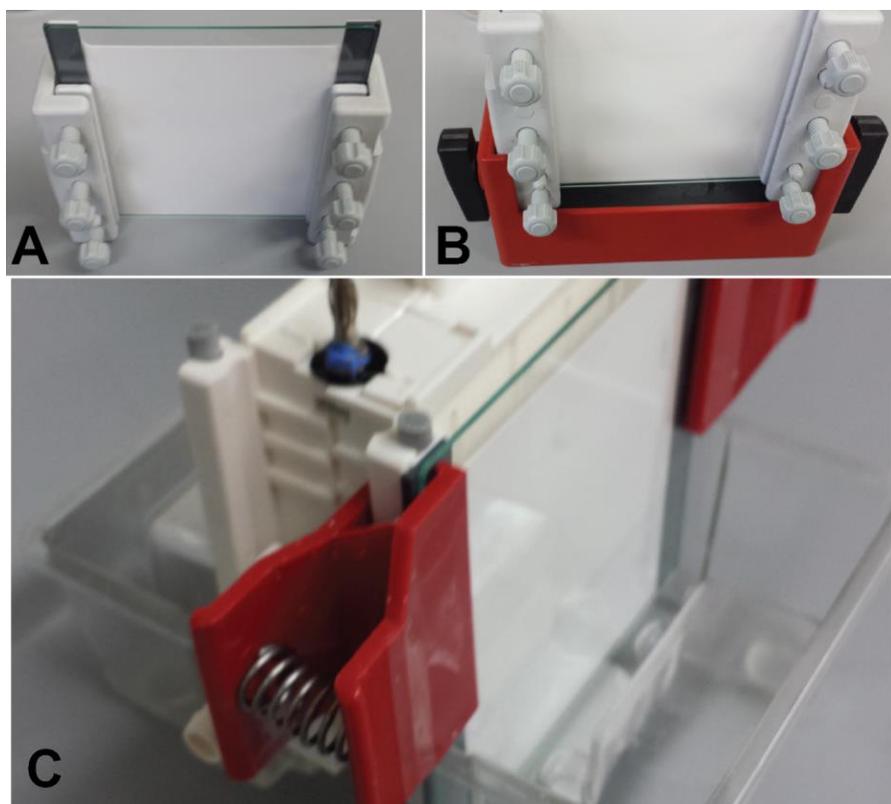


Figure 2.4 Picture illustration of gel casting setup.

(A) Glass and alumina plate with spacer in between. (B) Casting cradle on sealing gasket. (C) Glass and alumina plates in buffer chamber before buffer added (gel is not formed in this picture).

- 2) Put the casting cradle on top of the sealing gasket. Put one cam on each side of the gasket (Figure 2.4B). Twist cams from pointing bottom to pointing top (twist both at the same time).
- 3) Test the setup with H₂O (optional). Add H₂O into the space between the alumina plate and the glass plate. If the water level does not change within 2 min, the assembly is considered water seal. Pour the water out and dry the inside of the assembly with filter paper that is folded once.
- 4) There are two layer of the SDS-PAGE gel. The bottom layer is called the “resolving gel”, which is used to separate the proteins. This layer uses around 10-20% acrylamide and

contains small pores that allow shorter protein to pass faster than the longer ones. Higher concentrations of acrylamide in the resolving gel gives better separation of smaller proteins. For example, 15% acrylamide gel is good for proteins 5-100 kDa, while 10% gel is good for 20-300 kDa. Here I am using a 15% resolving gel recipe (for one gel). If two gels are desired, double the amount of chemicals for resolving gel but use the same amount of chemical for stacking gel (i.e. the recipe for stacking gel here is good for two gels).

- a) Weight out around 0.02 g of ammonium persulfate in a 1.7 mL microcentrifuge tube. Dissolve the solid with 100 μ L H₂O to make 10% NH₄(SO₄)₂. In a 50 mL centrifuge tube, mix the ingredients for 15 % resolving gel following the order in Table 2.3 below. Carry out this reaction in the hood. Mix the solutions gently by reversing the tube several times. Avoid bubbles.
- b) Use a Pasteur pipette to transfer the solution into the space between the alumina plate and the glass plate until ~1 cm below the top of the alumina plate. Avoid bubbles.
- c) Top the polymer solution right away with butanol. Butanol does not mix or react with the polymer solution. Let the gel sit for 30 min in the hood.

Ingredient	15% resolving gel	4% stacking gel
H ₂ O	1.65 mL	1.65 mL
1.0 M Tris, pH 8.8	2.55 mL	-
0.5 M Tris, pH 6.8	-	0.63 mL
40% Acrylamide:Bis-Acrylamide (37.5:1)	2.55 mL	0.26 mL
20% SDS in H ₂ O	34.5 μ L	12 μ L
10% ammonium persulfate	34.5 μ L	19 μ L
TEMED	4.5 μ L	2.4 μ L

Table 2.3 Recipe for SDS-PAGE gel.

- 5) The top layer is called the “stacking gel”, which is to make sure that all the proteins in the sample have entered the gel and share the same starting point.

- a) Pour out the butanol and rinse the space with H₂O. Dry carefully with folded filter paper. Insert the 10-well comb in the space and make sure it is even on both sides.
 - b) In a 15 mL centrifuge tube, mix the ingredients for 4% stacking gel following the order in Table 2.3. Use a new Pasteur pipette, transfer the polymer solution from one end of the comb while tilting the setup towards the same end. This practice is to avoid bubbles trapped under the comb. As the amount of solution in the comb increases, tilt the setup back slowly to fit more solution. Set the setup straight and fill the wells with more polymer solution when every well under the comb is filled with some polymer solution.
 - c) Let gel sit for another 30 min before use.
- 6) Release the cams to take the casting cradle out of the sealing gasket. Loosen the screws evenly to take out the assembly of alumina plate, spacer, and glass plate.
 - 7) Insert the assembly with gel into the buffer chamber. Use two clamps to fix the position of the plates against the sides. There will be a space between the plates and the back side of the buffer chamber. Fill this space with SDS-PAGE running buffer (PAGERB) and then fill the rest of the buffer chamber with running buffer too (Figure 2.4C).
 - 8) Mark the position of the wells with a sharpie on the glass plate. Pull the comb out of the plates using a straight up force. The wells should be filled with running buffer (PAGERB) automatically.
 - 9) The gel is now ready for loading samples and running.
 - 10) The samples for SDS-PAGE can be purified protein in solution, protein aggregates, or lysates. Our Hoefer gel setup allows loading of maximum 20-25 μ L of solution and no

more than 1 μg of target protein per well for best resolution. For lysate sample, protein concentration is hard to estimate but 10-20X dilution usually works for me.

- 11) Stock 5X loading buffer (PAGELB) does not contain thiol. In the hood, mix 95 μL loading buffer and 5 μL β -mercaptoethanol. For every 8 μL of the sample, add 2 μL of 5X loading buffer with β -mercaptoethanol. Boil the sample at 95 $^{\circ}\text{C}$ for 5-10 min.
- 12) Load 10-20 μL (depends on protein concentrations) of samples in each well carefully. Samples should automatically sink to the bottom of the wells because the loading buffer contains glycerol that is heavier than H_2O . Load 3-5 μL of molecular weight marker in a well without samples.
- 13) Put the safety lid with high voltage leads on top. With one hand, insert the leads into the power supply. Match the colors.
- 14) With one hand, turn on power supply. At constant (K) voltage mode, run the gel at 100-120 V, 400 mA for an initial 60 min. At constant voltage mode, the ampere setting is the maximum value allowed for ampere. The loading buffer containing dyes that marks where the front of the samples are. Allow the dyes to travel to about 1 cm above the bottom of the gel for good separation. If the dyes have not traveled to correct position at the end of 60 min, run the gel for longer. Gel and running buffer composition can affect the length required for the electrophoresis.
- 15) When the electrophoresis is finished, turn off the power supply and unplug the high voltage leads with one hand. Take off the lid and pour out the running buffer. Remove clamps. Gently open the glass plate and the alumina plate. Rinse the gel off with water into a plastic box. Pour out the water from the box and soak the gel with gel destaining solution (PAGEDS) for 30 min. The acid in PAGEDS can help fix the proteins in the gel. Replace

PAGEDS with staining solution (PAGESS). Let the gel stain for 2-12 hours. Pour out the staining solution and rinse the gel with water. Destain the gel with PAGEDS. This process takes around 1-5 hours.

16) The SDS-PAGE gel is now finished and ready for picture and analysis.

2.3.7 Point mutation using a commercial kit

In biochemistry, mutations are often used to study the function of particular residues in the sequence. Mutations may perturb the native protein structure and stabilize or destabilize the structure. A single residue mutation in a protein can usually be created by replacing one or two base pairs at the desired site. These mutations can be created in house when the mutation site is close to a restriction enzyme site. This method is usually referred to as “traditional PCR with modified primers”. The section below on “Adding a long piece of DNA into a sequence” uses the same principles. As long as the correct primers are chosen, you can generate mutation sites using the protocol in the following section. However, most of the time the mutation site is far away from the restriction enzyme site.

Another method of site-directed mutation is called “primer extension”. This method uses two overlapping primers that contain the base pair mutations, as well as two primers that cover a restructured enzyme site. The drawback of this method is that it requires two PCRs, and thus more time. Both “traditional PCR with modified primers” and “primer extension” are very powerful mutagenesis methods because they can introduce or delete a long sequence as well as a point mutation.

Sometimes, only a point mutation in a plasmid is desired. Unfortunately, most of the time, the mutation site is not close to the restriction enzyme sites. A Site-directed mutagenesis kit is a good option in this situation. I used the QuikChange Multi Site-Directed Mutagenesis Kit

(QCMSDM, Catalog # 200515) to make single and double mutations for our hIAPP plasmid. Unlike the QuikChange Mutagenesis Kit that requires primers on both sequence of the mutation site, the QCMSDM only requires one primer per mutation site. The actual enzyme used in this kit is a trade secret but the chemistry behind it is very interesting. Using a plasmid DNA as the template and a primer containing the mutation, the PfuTurbo DNA polymerase amplifies the whole mutated plasmid. A secret ingredient in the enzyme blend provided by the company seals the nicks in the mutated single strand-DNA.

If a transformation is made at this point, the mutation yield will be very low. Template DNA is methylated and is preferred by the cells, while the mutated DNA generated by PCR is not methylated. To increase mutation yield, an enzyme called Dpn I is used to treat the PCR products. This restriction endonuclease targets specifically methylated DNA, which is the original template DNA in the PCR product. The template DNA will be digested by Dpn I and thus will not be transformed into the cells. This step largely increases the yield of mutagenesis. The QCMSDM comes with a detailed protocol that you can follow. Here I provide a modified protocol that is suited to our facilities.

2.3.7.1 Day 1

- 1) Sequence plasmid template (A protocol for sequencing is described in this chapter).
- 2) Design primer:
 - a) Find the mutation site in the template sequence. With the mutation site of 3 base pairs (bp) as the center, copy 17 extra bp on both sides of the mutation sites (total of 37 bp) so the T_m can be equal or higher than 75 °C.
 - b) Calculate the melting temperature of the primers according to the following equation:

$$T_m = 81.5 + 0.41(\%GC) - (675/N) - \% \text{ mismatch}$$

For example, if I want to mutate 2 bp and use a primer that is 33 bp long, with 50 % GC.

$$T_m = 81.5 + 0.41 * 50 - 675/37 - 2/37 * 100 = 78.5$$

- c) It is good to start with more bp when designing the primers because it is easier to delete bp than add them. If the T_m is too high, delete bp from both termini. If the terminus is ended with one or multiple A or T, delete them until the terminal bp is a C or G. A primer with C or G at the terminus will bind better than one ends with A or T.
 - d) Double check T_m after eliminating the extra bp.
- 3) Order primers (usually takes 1-3 days). There are multiple companies available. I have been using <https://www.biotech.wisc.edu/>. It is on campus and provides good service.
- a) Visit <https://www.biotech.wisc.edu/>
 - b) Go to “DNA synthesis”. “Submit an order”
 - c) Login with your UW logins. New users need to setup account at Biotechnology (address is on the website).
 - d) Follow the website instructions. Input the number of primers you want to order in the “50 nmole cartridge” box. Continue onto the next page.
 - e) Input the name of the primer and the sequence of the primer (5’-3’).
 - f) Submit order

2.3.7.2 Day 2

- 4) Prepare primers for PCR

- a) After obtaining primers from Biotech, read the label on the tube and look for the weight. It should be around 500-700 μg .
 - b) Spin the primers tubes at maximum speed in the bench top centrifuge for 10 min.
 - c) For every 1 μg , add 2 μL of sterile H_2O to make final concentration 500 $\text{ng}/\mu\text{L}$. Let the tube sit in dark for at least 10 min and no longer than 2 hours at room temperature.
 - d) Vortex the tube for 10 sec to mix. This solution can be stored at $-20\text{ }^\circ\text{C}$ if not used in the same day.
- 5) Quantify the plasmid with Nanodrop. Each PCR reaction needs 50-100 ng of template DNA. The equation for calculating plasmid concentration is:
- $$\text{Conc } (\mu\text{g}/\text{mL}) = (\text{A}_{260\text{nm}} - \text{A}_{320\text{nm}}) * 50 \mu\text{g}/\text{mL}$$
- 6) In a PCR tube on ice, setup reaction as following (with the same order) for total 25 μL .
 - a) 5 μL of Template DNA (to get 50-100 ng)
 - b) 0.2 μL of Primer
 - c) 1 μL of dNTP mix (thaw on ice)
 - d) 15 μL of H_2O
 - e) 2.5 μL of 10X QuikChange Rxn Buffer
 - f) 0.2 μL of QuikSolution (Optional. This solution is basically DMSO. It helps reducing secondary structures in GC rich sequences.)
 - g) 1 μL of enzyme blend (thaw on ice)
 - 7) Mix the solution well and try to avoid bubbles. Bubbles bursting can ruin enzymes.
 - 8) Place the PCR tube in the G-Storm PCR machine and set up the cycles as following. This reaction may take 4-6 hours.

# of Cycles	Temp (C)	Time	Note
1	95	2 min	Initial melting of the template DNA
30	95	20 sec	Detaching polymerase and melting of the DNA
	55	30 sec	Binding of polymerase
	65	30 sec*6	(30 sec per 1k bp) extension of primers. 3-4 min for 6k bp plasmid
1	65	5 min	Final extension of the primers
1	37 or 4 C	Infinity	Storage

Table 2.4 PCR program for plasmid amplification, mutagenesis.

- 9) Digestion of the methylated template DNA
 - a) Add 1 μ L DpnI to each reaction tube and mix gently
 - b) Spin down the mixture for 1 min
 - c) Incubate the mixture at 37 °C for 30-60 min.
 - d) Kill the enzyme by incubating at 80 °C for 3 min.
- 10) Electroporation of the PCR product into XL10 Gold electrocompetent cells (a protocol for electroporation is described in this chapter)

2.3.7.3 Day 3

- 11) Pick colonies (I usually pick 4-8 for each electroporation) from the plates and grow in 10 mL LB with appropriate anti-biotics. Grow overnight with shaking.

2.3.7.4 Day 4

- 12) Save aliquots of the LB overnight culture. Mix the culture 2:1 with sterile 50% glycerol in H₂O. Store aliquots in -80 °C freezer.
- 13) Mini prep the rest of the cells to harvest mutated plasmid.
- 14) Quantify plasmid and submit sequencing samples. Use T7 primer and T7 terminator (These two sites are in most commercially available vectors).

2.3.7.5 Day 5

- 15) Electroporate the correct sequences (according to sequencing results) into BL21 cells to test expression.

2.3.7.6 Day 6

- 16) Pick colonies in the morning and grow the cells in 10 mL LB.
- 17) Stack a SDS-PAGE gel (a protocol for SDS-PAGE is described in this chapter).
- 18) When the cells reach OD = 0.6, save aliquots mixed in with 50% sterile glycerol at -80 °C. Take out 1 mL in a new Falcon tube. Add IPTG to the rest of the solution to get to final concentration 0.5-2 mM. Incubate both at 37 °C with shake for another 3 hours.
- 19) Spin down 1 mL of all cells. Decant the supernatant.
- 20) Run SDS-PAGE gel. For example, I usually pick 8 colonies, 2 from each miniprep plasmid aliquots, labeled as #1-#8. The gel can be run as following.

Wells	1	2	3	4	5	6	7	8	9	10
Samples	#1	#2	#3	#3 uninduced	MW	#4	#5	#6	#7	#8

- 21) Look for the correct mass in the induced samples, which should not appear in the uninduced sample (well 4). The cells from the same miniprep plasmid should show same results.
- 22) The cell aliquots saved in step 18) that express the correct protein can be used for future expression. The plasmid aliquot corresponds to these cell aliquots can be used for expression.

2.3.8 Addition of long DNA pieces into a plasmid

This protocol describes how to add a fragment of DNA into a plasmid. This protocol can be used for adding a tag to a protein which gene is already incorporated into a plasmid. In this section, I use hIAPP as an example and add a Histag to the N-terminus of the protein. Note that

Histag can also be added to the C-terminus of a protein but then it cannot be cleaved using existing methods.

2.3.8.1 PART 1. Primer Design

- 1) Identify the insertion location and look for restriction enzyme sites. (You only need to look at one strand because the other side is complementary)
- 2) Below shows part of the plasmid sequence of hIAPP in vector pTXB1. The expression starts from the Bolded M.

```

gannaaattccctctagaataatTTTgtTTtaactTTtaagaaggagatatacatatgaaatgc
  X N S L - N N F V - L - E G D I H M K C
aacactgccacatgtgcaacgcagcgcgcccgcaaatttttagttcattccagcaacaac
  N T A T C A T Q R P A N F L V H S S N N
tttggtgccattctctcatctaccaacccgggatccaatacatactgcatcacgggagat
  F G A I L S S T N P G S N T Y C I T G D
gcactagattgccctacccgagggcgagtcggtacgcatcgccgacatcgtgccgggtgc
  A L D C P T R G R V G T H R R H R A G C
gcggcccaacagtgacaacgccatcgacctgaaagtccttgaccggcatggcaatcccg
  A A Q Q - Q R H R P E S P - P A W Q S R
gctcgccgaccggctgttccactccggcgagcatccgggtgtacacgggtgcgtacgggtcga
  A R R P A V P L R R A S G V H G A Y G R
aggtctgcgtgtgacgggcaccgcgaaccacccggttgttgtggttggtcgacgtcgccgg
  R S A C D G H R E P P V V V F G R R R R

```

- 3) In this case, I want to add a Histag (usually 6 histidines) at the N-terminus of the peptide. Sequence can be uploaded online to search for the restriction enzyme sites. Here, because it is the pTXB1 plasmid, I already know that catatg is the recognition site of NdeI. This enzyme cuts as:

```

5'... C A^T A T G ...3'
3'... G T A T^A C ...5'

```

- 4) Current the translated sequence is ...EGDIHMKCNTATC..., and the desired sequence is ...EFDIHMHHHHHHMKCNTATC..... If Histag will be cleaved using Factor Xa or other enzymes, an enzyme recognition site, such as IEGR, should be added before the target

protein. The second M in this sequence is not necessary. I added it into the sequence to cleavage the Histag with cyanogen bromide (CyBr).

- 5) Design primers for plasmid amplification. Two primers are required with one forward (5'-3') and one reverse (5'-3' on the compliment sequence). I like to add the extra base pairs (the Histag) with the forward primer but either way works.
- 6) Things to keep in mind while designing primers:
 - a) Melting temperature for both primers should be similar (within 1 °C preferred)
 - b) Overlap the primer with as much as the sequence, especially when there is a long sequence with no overlaps on the end.
- 7) Design forward Primer:
 - a) Copy the sequence from the restriction enzyme site and on:

```
catatgaaatgcaacactgccacatgtgcaacgcagcgcccggaatttttagttcattcc
H  M  K  C  N  T  A  T  C  A  T  Q  R  P  A  N  F  L  V  H  S
agcaacaac
S  N  N
```

- b) Translate the desired added sequence (you can use online tools or check with an *E. coli* codon usage chart):
 - i) Picking a high usage codon is good for expression in general.
 - ii) Met: atg (there's only one codon for Met)
 - iii) His: cat or cac
 - c) On 5' end of restriction enzyme cutting site, add at least 6 base pairs, in this example, I used ggtggt.
 - d) Check sequence and make sure you didn't introduce any other restriction enzyme sites that you might use in the future.
 - e) here's the final forward primer:

ggtggtcatatgcaccaccatcatcacca**catgaaatgcaacactgccacatg**

G G H M H H H H H H M K C N T A T

The bolded base pairs are the ones that overlap with the original sequence.

8) Design reverse Primer:

a) Copy the sequence before desired sequence:

tgtttaactttaagaaggagatatacatatg

V - L - E G D I H M

All the base pairs in this sequence will contribute to the melting temperature. The numbers of base pairs you pick will depend on the desired melting temperature.

You need to include the restriction enzyme site and add ggtggt after it for restriction enzyme binding:

tgtttaactttaagaaggagatatacatatg**ggtggt**

V - L - E G D I H M **G G**

b) get the reverse compliment of this sequence:

accacc catatgtatatctccttcttaaagttaaaca

c) Double check the melting temperatures.

For this case, I have 53.5 °C for the forward and 53.8 °C for the reverse primer.

9) Order the primers at biotech (<https://www.biotech.wisc.edu/>) and expect them to be ready in 2 days.

2.3.8.2 PART 2. Plasmid construct

10) When receive primers, first spin them at max speed (17,000 G) for 10 min. Add sterile H₂O to rehydrate the DNA. The amount of H₂O varies for each sample. I usually add 20* the nmol number to make 50 μM stock concentration. For example, for a sample with 61 nmol, I add 1220 μL of H₂O. Let the primers sit for more than 10 min after the addition of H₂O and then vortex for 5 sec (3 times) to mix the solution.

Plasmid PCR recipe (may vary for different plasmids depends on the GC content and plasmid length. Here's the recipe that has been working for me):

- a) 10 μ L HF Phusion or GC Phusion buffer (has been giving me the same yield)
- b) 1 μ L of dNTP
- c) 1.5 μ L of 10 μ M forward primer
- d) 1.5 μ L of 10 μ M reverse primer
- e) 0.5 μ L of plasmid template (around 100 ng/ml concentration)
- f) 1 μ L DMSO
- g) 34 μ L of H₂O (or however much to bring the total volume to 50 μ L)
- h) 0.5 μ L of Phusion enzyme

11) Run PCR program on G-Storm as following:

# of Cycles	Temp (C)	Time	Note
1	95	5 min	Initial melting of the template DNA
30	95	30 sec	Detaching polymerase and melting of the DNA
	48.5	30 sec	Binding of polymerase
	72	8 min	(or 30 sec per 1k bp) extension of primers.
1	72	10 min	Final extension of the primers
1	4	Infinity	Storage

Table 2.5 PCR program for plasmid amplification

- 12) Make a 0.3% agarose gel.
- 13) Run gel at 100V until good separation is achieved. Use DNA base pair ladder to determine the right size of DNA (two different ladders are used in Figure 2.5). Cut the right DNA band out (the brightest band in lane 2,3,5,6 in Figure 2.5) and purify with Gel extraction kits. Elute with 44 μ L of H₂O.

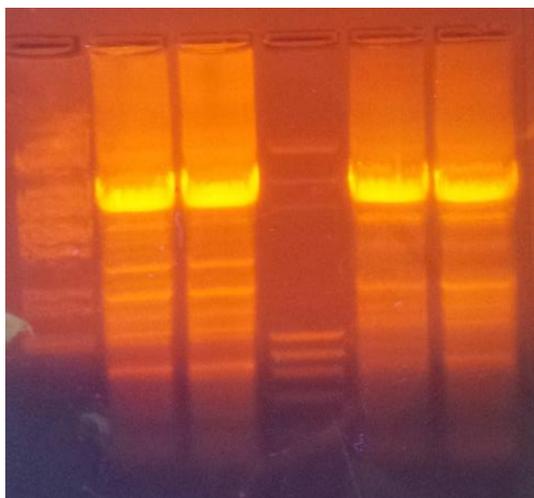


Figure 2.5 DNA agarose gel illuminated with UV lamp.

- 14) Set up NdeI cutting reaction:
 - a) 44 μ L DNA
 - b) 5 μ L of restriction enzyme cutting buffer (usually supplied with the restriction enzyme)
 - c) 1 μ L of restriction enzyme

Let reaction proceed at 37 °C for 1.5 hours. Kill the enzyme by heating at 65 °C for 20 min.
- 15) Clean up the reaction with PCR purification kits (same as the gel extraction kit). Elute with 35 μ L supplied buffer.
- 16) Setup T4 ligase reaction (or follow the user manual):
 - a) 18 μ L DNA (from previous step)
 - b) 2 μ L ligation buffer (vortex before use)
 - c) 0.3 μ L of T4 ligase
- 17) The user manual of T4 ligase will tell you to wait for only 45 min. However, the yield can be 10 times better if you let the reaction go overnight.

- 18) Kill the enzyme by heating at 65 °C for 20 min. You can store the DNA in -20 °C or electroporate after cooled on ice.

2.3.8.3 Part 3. Electroporation and test expression

- 19) Electroporate to XL1 cells (these cells grow slowly, might need 20-24 hours).
- 20) Pick colonies and mini prep
- 21) Submit samples for sequencing using T7 primers.
- 22) Electroporate the correct plasmids to BL21 cell for growth (~16 hrs is enough)
- 23) Pick colonies and grow in Falcon tubes (5 ml). You can pick up to 16 colonies for two gels. Save cell aliquots in 15% glycerol and store in -80 °C
- 24) Induce expression when OD600~0.5 by adding IPTG to 0.5 mM final concentration
- 25) Run SDS-PAGE gel with one uninduced sample, MW marker, and all the induced cells.
- 26) The cells that are able to express should have a band that is the correct MW (32K for HIAPP) that only shows up after induction.

2.3.9 Thioflavin T (ThT) fluorescence plate reader

ThT fluorescence assay is a traditional method for measuring amyloid fiber formation. ThT is a dye molecule that binds preferably to amyloid fiber or similar secondary structures. Upon binding, the fluorescent frequency of ThT red shifts from ~405 nm to 450 nm and its fluorescence intensity increases. By monitoring the fluorescence intensity of the ThT emission, we can monitor the amount of amyloid fiber presence in a solution.

I have used two different ThT plate readers. Although the instruments are 20 years apart in terms of production date, the usages are basically the same. Optimization is crucial to ThT plate reader usages because a bad calibration of the intensity can result in either saturation of detection or low sensitivity for the lower intensity wells.

Here I describe how to use the instrument EnVision Xcite by PerkinElmer in the Keck center. Training by a Keck Center TA is required before using this instrument. For a kinetics that is longer than 5 hours, I recommend testing the calibration samples and the setup of the instrument a day before the experiment.

- 1) Open operating software “Wallac-EnVision Manager” on the desktop (Figure 2.6).
- 2) The instrument should be in “idle” (Figure 2.6).
- 3) There are two tabs we will work with: Protocols and Labels. Labels are basically the data collection methods. Under “Labels”, find “Fluorescence Intensity”. Here shows the old methods I built, labeled as “Zanni Group”. You can either create one from scratch or create a copy of an existing label (Figure 2.6).
- 4) In side of the instrument, there are two filter holders (one for excitation and one for emission) that can be pulled out. Check the slot number of the desired filter. If the filter is not on the holder, it can be found in the drawers below the instrument. For ThT, we can use 430 nm for excitation and 486 nm for emission. They are not the ideal wavelengths for ThT but they will work for our purposes.

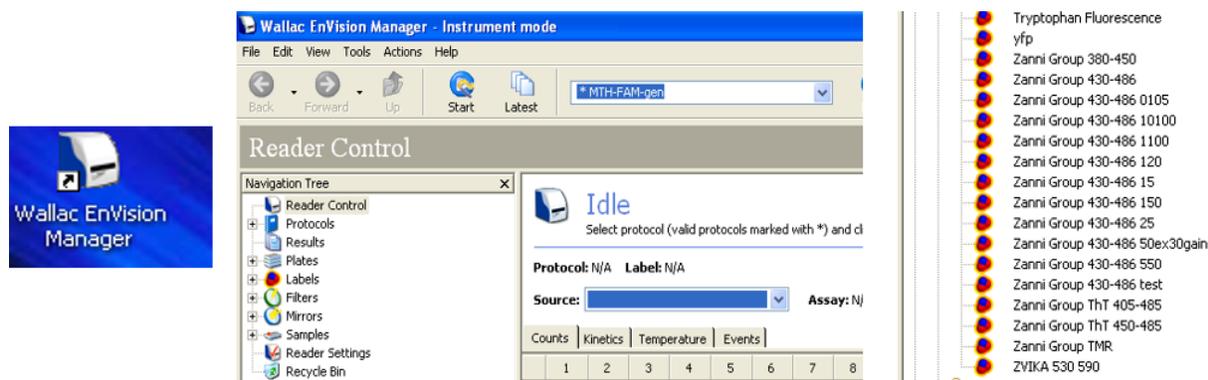


Figure 2.6 Screen print illustration of ThT plate reader program

- 5) In each label, there are many parameters we can use to optimize our measurement. Below is an example for method “Zanni Group 430-486”. Both the “Excitation” and “2nd excitation” are from the top because we will be using opaque plates. Top mirror should always be in Slot 1. The “Excitation filter” should be labeled CFP 430 – Ex. The slot number can be different each time.
- 6) The Emission Filter should be CFP – 486. There is no slot number shown in the example because the correct filter is not in the holder.
- 7) The rest of the parameters are crucial for the fluorescent result (Figure 2.7). Measurement height is related to the sample volume. When small sample volume is used, the measurement height should be changed according. If the measurement height is not correct, the fluorescence reading can be extremely high or low. Excitation light % and detector gain decide the range of detection. If both values are too low, samples with low fluorescence intensity will all read as 0. If both values are too high, the detector can saturate when the fluorescence intensity increases.
- 8) “Number of flashes” can affect the quality of the data. This value can be considered as the number of data averaged. The more number of flashes, the more reproducible each measurement gets, but also take more time for data collection. Same concept applies to the “number of flashes per A/D conversion”.

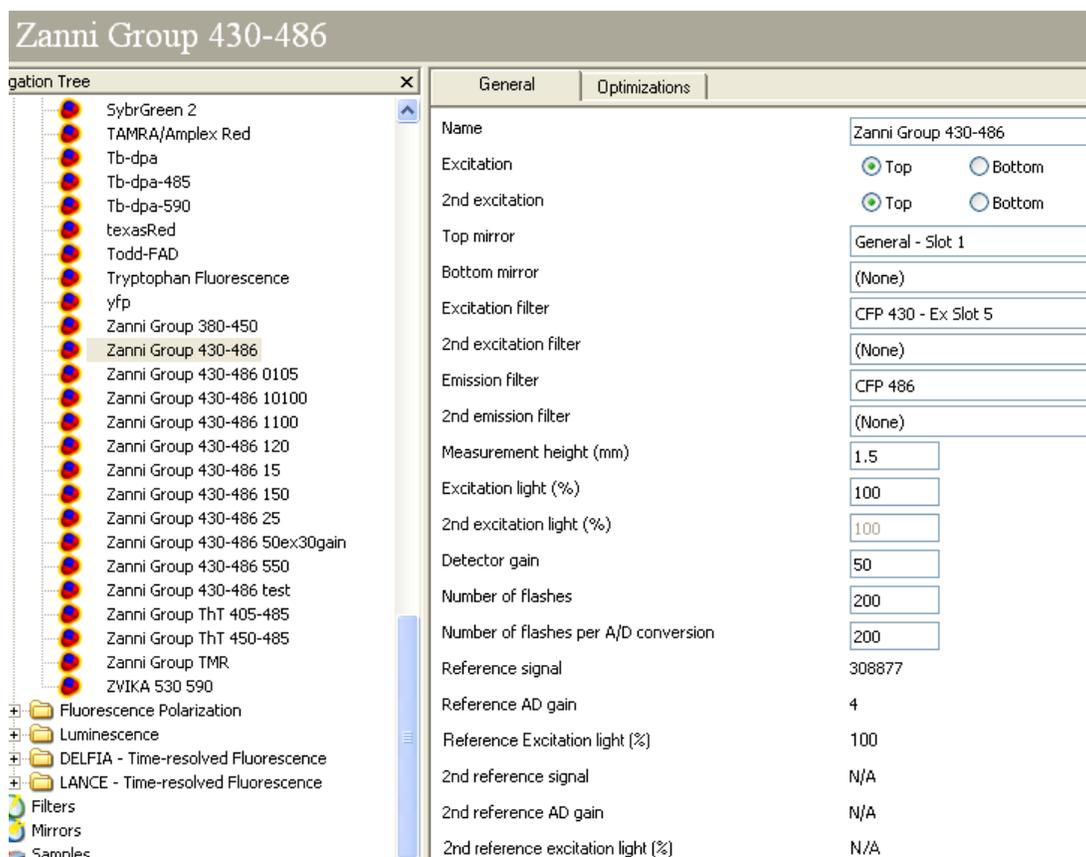


Figure 2.7 Screen print illustration of plate reader parameters

- 9) For each sample, the parameters for the label might be different. Thus, it is important to optimize the label. A 384-well plate is needed for the optimization and the same plate should be used for the measurement. Do not reuse these plates because ThT is quite difficult to clean off. Two solutions are required for optimization. One without ThT as the blank (BLANK) and one with ThT and amyloid sample (SAMP) as the highest intensity well. Prepare the amyloid sample with ThT ahead of time to make sure the peptides are aggregated (for spontaneous aggregating peptides). Add 50 μ L solution to the specified wells in the plate (follow instruction on the instrument. The wells should be 4 corners and one center).

- 10) For γ D-crystallin, the SAMP is made by UV-irradiated sample with ThT added. Because ThT is a fluorescence dye, I irradiated the sample without the presence of ThT. The ThT buffer was added after the irradiation.
- 11) Go to the Optimization tab and follow the optimization instruction. This step might take a long time.
- 12) Several labels can be added to one protocol. This means the instrument will measure the same sample multiple times using different parameters. I usually use at least 3 labels (each in a group as shown below) to make sure one of them can capture the correct range of fluorescence intensities. It is not necessary to optimize for every single label. I optimize one with moderate parameters and another two labels based on the optimized one. One of the two labels is more sensitive than the original one (higher excitation light) and one is less sensitive (lower excitation light and gain).
- 13) Figure 2.8 shows a typical protocol with several groups. One label can be applied to each group and multiple groups can be added to one plate. For each “Run”, each group will be measured once the same order they are in the assay.
- 14) There is an option for repeats. This option can be used for kinetics measurement. There are two concerns with using the repeats function. One is that the instrument is a shared equipment so there are other users might want to use it during the kinetics. The other concern is the evaporation of liquid. This particular instrument does not have temperature and humidity control so the solution can evaporate eventually if the plate is left open to air.

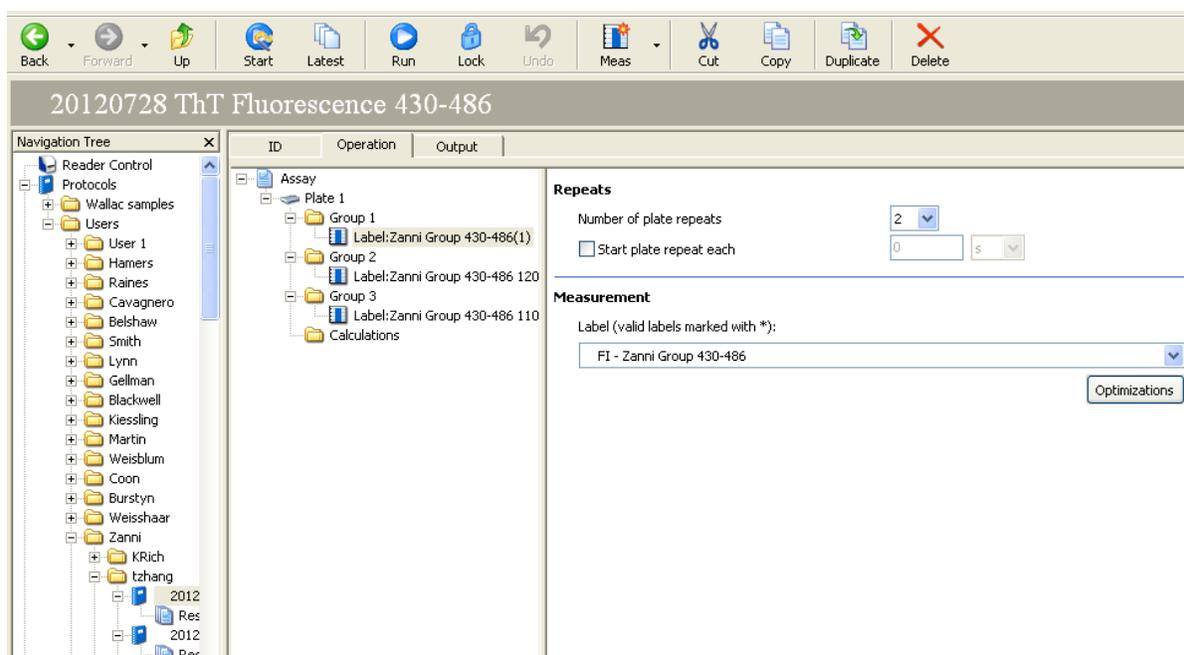


Figure 2.8 Setup of Assays and Groups

- 15) Figure 2.9 shows a typical result from a plate reader. Colors correspond to different intensities. The color scale can be set to fit a particular range of interest. The actual fluorescence intensity readings are reported as numbers (Table 2.6) and exported as an excel file. Note that not every single well is scanned in this protocol to speed up the scanning time. Wells to be scanned can be specified in the protocol page.

- 16) The sample preparation is important to ThT fluorescence experiments. The ideal ratio of peptide and ThT is calculated and established in the literature. However, these conditions mostly apply only to samples with μM level concentrations. Because concentration play a role in the kinetics of peptide aggregation, especially in hIAPP, we want to measure ThT fluorescence of hIAPP using the same concentration as the 2D IR samples (0.5-1 mM). I have measured ThT fluorescence for 1 mM hIAPP. It is possible with very low ThT concentration but the ideal ratio is not achieved. Here are some concentrations I used before: 10 μM of γD -crystallin with 25 μM ThT; 0.4 mM hIAPP with 4 μM ThT.
- 17) Finally, I want to show some examples of a kinetics measurement. In Figure 1.1Figure 2.10A-C are the measurement of the same sample. The difference of the traces come from the difference parameter settings for “excitation light (%)” and “detector gain”.

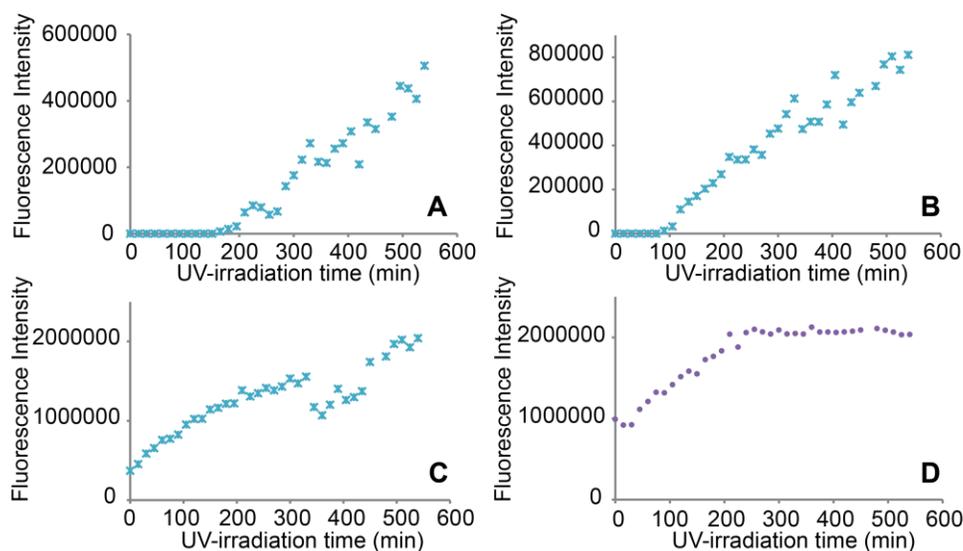


Figure 2.10 Kinetics plots of UV-irradiation induced γD -crystallin amyloid fiber formation.

Plots (A-C) are for the same sample with different parameters. (A) Low excitation light (%) with low detector gain. (B) Low excitation light (%) with medium detector gain. (C) Medium excitation light (%) with medium detector gain. Plots (A) and (B) are not sensitive enough for the lower intensity samples. (C) The parameters are good for the samples. (D) Kinetics of UV-irradiated WT γD -crystallin. Saturated ThT fluorescence kinetics around 200 min due to high values of the parameters.

2.3.10 ThT fluorescence of tissue samples

Measurement of the ThT fluorescence of tissue samples is done in the Professor Jim Weisshaar's lab. The lens tissues are sliced to 25 μM thick and deposited onto a coverslip. A hybridization chamber (9.8 mm x 20 mm by Grace Bio-labs) is attached onto the coverslip so that the lens tissue is placed in the center of the chamber. For UV irradiation, add 100 μL phosphate buffer or H₂O in the chamber before irradiation. The solution can evaporate after the UV irradiation because the hole on the chamber are left open. Add the same solution to the native lens as a control. Acid can quench ThT fluorescence so a pH 3 buffer with small buffering capacity (1-10 mM of phosphate or tris) should be used to treat the sample. After acid treatment, the solution will evaporate. Before measurement, add 100 μL phosphate buffer containing 25 μM ThT (pH 7.4) to all samples. Measure pH of the acid treated sample after add ThT buffer. If the pH is lower than 5, the fluorescence intensity of the sample can be quenched (larger than native samples but will not correctly reflect the amount of fibers). A 405 nm laser is used for excitation and a 482 nm filter is used to measure the emission. For each sample, a phase contrast image is recorded with correct focus before switching to the fluorescence mode. Collect at 3 images at different spots for each sample.

2.3.11 Columns

This section is a general introduction of different types of columns I have worked with. One distinction between columns is that some of the column works in denaturing conditions (reverse phase) so these columns are good to use when purifying small peptides that refold easily or are intrinsically disordered (hIAPP and A β). Some of the columns can be used to purify proteins without denaturing them, for examples ion exchange columns and size exclusion columns. The

drawback of these columns is that they usually do not have a good peak resolution so it is common to combine the purification power of two or three different columns for one protein.

Before introducing the columns, there are some tips of using and maintaining the columns that apply to basically all of them. First, never run the column dry. This does not mean it is the end of world if this that happens. Most of the column we use can be regenerated if they are run dry but it is always good to avoid such situations. However, sometimes regeneration cannot bring a column back if the bedding is disrupted (mostly apply to the reverse phase columns). Second, always check the compatibility before using the columns with a new buffer. Third, always spin down or filter solutions before injecting onto the columns. Solids and insoluble material can clog the column. This practice can elongate the lifetime of a column and prevent contaminations from aggregates leftover. Last, only run the columns at or lower than the recommended flow rate, as well as monitor the pressure levels. It is possible to ruin the columns' packing by running them too fast or with high pressure.

2.3.11.1 Immobilized metal ion affinity chromatography (IMAC)

IMAC may also be referred to as Ni-NTA columns, although other metals like cobalt and copper can also be used instead of nickel. NTA is the short hand for nitrilotriacetic acid. It is a strong chelating agent that binds to metal ions. In IMAC, NTA is attached onto agarose beads that function as immobilized support for the column. The Ni-NTA column can be used to purify proteins containing a tag of 6-10 histidine. Protein containing the Histag will bind to the metal ions that are chelated to the NTA resin, while the rest of the lysate flow though the column. After extensive wash, the target proteins can be eluted with buffer containing imidazole which competes with the Histag binding to the metal ions. Because of its high binding capacity and efficient purification, Ni-NTA is very often used as a single step protein purification.

Both commercially packed Ni-NTA column or Ni-NTA resins are available for purchasing. The packing of a Ni-NTA column does not affect the purification efficiency so either approach should provide identical yield and purity. Other metals can be used if there are Ni-specific binding proteins other than the target protein in the lysate. Ni can be stripped with a solution containing EDTA. Different metal ions can be reloaded onto the column in a solution when all the EDTA molecules are washed out from the column. Below is a general protocol for using a Ni-NTA column to purify proteins with Histag.

- 1) Equilibrate column with NiRB for at least 10 CV (column volume)
- 2) Adjust lysate pH to 7.5 or 8. Save one aliquot of lysate for SDS-PAGE gel. Load lysate slowly at <math><1\text{mL}/\text{min}</math>. Collect the flow through (NiFT) in case the binding is not sufficient. Save one aliquot of NiFT for gel.
- 3) Wash column with NiRB for at least 10 CV. If a detector is available, monitor 280 or 220 nm OD intensity until the intensity reaches a flat line.
- 4) To elute the target protein, you can either use gradient (if pump allows) or step gradient. When using a gradient, run the column using 0-100% NiEB in 100 min. Most target proteins will elute around 50 min. When using a step gradient, you can pick particular percentage of NiEB for each step depending on the protein and the impurity. For example, you can use 40 mL of 10%, 25%, 40%, 60% and 100% for Histagged γ D-cyrstallin. 10 and 25% NiEB can further clean the column while most of the target protein will elute with 40 or 60% NiEB. Use 100% NiEB to clean the column.
- 5) Save an aliquot from each percentage elution collected for a gel. Use a SDS-PAGE gel to decide which collection contains the purified target protein.

- 6) There is still high imidazole concentration in the collection with the target protein. Now you have many choices of what to do next depending on the protein you are working with. First, you can simply dialyze the protein into a desired buffer and store it frozen or lyophilized. Second, you can concentrate it down and perform buffer exchange using a spin column if the protein is large enough. This method is more violent so some protein might precipitate out. The perk of this method is that you can obtain protein with much higher concentration than what you started with. Finally, you can also inject the solution with imidazole onto a size exclusion column. This column can both further purify your protein as well as putting your protein in a new desired buffer (see SEC section below).

The metal ions can be reduced or washed away after extensive uses. Regenerating NTA columns is easy and efficient. This following protocol describes how to regenerate a Ni-NTA column but it can also be applied to other metal ions.

- 1) Run NiSS through column for around 5 CV. Collect the solution for special Ni wastes.
- 2) Rinse the column with H₂O for at least 20 CV.
- 3) Load the column with NiSO₄ solution (or NiCl₂) for at least 5 CV. Collect the solution for Ni waste.
- 4) Rinse the column again with 5 CV of H₂O and then equilibrate the column with NiRS. The Ni-NTA column is now ready to use.
- 5) If another metal ion is desired, you can simply replace the NiSO₄ solution with the solution of the desired metal ion. Note that heavy metals should not be poured down the drain.

2.3.11.2 Chitin binding chromatography (CBC)

The CBC is also a kind of affinity column mainly sold by the New England Biolabs (NEB). Chitin is a polymer of N-acetylglucosamine. It can be made from the shell of crabs. There

are proteins that has a chitin binding domain (CBD) that binds specifically to chitin. The pTXB1 vector (from NEB) has the CBD in the expressed protein to aid the purification. CBD allows on-column cleavage that simplifies protein purification. However, there are some drawbacks of CBC. First, the CBD has relatively weak affinity to chitin, comparing to Ni-NTA affinity for Histag. Thus the purification from CBC is not going to be perfect. Second, along with the low affinity, the chitin resin has much lower binding capacity than Ni-NTA. For comparison, with the same volume of resin, Ni-NTA can bind around 20-100 times more protein than chitin resin. Finally, chitin resin has much shorter shelf life. The stripping of chitin resin is usually not complete. It is recommended that the same batch of chitin resin should be only used for the same peptide to avoid cross contamination. The same batch of chitin resin can be used for around 5 times before the purification yield drops dramatically. Below is a general protocol for CBC that is used to purify hIAPP in the Zanni lab. You can visit NEB's website for more information. Note that the buffers used to purify γ D-crystallin are different.

- 1) Equilibrate column with 10 CV of H₂O followed by 10 CV of ChihIAPPB.
- 2) Adjust lysate pH to 8 and load slowly at <1mL/min. Save an aliquot of lysate for SDS-PAGE gel. Collect the flow through and save an aliquot for gel.
- 3) Wash column with 10 CV of ChihIAPPB. Collect wash solution and save an aliquot for gel.
- 4) Rinse the column of 2 CV of ChihIAPPC without DTT. (This step is only required when you need an amidated C-terminus on the peptide. For peptides without the C-terminus amidation, use CBCCB without ammonia bicarbonate.)
- 5) Flush column with ChihIAPPC with DTT for 5 CV (save this solution for step f). Leave the column in fridge with the top open.

- 6) After overnight cleavage (the cleavage time varies depending on peptides and reagents), elute the column with the 5 CV of ChihIAPPC used to flush the column. Use an additional 2 CV of ChihIAPPB to elute the column.
- 7) The elution should contain the target peptide with an amidated C-terminus.
- 8) The chitin column now can be cleaned with 0.1 M NaOH solution. Rinse the column with 5 CV of ChiStrip and let it sit in the fridge for 30 min. Rinse again with 5 CV of ChiStrip followed by 10 CV of H₂O.
- 9) You can either store the column as is or store it in 20% ethanol if it will not be used again for a while.

Cleaving on column with ammonia bicarbonate generates bubbles over time. It is important to leave the column open during incubation. This is also why chitin columns are usually not sold as prepacked columns. You can use disposable gravity columns for chitin resin.

2.3.11.3 Ion Exchange Chromatography (IEC)

Ion exchange chromatography separates amino acids, peptides and proteins based on the total charge on these molecules. Side chains on the amino acids can be charged or not charged. Thus, the net charge on a protein or a peptide can be positively charged, negatively charged or neutral given the pH of the solution. The pH that a molecule can be neutral is called isoelectric point (pI). When pH of the solution is greater than the pI, the protein is negatively charged. There are two kinds of IEC, anion (positively charged stationary phase) and cation (negatively charged stationary phase). The charges come from the functional group from the resin beads (usually sepharose). These two types of columns have different selectivity towards the proteins and work at different pH values. Usually one type is selected against the other based on the stability of the target protein. To use an anion column, the pH of the solution should be higher than the pI of the target protein;

for cation columns, the pH should be lower. If the protein is more stable in higher pH than lower ones, then anion column should be used.

Different functional group for the IEC can provide various binding affinities. The ideal functional group should bind tighter to the target protein at the running pH than to the other proteins. However, this property is difficult to predict. I usually start with high binding affinity resins and move to weaker affinities if the first try does not work.

The purification yield and efficiency of IEC highly depend on the elution method. There are two ways to elute IEC column, by changing the pH value or by changing the salt concentration. If an anion column is used, proteins that bind to the column elute when the pH value of the running buffer decreases. When the pH decreases, the charges on the proteins become less negative and bind less tightly to the positive resin. However, eluting the IEC by changing salt concentration is used more frequently because some protein may precipitate when pH is at their pI. In a solution, the Na^+ and Cl^- ions are attracted to the negatively and positively charged groups on the biomolecules, respectively. When the salt concentration increases, the ions around the charged group increases. Ultimately, the ions are surrounding the charged group so much that a negatively charged group will appear neutral or positively charged to the positive stationary phase. Thus, proteins can be eluted as a function of increasing salt concentration. This protocol below is used to purify αB -crystallin using Q sepharose beads (anion column). For different proteins, the buffer choice may be different but the overall concept should be the same.

- 1) Clean column with IECC for at least 5 CV at 5 mL/min (the flow rate maximum is set the manufacture). Due to the binding mechanism, IEC takes long times to clean and run.
- 2) Wash the IEC with H_2O for 5 CV at 5 mL/min. Equilibrate with IECA for at least 20 CV.

- 3) Adjust the pH of the lysate to 8.5. The pI value for α B-crystallin is around 6.3. Both the pH of the running buffer and the lysate should be at least 2 points higher than the pI value of the target protein. This step is crucial for binding of the protein.
- 4) Load lysate onto the IEC at 1 mL/min. Save one aliquot of the lysate for SDS-PAGE. Collect the flow through of the column, save an aliquot for gel.
- 5) Wash the column with IECA at 5 mL/min for at least 1 hour. If the column is connected to a detector, the 220 nm intensity should decrease and become a flat line.
- 6) Elution of IEC can be done with a gradient or stepwise. If the column is connected with an HPLC system (it can be), then a gradient elution is possible. Use 0-40% IECB in 2 hours. The target protein should be eluted at 10-20% IECB. If an HPLC system is not available, the elution of IEC can be treated similar to a Ni-NTA column. Use 30-50 mL of 5%, 10%, 20% and 30% IECB for step elution. The gradients and steps can be extended to higher percentage if the elution point of the protein is unknown. Collect all elutes in aliquots. Check the aliquots with SDS-PAGE. I personally find step elution more efficient.
- 7) Flow rate can affect the resolution of the elution. Start with 5 mL/min and modify accordingly.
- 8) Elute all protein with 100% IECB after eluting target protein. Collect the elution and save an aliquot for gel.
- 9) Wash column with IECC and H₂O as in step a). For long-term storage, wash column with 5 CV of IECS and store the column at 4-25 °C.

2.3.11.4 Size Exclusion Chromatography (SEC)

SEC is also known as gel filtration chromatography. It separates molecules base on their sizes. Different from SDS-PAGE, SEC is a mild enough technique that the biomolecules remain

in their native conformations while in the SEC column. The size dependent separation is referring to the three-dimensional sizes of the molecules. Because this method is so mild, the weakly associated protein dimer, trimer and other oligomers can be separated from the monomers. The mechanism of separation using SEC is different from IEC. Biomolecules do not bind to the resins in SEC. Instead, the resins used in SEC are porous so small molecules can get trapped in the pores and interact longer with the beads. Large proteins, however, cannot fit into the pores so that they simply pass among the beads. Thus, larger molecules elutes faster than smaller molecules. In fact, the last peak on a SEC, also a mark of the end of a run, is usually the “salt peak” (Figure 2.11, all of the salt molecules that are not in the running buffer).

Because the interaction between molecules and the resin of SEC is not due to binding, a gradient using a different buffer will not affect the elution. For SEC, usually only one buffer is needed. The buffer pH, salt concentration, flow rate and running temperature can all be optimized to improve the peak resolution in elution. Note that SEC is not a high-pressure column. The flow rate for this column should never exceed 1 mL/min. High flow rate or pressure can destroy the column packing, which results in bad resolution, long tailing and asymmetric peaks. Repacking is the only way to fix such problems.

SEC has a narrow range for purification due to the column mechanism. Resin grades decide the range of the molecules that a SEC can separate. I use Superdex resin, which is a matrix of dextran and agarose. Superdex 75 can be used to purify proteins (3K -70K Da). Superdex 30 purifies smaller peptides up to 10 K Da but it can be used for desalting. Superdex 200 has a range of 10K to 600K Da.

Here is a general protocol for purification on HPLC using Superdex 75 packed column. For a new protein system, this protocol can be used as a starting point to find the optimal purification conditions.

- 1) Equilibrate SEC column with SECR for 2 hours at 0.4 mL/min. Monitor the 220 nm and 280 nm until they are both flat.
- 2) The injection for SEC cannot be lysate. The injection solution should be the elution from at least one round of purification (Ni-NTA or IEC). The injection can be very concentrated but not cloudy. Make sure to spin down (max speed for 5 min) before injection on to the column. This step is crucial. Precipitates in the injection can ruin the SEC column.
- 3) With 0.4 mL/min, inject at most 0.2 mL onto the column per run (unless the run's purpose is desalting). Larger injection volumes can hurt resolution of the peaks.
- 4) Run column at 0.4 mL/min with SECR. Collect all peaks for the first run and analyze the peaks with MALDI. As long as the injection volume remains the same, the SEC column traces are reproducible (same peak should appear at most +/- 2 min for each run). If the peaks from several injections differ by too much, it is necessary to check the HPLC or the column for leaks.
- 5) The SEC column can be stored in H₂O at room temperature.

2.3.11.5 Reverse Phase Chromatography (RPC)

In RPC, the stationary phase is hydrophobic (non-polar), which is opposite to the “normal phase” that was invented earlier. Commonly used bedding materials are C18, C8 and C4 on silica beads. Theoretically the longer the carbon chain, the more hydrophobic the stationary phase. For the same samples, a C4 column might have a shorter retention time than C18 column. It is hard to

predict which column works better on what sample so the choice of column is usually made empirically.

The protocol for purifying specific peptide is written in many theses. Here is a general guide for deciding the purification condition for a new peptide or protein system, as well as some explanations for our current purification method.

- 1) When working a new sample, it is important to first do a solubility test. Try PRA (mostly H₂O) first. This is the best option if works. PRA with TFA (small amount) can also be used for dissolving the sample. DMF and DMSO are good solvents but they are rather viscous (DMF more viscous than DMSO). If there is free cysteine in the sample (not in a disulfide bond), DMF should be used instead of DMSO because DMSO is oxidative. For hIAPP, DMSO is used because it oxidizes the cysteine residues to form desired disulfide bonds.
- 2) The commonly used solvents for HPLC are: RPA (H₂O with 0.05% HCl) and PRB (80% ACN in H₂O with 0.05% HCl). PRB is 80% ACN and not 100% because the mixing of ACN and H₂O produces heat, which is not ideal for the purification. Mixing some H₂O in ACN can dramatically reduce the heat generated in the column. TFA can be used in place of HCl in the solvents but TFA is extremely difficult to remove from the purified samples. TFA has absorption band in the mid IR region that the sample signal can appear. Thus, I generally do not use TFA unless it is absolutely necessary.
- 3) When working with a new biomolecule, it is hard to predict the elution condition. A C18 analytical column can be used to empirically determine the elution condition, and thus determine an ideal gradient. Inject 20-50 μ L of sample onto the column. Run this column with an initial gradient of 0-100% B in 50 min, with a 10 min at 0%B loading time. Most peptide and protein samples elutes before 60% B. Collect all peaks for MALDI.

- 4) After identifying the peak with desired sample, refine the gradient. For example, if a sample elutes around 30% with the previous gradient, refine the gradient to 10%-40% in 50 min with a 10 min at 10% B loading time. With this gradient, the sample should be eluted before 30% B but after 15% B. When the initial column condition is moved from a lower percentage B buffer to a higher percentage, the elution point is usually moved to a smaller percentage. Refine the gradient further following the sample principle if needed. For a dirty sample (with many different fragments), the ideal gradient is around 0.2% B/min change in buffer composition.
- 5) For large-scale purification (prep column), as long as the bead size and type are the same, the elution point should be similar. For some columns, the manufacture provides a conversion sheet that converts a gradient on the analytical column to a prep column. I usually take the best gradient on analytical column and broaden the gradient by 5% B on both sides of the gradient. For example, if analytical column runs at 20-30% in 50 min, I use 15%-35% in 50 min for the first prep column run. The gradient can be refined from this point.
- 6) The gradient can be much longer but ideally the sample elutes between 30-60 min after injection. There is a tradeoff between diffusion and separation. At later retention time peaks usually have better separation but become broader due to diffusion. 30-60 min after injection is (empirically) a sweet spot for the tradeoff.
- 7) Always wash the reverse phase column with 100% B at the end of a run for cleaning. Always rinse the HPLC and the column with 100% methanol at the end of the day after use. The acid in the solvents can cause leaks in the system.
- 8) Finally, the RPC cannot be repacked and are expensive. Use with care.

2.4 Expression and purification of peptide and protein

All the buffers used for expression and purification are described at the beginning of this chapter. I describe the general protocol for each expression that I have worked with. These protocols could be the starting point when a new system is studied. Some of the following protocols might not be the optimal so they can be modified and improved by the user. All plasmids we currently own in the Zanni group are in the appendix.

2.4.1 Expression and purification of γ D-crystallin

We have wild type, S84C mutation and 4-proline mutation γ D-crystallin gene in the Zanni group. The expression and purification of all these plasmids can follow the same protocol.

- 1) Grow freshly transformed cells and incubate at 37 °C overnight on ampicillin agar plate. It is also ok to streak an ampicillin agar plate with stock BL21 cells and incubate at 37 °C overnight.
- 2) Pick colonies and grow in 10 mL LB with ampicillin overnight.
- 3) Grow in 1L LB and till $OD_{600nm} = 0.6-0.8$. Induce with 0.5 g IPTG and incubate for another 2.5-3 hrs at 37C with shaking.
- 4) Harvest the cells by spinning for 10 min at 5,000 g.
- 5) Resuspend cells in NiRB and lyse the cells by French press. Spin down the debris at 50,000 G for 30-45 min. Pour out the supernatant and store in 4 °C if not use immediately.
- 6) Adjust pH of the lysate to ~7.4. Purify the protein with Ni-NTA column (the usage of Ni-NTA is described in this chapter). Elute the target protein with step elution at 10%, 20% and 60% NiEB. The correct proteins should be in the fraction of 60% NiEB. Check all elution fractions and lysate with SDS-PAGE.
- 7) At this point, the protein solution can be dialyzed against a desired buffer at pH 5.5.

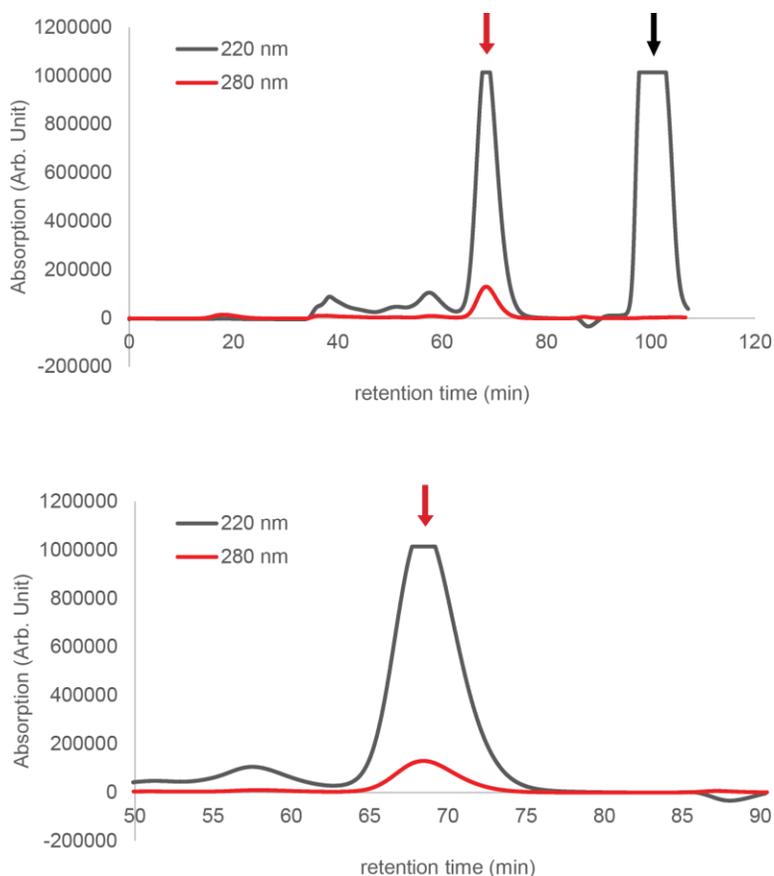


Figure 2.11 Size Exclusion Chromatograph of γ D-crystallin

Top Spectrum shows the entire trace and the γ D-crystallin protein elutes around 68 min (red arrow). The black arrow shows the “salt peak”. Bottom spectrum zooms in at the γ D-crystallin peak.

- 8) If higher purity is needed, further purify the protein solution using SEC with IECB (usage is described in this chapter).
- 9) Collect the correct peak from SEC. Dialyze the purified sample against buffer of choice. Aliquot the proteins in solution and store at $-20\text{ }^{\circ}\text{C}$ or $-80\text{ }^{\circ}\text{C}$. To use the proteins samples, thaw them at room temperature and spin down the aggregates. Measure concentration of the supernatant before use.
- 10) A reverse phase HPLC can be used to check the purity and mass for γ D-crystallin. The HPLC chromatograph is shown in Figure 2.12.

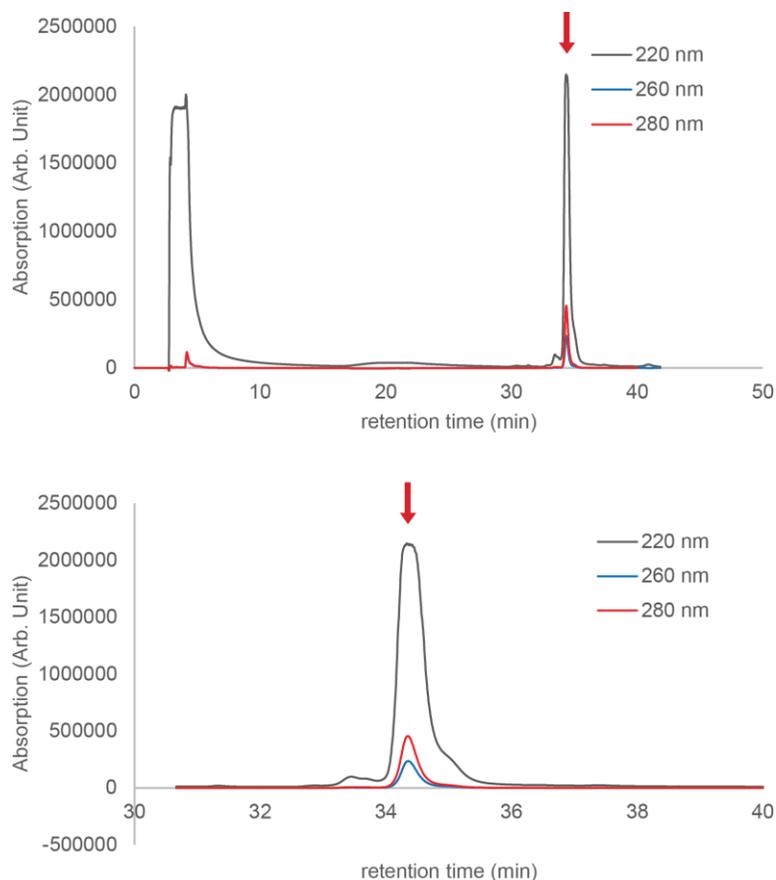


Figure 2.12 HPLC retention time trace of γ D-crystallin

Top figure is the entire trace. Bottom figure shows the zoomed in image of the γ D-crystallin peak. The correct peak is shown with a red arrow.

- 11) Matrix-assisted laser desorption/ionization (MALDI) can be used to check the molecular weight of the protein. A MALDI spectrum of unlabeled γ D-crystallin is shown in Figure 2.13. Note that the MALDI spectrum of large protein is much broader than the peak for small peptide. The correct mass calculated for Histagged γ D-crystallin should be 23121.65 but for large proteins it is sometimes difficult to pick the correct monoisotopic peak even with correct calibration of the range. In Figure 2.13, we can see the single charged peak (23312.9), the double charged peak (11641.3), and the peak for a dimer sharing one charge (47143.5). This is also commonly observed for large proteins.

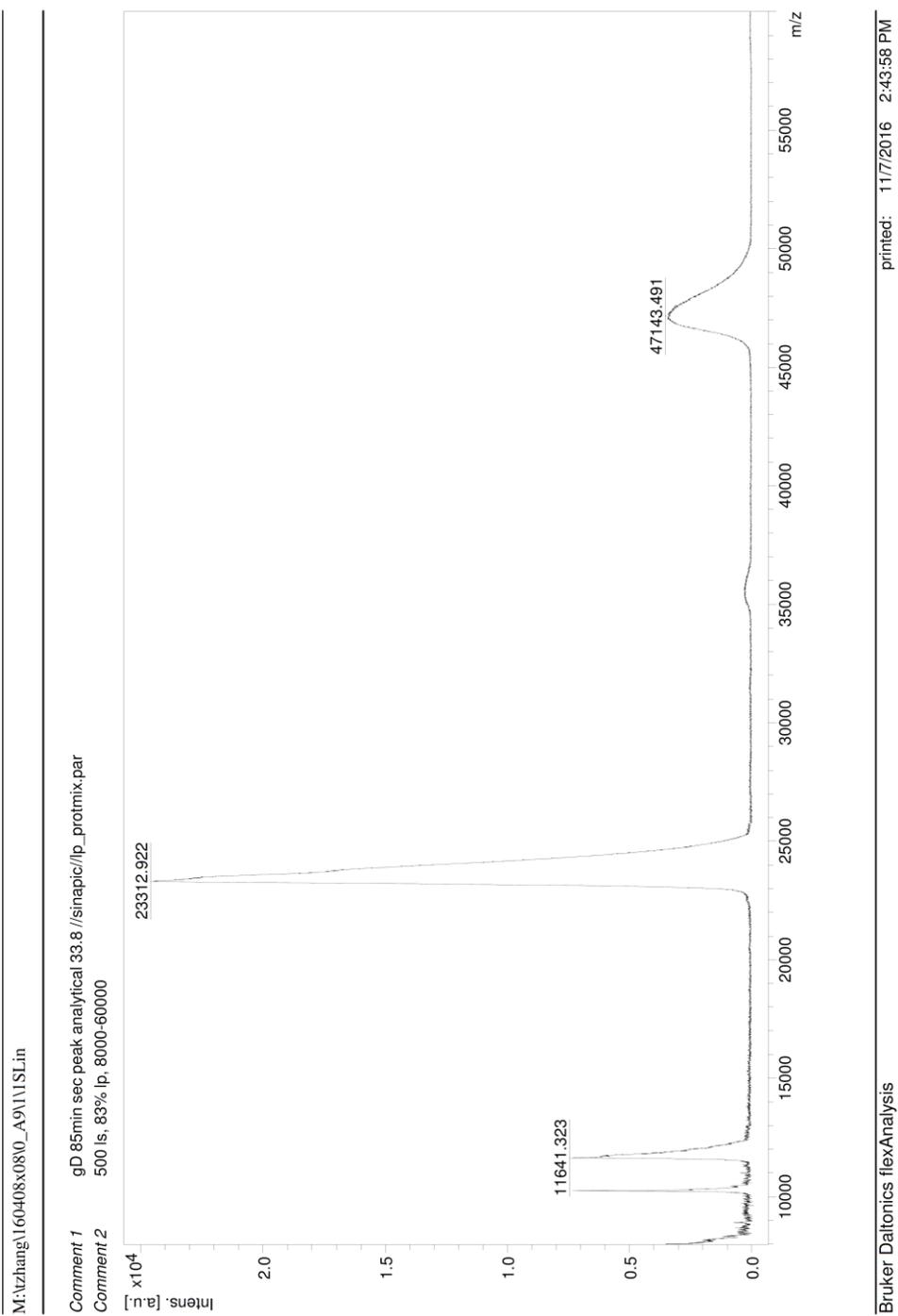


Figure 2.13 MALDI spectrum of γ D-crystallin.

2.4.2 Expression and purification of γ D-crystallin CTD and NTD

The two domains of γ D-crystallin need to be expressed separately and purified before being ligated together. This protocol can be used for other proteins after optimizations but the principle is basically the same. To make segmentally labeled protein, express either the N-terminal domain (NTD) or the C-terminal domain (CTD) in ^{13}C labeled media. The CTD has a His tag at the C-terminus with a Factor Xa site for cleavage. The NTD is coded in a pTXB1 vector and co-expressed with an intein and a chitin-binding domain. The intein can be cleaved with thiol. The ligation site is at serine (residue 84), which is chemically similar to cysteine. In the protocol below, I describe the expression and purification of ^{13}C CTD and ^{12}C NTD. The protocol in the next section will discuss the ligation of these two domains.

2.4.2.1 γ D-crystallin CTD

The purification of γ D-crystallin CTD is basically the same as the full length γ D. A pet16b vector is used for the CTD of γ D-crystallin. The expressed protein is fused to a Histag and a factor Xa cleavage site IEGR. The actual protein sequence is MGHHHHHHHHHSSGHIEGRCGSH... (full sequence of the gene in appendix). Factor Xa cleaves after IEGR and leaves a cysteine residue at the N-terminus of the protein. The cleavage using Factor Xa is described in the next section. Here I emphasize the difference between expression in LB and expression in minimal media.

- 1) Grow freshly transformed cells and incubate at 37 °C overnight on ampicillin agar plate.
- 2) Pick colonies and grow in Falcon tubes with 10 mL LB.
- 3) Spin down the Falcon tubes at 3500 G using the Beckman centrifuge for 5-10 min. Decant the supernatant. Resuspend the cells with 5-10 μL of minimal media buffer (no carbon source added). Incubate the cells for 30 min with shaking.

- 4) Add the filtered stock solution and 500 μ L trace element into 0.5 L minimal media buffer. Shake this solution at 37 °C for 20-30 min.
- 5) Add the resuspended cells to the warm 0.5 L minimal media (incubated in 37 °C with shaking). Grow at 37 °C with shaking until $OD_{600\text{ nm}} = 0.6$. The time it takes to get to $OD_{600\text{ nm}}$ depends on the cell type, initial cell concentration and whether or not base powder is added.
- 6) Induce protein expression with 0.25 g IPTG. Grow cells for additional 7 hrs (without base powder) or 3-4 hrs (with base powder).
- 7) The rest of the expression and purification is the same as for expression and purification of full length γ D-crystallin.
- 8) This protein cannot be used directly for ligation. The cleavage of the Histag is described in the section below.

2.4.2.2 γ D-crystallin NTD

The gene of γ D-crystallin NTD is coded in the same vector (pTXB1) as hIAPP. The expression of γ D-crystallin N-terminal domain is similar to the expression of hIAPP. This vector is interesting because the C-terminal modification for the target protein can be different depends what molecule is used for cleavage. When cleaved with DTT, the target protein will have a free COO- on the C-terminus. When cleaved with DTT with the presence of NH_4HCO_3 , the C-terminus will be amidated. We used this method for amidated hIAPP expression. For the proteins used in ligation, a small molecule call MESNA should be used in place of DTT.

- 1) Grow freshly transformed cells and incubate at 37 °C overnight on ampicillin agar plate.
- 2) Pick colonies and grow in Falcon tubes with 10 mL LB.

- 3) Grow the cells in 0.5 L LB with ampicillin until $OD_{600nm} = 0.6$. Induce with 0.25 g IPTG and incubate for another 3 hrs.
- 4) Harvest cells by centrifuge at 5000 G for 10 min.
- 5) Resuspend the cells in 20-25 mL Chi γ DB and lyse the cells by French press.
- 6) Equilibrate the chitin column with Chi γ DB. Load and wash cell lysate.
- 7) To cleave the intein, use Chi γ DC buffer (0.32 g of MESNA in 40 mL Chi γ DB, pH to 8.5 at 4 °C). Flash the chitin column with 5 CV of Chi γ DC buffer. Incubate the chitin column at 4 °C for an extended amount of time. Take an aliquot every 12 hrs for up to 3 days to examine the cleavage progress using MALDI.
- 8) Note that the MESNA group is not very stable against hydrolysis so the MESNA modified protein should be used freshly to maximize yield. However the protein can be stored in the Chi γ DS buffer (5 mM Bis-Tris, pH 6.5, 250 mM NaCl) at -80 °C if must.

2.4.3 Native chemical ligation

Native chemical ligation is a novel method for connecting two pieces of proteins together through a peptide bond. The word “native” means that there is no non-natural bond introduced in this ligation. Although powerful, this method is rather difficult. Sean successfully made ligated γ D-crystallin. I have tried ligation of α B-crystallin but it did not work. Here, I want to present the general idea of native chemical ligation and show case of γ D-crystallin. I will discuss the difficulties and the problems with this ligation. I hope these problems can be avoided for future researches or serve as a guide for future ligation designs.

As shown below in Figure 2.14, a cysteine residue at the C-terminus of one protein and the MESNA group react to form a peptide bond. Thus, when using native chemical ligation, there is always a cysteine mutation to the original residue. The efficiency of ligation varies largely for each

protein and residue before the ligation site. I describe a general protocol for this ligation. This protocol should be modified for each protein.

- 1) Factor Xa cleavage of Histag on the CTD protein to reveal the cysteine residue.
 - a) Dialyze 20-30 mL of CTD protein solution against 5 L of Factor Xa cleavage buffer three times (2, 2, and 3 h).
 - b) Use spin column (3.5K or 5K Da MWCO) to spin down the protein solution to 4 mL (to obtain protein concentration greater than 1 mM) and transfer to a centrifuge tube. Add 20 μ L of Factor Xa (enzyme in glycerol) and 5 μ L CaCl₂ stock solution (1.0 M) slowly while shaking to make 1.0 mM in the final concentration.
 - c) Some proteins bind to Ca²⁺ and form precipitate so Factor Xa cleavage does not work for all the proteins. Other cleavage sites and enzymes can be used as alternatives.
 - d) Incubate the solution in the dark overnight. Check cleavage progress with MALDI every 12 hrs. After full cleavage (might not be 100%), load the protein solution through a Ni column. Collect the flow through.
 - e) Dialyze the flow through (γ D-crystallin C-terminal domain with N-terminal cysteine) against γ D-crystallin column buffer.
- 1) Mix the NTD (with MESNA) and CTD (with N-terminal cysteine) at 1:4 volume ratio. Higher concentration will result in better ligation yield. Use all products from an one liter growth (Check New England Biolab protocol for the control experiments). The final MESNA concentration should be 10 mM to create a suitable condition for ligation. Incubate the solution at 4 °C overnight.

- 2) Check ligation reaction with MALDI every 4 hr after the first 8 hr if possible. The reaction might take days.
- 3) Dialyze the reaction solution into IEC buffer A and purify the protein solution with IEC using IEC buffer A and B with a 0-50 % gradient.
- 4) Purify the correct peak of ligated protein from IEC with SEC using IECB buffer.
- 5) After dialyzing the protein into the desired buffer, aliquot the ligated protein and store at -80 °C. Only thaw the sample before use. Frequent freeze-and-thaw cycles can cause precipitation of the protein.

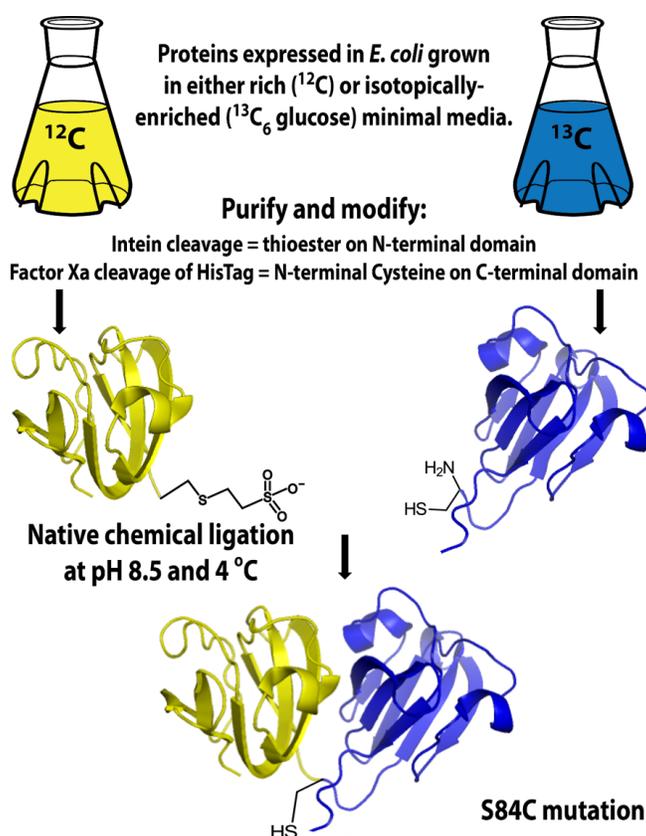


Figure 2.14 Native chemical ligation schematic for γD -crystallin.

2.4.4 Expression and purification of α B-crystallin

The plasmid of non-tagged α B-crystallin is a generous gift from the King Group at MIT. The expression and purification of α B-crystallin has been published previously. However, following those previous protocols does not produce good purification result based on my experience. Here I describe the expression and purification of α B-crystallin I modified based on my experience.

- 1) Grow freshly transformed cells and incubate at 37 °C overnight on ampicillin agar plate. It is also ok to streak an ampicillin agar plate with stock BL21 cells and incubate at 37 °C overnight.
- 2) Pick colonies and grow in 10 mL LB with ampicillin overnight.
- 3) Grow in 1L LB and till $OD_{600nm} = 0.6-0.8$. Induce with 0.5 g IPTG and then transfer the flask to 18 °C with shaking. Let the induction carry on overnight.
- 4) Harvest the cells next morning by spinning for 10 min at 5,000 G.
- 5) Resuspend cells in IECA buffer and lyse the cells by French press. Spin down the debris at 50,000 G for 30-45 min. Pour out the supernatant and store in 4 °C if not use immediately.
- 6) Adjust pH of the lysate to 8.5. Purify the protein with IEC (the usage of IEC is described in this chapter). Elute the target protein with step elution at 5%, 10% and 40% IECB. The correct proteins should be in the fraction of 40% IECB.
- 7) Use IECB buffer for SEC column equilibration and protein purification. Concentration the elution from IEC with 10K or 3K MWCO spin column (optional). Take 0.2 mL of concentrated solution and inject onto the column after spin down at maximum speed for 5

min. Automated sampler on the HPLC can be used for this purification. The figure below shows a typical HPLC trace for α B-crystallin on SEC.

- 8) Collect the correct peak from SEC. Dialyze the purified sample against H₂O or a buffer of choice. Aliquot the solution and lyophilize. Store the protein powder dry with at -20 °C.
- 9) Comparing the SEC trace of α B-crystallin and γ D-crystallin, it is interesting to note that γ D-crystallin elutes as monomers and α B-crystallin elutes as oligomers.

Reverse phase HPLC can be used to test the purity of α B-crystallin and use MALDI to verify the molecular mass (shown in

- 10) Figure 2.16).

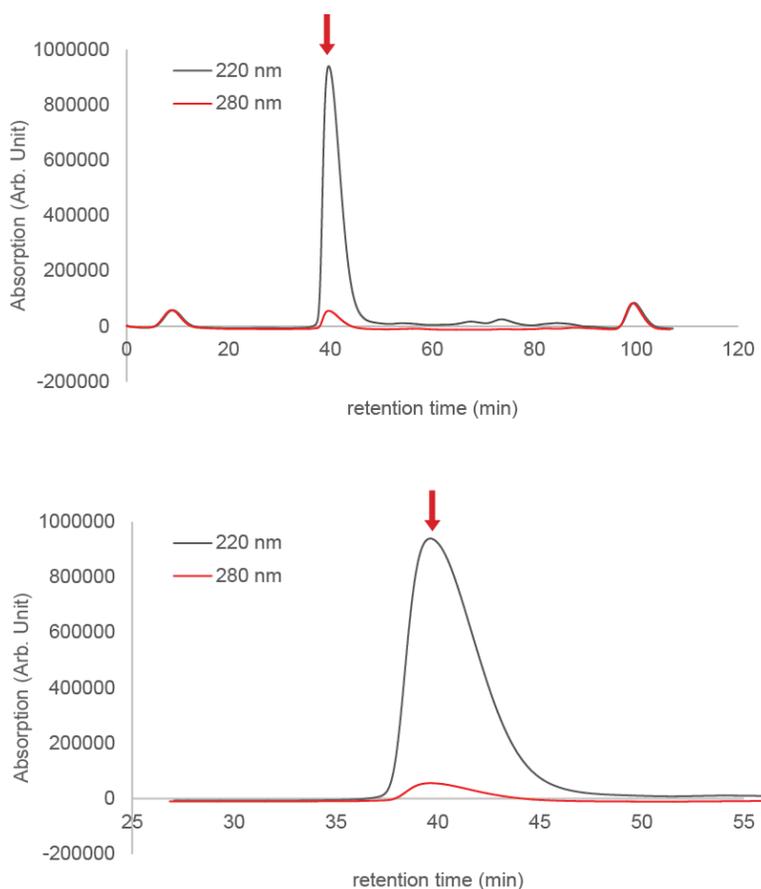


Figure 2.15 Size Exclusion Chromatograph of α B-crystallin
Top graph shows the full trace and bottom graph shows the peak zoomed in.

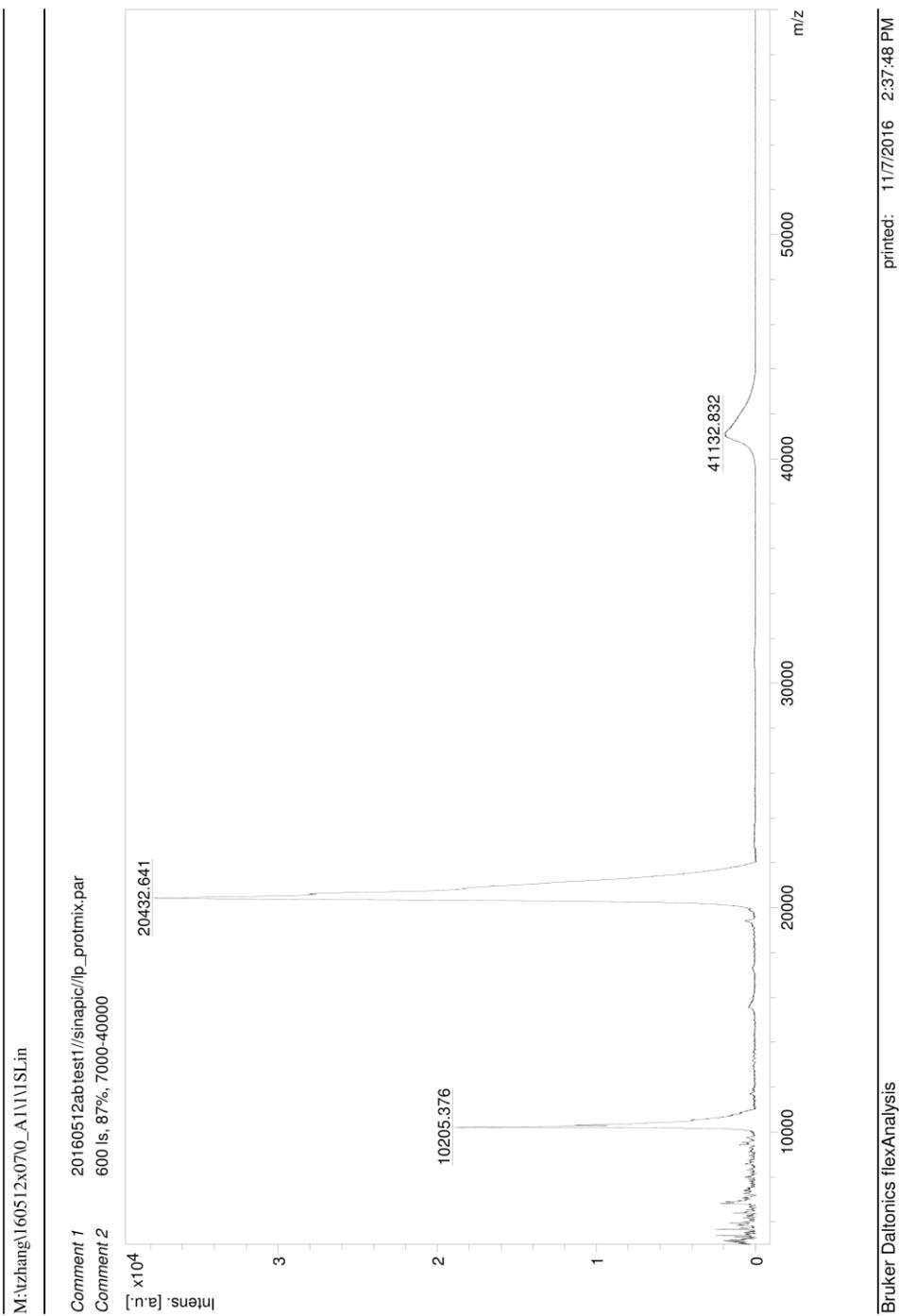


Figure 2.16 MALDI spectrum of α B-crystallin

2.4.5 Expression and purification of hIAPP and its variations using Chitin binding domain

This protocol is developed to make uniformly labeled hIAPP that can be used for both 2D IR studies and NMR studies. The hIAPP found in the body has a free N-terminal domain and an amidated C-terminal domain. There are many other existing protocols for expressing hIAPP in *E. coli* but this is the only published protocol that produces hIAPP with the correct modifications.

2.4.5.1 Day 1

- 1) Try to grow from freshly transformed cells. If not, streak frozen stock on a plate and grow from a single colony

2.4.5.2 Day 2

- 2) Grow in 0.5 L LB or minimal media at 37 °C. An overnight 10 mL growth can make the day shorter but it is not necessary.
- 3) Induce at $OD_{600\text{ nm}} = 0.6$ with 0.5 mM IPTG (0.06 g), save an aliquot of uninduced cells for gel.
- 4) Harvest cells after incubation at 3 hrs at 37 °C or 20 °C overnight. (6 hrs at 37 °C or 18 °C overnight if minimal media is used)

2.4.5.3 Day 3

- 5) Harvest cells. Resuspend pellet with ChihIAPPB, then French press. Save an aliquot of pellet for gel
- 6) Adjust pH of the lysate to 8.0 before load on chitin column.
- 7) Load the lysate slowly on a chitin column. Save the flow through (An aliquot for the gel). Wash with ChihIAPPB for at least 3 CV.

- 8) Flush the column with ChihIAPPC. Incubate the column in 4 °C overnight with the column open to air.
- 9) Elute chitin column with ChihIAPPB the next morning. Total elution volume should be around 30 mL.
- 10) Use 1KDa MWCO dialysis tubing for dialysis the elution from previous step. Dialysis round 1 with 5 L H₂O (one Aliquot for gel)
- 11) Dialysis round 2 and 3 with 5 L H₂O, 0.1 M HCl (save an aliquot for MALDI)
- 12) Transfer dialyzed solution to 50 mL centrifuge tubes (25 mL maximum in each tube). Add Cyanogen Bromide and incubate in dark for 24 hrs (5 mg per 1 mL works, 30X excess Cyanogen Bromide).
- 13) Remove the intein on the Chitin column with ChiStrip. Save an aliquot of this flow through for gel. The MW of the intein should be around 25 KDa.

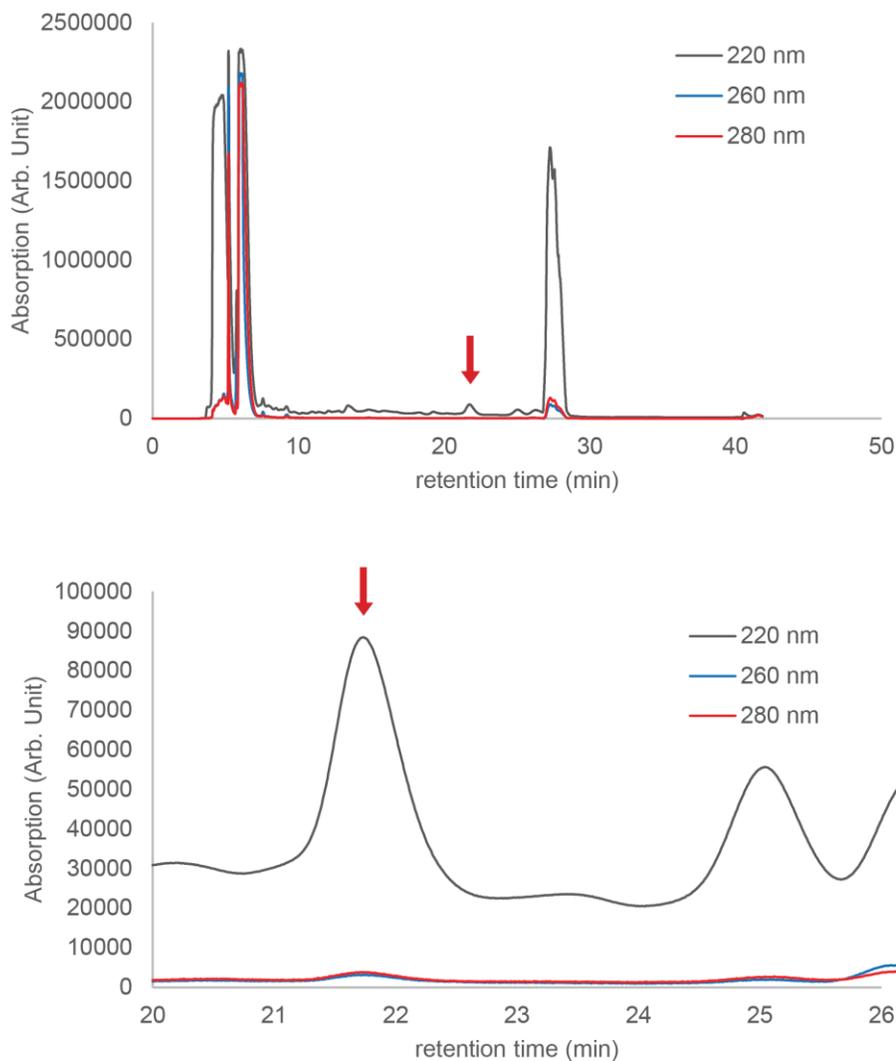


Figure 2.17 HPLC trace of hIAPP L12P V32P purification.

Top figure is the entire trace. Bottom figure shows the zoomed out image of the hIAPP L12P V32P peak. The correct peak is shown with a red arrow.

2.4.5.4 Day 4

- 14) Dry aliquots (under N₂ or Lyophilize). 20% total solution at a time
- 15) Purify with reverse phase HPLC.
- 16) 2P hIAPP elutes around 29% B (B: 80% ACN) shown in Figure 2.17. The correct peptide peak is indicated with a red arrow. WT hIAPP elutes around 35% B. The percentage of elution depends highly on the column, HPLC, and injection size.
- 17) See Figure 2.18 for ¹³C-¹⁵N labeled L12P V32P hIAPP.

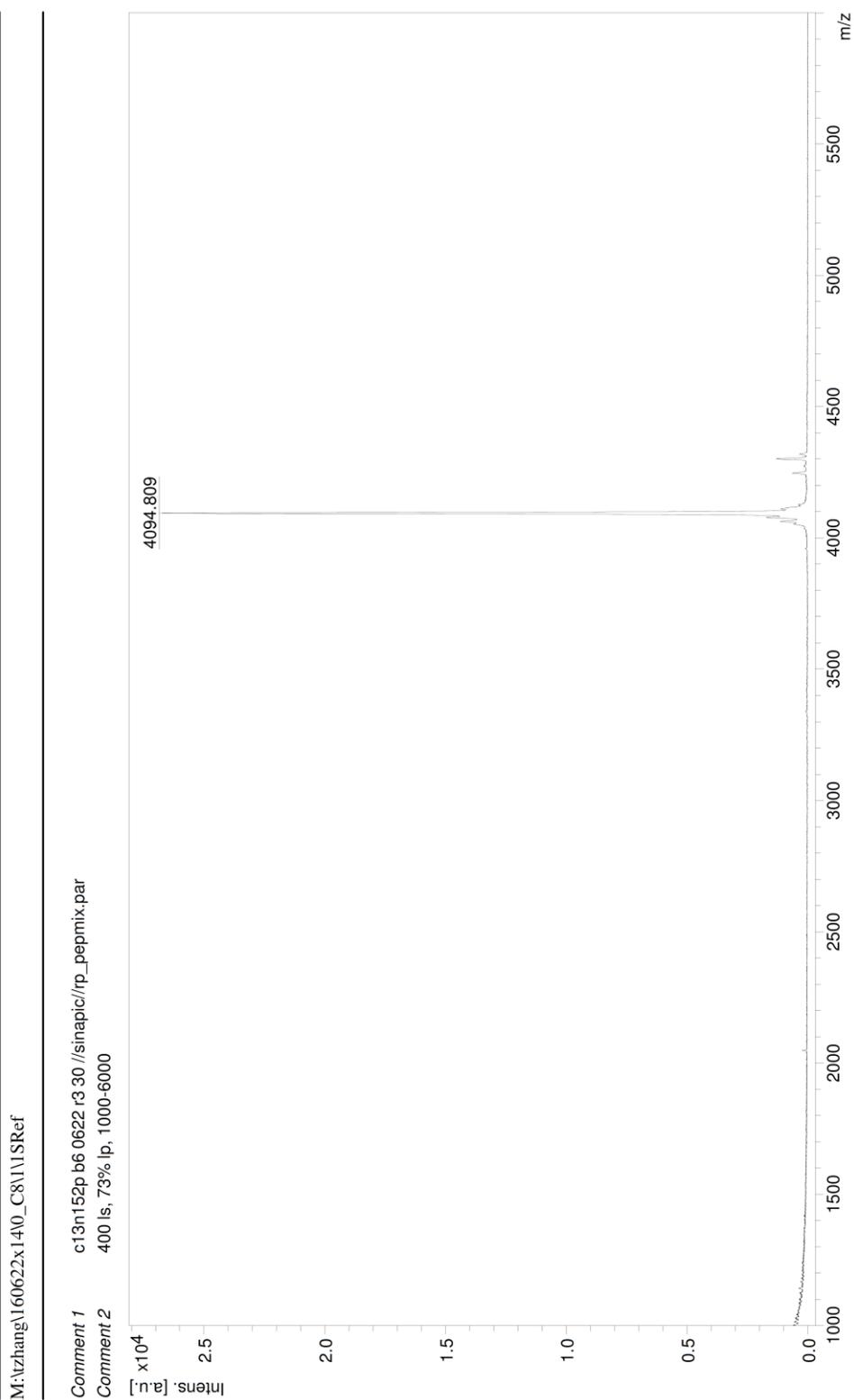


Figure 2.18 MALDI spectrum of ^{13}C - ^{15}N labeled L12P V32P hIAPP

2.4.6 Expression and purification of hIAPP and its variations using Histag

This protocol is for hIAPP that has a Histag added at the N-terminus of the protein. I added 6 histidine residues and a methionine in front of the original hIAPP sequence. The peptide sequence expressed in the protocol is:

MHHHHHMKCNTATCATQRLANFLVHSSNCFGAILSSTNVGSNTY (+Intein).

After cleavage using DTT and cyanogen bromide, the peptide produced should be:

KCNTATCATQRLANFLVHSSNCFGAILSSTNVGSNTY (amidated N-terminus)

2.4.6.1 Day 1 and Day 2

- 1) For Day 1 and Day 2, follow the same steps in protocol for hIAPP without Histag (Section 0).

2.4.6.2 Day 3

- 2) Harvest cells. Resuspend pellet with NiRBhIAPP (2M urea, 100mM NaH₂PO₄, pH 8.0), then French press. Save an aliquot of pellet for gel.
- 3) Adjust pH of the lysate to 8.0 before load on Ni column.
- 4) Equilibrate Ni Column with column buffer. Load the lysate slowly. Save the flow through (An aliquot for the gel). Wash with column buffer for at least 3 CV.
- 5) Elute with NiEBhIAPP (2M urea, 100 mM NaH₂PO₄, and 500 mM imidazole) (An aliquot for gel).
- 6) Use spin column (MWCO <10K) for buffer exchange. Spin down the old solution and then add new buffer (in 500 mL, 2.38 g HEPES, 1.46 g NaCl, 60.06 g urea). Do this step for at least two times. No need to adjust pH. It should be around 6.5. The spinning speed can be slow due to the high concentration of urea.

- 7) A small amount of the solution can be examined with reverse phase HPLC to see if the correct protein is expressed. The mass we expect to see here is around 32849 Da, which is the combined mass of Histagged hIAPP and the intein (27858.3 Da, see sequence in Appendix 4).
- 8) After the last spin, pipet out the solution from the bottom of the spin filters into a 50 mL centrifuge tube. Quantify the protein concentration if possible using Nanodrop. Dilute with the new buffer to 0.5 mM-0.05 mM (2 mg/mL - 0.2 mg/mL). Use an extinction coefficient of 37,000.
- 9) Work in the fume hood from this step onwards. Add a small stir bar to the solution and slowly add NH_4HCO_3 (3.84 g per 30 mL. It's ok to add more) while stirring. Add DTT (0.44 g) 5 min after dissolving all NH_4HCO_3 .
- 10) Leave solution stirring at room temperature overnight. Punch holes on the cap and put it on.

2.4.6.3 Day 4

- 11) Use 1 kDa MWCO dialysis tubing for dialysis. Dialysis round 1 with 5 L H_2O
- 12) Dialysis round 2 and 3 with 5 L H_2O , 0.1 M HCl (save an aliquot for MALDI)
- 13) For unlabeled expression, the major MALDI peaks seen at the step should be around 4991 Da and 27858.3 Da. With ^{13}C - ^{15}N labels, the mass should be around 5270 Da (for WT hIAPP with Histag), 5251 Da (for L12P, V32P hIAPP with Histag) and 29438 Da (Intein). Sample MALDI spectra are shown in Figure 2.19 and
- 14) Figure 2.20.

- 15) Transfer dialyzed solution to 50 mL centrifuge tubes (25 mL maximum in each tube, usually 3-4 tubes). Add Cyanogen Bromide and incubate in dark for 24 hrs (5 mg per 1 mL works, 30X excess Cyanogen Bromide)

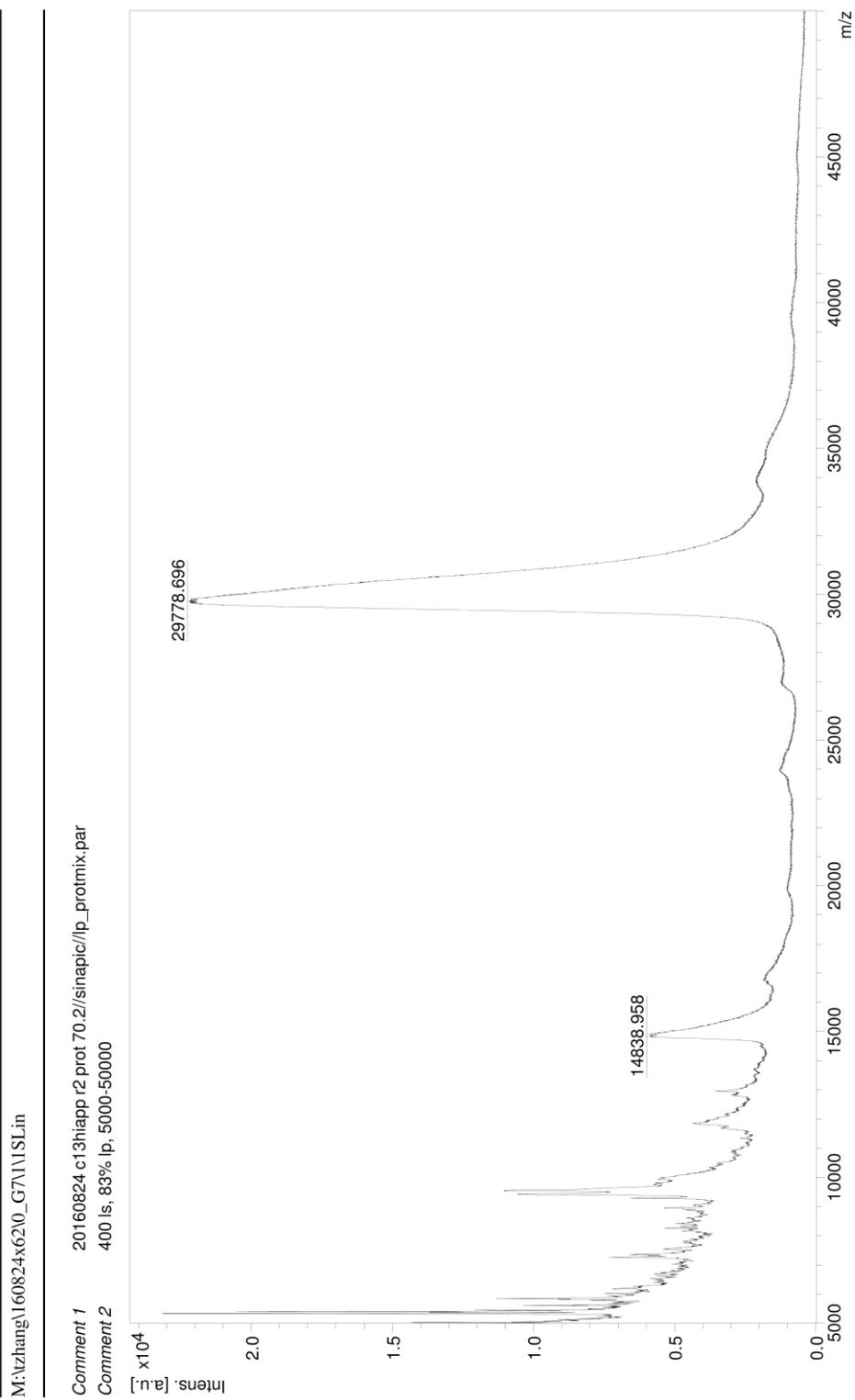


Figure 2.19 MALDI spectrum of ^{13}C - ^{15}N labeled intein

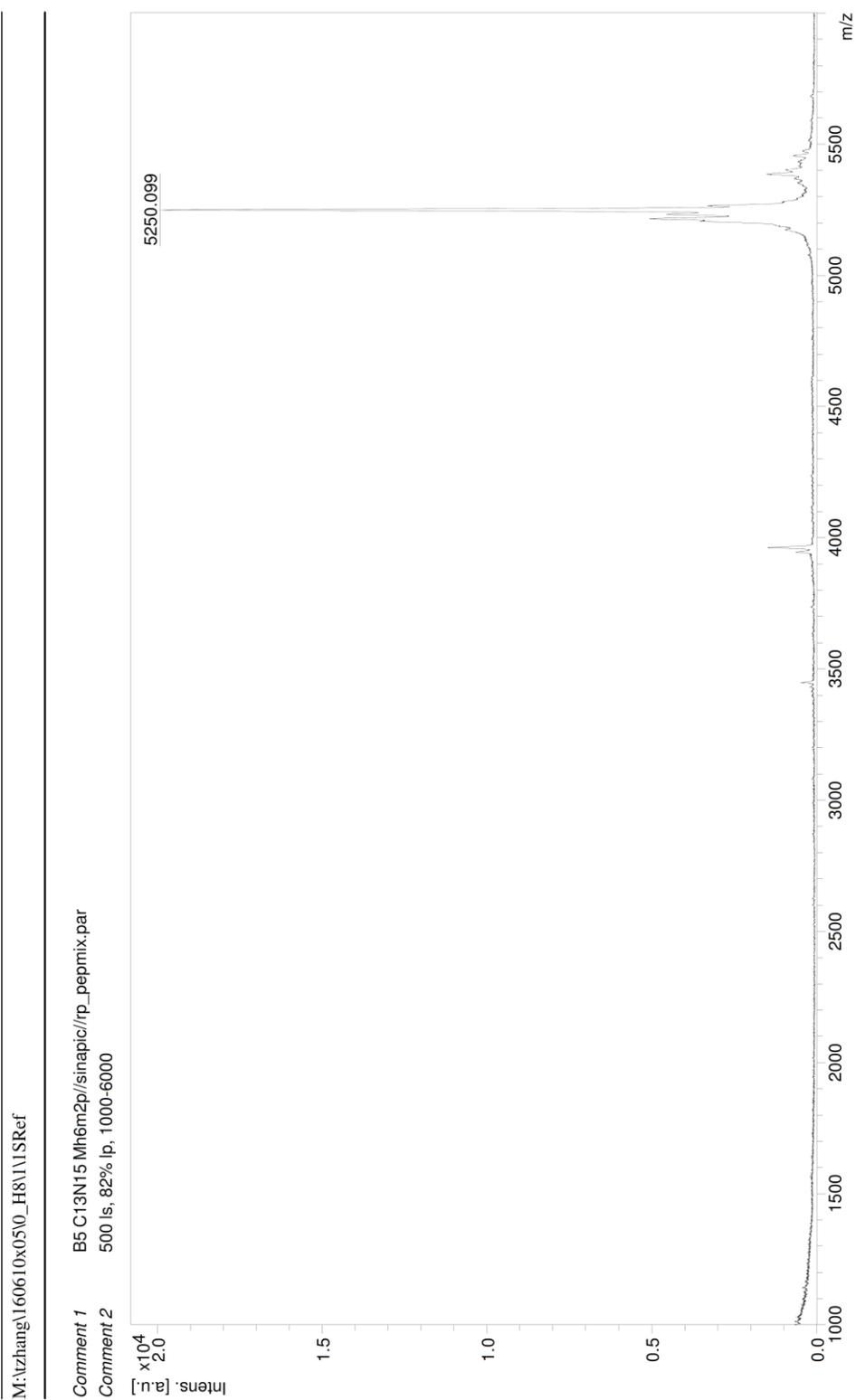


Figure 2.20 MALDI spectrum of ¹³C-¹⁵N labeled L12P V32P hIAPP with Histag

2.4.6.4 Day 5

16) Follow the same procedure as in Section 2.4.5.4.

2.5 Manual synthesis of isotope labeled peptides using Fmoc chemistry

There are two major types of chemistry used in peptide synthesis: Boc and Fmoc. While both produce high purity peptides, Fmoc requires much milder conditions that are more suitable for a research lab. Boc chemistry is generally cheaper but its final protection step requires the use of HF which is highly toxic. Therefore, we use Fmoc chemistry for the peptide synthesis in the Zanni group.

Solid phase peptide synthesis (SPPS) using Fmoc chemistry can produce peptides up to 60 amino acids with high yield when aided with microwave (not the household kind). Without a microwave, the yield is much lower and the reaction time increases significantly. A modern solid phase synthesizer coupled with a microwave can make a 40 residue peptide within 24 hours. One huge advantage of SPPS is the freedom and easiness of incorporating non-natural amino acids into the sequence. Heavy isotope labeled residues show up at a lower frequency than the natural abundant isotopes due to their mass difference. We use SPPS to make peptides with a $^{13}\text{C}=^{18}\text{O}$ isotope (only on the amide bond) label on the amino acid of interest to obtain residue specific resolution on infrared spectroscopy. These isotope labeled amino acids can be expensive. A synthesizer usually consumes more reagents than manual delivery due to the mechanical design of an instrument.

The ^{18}O labeling applies to the amino acids with hydrophobic side-chains (Ala, Gly, Ile, Leu, Phe and Val). In Fmoc synthesis, many other amino acids sidechains require a protecting group during the synthesis. These protecting groups are acid-labile and should only be removed

during the final cleavage step. Because the ^{18}O labeling protocol is acid-catalyzed, any acid-labile sidechain protecting groups will be removed. The Fmoc group, however, is base-labile so it will not be affected during ^{18}O labeling. Sometimes, when an Fmoc $1\text{-}^{13}\text{C}$ amino acid (^{13}C isotope label on the backbone carbonyl) is too expensive, a $1\text{-}^{13}\text{C}$ amino acid without the Fmoc group can be purchased. We add the Fmoc group before ^{18}O labeling. The ^{18}O labeling protocol is published previously¹ so I will not include it here. The protocols below can be used for manual synthesis. The reagents and concentrations are essentially the same as what we use on a CEM synthesizer. Note that for Fmoc chemistry, amino acids are added from the C-terminus to the N-terminus (reversed from protein expression in bacteria). Protein sequences are usually written from N- to C-terminus.

2.5.1 Manual SPPS without microwave

Solid phase synthesizer has high yield and fast speed when comes to peptide synthesis. However, the instrument requires rather lengthy maintenance. Or sometimes the instrument requires a service or is waiting for parts. If a rather short peptide is desired at the time, it can actually be synthesized manually with the help of a Schlenk line. Synthesis without a microwave has lower yield for each step so the length of the sequence cannot be too long. In this protocol, I describe the synthesis of FGAIL fragment with amidated C-terminus manually without a microwave. Every residue is double coupled to ensure yield.

- 1) Set up the glassware as in Figure 2.21. The top funnel has fritted layer that allows solvents and air to pass through it, but not resins. Connect the arm with a Schlenk line and let N_2 gas flow through gently.

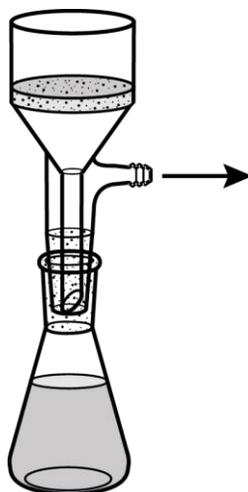


Figure 2.21 Manual peptide synthesis setup

- 2) For 0.1 mM scale (0.1 mmole of peptide assuming 100% yield), weigh out 0.625 g of Fmoc-protected PAL-PEG-PS resin (loading capacity 0.16 mmol/g). The loading capacity for each resin is different so for each resin, the amount used should be calculated accordingly. Swell resin in this setup 7 mL of DMF for two hours at room temperature before synthesis.
- 3) Weight out each amino acid for 6 mL of 0.2 mM solution. Leave them as solids in small centrifuge tubes until use.
- 4) Make activator and activator base fresh daily. Add 1.896 g of HBTU in 10 mL DMF to get final concentration of 0.5 M.
- 5) Make activator base of 1.7 mL DIEA and 3.3 mL of NMP.
- 6) Make deprotection solution by dissolving 103.3 g of HOBT in 480 mL of DMF and mix well with 120 mL piperidine.
- 7) Use double coupling for the first several amino acids because it is difficult to add onto the resin. The following steps are for each amino acid.

- a) Drain solution in the funnel by switch outlet to vacuum on the Schlenk line. Turn on N₂ bubbles and wash resin with 7 mL of DMF. Drain with vacuum. With N₂ bubbling, add 7 mL of deprotection solution (DS).
 - b) Let this solution bubble for 20 min and then drain with vacuum.
 - c) With N₂ on, add another 7 mL of DS. Incubate for another 20 min.
 - d) Wash resin with 7 mL DMF for 6 times.
 - e) Add 6 mL of DMF to the first amino acid Leucine.
 - f) With N₂ gas on, add 3 mL of amino acid solution, 1 mL of activator, and 0.5 mL activator base. Let this solution bubble for 2-3 hours. Drain.
 - g) Repeat f) with the same amino acid and let the solution bubble for another 2-3 hours. Drain.
 - h) Wash resin 3 times with DMF. Resin can be stored at this point with DMF at 4 °C overnight.
- 8) Repeat step 7) for the rest of the amino acids.
- 9) After all the amino acids are added to the sequence, the last step should be deprotection (as described in step 7a)) to produce a free NH₃ at the N-terminus. Resin beads can be rinsed with dichloromethane, dried and stored at -20 °C until cleavage.

2.5.2 Manual SPPS with microwave

When using the solid phase peptide synthesizer, the instrument adds a certain volume of extra solution to make sure of the delivery of sufficient solution. By adding the reagents manually, we can reduce the reagents used. Some settings of the microwave are shown in Table 2.7. The actual setting may vary for different protocols. The best practice is putting the microwave in the hood if possible. When programming the instrument, add a pause to every reagent adding step.

Take the reaction vessel out of the microwave and add or drain the reagent manually. The tubing for the reaction vessel has a Luer-Lok connector that can be tightly screwed into a syringe. Pulling the syringe to drain the reagent creates the same effect as the draining under vacuum. Reagent concentration is the same as in the manual synthesis. Make sure to rinse the reaction vessel well after each reaction. Residual reagents can lower the yield.

Settings	Microwave power	Temperature	Time
Basic coupling	20 W	75 C	5 min
Initial deprotection	40 W	75 C	0.5 min
Deprotection	40 W	75 C	3 min
Cysteine coupling	20 W, single	50 C	5 min
Histidine coupling	20 W, double	50 C	5 min

Table 2.7 Microwave setting for Fmoc chemistry solid phase synthesis

2.6 Sample storage and preparation for 2D IR spectroscopy measurements

2.6.1 Peptides and proteins in solution

The preparation of disordered peptides and folded proteins are slightly different. Either of the preparations have its difficulties. For small peptides, it is hard sometimes to see the all random coil state (no aggregation). The sample preparation should be fast and the dead time between when the sample is prepared and when the first spectrum is measured should be as short as possible.

- 1) hIAPP
 - a) Dry 2 μ L hIAPP (1 mM) in HFIP-d under N₂ in a fume hood. A speed vacuum attached to a lyophilizer (concentration centrifuge) is also recommended.
 - b) Aggregation is initiated by dissolving hIAPP in 2 μ L of 20 mM Tris D₂O buffers, pH 7.4 to make 1 mM final concentration of peptides (other buffers can also be used).

- c) Place 2 μL of peptide solution immediately between two CaF_2 windows after dissolving in buffer. Use a 56 μm spacer to fix the path length. Measure the sample as soon as possible.
 - d) For FTIR, the kinetics measurements can be set up with the instrument settings. It is important to collect a background spectrum without solution and a D_2O buffer background spectrum before measuring the peptide sample. A HeNe beam can be used to ensure that light is going through the sample.
 - e) For 2D IR spectroscopy, the same sample cell and assembly can be used in both our FTIR and 2D IR setups. Before measuring the sample, the laser setup is optimized using a standard calibration molecule, *N*-acetyl-proline (NAP). The peptide sample is placed where the pump and probe pulses are overlapped and focused. We collect frequency domain data on the probe axis with a mercury cadmium telluride (MCT) detector. With a pulse shaper, time t between the two pump pulses is scanned (time domain data) and a Fourier transformation gives us the pump frequencies. 2D IR spectroscopy with a pulse shaper is suitable for kinetics measurements of protein aggregation because a single spectrum can be collected in less than 1 min.
- 2) γD -crystallin
- a) Freeze protein solution with liquid N_2 and lyophilize.
 - b) Redissolve protein with D_2O . Leave the solution at room temperature for 12 hrs. Freeze protein solution again and lyophilize.
 - c) Redissolve protein with small amount of D_2O to reach a concentration that is higher than desired concentration. Check sample pH (try to get to pH 5.5) and centrifuge

down the precipitates. Take the supernatant and measure concentration with the Nanodrop. Dilute the stock solution to reach the desired concentration. The protein solution is now ready for 2D IR.

- d) For acid treatments, dialyze protein samples with low concentration of phosphate buffer (1 mM). After performing step a) and b), redissolve the protein in a buffer with higher phosphate concentration of pH 3.0. Spin down the solution before 2D IR measurement.
- 3) α B-crystallin
 - a) α B-crystallin is more stable when stored in the powder form. It dissolves immediately in D₂O. Add 200 μ L of D₂O to an aliquot of α B-crystallin. Lyophilize the solution after 2-4 hrs of incubation in D₂O.
 - b) Redissolve α B-crystallin in desired buffer or D₂O (if the stock solution already has salt in it) and measure 2D IR as soon as possible.

2.6.2 Preparation and measurement of transmembrane peptides

I learned the preparation of transmembrane peptides from Dr. Yoshiaki Yano. The peptide we work with is a short (21 AA) transmembrane peptide with isotope labels. Transmembrane membrane peptides are usually not soluble in H₂O so the preparation is different from other peptides. The preparation of transmembrane peptide samples and peptide on lipid bilayer samples differ if the vesicles are formed before the peptide is mixed with the lipids.

We used 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) only but the mixture of POPC and other lipids can be used.

- 1) Dissolve POPC in CHCl₃ with a final concentration of 50-200 mM. Vortex the solution for 1 sec before taking any solution out.

- 2) Use an air tight syringe, take out 40 μL of POPC solution (we used 50 mM) dry under N_2 . Store the stock solution under N_2 .
- 3) Dissolve peptide in HFIP-d for H/D exchange. Dry the peptide solution under N_2 in the hood. Redissolve peptide in HFIP-d (1mM final concentration).
- 4) Add 40 μL peptide solution in HFIP-d into the aliquots of dried lipid samples. Dissolve lipids and mix the sample by pipetting up and down.
- 5) Dry the samples as quickly as possible initially under N_2 in the hood and put the sample in the lyophilizer.
- 6) Note that the concentration of peptides inserted in the vesicles has a threshold. For each peptide, there should be at least 50 lipid molecules (1:50).
- 7) Add 10 μL of buffer to the dried samples to get final concentrations of 4 mM peptide and 200 mM lipids.
- 8) Buffer with or without sucrose can be used (must be deuterated for IR measurements). This is an invention of Dr. Yoshiaki Yano. Adding sucrose does not change the coupling or structure of the peptides (we tested that in lab) but makes the lipid much more soluble and dramatically reduces scattering of the sample.
- 9) Incubate samples with buffer at room temperature for 30 min. Make sure all peptides are dissolved by tapping on the tube. Sonicate sample for 1 hr without heating.
- 10) If the lipid does not dissolve very well, additional sonication and vortexing can be done.
- 11) Centrifuge for 10 min at 5000 G before taking samples out of each aliquot. Use a 25 μM spacer and around 1 μL per 2D IR/FTIR sample.
- 12) The rest of the sample can be stored at room temperature under N_2 . Short sonication and centrifugation can be added if the lipids are crushing out of the solution.

2.7 Sample preparation for TEM samples

There are at least two TEM facilities on campus. One is the medical school electron microscope facility and the other one is the materials science center at UW. The sample preparation is the same for both places. We can be trained to use the TEM independently. Randy Massey at the medical school facility will also run TEM samples for us if we are not trained.

Here is a brief description of TEM sample preparation. Note that the perfect dye might differ for different samples. The quality of the TEM result highly depends on the sample itself.

- 1) A good TEM sample should be around 1 mM to 10 μ M in concentration (also depends on the size of the proteins). The salt concentration should be relatively low (up to 10 mM total salt concentration is desired) since salt can crystallize on TEM grids.
- 2) Commonly used TEM dyes are ammonium molybdate, methylamine tungstate, and uranyl acetate. They come in powder form or premade solutions. Note that some dyes have short shelf life in solution so fresh dyes should be prepared in that case.
- 3) TEM grids with or without supporting film can be purchased. For our usage, it is more convenient to purchase the grids with supporting film. The ones we have current are “200 mesh Cu with carbon support films” from Ted Pella, Inc. (Product # 01840-F).
- 4) Add 1-2 μ L of protein solution on the TEM grid (make sure to use the correct side). Use a wedge of filter paper to dry the solution from the very edge of the grid. Let the film dry for another 1 min or so.
- 5) Add 1-2 μ L of dye solution directly on the film. Use a filter paper wedge to absorb the extra dye from the very edge of the grid. Let the sample dry.
- 6) The grid is now ready for measurement. Put the grid onto the sample holder. It does not matter which side of the grid is on top.

2.8 References

- (1) Middleton, C. T., Woys, A. M., Mukherjee, S. S., and Zanni, M. T. (2010) Residue-specific structural kinetics of proteins through the union of isotope labeling, mid-IR pulse shaping, and coherent 2D IR spectroscopy, *Methods* **52**, 12-22.

3 Chapter 3

Amyloid Formation in Human γ D-crystallin induced by UV-B Photodamage*

3.1 Abstract

γ D-crystallin is an abundant structural protein of the lens that is found in native and modified forms in cataractous aggregates. We establish that UV-B irradiation of γ D-crystallin leads to structurally specific modifications and precipitation via two mechanisms: amorphous aggregates and amyloid fibers. UV-B irradiation causes cleavage of the backbone, in large measure near the interdomain interface, where side chain oxidations are also concentrated. 2D IR spectroscopy and expressed protein ligation localize fiber formation exclusively to the C-terminal domain of γ D-crystallin. The native β -sandwich domains are not retained upon precipitation by either mechanism. The similarity between the amyloid forming pathway when induced by either UV-B irradiation or low pH suggests that the propensity for the C-terminal β -sandwich domain to form amyloid β -sheets determines the misfolding pathway independent of the mechanism of denaturation.

* The material in this chapter was originally published as: Sean D. Moran, Tianqi O. Zhang, Sean M. Decatur, Martin T. Zanni. "Amyloid Fiber Formation in Human γ D-Crystallin Induced by UV-B Photodamage." *Biochemistry*. 2013, 52, pp 6169-6181. DOI:10.1021/bi4008353. I contributed to this work in the experimental design, sample preparation, data collection and analysis.

3.2 Introduction

Cataracts are a common protein misfolding disease of the ocular lens, which affects approximately 50% of the population over the age of 65.^{1,2} This disease results from accumulated damage to lens crystallin proteins, which destabilizes their folds and causes them to aggregate, resulting in the blurring of vision.¹ Currently, the only treatment for cataracts is invasive surgical extraction that is carried out in the advanced stages of the disease. As a result, there is much interest in understanding the cause of cataracts and the mechanism by which they form. However, this task is hindered by the difficulty of experimentally probing the structures of protein aggregates, which has also been an issue of the study of other diseases such as Alzheimer's disease and type II diabetes.

A complicating factor for cataracts is the likely diversity of destabilizing factors that induce the aggregation and could result in complex mixtures of different structural states. A wide variety of processes such as oxidation, cross-linking, cleavage, and deamidation of crystallins have been identified as factors in their formation based on the distribution of modifications observed in proteins extracted from tissue.^{1,3,4} These modifications arise from environmental factors such as exposure to ultraviolet light⁵⁻¹⁰ and oxidative stress.^{3,4,7,8,11-13} Because lens crystallins are not regenerated after development, damage to the proteins can accumulate over multiple years leading to the formation of age-onset cataracts.^{1,2}

In order to understand the structures of lens crystallin aggregates and their mechanisms of formation, classes of crystallin proteins have been examined under a variety of denaturation conditions *in vitro*. Typically, proteins are denatured using heat, chemicals, or pH shifts.¹⁴⁻¹⁹ Although these conditions have been utilized extensively, the resulting crystallin molecular structures are still poorly understood. Two broad classes of aggregates have been observed:

amorphous aggregates and amyloid aggregates. Amyloid aggregates result from a massive conformational shift in the protein upon precipitation.^{14,15,17,19,20} Recently, we applied segmental ¹³C labeling and ultrafast two-dimensional infrared (2D IR) spectroscopy to examine the acid-induced amyloid fiber structure of the abundant human structural lens protein γ D-crystallin.¹⁵ By isotope labeling the domains individually, we were able to spectroscopically monitor the structures and dynamics associated with the C-terminal and N-terminal domains of γ D-crystallin. We found that β -sheets of the amyloid fiber originate from residues in the protein's C-terminal domain, while its N-terminal domain becomes highly disordered.¹⁵ Using this information, along with spectral simulations as well as enzyme digests and mass spectrometry,²¹ we built and refined a structural model of the acid-induced fibers.

While acid-induced denaturation is a useful way to initiate crystallin aggregation, radiation induced aggregation may be more relevant to the disease. Exposure to ultraviolet irradiation in sunlight has been linked to the formation of age-related cataracts.^{1,5,7,13} Human γ D-crystallin absorbs ultraviolet light and has a large cross section in the UV-B range due to the presence of aromatic amino acids, particularly tryptophan and tyrosine.²² The four tryptophans have been shown to efficiently funnel UV-B excitation to thermal energy through a process involving rapid energy transfer and internal conversion, thereby protecting the protein from UV initiated photochemistry.²²⁻²⁴ However, rare quenching events may lead to ultraviolet photodamage, and it has recently been shown that the conversion of tryptophans in γ D-crystallin to kynurenine (a known marker of cataract formation) destabilizes its fold.⁶ Here, we induce aggregation of human γ D-crystallin by exposure to UV-B irradiation. We show that this results in oxidative damage to side chains, as well as cross-linking and regiospecific scission of the polypeptide backbone. We demonstrate that the polypeptide backbone is cleaved photochemically, and use mass spectrometry

to show that two regions of the protein, in the vicinity of specific tryptophan residues, are highly susceptible to peptide bond cleavage. Using Thioflavin T binding, transmission electron microscopy, segmental isotope labeling, and 2D IR spectroscopy, we conclusively demonstrate that the aggregates formed by this process contain amyloid fibers, and that the fiber forming residues originate from the C-terminal domain. Using this information, we propose a mechanism for the initiation of aggregation by UV-B irradiation. Since UV irradiation is known to correlate with cataract formation in adults,^{1,5,7,13} our results suggest that amyloid fibers of γ D-crystallin could form in the lens, via side chain damage, polypeptide cross-linking, or fragmentation, all of which can destabilize the protein enough to allow entry into an amyloid misfolding pathway.

3.3 Material and Methods

3.3.1 Preparation of native and denatured γ D-crystallin samples

All chemicals were purchased from Sigma-Aldrich and used as received, unless otherwise noted. His₆-tagged human γ D-crystallin, γ D-crystallin (S84C), and its isolated domains were expressed in *E. coli* as described previously.¹⁵ Segmentally-labeled proteins were prepared by expressing one of the two domains in ¹³C labeled media and linking it to the other through expressed protein ligation.¹⁵ A bank of four 8 W medium pressure mercury UV-B lamps were used to irradiate samples at a distance of approximately 30 cm, with an irradiance of 35 W/m². Samples, at approximately 5 μ M concentration, were agitated gently using an orbital shaker during irradiation. Oxygen-free samples were prepared by lyophilizing the protein to dryness, purging with nitrogen, and redissolving the sample in deoxygenated buffer prepared using three freeze-pump-thaw cycles on a Schlenk line. A modified Fenton reagent composed of 0.1 mM (NH₄)₂Fe(SO₄)₂, 0.2 mM EDTA, 1 mM ascorbic acid, and 0.03% H₂O₂ was used to prepare samples damaged by hydroxyl radicals.²⁵

3.3.2 Sample preparation for mass spectrometry

Samples for mass spectrometry were prepared by illuminating γ D-crystallin for 6 hrs, followed by digestion with bovine trypsin in the presence of 0.01% ProteaseMax surfactant (Promega). Digestions were performed at 42 °C for three hours, and the samples were acidified with 0.05% TFA. Samples were centrifuged to sediment degraded surfactant and insoluble materials. LC-MS/MS data was collected using a Thermo Scientific LTQ Orbitrap XL mass spectrometer. Collision-induced dissociation (CID) spectra of the tryptic peptides were used to generate sequence data. The data was analyzed as described previously²¹ against the *E. coli* protein sequence database, containing 4205 entries, with the sequence of the recombinant human γ D-crystallin appended to it. Sequence matches were identified and semitryptic fragments and amino acid adducts were identified from the sequence data.

3.3.3 Trypsin digestion of γ D-crystallin

UV-vis samples were prepared by treating aggregated γ D-crystallin samples with bovine trypsin (1:50 ratio), resulting in resolubilization. Spectra were measured in a 1 cm quartz cell (Starna). Undigested samples were prepared for SDS-PAGE by dissolving the samples in Laemmli sample buffer containing 20 mM 2-mercaptoethanol. SDS gels were analyzed using ImageJ software.

3.3.4 Sample preparation for transmission electron microscopy

Transmission electron micrographs were taken at the University of Wisconsin Medical Sciences Electron Microscopy Facility, using a Philips CM 120 transmission electron microscope. Samples were negatively stained using methylamine tungstate for imaging. The size distributions of structures in the TEM images were determined via random selection and diameter measurement of 50 structures.

3.3.5 Thioflavin T assay of UV-irradiated γ D-crystallin

ThT binding assays were performed by removing aliquots of γ D-crystallin at a series of time points, adding 2.5 equivalents of ThT, and incubating for 15 minutes. Fluorescence emission at 486 nm was measured after excitation of ThT at 430 nm. The measurement were performed for three samples prepared in parallel, and averaged.

3.3.6 Sample preparation and data collection for 2D IR spectroscopy

2D IR spectra were collected on UV-B damaged γ D-crystallin after 6 hours of illumination in deuterated buffers. Samples were then placed between CaF_2 plates separated by a 56 μm Teflon spacer. Spectra were collected with perpendicular pump-probe polarization, and processed as described previously.^{15,26}

3.4 Results

Aggregation of human γ D-crystallin variants was induced by illuminating dilute samples (5-50 μM in 20 mM sodium phosphate, 100 mM NaCl, pH 7.0) with UV-B radiation from a bank of four unfiltered 8 W medium-pressure mercury vapor lamps (Spectroline, Inc.) The maximum output of the lamp was in the 275-325 nm range, overlapping with the UV absorption spectrum of human γ D-crystallin (Figure 3.1A). The total power is 35 W/m^2 at a distance of 30 cm. Under these conditions, the total energy incident on the sample in 1 hour is approximately equal to that from sea-level solar radiation over the course of 3 years assuming 1 hour per day of direct exposure.²⁷ Although the power of the UV-B light bank is roughly 10^3 times that of sunlight on the earth's surface, multiphoton processes are still unlikely. Thus, these conditions are representative of exposure to UV-B solar radiation, albeit on a compressed timescale.

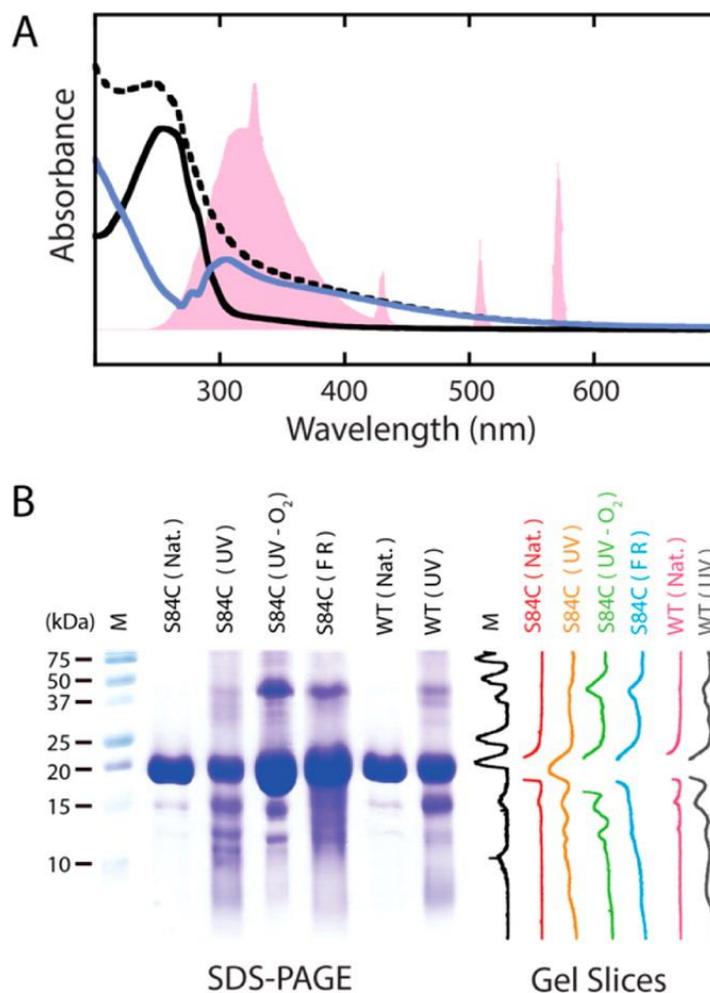


Figure 3.1 UV-B photodamage products of human γ D-Crystallin.

A. UV-vis spectra of γ D-Crystallin before and after UV-B illumination. The output spectrum of the mercury vapor lamp, normalized to the protein absorption, is shown in pink. The spectrum of the protein before illumination (solid, black) shows absorption below 300 nm consistent with the tyrosine and tryptophan content of the protein. After exposure to UV-B light, the spectrum (dashed, black) becomes broadened into the visible range. Subtraction of the pre-UV-B spectrum from the spectrum of the photodamaged protein yields a difference spectrum (solid, blue) with a bleach between 250 and 305 nm. The band structure of this bleach suggests a loss of tryptophan absorption. B. SDS-PAGE characterization of γ D-Crystallin degradation via UV-B photodamage and chemical methods. From left to right are molecular weight marker (M), undamaged γ D-Crystallin (S84C) (S84C (Nat.)), UV-B photodamaged γ D-Crystallin (S84C) (S84C (UV)), UV-B photodamaged γ D-Crystallin (S84C) in deoxygenated buffer (S84C (UV O₂)), γ D-Crystallin (S84C) exposed to Fenton's reagent (S84C (FR)), undamaged wild type γ D-Crystallin (WT (Nat.)), and UV-B photodamaged wild type γ D-Crystallin (WT (UV)), respectively. Vertical slices of the gel image are shown for comparison of each sample.

3.4.1 UV-vis spectra of γ D-crystallin before and after UV-B irradiation

Exposure of γ D-crystallin to UV-B light resulted in visible turbidity in the samples within 1 hour, and the continued production of precipitates for as long as we measured (12 hours). For the purposes of this paper, exposure times of 6 hours is examined. The precipitation of γ D-crystallin is accompanied by the formation of a pale yellow color. In order to solubilize the precipitates without chemically altering the side chains, we used bovine trypsin to digest the aggregates. Figure 3.1A shows the UV-vis spectra of otherwise identical samples with and without exposure to UV-B light. A clear extension of the spectrum into the visible is present in the UV-B exposed sample, accounting for the color change. This extension of the absorbance into the visible range is consistent with the photooxidative modification of aromatic side chains in proteins,²⁸ and is similar to the color changes that are sometimes observed in the lens with the formation of mature cataracts. Additionally, an apparent bleach occurs between 250 and 305 nm in the difference spectrum between the two samples, consistent with the loss of absorption from aromatic amino acids after UV-B exposure. The sequence of γ D-crystallin contains 14 tyrosine and 4 tryptophan residues,²⁹ both of which have been implicated in UV-B photodamage of proteins.^{9,30-32} Tyrosine and tryptophan absorb between 250 and 300 nm. The bleach in the difference spectrum in Figure 3.1A has a band structure comparable to the vibronic bands of the 1L_b mode of tryptophan, indicating that tryptophan side chains are damaged.^{30,33} From the magnitude of the bleach, we estimate that at least 75% of the tryptophans in the sample have been photodamaged within 6 hours.

3.4.2 UV-B irradiation cleavages the backbone of γ D-crystallin

In addition to damage to aromatic side chains, UV-B exposure also results in the cleavage of the γ D-crystallin polypeptide backbone. SDS-PAGE analysis (Figure 3.1B) shows that insoluble

high molecular weight aggregates, covalent cross-linked products, and discrete cleavage products are produced. Slices of the gel image shown stain intensity profiles of each lane. Approximately 70% of the protein components of this sample exist in cross-linked and cleaved photoproducts. The linkages that produce the high molecular weight products are not disulfide bonds because samples were reduced with 2-mercaptoethanol prior to electrophoresis. Thus, they must arise from other covalent linkages such as dityrosine adducts,²⁸ which are a known products of the irradiation of proteins with UV-B light. Four major cleavage products are observed, with molecular weights estimated at 6.4, 10.5, 12.4, and 14.9 kDa. The small number of cleavage products indicates that the cleavage sites are sequence or structure dependent. Because the S84C mutation, which is required for our isotope labeling and 2D IR experiments, introduces an additional redox-active amino acid into the γ D-crystallin sequence, we also performed this experiment on the wild type protein. A similar cleavage pattern is observed in the wild type protein (Figure 3.1B) indicating that the UV-B induced cleavage is not dependent on the presence of an extra redox-active cysteine. We also performed the UV-damage experiment in deoxygenated water, and the resulting cleavage pattern observed by SDS-PAGE shows similarly sized fragments but with a lower yield, indicating that molecular oxygen, or reactive oxygen species, may be involved in the photodamage mechanism. To test whether the product distribution results from direct photoexcitation of the protein or its interaction with photochemically produced reactive oxygen species, we also used a modified Fenton reagent, which produces hydroxyl radicals *in situ*, to damage the protein.²⁵ The highly specific UV-B cleavage pattern is not observed in the protein after damage by hydroxyl radicals; instead, a broad distribution of products is observed on the gel. Thus, we conclude that the mechanism of cleavage is initiated by UV-B photoexcitation of the protein itself, but is facilitated by molecular oxygen. This result is interesting in light of the fact that molecular oxygen

is indeed present in the lens, but at smaller levels in the lens nucleus than in the outer portions of the lens.³

3.4.3 Mass spectrometry of UV-B irradiated γ D-crystallin identifies semitryptic fragments

To identify damaged side chains and determine the damage sites that lead to the UV-B induced fragments, we digested the UV-B damaged protein with bovine trypsin and analyzed the resulting cleavage products using LC-MS/MS. This method is similar to the one we used to identify the amyloid fiber core sequence of acid-induced γ D-crystallin fibers in a previous publication.²¹ A sequence coverage of 94% was obtained from the analyzed tryptic fragments generated by this method. Representative MS/MS spectra of tryptic and semitryptic fragments are shown in Figure 3.2A-F. In the mixture of peptides generated by sequential UV-B photodamage and trypsin digest, we observe a variety of species generated from the same sequence. Figure 3.2A shows the CID spectrum and assignments of an unmodified peptide covering residues 117–139. In addition to this peptide, a variety of others corresponding to the same sequence region were also observed. For example, multiple oxidation products were identified in the 117–139 tryptic peptide, with Trp130 and His121 residues being oxidized (Figure 3.2B). The locations of oxidation are determined from +16 m/z shift in β -series ions compared to the control (Figure 3.2A). A number of residues have been shown to be similarly modified in natural cataracts.^{4,7,11} This damage is consistent with the loss of aromatic amino acid absorbances in the UV-vis difference spectrum shown in Figure 3.1A, and may also act to destabilize the protein fold resulting in aggregation.

In addition to oxidized products, smaller semitryptic peptides were observed. In these peptides, one end does not correspond to a tryptic site. It is unlikely that these result from residual chymotrypsin activity in the bovine trypsin we used for the cleavage reaction, because they do not all correspond to chymotryptic sites and the trypsin was treated with L-1-Tosylamide-2-phenylethyl chloromethyl ketone to inhibit chymotryptic activity. Furthermore, these sites are not observed in tryptic digests of γ D-crystallin that have not been treated with UV-B radiation.²¹ Thus, they most likely arise from cleavage of peptide bonds as a direct result of UV-B radiation. Two such peptides are shown in Figure 3.2C-D: one containing residues 117–130 and another containing residues 131–139. These peptides arise from backbone cleavage between Trp130 and Val131. Together, they cover the entirety of the tryptic peptide 117–139. A variety of other semitryptic peptides and peptide pairs were observed for this region and others in γ D-Crystallin. Notably, many of the semitryptic peptides also contain internal oxidations, indicating that these modifications are closely linked. An example of such a peptide is the semitryptic peptide 130–139, described in Figure 3.2E,F. In Figure 3.2E, this peptide is identified without oxidative modification. Figure 3.2F shows a similar fragment pattern, but a number of the β -series ions are shifted by +16 m/z. Again, from the full ion assignment (Table A1. 5-6) we determine that the oxidation is on the tryptophan residue Trp130.

By mapping the sites of oxidized and nontryptic cleavage to the primary structure of γ D-crystallin (Figure 3.3A), we obtain a picture of the sequence context of both kinds of damage sites. There appears to be a close association between the positions of semitryptic sites and tryptophan residues in the sequence, while other redox-active amino acids (Met, Cys, Tyr, His) are not obviously correlated with the cleavage sites. Perhaps the most striking observation is that there are two stretches of sequences, containing about 15 amino acids, which are particularly susceptible to

cleavage. The locations of these sites in the protein's primary structure, and the approximate lengths of expected cleavage fragments, are represented in schematic form in Figure 3.3B. Each domain contains one such sequence, centered around analogous tryptophans (W42 and W130). Cleavage of the γ D-crystallin backbone within one or both of these would result in fragments with molecular masses of approximately 15.4 (17.8), 15.3, 9.9, 5.6 (8.), and 5.3 kDa (His₆-tagged molecular masses in parentheses), consistent with the SDS-PAGE results presented above. Unlike the backbone cleavage sites, which occur in clusters in the peptide primary structure, the oxidized amino acid side chains we identified appear to be distributed throughout the protein with no obvious sequence association. However, a sequence alignment of the N-terminal and C-terminal domains (Figure 3.3C) shows that the oxidized amino acids occur in similar regions of both domains, even when their identities are different.

Visualization of side chain oxidation and nontryptic cleavage sites in the native crystal structure²⁹ of γ D-crystallin (Figure 3.3D) reveals that they occur in β -hairpins located near the domain interface, and are quasi-symmetric about the interface. These interfacial β -hairpins are relatively occluded from solvent, and it is therefore unlikely that they form primary attack sites for radical species in solution.^{34,35} Together with the comparison of photochemically generated and hydroxyl radical generated fragmentation in Figure 3.1B, this result suggests that the oxidation and fragmentation occurs as a result of the absorption of UV-B light by γ D-crystallin itself and not indirect processes arising from the radiolysis of solvent species. Furthermore, the close structural association between oxidation and nontryptic cleavage sites suggests that they may arise from a common mechanism.

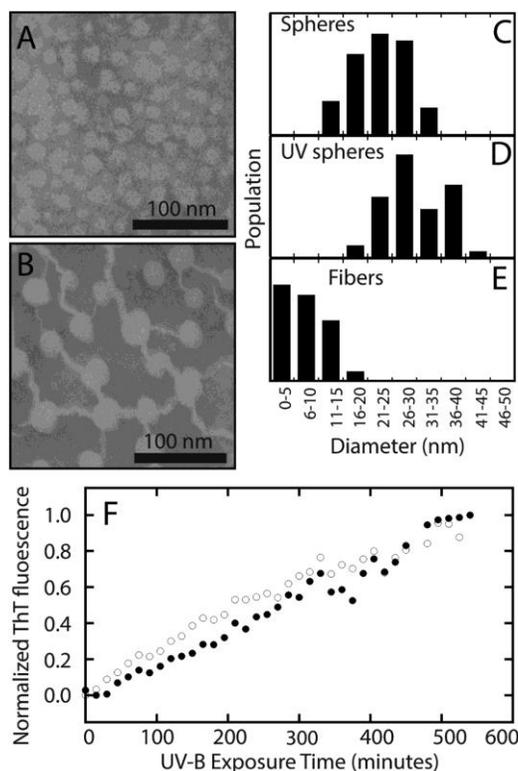


Figure 3.4 Aggregates of human γ D-Crystallin.

TEM images of undamaged (A) and UV-B photodamaged (B) γ D-Crystallin (S84C) show the presence of spherical bodies in both samples and the formation of fibers in the photodamaged sample after 6 h of illumination. Analysis of the diameters of spheres and fibers in both samples (C-E) shows a slight increase in sphere sizes from ~ 20 nm to ~ 30 nm and the formation of fibers with a mean diameter of ~ 6 nm. The aggregation of wild type (open circles) and S84C (closed circles) γ D-Crystallin is accompanied by an increase in ThT fluorescence over the course of 10 h (F).

3.4.4 Structural analysis of UV-B irradiated γ D-crystallin

The above analysis provides no structural information on the aggregates, so we turn to transmission electron microscopy, thioflavin T (ThT) fluorescence, and two-dimensional infrared (2D IR) spectroscopy. Samples of aggregates taken prior to illumination and after 6 hours of illumination were negatively stained with methylamine tungstate. Representative TEM images are shown in Figure 3.4A-B, and the diameters of the observed structures are shown in Figure 3.4C-E as histograms. Prior to UV-B damage, spherical structures with a diameter of ~ 20 nm are seen in γ D-crystallin samples. Such structures were also observed in early acid-induced aggregates of γ D-

crystallin, prior to fiber formation.¹⁴ After irradiation for 6 hours, fibers are observed in addition to the spheres. The fibers have mean diameters of approximately ~6 nm, similar to those observed in TEM images of acid-induced γ D-crystallin amyloid fibers,²¹ but have a relatively wide distributions of diameters. These results are in striking contrast to previous studies on UV-induced aggregates of mammalian γ -crystallins, which did not show the presence of fibers in TEM images despite FTIR evidence for extended vibrational coupling in β -sheets.¹⁶

ThT is a fluorescent dye that acts as an indicator for amyloid-like β -sheets, based on an increase in fluorescence quantum yield at 486 nm after excitation at 430 nm.³⁶ We tested for the presence of amyloid-like structures in UV-B induced γ D-crystallin (S84C) and wild type γ D-crystallin aggregates by removing aliquots over the course of a 10 hour illumination, and adding ThT in a 2.5:1 dye:protein ratio. The normalized fluorescence intensity change is plotted in Figure 3.4F. The aggregation kinetics show a slow rise in fluorescence over the course of 10 hours, after which the heterogeneity of the samples prevented the collection of ThT fluorescence data. ThT binds to amyloid β -sheets and thus the increase in fluorescence is evidence for the formation of amyloid fibers.³⁶ The kinetics of the ThT fluorescence increases in S84C mutant and wild type samples are nearly identical, indicating that the mutation is not responsible for ThT binding. No clear lag phase is observed, indicating that nucleation events are not the rate limiting step in UV-B induced fiber formation. Moreover, the steady rise in ThT binding with exposure time suggests that only damaged proteins are contributing to amyloid formation; if damaged proteins were acting as seed to recruit the aggregation of undamaged proteins, then we would expect a sigmoidal dependence on fluorescence with exposure time. Thus, ThT binding provides evidence for amyloid formation. However, ThT does not always accurately reflect structural changes in fibers^{37,38} and it is not known whether ThT can also bind to the spherical morphologies observed in the TEM and

that increase in size upon UV-B irradiation (Figure 3.4A-B). For a more structurally specific probe, we turn to 2D IR spectroscopy.

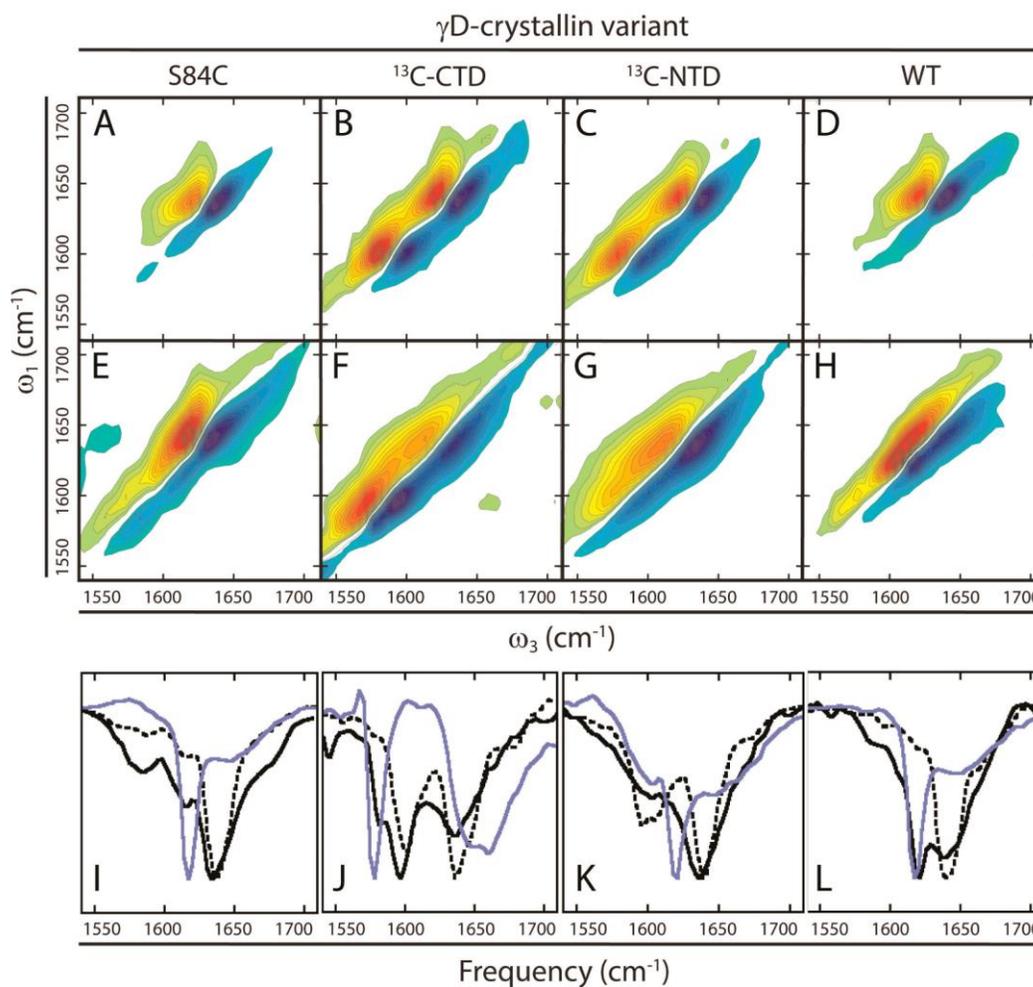


Figure 3.5 2D IR spectra of γ D-Crystallin variants.

In native, undamaged γ D-Crystallin (A-D), unlabeled β -sandwich domains appear at $\sim 1638\text{ cm}^{-1}$ along the diagonal, and labeled domains appear at $\sim 1598\text{ cm}^{-1}$. Segmentally labeled S84C proteins (B,C) have highly similar spectra with two major diagonal peaks. UV-B photodamaged samples of these proteins (E-H) reveal shifts in the spectra consistent with denaturation of the N-terminal domain and the formation of low-frequency amyloid β -sheet signals arising from the C-terminal domain. Diagonal slices (I-L) of native (dashed, black), UV-B photodamaged (solid, black), and acid-induced amyloid fibers (solid, blue) of each variant reveal the appearance of signals consistent with amyloid fiber formation in the UV-B photodamaged samples.

2D IR spectroscopy has recently emerged as a useful technique in structural analysis of proteins, especially in cases involving complex, heterogeneous, or disordered samples such as membrane proteins and protein aggregates.^{15,21,39-48} 2D IR spectroscopy is sensitive to protein secondary structure, vibrational coupling, and solvent exposure based on frequencies and cross-peaks. Unlike FTIR spectroscopy, 2D IR signals scale with the fourth power of the transition dipole moment, leading to better resolution of peaks in the spectra.⁴⁹ Isotope labeling strategies, including segmental ¹³C labeling,^{15,21} allow further resolution of signals from specific residues within the protein.²⁶ Here, we use 2D IR spectroscopy and segmental ¹³C labeling to examine the secondary structure content of UV-B induced γ D-crystallin aggregates and the domain origin of the fiber β -sheets.

Figure 3.5A-C shows 2D IR spectra of unlabeled γ D-crystallin (S84C) with a uniformly ¹³C labeled C-terminal domain (CTD), γ D-crystallin (S84C) with a uniformly ¹³C labeled N-terminal domain (NTD), and unlabeled wild type γ D-crystallin. The expressed protein ligation protocol used to isotope label the domains results in an S84C mutation, which was replicated in the unlabeled protein so that the unlabeled protein sequence does not differ from the labeled protein sequence. The S84C mutation has no effect on the protein stability, spectra or aggregation.¹⁵ Figure 3.5G-I shows slices through the diagonal of these 2D IR spectra (black, dashed line). The features of these spectra were discussed in detail in previous publications^{15,21} and have been calculated from molecular dynamics simulations.⁵⁰ Briefly, the most intense signals along the diagonals in these spectra arise from the short, antiparallel β -sheets of the protein's β -sandwich motifs. In the segmentally labeled spectra (Figure 3.5B, C), the signal from the labeled domains is red-shifted by approximately 40 cm⁻¹, consistent with replacement of ¹²C with ¹³C in an entire domain. Because vibrational couplings between residues within the domains are conserved,¹⁵ the lineshapes

of the individual β -sandwich domains in the labeled proteins are similar to those of the unlabeled proteins. Thus, by isotope labeling, we can separately resolve the structure changes associated with residues in the CTD or NTD.

Protein	β -sheet fold	Amide I frequency (^{12}C)	Amide I frequency (^{13}C)	References
γ D-crystallin (UV-B induced aggregates)	Amyloid	1624 cm^{-1}	1584 cm^{-1}	This publication
γ D-crystallin (native)	Greek key	1638 cm^{-1}	1598 cm^{-1}	(14, 15, 50)
γ D-crystallin (acid induced fibers)	Amyloid	1617 cm^{-1}	1575 cm^{-1}	(14, 15)
hIAPP	Amyloid	1617 cm^{-1}	N/A	(44, 47, 49)
A β	Amyloid	1630 cm^{-1}	N/A	(42)
TrpZip	β -hairpin	1635 cm^{-1}	N/A	(52)
Concanavalin A	Antiparallel β -sheet	1635 cm^{-1}	N/A	(53)
Parallel β -dimer	Non-natural	1640 cm^{-1}	N/A	(54)

Table 3.1 Representative IR frequencies of β -sheet amide I modes.

2D IR spectra of the three protein variants after 6 hours of UV-B exposure are shown in Figure 3.5E-H, along with slices in Figure 3.5I-L (black, solid line). It is clear from these spectra that UV-B photodamage results in conformational changes in γ D-crystallin based on changes in 2D frequencies and lineshapes compared to the spectra of the native proteins. The spectrum of UV-B aggregated unlabeled protein displays a new, narrow diagonal feature at approximately 1624 cm^{-1} that does not exist in the native protein spectra (Figure 3.5E vs A). We assign this feature to extensively coupled peptide bonds in an amyloid-like β -sheet conformation. Although both the native and amyloid states of γ D-crystallin contain β -sheets, they appear at different frequencies due to the differences in β -sheet architecture, which in turn influence the vibrational coupling between peptide bonds. Because neighboring peptide bonds in adjacent β -strands have large negative vibrational coupling constants, large shifts to lower frequency can occur when multiple β -strands form a β -sheet, delocalizing vibrations. The magnitude of this effect is demonstrated by a survey of reported values in the literature for the frequencies of β -sheets in soluble proteins and amyloid fibers (Table 3.1). In soluble proteins such as native γ D-crystallin, concanavalin A, and

small β -hairpins, the β -sheet frequency appears between 1630 and 1640 cm^{-1} . Reported amide I frequencies for the β -sheets in well-ordered amyloid fibers, on the other hand, fall between 1617 and 1620 cm^{-1} ,^{15,41,42,47,49,51} much closer to the 1624 cm^{-1} reported here.

For a comparison of the UV-B induced aggregates to well-ordered amyloid fibers, we plot slices through previously published 2D IR spectra¹⁵ of γ D-crystallin that was induced to aggregate by acid denaturation (blue). Acid denaturation causes amyloid fibers to form,^{14,15,21} and the resulting peak matches that for UV-B induced aggregation. In addition to this amyloid β -sheet feature, both spectra show significant broadening of the amide I peak near 1640 cm^{-1} , indicating a disordering of the non-amyloid components of the aggregates. Thus, the conformational changes in UV-B induced aggregates are not as severe as those in the acid-induced amyloid fibers, but it is clear that the UV-B aggregates share the main features of the acid induced amyloid fibrils, indicating broad similarities in their aggregate structures. UV-B induced aggregates of the unlabeled wild type protein (Figure 3.5H,L) show an amyloid peak in a similar position to that of the S84C mutant (Figure 3.5E,I), indicating again that the mutation itself does not cause amyloid fiber formation.

To study the structure of the UV-B aggregates of γ D-crystallin in greater detail, we turn to the spectra of proteins with ^{13}C labeled domains.¹⁵ The 2D IR spectra of the C-terminally labeled aggregates are shown in Figure 3.5F with diagonal slices in Figure 3.5J. These spectra allow us to assign the ordered fiber β -sheet signal to residues derived from the protein's two domains. If well-formed amyloid β -sheets are formed from NTD residues, then a sharp peak will appear at 1624 cm^{-1} like in the wild type spectrum above. If the sharp peak is instead observed at 1624 - 40 = 1584 cm^{-1} , then ^{13}C labeled residues in the CTD are forming the amyloid β -sheets since isotope labeling causes a 40 cm^{-1} frequency shift. In Figure 3.5F, a narrow feature is visible on the low-

frequency side of the ^{13}C labeled diagonal amide I peak, at 1584 cm^{-1} . A comparison to our previously published acid-induced denaturation of γD -crystallin,¹⁵ using the same isotope labeling scheme, shows that the sharp amyloid features appear at similar frequencies. Thus, residues from the C-terminal domain are the source of well-ordered amyloid β -sheets. These amyloid features are small compared to the majority of the CTD features, which absorb at 1595 cm^{-1} . 1595 cm^{-1} is about 10 cm^{-1} lower frequency than before UV-B damage (compare solid to dashed lines). The only reasonable way that a lower frequency can be created is by additional negative coupling constants – dehydration would produce a higher frequency shift and hydration caused by unfolding would create a peak no lower than 1610 cm^{-1} .⁵⁵ This frequency range is most consistent with a β -sheet structure that is more strongly coupled than the β -sandwich domains in native γD -crystallin, but less ordered than fully formed amyloid fibers (see Table 3.1). Kinetics experiments following amyloid fiber formation have revealed a progressive shift to lower frequency as fibers form.¹⁵ Thus, we postulate that in addition to well-formed amyloid fibers, there are also β -sheet intermediates. In contrast, the ^{12}C amide I signal of the NTD is centered at 1640 cm^{-1} , which matches that of the native protein, but is significantly broadened compared (the broadening is more apparent in the 2D IR spectra themselves, Figure 3.5F, which shows an elongation along the diagonal). Thus, the unlabeled N-terminal domain becomes disordered upon UV-B exposure.

No sharp peak is observed at 1624 cm^{-1} in Figure 3.5F-G, implying that the NTD does not contribute to the amyloid β -sheets, but since this region is spectrally congested, we also swapped our labeling scheme. Shown in Figure 3.5G is the 2D IR spectrum of γD -crystallin with a ^{13}C labeled NTD. No narrow β -sheet feature is visible at 1584 cm^{-1} , confirming that the NTD does not contribute to the fibers. As above, the NTD features are broadened, consistent with the

conclusion that the NTD becomes disordered upon aggregation. This spectrum also emphasizes a broadening to the width of the CTD, in addition to the lower frequency. Thus, we conclude that fiber forming residues are exclusively derived from the C-terminal domain and that the β -sandwich motifs are not conserved upon UV-B exposure.

3.5 Discussion

Our results provide the most detailed analysis of the influence of UV-B irradiation on lens crystallin structure to date. This information is important because exposure to ultraviolet radiation (including UV-B) from sunlight is a known cause of cataracts.^{1,5,7,13} Based on SDS-PAGE, mass spectrometry, transmission electron microscopy, ThT binding, and 2D IR spectroscopy, we identify the location of side chain oxidation and polypeptide cleavage induced by UV-B radiation, describe the morphology of the resulting aggregates, and obtain unprecedented details of the molecular structures of proteins within the aggregates. Among the structures observed are amyloid fibers that form from residues in the C-terminal domain of γ D-Crystallin, a finding that is the first definitive proof that amyloid fibers can form under conditions relevant to cataract formation. This finding is particularly interesting because the presence of amyloid fibers in natural cataract material has not been conclusively established, possibly due to the structural complexity of the aggregates and the difficulty associated with obtaining detailed structural information about them.

In contrast to acid-induced amyloid fiber formation,²¹ the UV-B induced aggregation of γ D-crystallin is accompanied by clear changes in the covalent structure of the protein, in the form of cross-linking, polypeptide cleavage, and side chain damage (Figure 3.1). However, these modifications are similar in nature to those observed in cataractous lenses where a variety of protein fragments and post-translational modifications have been characterized.^{1,3,4} From our LC-MS/MS results (Figure 3.2), it is clear that UV-B photodamage of γ D-crystallin is structurally

specific (Figure 3.3), with both oxidative damage and backbone cleavage occurring at sites near the interdomain interface.

The locations of these damage sites within the sequence and structure of γ D-crystallin provide clues to the mechanism by which the protein is initially modified and destabilized to induce aggregation. First, backbone cleavage occurs largely in two specific stretches of sequence, between residues 40–53 and 120–134 where cleavage of nearly every polypeptide bond is observed. These sequences contain tryptophan residues W42 and W130, which have been shown to participate in photoprotective energy transfer mechanisms that are thought to prevent damage from UV radiation.^{22–24} Interestingly, they are a homologous pair that has much higher fluorescence quantum yields than W68 and W156 in single tryptophan mutants.²² This may also indicate that they are more susceptible to other excited state quenching processes including side chain oxidation and backbone cleavage. Because we showed that the specific polypeptide cleavage patterns are light-dependent (Figure 3.1B), we therefore invoke a model in which the failure of the hypothesized photo-protection mechanism of γ D-crystallin results in distributed damage to the protein in regions near photochemically active tryptophan residues. In previous reports on the UV-B photo-induced aggregation of chicken egg white lysozyme, tryptophan and tyrosine photochemistry induced the reduction of native disulfide bonds, thereby inducing aggregation.^{56,57} The close tertiary structural association of backbone cleavage and side chain oxidation sites in γ D-crystallin suggests that the two kinds of damage arise from this initial tryptophan photochemistry, and the presence of oxidized residues in both tryptic and semitryptic peptides (Figure 3.2) indicates that both kinds of damage can occur in the same molecule.

The details of the chemical mechanism of UV-B photodamage, while currently unknown, are consistent with a radical transfer mechanism initiated by tryptophan photooxidation. In

previous reports on other proteins, oxidative photodamage of aromatic amino acid side chains (Trp and Tyr) was shown to be capable of initiating peptide bond cleavage.^{35,58} For example, the formation of N-formylkynurenine from tryptophan results in increased peptide bond hydrolysis.^{58,59} Additionally, the formation of side chain radicals via photooxidation can result in radical migration to the main chain (C α) and subsequent peptide bond cleavage.³⁵ Based on the broad distribution of side chain oxidation and main chain cleavage products within specific regions of sequence, as well as the dependence of backbone cleavage on UV-B light, we believe that such a radical based mechanism is the most plausible explanation for our results. This mechanism, initiated by the failure of the hypothesized energy transfer mechanism of photo=protection, would likely become more pronounced as the protein is destabilized by accumulated photodamage. The cleavage of the polypeptide backbone in γ D-crystallin is an obvious mode of protein destabilization, since it compromises the integrity of the molecule that is required for folding. However, the presence of side chain oxidation in the region near the interdomain interface suggests another mode of protein destabilization in the fraction of protein that is not cleaved by UV-B radiation. The interface region has been shown to be particularly important in the stabilization of γ D-crystallin,⁶⁰ and modification of amino acid side chains at the interface is likely to facilitate unfolding of the protein. Because only local unfolding is required to nucleate fibers,⁶¹ even relatively small changes in stability due to oxidation can result in the entry of the protein into the amyloid aggregation pathway.

Although these results suggest mechanisms for aggregation initiation *in vitro*, they do not necessarily describe the aggregation of γ D-crystallin *in vivo* despite the consistency of the types of modification we observe with lens proteomics results that identified a variety of oxidation and cleavage products in lens crystallins. Unfortunately, only a few studies have examined these

modifications, and their role in cataract formation, in detail. One such study identified the locations of side chain modifications in three isoforms of a γ D-crystallin fragment covering residues 87–173, or the entire C-terminal domain, using antibodies raised against six peptides contained within this region.⁶² While the epitopes in that study did not cover the entire C-terminal sequence (see regions highlighted in yellow in Figure 3.3A), large changes in reactivity to the antibodies were observed at residues 114–120, as well as 137–143, 149–154, and 165–173, suggesting side chain modifications in the natural fragments.⁶² Our MS/MS results, which cover a much larger fraction of the γ D-crystallin sequence, also show side chain oxidations in this region (Figure 3.3A), drawing a link between natural modifications in cataracts and our *in vitro* results. In fact, our results show that the region of the C-terminal domain that is most susceptible to damage lies in a gap between antibody recognition sites in the previous study,⁶² indicating that damage in the lens may in fact be much more extensive. Despite the similarities in the locations of side chain oxidations to the natural γ D-crystallin 87–173 fragment, our results suggest an alternative polypeptide cleavage pattern than that which was observed in the previous study on human lens material. However, it must be noted that characterization of a single, high abundance protein fragment does not preclude the presence of other fragments that exist at lower concentration or that have such a large degree of modification that they are difficult to identify using immunoassays.

From TEM images, we learn that γ D-Crystallin samples contain aggregates with spherical morphology both before and after UV–B illumination, with a slight increase in diameter after illumination. However, the UV–B exposed samples also contain fibers that are not present in samples not exposed to UV–B light. These fibers are predominantly ~5–10 nm across, within the range typical for amyloid fibers and similar to those produced by acid denaturation of γ D-Crystallin.^{14,21} Perhaps the most striking aspect of these fibers is that they are closely associated

with spheres, and appear to terminate at the spheres. On the other hand, spheres without fibers are also observed in the TEM images. This suggests that the fibers are growing from the spheres. Such a fiber formation mechanism is possible if two aggregation processes are at work: initial precipitation and subsequent fiber formation. It is known that single point mutations can cause crystallins to precipitate very quickly, which is thought to be the origin of congenital juvenile cataracts that form amorphous aggregates.^{1,63-66} A similar change in solubility may be expected upon covalent modification by UV-B irradiation. Therefore, we hypothesize that the initial precipitation into spherical aggregates occurs as a result of UV-B photodamage, and the spherical aggregates then reorganize into fibers. The ThT binding kinetics study (Figure 3.4F) supports this model. Typical amyloid aggregation is characterized by the presence of a lag phase due to the fact that fiber nucleation is the rate determining step in fiber formation.⁶⁷ The UV-B induced aggregation, monitored by ThT fluorescence, is characterized by a steady rise in aggregate concentration with no discernible lag phase. Thus, it is likely that the initial aggregation of the protein is fast and therefore not rate limiting, and the steady rise in ThT fluorescence occurs as precipitated proteins reorganize into fibers and additional proteins precipitate as photodamage is continued.

2D IR spectroscopy, which is sensitive to protein secondary structure,^{15,21,39-48,68} provides the molecular structure information required to link the observed UV-B photodamage pattern in γ D-Crystallin to the morphology of the aggregates. From the 2D IR spectra of unlabeled γ D-Crystallin (S84C and wild type, Figure 3.5E,H) we learn that the UV-B induced aggregates contain both a narrow β -sheet feature at $\sim 1624\text{ cm}^{-1}$ and a broadened feature at $\sim 1640\text{ cm}^{-1}$. The samples used to collect these spectra, like those shown in TEM images, contain both fibers and spheres that cannot currently be analyzed separately. However, it is clear from a comparison of these spectra

to those of the corresponding native proteins (Figure 3.5A,D) that the native β -sandwich fold is not retained in the aggregates. This result is also important because it rules out domain swapping as the predominant mechanism for UV-B induced γ D-crystallin aggregation. In γ D-crystallin, the two domains are connected by a flexible loop, and make contact with each other via amino acid side chains at the interdomain interface.⁶⁰ Mutation studies of this interface have shown that it contributes to protein stability,⁶⁰ but the individual domains are also quite stable and the presence of both domains is not necessary for their proper folding.^{14,69} The β -crystallins also contain two domains, but associate into domain-swapped oligomers.^{1,2} Thus, one postulate for γ D-crystallin aggregation is that damage destabilizes the interface leading to propagation of domain swapping that ultimately causes enough protein clusters to induce precipitation.⁷⁰ Although this mechanism is not consistent with the amide I lineshapes in the 2D IR spectra of the final UV-B aggregates, the spectra are congested enough that domain swapped aggregates might constitute a minority population. Additionally, they may play a role in an early step in the aggregation mechanism that is not observed in our current data, as has been observed in molecular dynamics simulations.⁷⁰ However, due to the degree of side chain damage observed at the interface, specific contacts are likely lost due to the same damage that destabilizes the protein, thus preventing domain swapping from occurring.

The appearance of the amide I signal at $\sim 1624\text{ cm}^{-1}$ in unlabeled UV-B photodamaged γ D-crystallin allows us to examine the architecture of β -sheets in the resulting aggregates. β -sheets in proteins have infrared signatures that cover a relatively broad range of frequencies, from $\sim 1617\text{ cm}^{-1}$ to $\sim 1640\text{ cm}^{-1}$, depending on their size and organization.^{15,42,47,49,52-54,71} From a survey of frequencies reported in the literature for both soluble and fibrillar proteins containing β -sheets (Table 3.1, and references therein), we know that β -sheets in amyloid fibers have frequencies near

1620 cm^{-1} , while those in soluble proteins have frequencies greater than 1630 cm^{-1} . In acid induced fibers of γD -crystallin, the β -sheet frequency is 1617 cm^{-1} , and the native β -sandwich frequency is 1638 cm^{-1} .^{14,15} Thus, we conclude that the fiber components of the UV-B induced aggregates are indeed amyloid fibers. The frequency difference of $\sim 6\text{--}8$ cm^{-1} between UV-B induced and acid induced γD -crystallin amyloid fibers reflects a decreased degree of vibrational coupling in the UV-B induced fibers.^{15,49,51,68} It has been shown that the IR frequency of the β -sheet amide I feature depends on the number of residues in the β -sheet.^{49,51} Early intermediates in the acid-induced pathway, which are presumably made up of small aggregates in the β -sheet conformation, also have higher β -sheet frequencies.¹⁵ During aggregation, these frequencies progressively shift to lower values as the fibers grow.¹⁵ However, the fibers we observe in the TEM images of UV-B induced aggregates are already longer than 10 nm and up to >100 nm along their axis (Figure 3.4B). Thus, it is unlikely that the amide I frequency difference between acid and UV-B induced fibers results from differences in fiber length. Instead, the higher β -sheet frequency in the UV-B fibers probably arises from structural disorder, which can also reduce vibrational coupling. Based on our results, this disorder may arise from either of two sources. First, the assembly of polypeptide fragments of different lengths may disrupt normal fiber packing. A more likely source of disorder, though, is oxidative damage to side chains, which can prevent efficient side chain packing between fiber β -sheets or recruit water into hydrophobic interfaces, thereby disrupting local secondary structure and vibrational coupling.

Using segmental ^{13}C labeling, which resolves the two domains in 2D IR spectra,^{15,21} we previously showed that in the acid-induced fibers of γD -crystallin, only the C-terminal domain forms β -sheets while the N-terminal domain becomes disordered.¹⁵ This result was surprising because both domains, which are highly similar, can form amyloid fibers in isolation.¹⁴ Careful

consideration of the sequences of the domains, however, reveals that the C-terminal domain contains five additional carboxylic acids, which are protonated at low pH (≤ 3) and are a likely source of destabilization. This is in contrast to the UV-B induced photodamage in γ D-crystallin, which is nearly symmetrical between the two domains (Figure 3.3C,D). Based on this, and the fact that backbone cleavage destabilizes and separates the domains, one might conclude that both the N-terminal and C-terminal domains are likely to form fibers in UV-B induced aggregates. However, our 2D IR spectra of segmentally ^{13}C labeled γ D-crystallin (Figure 3.5F,G) show that, like in the acid induced fibers, only residues from the C-terminal domain form amyloid fibers in UV-B induced aggregates. Thus, we conclude that the C-terminal domain has a higher intrinsic amyloidogenicity than the N-terminal domain, and forms fibers more readily regardless of the method of denaturation used. It should be noted that there are some conditions under which the N-terminal domain can in fact form fibers, as was demonstrated by previous experiments.¹⁴ Within the product mixture generated by UV-B photodamage, though, the amorphous aggregation pathway appears to be favored over the amyloid aggregation pathway for N-terminal domain sequences. From these results, we learn that the destabilization of the C-terminal domain induces its entry into the amyloid misfolding pathway. However, a number of questions remain unresolved about the mechanism of fiber formation and the structures of the proteins in the fibers. The 2D IR spectra of segmentally labeled proteins show that some fraction of the C-terminal domain remains outside the fiber cores in a denatured conformation. While it is clear that only the C-terminal domain forms amyloid β -sheets, our data does not show which components of the complex photodamage product mixture form fibers, spheres, or soluble components. The fibers may contain full-length proteins destabilized by side chain oxidation, fragments containing some C-terminal residues, or both.

3.6 Conclusion

Despite the numerous unresolved structural and mechanistic details of UV-B induced γ D-crystallin aggregation, the incorporation of only the C-terminal domain of γ D-crystallin into UV-B induced amyloid fibers has important implication for the formation of cataracts in vivo. The γ -crystallins constitute $\sim 25\%$ of lens proteins,⁷² and both γ -crystallins and other crystallin proteins have been found to form amyloid fibers under relatively mild denaturing conditions.^{17,19,20,73} Because both backbone cleavage and side chain oxidation are observed in UV-B damaged γ D-crystallin as well as in natural cataracts,^{1-13,62,74-76} it is reasonable to assume that the mechanisms we have identified here could also apply in the lens. However, the presence of amyloid fibers in natural cataracts has not been conclusively established, and their role in cataract formation remains a subject of debate. Only a few studies have provided evidence that the amyloid misfolding pathway may be physiologically relevant. For example, FTIR spectra of decapsulated cataract tissue exhibits an absorption at 1620 cm^{-1} , which is especially pronounced in cataract tissue taken from patients that suffer from glaucoma.⁷⁷ This amide I frequency is indicative of the extended vibrational coupling in amyloid β -sheets (Table 3.1).^{49,51} Additionally, the presence of heavily modified C-terminal fragments of γ D-Crystallin, such as the 9 kDa fragment discussed above,⁶² suggests that highly amyloidogenic protein fragments are produced in the lens. This raises the question: if the propensity of crystallins is to form amyloid fibers in vitro, and amyloidogenic species are present in the lens, why is there so little evidence for amyloid species in cataract material? One suggestion has been that there is a competition between fast precipitation into unfolded structures, such as the ones reported here as spheres in the TEM images, and amyloid fiber formation.⁷⁸ But in this report, our results suggest that disordered precipitates can reorganize into fibers, so presumably a similar mechanism could occur in vivo. It is also possible that protein

crowding, or interactions with α -crystallins (which are chaperone proteins) prevent such fragments from forming fibers;^{79,80} in vitro studies have shown that α -crystallins do indeed bind fibers formed by proteins such as α -synuclein⁸¹ and modulate the amyloid aggregation of γ -crystallins.⁸⁰ Finally, it is possible that fibers are indeed present in vivo, but standard techniques for studying cataract material cannot detect them due to the congestion of the samples with other structures that prevent resolution in TEM images or fiber-specific fluorescent dye binding. This latter point emphasizes the important role that 2D IR spectroscopy may play in the study of cataract formation. Due to its ability to resolve secondary structures in complex, heterogeneous samples, it is a unique and powerful tool for investigating protein aggregate architecture, both *in vitro*, and potentially, in samples extracted from tissue.

3.7 References

- (1) Wang, Y., and King, J. A. (2010) Cataract as a protein aggregation disease, *Protein Misfolding Diseases*, pp 487–515, John Wiley & Sons, Inc.
- (2) Bloemendal, H., de Jong, W., Jaenicke, R., Lubsen, N. H., Slingsby, C., and Tardieu, A. (2004) Ageing and vision: structure, stability and function of lens crystallins. *Prog. Biophys. Mol. Biol.* **86**, 407–485.
- (3) Truscott, R. J. W. (2005) Age-related nuclear cataract-oxidation is the key. *Exp. Eye Res.* **80**, 709–725.
- (4) Hanson, S. R. A., Hasan, A., Smith, D. L., and Smith, J. B. (2000) The major in vivo modifications of the human water-insoluble lens crystallins are disulfide bonds, deamidation, methionine oxidation and backbone cleavage. *Exp. Eye Res.* **71**, 195–207.
- (5) Varma, S. D., HeyDe, K. R., and Kovtun, S. (2008) UV-B-induced damage to the lens in vitro: prevention by caffeine. *Journal of ocular pharmacology and therapeutics: the official journal of the Association for Ocular Pharmacology and Therapeutics* **24**, 439–444.
- (6) Xia, Z., Yang, Z., Huynh, T., King, J. A., and Zhou, R. (2013) UV-radiation Induced Disruption of Dry-Cavities in Human γ D-Crystallin Results in Decreased Stability and Faster Unfolding. *Sci. Rep.* **3**, 1560.
- (7) Davies, M. J., and Truscott, R. J. W. (2001) Photo-oxidation of proteins and its role in cataractogenesis. *J. Photochem. Photobiol. B: Biol.* **63**, 114–125.

- (8) Hott, J. L., and Borkman, R. F. (1992) Analysis of photo-oxidized amino acids in tryptic peptides of calf lens γ -ii Crystallin. *Photochem. Photobiol.* **56**, 257–262.
- (9) Schauerte, J. A., and Gafni, A. (1995) Photodegradation of tryptophan residues and attenuation of molecular chaperone activity in α -crystallin are correlated. *Biochem. Biophys. Res. Commun.* **212**, 900–905.
- (10) Wang, S. S.-S., and Wen, W.-S. (2010) Examining the influence of ultraviolet C irradiation on recombinant human γ D-Crystallin. *Mol. Vis.* **16**, 2777–2790.
- (11) Hains, P. G., and Truscott, R. J. (2007) Post-translational modifications in the nuclear region of young, aged, and cataract human lenses. *J. Proteome Res.* **6**, 3935–3943.
- (12) Kubiak, K., Kowalska, M., and Nowak, W. (2003) Molecular dynamics study of early events during photooxidation of eye lens protein γ B-Crystallin. *THEOCHEM* **630**, 315–325.
- (13) Balasubramanian, D., Du, X., and Zigler, J. S. (1990) The reaction of singlet oxygen with proteins, with special reference to crystallins. *Photochem. Photobiol.* **52**, 761–768.
- (14) Papanikolopoulou, K., Mills-Henry, I., Tho, S. L., Wang, Y. T., Gross, A. A. R., Kirschner, D. A., Decatur, S. M., and King, J. (2008) Formation of amyloid fibrils in vitro by human γ D-Crystallin and its isolated domains. *Mol. Vis.* **14**, 81–89.
- (15) Moran, S. D., Woys, A. M., Buchanan, L. E., Bixby, E., Decatur, S. M., and Zanni, M. T. (2012) Two-dimensional IR spectroscopy and segmental ^{13}C labeling reveals the domain structure of human γ D-Crystallin amyloid fibrils. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 3329–3334.
- (16) Fatima, U., Sharma, S., and Guptasarma, P. (2010) Structures of differently aggregated and precipitated forms of γ B-Crystallin: an FTIR spectroscopic and EM study. *Protein Pept. Lett.* **17**, 1155–1162.
- (17) Ecroyd, H., and Carver, J. A. (2009) Crystallin proteins and amyloid fibrils. *Cell. Mol. Life Sci.* **66**, 62–81.
- (18) Kosinski-Collins, M. S., and King, J. (2003) In vitro unfolding, refolding, and polymerization of human γ D Crystallin, a protein involved in cataract formation. *Protein Sci.* **12**, 480–490.
- (19) Meehan, S., Berry, Y., Luisi, B., Dobson, C. M., Carver, J. A., and MacPhee, C. E. (2004) Amyloid fibril formation by lens Crystallin proteins and its implications for cataract formation. *J. Biol. Chem.* **279**, 3413–3419.
- (20) Wang, Y., Petty, S., Trojanowski, A., Knee, K., Goulet, D., Mukerji, I., and King, J. (2010) Formation of amyloid fibrils *in vitro* from partially unfolded intermediates of human γ C-Crystallin. *Invest. Ophthalmol. Vis. Sci.* **51**, 672–678.
- (21) Moran, S. D., Decatur, S. M., and Zanni, M. T. (2012) Structural and sequence analysis of the human γ D-crystallin amyloid fibril core using 2D IR spectroscopy, segmental ^{13}C labeling, and mass spectrometry, *J. Am. Chem. Soc.*

- (22) Chen, J., Callis, P. R., and King, J. (2009) Mechanism of the very efficient quenching of tryptophan fluorescence in human γ D- and γ S-crystallins: the γ -crystallin fold may have evolved to protect tryptophan residues from ultraviolet photodamage. *Biochemistry* **48**, 3708–3716.
- (23) Chen, J., Flaugh, S. L., Callis, P. R., and King, J. (2006) Mechanism of the highly efficient quenching of tryptophan fluorescence in human γ D-Crystallin. *Biochemistry* **45**, 11552–11563.
- (24) Xu, J., Chen, J., Toptygin, D., Tcherkasskaya, O., Callis, P., King, J., Brand, L., and Knutson, J. R. (2009) Femtosecond fluorescence spectra of tryptophan in human γ -Crystallin mutants: site-dependent ultrafast quenching. *J. Am. Chem. Soc.* **131**, 16751–16757.
- (25) Xu, G., and Chance, M. R. (2007) Hydroxyl radical-mediated modification of proteins as probes for structural proteomics. *Chem. Rev.* **107**, 3514–3543.
- (26) Middleton, C. T., Woys, A. M., Mukherjee, S. S., and Zanni, M. T. (2010) Residue-specific structural kinetics of proteins through the union of isotope labeling, mid-IR pulse shaping, and coherent 2D IR spectroscopy. *Methods*, **52**, 12–22.
- (27) Bilbao, J., and de Miguel, A. (2010) Estimation of UV-B irradiation from total global solar meteorological data in central Spain, *J. Geophys. Res.* **115**.
- (28) Dyer, J. M., Bringans, S. D., and Bryson, W. G. (2006) Characterisation of photo-oxidation products within photoyellowed wool proteins: tryptophan and tyrosine derived chromophores. *Photochem. Photobiol. Sci.* **5**, 698–706.
- (29) Basak, A., Bateman, O., Slingsby, C., Pande, A., Asherie, N., Ogun, O., Benedek, G. B., and Pande, J. (2003) High-resolution X-ray Crystal Structures of Human γ D Crystallin (1.25 Å) and the R58H Mutant (1.15 Å) Associated with Aculeiform Cataract. *J. Mol. Biol.* **328**, 1137–1147.
- (30) Creed, D. (1984) The photophysics and photochemistry of the near-uv absorbing amino acids—i. tryptophan and its simple derivatives. *Photochem. Photobiol.* **39**, 537–562.
- (31) Kerwin, B. A., and Remmele, R. L., Jr. (2007) Protect from light: photodegradation and protein biologics. *J. Pharm. Sci.* **96**, 1468–1479.
- (32) Vladimirov, Y. A., Roshchupkin, D. I., and Fesenko, E. E. (1970) Photochemical reactions in amino acid residues and inactivation of enzymes during U.V.-irradiation. A review. *Photochem. Photobiol.* **11**, 227–246.
- (33) Callis, P. R. (1997) 1La and 1Lb transitions of tryptophan: Applications of theory and experimental observations to fluorescence of proteins, *Methods in Enzymology* (Ludwig Brand, M. L. J., Ed.) pp 113–150, Academic Press.
- (34) Cheal, S. M., Ng, M., Barrios, B., Miao, Z., Kalani, A. K., and Meares, C. F. (2009) Mapping protein-protein interactions by localized oxidation: consequences of the reach of hydroxyl radical. *Biochemistry* **48**, 4577–4586.
- (35) Xu, G., and Chance, M. R. (2007) Hydroxyl radical-mediated modification of proteins as probes for structural proteomics. *Chem. Rev.* **107**, 3514–3543.

- (36) Biancalana, M., and Koide, S. (2010) Molecular mechanism of Thioflavin-T binding to amyloid fibrils. *Biochim. Biophys. Acta*, **1804**, 1405–1412.
- (37) Middleton, C. T., Marek, P., Cao, P., Chiu, C. C., Singh, S., Woys, A. M., de Pablo, J. J., Raleigh, D. P., and Zanni, M. T. (2012) Two-dimensional infrared spectroscopy reveals the complex behavior of an amyloid fibril inhibitor. *Nat. Chem.* **4**, 355–360.
- (38) Wolfe, L. S., Calabrese, M. F., Nath, A., Blaho, D. V., Miranker, A. D., and Xiong, Y. (2010) Protein-induced photophysical changes to the amyloid indicator dye thioflavin T. *Proc. Natl. Acad. Sci.* **107**, 16863–16868.
- (39) Woys, A. M., Mukherjee, S. S., Skoff, D. R., Moran, S. D., and Zanni, M. T. (2013) A strongly absorbing class of non-natural labels for probing protein electrostatics and solvation with FTIR and 2D IR spectroscopies. *J. Phys. Chem. B* **117**, 5009–18.
- (40) Dunkelberger, E. B., Buchanan, L. E., Marek, P., Cao, P., Raleigh, D. P., and Zanni, M. T. (2012) Deamidation accelerates amyloid formation and alters amylin fiber structure. *J. Am. Chem. Soc.* **134**, 12658–12667.
- (41) Londergan, C. H., Wang, J., Axelsen, P. H., and Hochstrasser, R. M. (2006) Two-dimensional infrared spectroscopy displays signatures of structural ordering in peptide aggregates. *Biophys. J.* **90**, 4672–4685.
- (42) Kim, Y. S., Liu, L., Axelsen, P. H., and Hochstrasser, R. M. (2009) 2D IR provides evidence for mobile water molecules in β -amyloid fibrils. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 17751–17756.
- (43) Remorino, A., Korendovych, I. V., Wu, Y., DeGrado, W. F., and Hochstrasser, R. M. (2011) Residue-specific vibrational echoes yield 3D structures of a transmembrane helix dimer. *Science* **332**, 1206–1209.
- (44) Wang, L., Middleton, C. T., Singh, S., Reddy, A. S., Woys, A. M., Strasfeld, D. B., Marek, P., Raleigh, D. P., de Pablo, J. J., Zanni, M. T., and Skinner, J. L. (2011) 2DIR spectroscopy of human amylin fibrils reflects stable β -sheet structure. *J. Am. Chem. Soc.* **133**, 16062–16071.
- (45) Manor, J., Mukherjee, P., Lin, Y. S., Leonov, H., Skinner, J. L., Zanni, M. T., and Arkin, I. T. (2009) Gating mechanism of the influenza A M2 channel revealed by 1D and 2D IR spectroscopies. *Structure* **17**, 247–254.
- (46) Mukherjee, P., Kass, I., Arkin, I. T., and Zanni, M. T. (2006) Picosecond dynamics of a membrane protein revealed by 2D IR. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 3528–3533.
- (47) Shim, S. H., Gupta, R., Ling, Y. L., Strasfeld, D. B., Raleigh, D. P., and Zanni, M. T. (2009) Two-dimensional IR spectroscopy and isotope labeling defines the pathway of amyloid formation with residue-specific resolution. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 6614–6619.
- (48) Woys, A. M., Lin, Y. S., Reddy, A. S., Xiong, W., de Pablo, J. J., Skinner, J. L., and Zanni, M. T. (2010) 2D IR line shapes probe ovispirin peptide conformation and depth in lipid bilayers. *J. Am. Chem. Soc.* **132**, 2832–2838.

- (49) Strasfeld, D. B., Ling, Y. L., Gupta, R., Raleigh, D. P., and Zanni, M. T. (2009) Strategies for extracting structural information from 2D IR spectroscopy of amyloid: application to islet amyloid polypeptide. *J. Phys. Chem. B* **113**, 15679–15691.
- (50) Lam, A. R., Moran, S. D., Preketes, N. K., Zhang, T. O., Zanni, M. T., and Mukamel, S. (2013) Study of the γ D-Crystallin Protein Using Two-Dimensional Infrared (2D IR) Spectroscopy: Experiment and Simulation, Submitted.
- (51) Hahn, S., Kim, S. S., Lee, C., and Cho, M. (2005) Characteristic two-dimensional IR spectroscopic features of antiparallel and parallel β -sheet polypeptides: simulation studies. *J. Chem. Phys.* **123**,084905.
- (52) Jones, K. C., Peng, C. S., and Tokmakoff, A. (2013) Folding of a heterogeneous β -hairpin peptide from temperature-jump 2D IR spectroscopy. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 2828–2833.
- (53) Cheatum, C. M., Tokmakoff, A., and Knoester, J. (2004) Signatures of β -sheet secondary structures in linear and twodimensional infrared spectroscopy. *J. Chem. Phys.* **120**, 8201–8215.
- (54) Woys, A. M., Almeida, A. M., Wang, L., Chiu, C.-C., McGovern, M., de Pablo, J. J., Skinner, J. L., Gellman, S. H., and Zanni, M. T. (2012) Parallel β -sheet vibrational couplings revealed by 2D IR spectroscopy of an isotopically labeled macrocycle: quantitative benchmark for the interpretation of amyloid and protein infrared spectra. *J. Am. Chem. Soc.* **134**, 19118–19128.
- (55) Manor, J., Feldblum, E. S., Zanni, M. T., and Arkin, I. T. (2012) Environment polarity in proteins mapped noninvasively by FTIR spectroscopy. *J. Phys. Chem. Lett.* **3**, 939–944.
- (56) Xie, J.-B., Cao, Y., Pan, H., Qin, M., Yan, Z.-Q., Xiong, X., and Wang, W. (2012) Photoinduced fibrils formation of chicken egg white lysozyme under native conditions. *Proteins: Struct., Funct., Bioinformatics* **80**, 2501–2513.
- (57) Xie, J., Qin, M., Cao, Y., and Wang, W. (2011) Mechanistic insight of photo-induced aggregation of chicken egg white lysozyme: The interplay between hydrophobic interactions and formation of intermolecular disulfide bonds. *Proteins: Struct., Funct., Bioinformatics* **79**, 2505–2516.
- (58) Previero, A., Coletti Previero, M. A., and Jollès, P. (1966) Nonenzymatic cleavage of tryptophyl peptide bonds in peptides and proteins: I. Cleavage of C-tryptophyl peptide bonds in model peptides through their conversion into N'-formylkynurenine derivatives. *Biochem. Biophys. Res. Commun.* **22**, 17–21.
- (59) Sakiyama, F. (1977) Selective chemical cleavage of the peptide bond at N-formylkynurenine in ozone-oxidized hen egg-white lysozyme. *J. Biochem.* **82**, 365–375.
- (60) Flaugh, S. L., Kosinski-Collins, M. S., and King, J. (2005) Interdomain side-chain interactions in human γ D Crystallin influencing folding and stability. *Protein Sci.* **14**, 2030–2043.
- (61) Brubaker, W. D., Freites, J. A., Golchert, K. J., Shapiro, R. A., Morikis, V., Tobias, D. J., and Martin, R. W. (2011) Separating instability from aggregation propensity in γ S-Crystallin variants. *Biophys. J.* **100**, 498–506.
- (62) Srivastava, O. P., and Srivastava, K. (1996) Characterization of three isoforms of a 9 kDa γ D-crystallin fragment isolated from human lenses. *Exp. Eye Res.* **62**, 593–604.

- (63) Sandilands, A., Hutcheson, A. M., Long, H. A., Prescott, A. R., Vrensen, G., Löster, J., Klopp, N., Lutz, R. B., Graw, J., Masaki, S., Dobson, C. M., MacPhee, C. E., and Quinlan, R. A. (2002) Altered aggregation properties of mutant γ -crystallins cause inherited cataract. *EMBO J.* **21**, 6005–6014.
- (64) Santhiya, S. T., Shyam Manohar, M., Rawlley, D., Vijayalakshmi, P., Namperumalsamy, P., Gopinath, P. M., Löster, J., and Graw, J. (2002) Novel mutations in the γ -Crystallin genes cause autosomal dominant congenital cataracts. *J. Med. Genetics* **39**, 352–358.
- (65) Stephan, D. A., Gillanders, E., Vanderveen, D., Freas-Lutz, D., Wistow, G., Baxevanis, A. D., Robbins, C. M., Van Auken, A., Quesenberry, M. I., Bailey-Wilson, J., Juo, S.-H. H., Trent, J. M., Smith, L., and Brownstein, M. J. (1999) Progressive juvenile-onset punctate cataracts caused by mutation of the γ D-Crystallin gene. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 1008–1012.
- (66) Evans, P., Wyatt, K., Wistow, G. J., Bateman, O. A., Wallace, B. A., and Slingsby, C. (2004) The P23T cataract mutation causes loss of solubility of folded γ D-Crystallin. *J. Mol. Biol.* **343**, 435–444.
- (67) Marek, P., Mukherjee, S., Zanni, M. T., and Raleigh, D. P. (2010) Residue-specific, real-time characterization of lag-phase species and fibril growth during amyloid formation: a combined fluorescence and IR study of p-cyanophenylalanine analogs of islet amyloid polypeptide. *J. Mol. Biol.* **400**, 878–888.
- (68) Hamm, P., and Zanni, M. T. (2011) Concepts and Methods of 2D Infrared Spectroscopy, *Cambridge University Press*, New York.
- (69) Mills, I. A., Flaugh, S. L., Kosinski-Collins, M. S., and King, J. A. (2007) Folding and stability of the isolated Greek key domains of the long-lived human lens proteins γ D-Crystallin and γ SCrystallin. *Protein Sci.* **16**, 2427–2444.
- (70) Das, P., King, J. A., and Zhou, R. (2011) Aggregation of γ crystallins associated with human cataracts via domain swapping at the C-terminal β -strands. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 10514–10519.
- (71) Demirdöven, N., Cheatum, C. M., Chung, H. S., Khalil, M., Knoester, J., and Tokmakoff, A. (2004) Two-dimensional infrared spectroscopy of antiparallel β -sheet secondary structure. *J. Am. Chem. Soc.* **126**, 7981–7990.
- (72) Su, S., Liu, P., Zhang, H., Li, Z., Song, Z., Zhang, L., and Chen, S. (2011) Proteomic analysis of human age-related nuclear cataracts and normal lens nuclei. *Invest. Ophthalmol. Vis. Sci.* **52**, 4182–4191.
- (73) Meehan, S., Knowles, T. P. J., Baldwin, A. J., Smith, J. F., Squires, A. M., Clements, P., Treweek, T. M., Ecroyd, H., Tartaglia, G. G., Vendruscolo, M., MacPhee, C. E., Dobson, C. M., and Carver, J. A. (2007) Characterisation of amyloid fibril formation by small heat-shock chaperone proteins human α A-, α B- and R120G α B-Crystallins. *J. Mol. Biol.* **372**, 470–484.
- (74) Srivastava, O. P., McEntire, J. E., and Srivastava, K. (1992) Identification of a 9 kDa γ -Crystallin fragment in human lenses. *Exp. Eye Res.* **54**, 893–901.

- (75) Srivastava, O. P., Srivastava, K., and Silney, C. (1994) Covalent modification at the C-terminal end of a 9 kDa γ D-Crystallin fragment in human lenses. *Exp. Eye Res.* **58**, 595–604.
- (76) Santhoshkumar, P., Udupa, P., Murugesan, R., and Sharma, K. K. (2008) Significance of interactions of low molecular weight crystallin fragments in lens aging and cataract formation. *J. Biol. Chem.* **283**, 8477–8485.
- (77) Lee, S., Lin, S., Li, M., and Liang, R. (1997) Possible mechanism of exacerbating cataract formation in cataractous human lens capsules induced by systemic hypertension or glaucoma. *Ophthalmic Res.* **29**, 83–90.
- (78) Kosinski-Collins, M. S., and King, J. (2003) In vitro unfolding, refolding, and polymerization of human γ D Crystallin, a protein involved in cataract formation. *Protein Sci.* **12**, 480–490.
- (79) Rekas, A., Jankova, L., Thorn, D. C., Cappai, R., and Carver, J. A. (2007) Monitoring the prevention of amyloid fibril formation by α -Crystallin. *FEBS J.* **274**, 6290–6304.
- (80) Acosta-Sampson, L., and King, J. (2010) Partially folded aggregation intermediates of human γ D-, γ C-, and γ S-Crystallin are recognized and bound by human α B-Crystallin chaperone. *J. Mol. Biol.* **401**, 134–152.
- (81) Shamma, S. L., Waudby, C. A., Wang, S., Buell, A. K., Knowles, T. P. J., Ecroyd, H., Welland, M. E., Carver, J. A., Dobson, C. M., and Meehan, S. (2011) Binding of the molecular chaperone α B-Crystallin to A β amyloid fibrils inhibits fibril elongation. *Biophys. J.* **101**, 1681–1689.

4 Chapter 4

An alternative structural isoform in amyloid-like aggregates formed from thermally denatured human γ D-crystallin*

4.1 Abstract

The eye lens protein γ D-crystallin contributes to cataract formation in the lens. *In vitro* experiments show that γ D-crystallin has a high propensity to form amyloid fibers when denatured, and that denaturation by acid or UV-B photodamage results in its C-terminal domain forming the β -sheet core of amyloid fibers. Here, we show that thermal denaturation results in sheet-like aggregates that contain cross-linked oligomers of the protein, according to transmission electron microscopy and SDS-PAGE. We use two-dimensional infrared spectroscopy to show that these aggregates have an amyloid-like secondary structure with extended β -sheets, and use isotop dilution experiments to show that each protein contributes approximately one β -strand to each β -sheet in the aggregates. Using segmental ^{13}C labeling, we show that the organization of the protein's two domains in thermally induced aggregates results in a previously unobserved structure in which both the N-terminal and C-terminal domains contribute to β -sheets. We propose a model

* The material in this chapter was originally published as: 8. Sean D. Moran [†], Tianqi O. Zhang [†], Martin T. Zanni. "An Alternative Structural Isoform in Amyloid-like Aggregates Formed from Thermally Denatured Human γ D-crystallin." *Protein Science*. 2014, 23, pp321-331. DOI: 10.1002/pro.2422 ([†]equal contribution). I contributed to this work in the experimental design, sample preparation, data collection and analysis.

for the structural organization of the aggregates and attribute the recruitment of the N-terminal domain into the fiber structure to intermolecular cross linking.

4.2 Introduction

Cataracts are a common protein aggregation disease of the eye lens, resulting from the formation of insoluble deposits of lens crystallin proteins that blur vision.^{1,2} Aggregation may result from mutations³ as well as accumulated damage from sources such as UV radiation,⁴⁻⁸ oxidative stress,^{5,9} and a variety of post-translational modifications^{1,10} that induce precipitation or misfolding of lens crystallins. Although the causes of damage are well known, the resulting structures of aggregated lens crystallin proteins are poorly understood. *In vitro* denaturation studies have shown that crystallins can form a variety of different kinds of aggregates, but that they have a high propensity for the formation of amyloid fibers.^{7,11-15} In fact, acidic conditions,¹¹⁻¹³ chemical denaturants, UV radiation,⁷ and thermal denaturation¹⁶ have been shown to result in amyloid fiber formation.

Human γ D-crystallin is a structural eye lens protein containing two domains with similar folds¹⁷ but very different stabilities and behaviors during aggregation.^{7,8,11-13,18-21} In the native state, each of the domains of γ D-crystallin is made up of a β -sandwich composed of two Greek key subdomains.¹⁷ Equilibrium unfolding and fluorescence studies have shown that the protein's N-terminal domain is less stable than its C-terminal domain, and that a folding intermediate with a disordered N-terminal domain and native-like C-terminal domain exists.¹⁹ The hydrophobic interface between the domains contains specific residue contacts that stabilize the native fold, and the presence of this interface allows the C-terminal domain to act as a template for the folding of the N-terminal domain.²¹ The specific association between N- and C-terminal domains can also result in the formation of dimers or polymers through a domain-swapping mechanism that has

been hypothesized as a possible route to precipitation and cataract formation.^{22,23} In addition to domain swapping, amorphous aggregation and amyloid fiber formation have been observed.^{7,11-13,24} Amyloid fiber formation is the dominant *in vitro* mode of aggregation for γ D-crystallin and a number of other crystallins.^{7,11-16,25} Both the isolated N- and C-terminal domains of γ D-crystallin, as well as the full-length protein, can form amyloid fibers *in vitro* upon denaturation with acid.¹¹⁻¹³ Previously, we found that the acid-induced fibers formed by the full-length protein behave differently than might be expected from examination of the isolated domains.^{11,12} Full-length, acid-induced γ D-crystallin aggregates only contain core β -strands from the C-terminal domain, while the N-terminal domain becomes highly disordered.¹² In fact, nearly the entire C-terminal domain forms amyloid β -sheets.¹¹ We also observed similar domain behaviors in UV-B induced aggregates of γ D-crystallin, in which amyloid fiber formation by the C-terminal domain is accompanied by extensive side chain oxidation, peptide bond cleavage, and covalent protein cross-linking.⁷ The results of these studies suggest that the C-terminal domain has a higher intrinsic amyloid propensity than the N-terminal domain, independent of the mode of denaturation.

In this paper, we show that γ D-crystallin can form alternative aggregate structures upon thermal denaturation. Using transmission electron microscopy and SDS-PAGE, we show that the aggregates formed by thermal denaturation have sheet-like morphology and may contain covalently cross-linked proteins. To study the structure of these aggregates, we use a combination of two-dimensional infrared (2D IR) spectroscopy²⁶ and segmental ¹³C labeling, which we implement through expressed protein ligation.^{7,11,12} 2D IR spectroscopy is a useful technique for determining protein structures in complex systems such as aggregates or membranes because spectra can be collected straightforwardly in heterogeneous samples.^{7,11,12,27-36} Like FTIR spectroscopy, 2D IR spectroscopy is sensitive to the secondary structure, solvent exposure, and

dynamics of a protein based on IR frequencies and line shapes.²⁶ Further structural resolution can be obtained using isotope labeling of a single amino acid,^{28–35,37–40} labeling of an entire sequence segment,^{7,11,12,41} or through the incorporation of non-natural vibrational probes.^{27,42–47} 2D IR spectroscopy has a number of important advantages over FTIR spectroscopy. First, it allows the resolution of cross-peaks, which reveal coupling between different vibrational modes.^{26,31,36,48} Second, 2D IR signal strengths scale with the fourth power of the transition dipole moment, which results in increased resolution of the various peaks in a 2D IR spectrum.^{26,38} Third, the relative contributions of inhomogeneous and homogeneous broadening can be resolved from 2D line shapes, revealing important information about structural order and dynamics.^{12,29,32,45,49,50} Finally, all of these advantages can be obtained in kinetics studies that take advantage of rapid-scan technology based on mid-IR pulse shaping.^{38,51} Here, we use 2D IR spectroscopy, segmental ¹³C labeling, and isotope dilution to show that the aggregates formed from thermal denaturation of γ D-crystallin contain β -strands from both the N-terminal and C-terminal domains, a result we tentatively attribute to intermolecular cross-linking. These results are strikingly different from those obtained from acid^{11,12} and UVB⁷-induced denaturation, and demonstrate that different structural isoforms may be produced depending on the mode of denaturation that induces aggregation. With this information, and comparisons to acid-induced amyloid fibers of γ D-crystallin,¹² we propose a model for the thermally induced aggregates of the protein. Finally, we discuss the implications of this model for the possible role of amyloid-like crystallin aggregates in cataracts.

4.3 Material and Methods

All chemicals were purchased from Sigma-Aldrich and used as received, unless otherwise specified. Human γ D-crystallin, and its segmentally labeled variants, were expressed in *E. coli* and purified as described previously.¹²

Samples of human γ D-crystallin were dissolved to $\sim 20 \mu\text{M}$ in 20 mM sodium phosphate buffer (pH 7.0) with 100 mM NaCl. For samples used in 2D IR experiments, buffers were prepared with D_2O instead of H_2O . All samples were thermally denatured by placing them in microcentrifuge tubes on a heating block set to 80 °C, and incubated for 3 hours. The resulting precipitates were gently resuspended by pipetting. Transmission electron micrographs of methylamine tungstate stained samples were collected at the University of Wisconsin Medical School Electron Microscope Facility, using a Philips CM 120 transmission electron microscope. SDS-PAGE samples were prepared by immediately dissolving the aggregates in Laemmli sample buffer (containing 20 mM 2-mercaptoethanol), and were heated for 10 minutes at 95 °C to disperse the aggregates. Band intensities on the SDS gel were integrated using ImageJ software, which is available free of charge from <http://rsbweb.nih.gov/ij/>.

Samples of γ D-crystallin were cooled to room temperature and placed between CaF_2 windows separated by a 56 μM Teflon spacer. 2D IR spectra were collected and analyzed as described previously.^{12,38} In addition to uniformly labeled and segmentally labeled samples, samples were prepared with mixtures of ^{12}C and ^{13}C proteins for an isotope dilution study. Proteins were mixed prior to denaturation, with compositions ranging from 10–90% ^{13}C labeled γ D-crystallin. Frequencies of the ^{13}C labeled amyloid β -sheet peaks were obtained from the spectra directly or through the subtraction of the unlabeled (^{12}C) aggregate spectrum from dilute (10-20%) ^{13}C aggregates.

4.4 Results and Discussion

In this section, we present the results associated with the aggregation of human γ D-crystallin (S84C) as induced by thermal denaturation. We also present a comparison of the thermally induced aggregates to acid-induced amyloid fibers of γ D-crystallin, for which we have developed a structural model in previous publications using 2D IR spectroscopy,¹² and which is consistent with mass spectrometry.¹¹ Thermal denaturation of human γ D-crystallin (S84C) resulted in visible turbidity of the samples within 30 min, followed by continued formation of visible precipitates over the course of 3 h of incubation at 80°C. Transmission electron micrographs showed aggregates as either twisted (Figure 4.1A) or flat sheets (Figure 4.1B). Analysis of the samples by SDS-PAGE (Figure 4.1D) reveals that the aggregates are composed not only of monomers (~20 kDa), but also dimers and a small amount of trimers and higher-order oligomers. Based on integration of the gel band intensities and the presence of some high molecular weight species that did not enter the gel, we estimate that a minimum of 25% of the total protein is cross-linked. Because SDS-PAGE samples were prepared by incubating the aggregates at 95°C in the presence of detergent and a thiol reductant, non-covalent and disulfide-based oligomers are unlikely to be present. Aggregates prepared in the presence of 2-mercaptoethanol (Figure 4.1D, Lane 4) also showed the formation of oligomeric or crosslinked species. Thus, we tentatively conclude that the oligomers are produced by covalent, nondisulfide cross-links between protein molecules. However, exceptionally stable disulfide-based oligomers⁵² and stable non-covalent assemblies⁵³ cannot be ruled out. Wild type protein was analyzed in the same manner and resulted again in identical band patterns, showing that the S84C mutation, which is required for expressed protein ligation in segmentally labeled proteins, does not influence the product distribution. For the purpose of comparison, we present a typical TEM image of acid-induced aggregates (Figure

4.1C), which form uniform suspensions of fibers with diameters of ~5–10 nm.¹¹ UV-B induced aggregates also form fibers with similar diameters.⁷ The amyloid fibers formed by acid and UV-B-induced denaturation are much narrower than the sheets formed from thermal denaturation, which have irregular dimensions that range from ~50 to 1,000 nm across. Previous SDS-PAGE analysis of acid-induced aggregates showed neither crosslinking nor polypeptide cleavage,¹¹ while UV-B induced aggregates showed both kinds of protein modifications.⁷

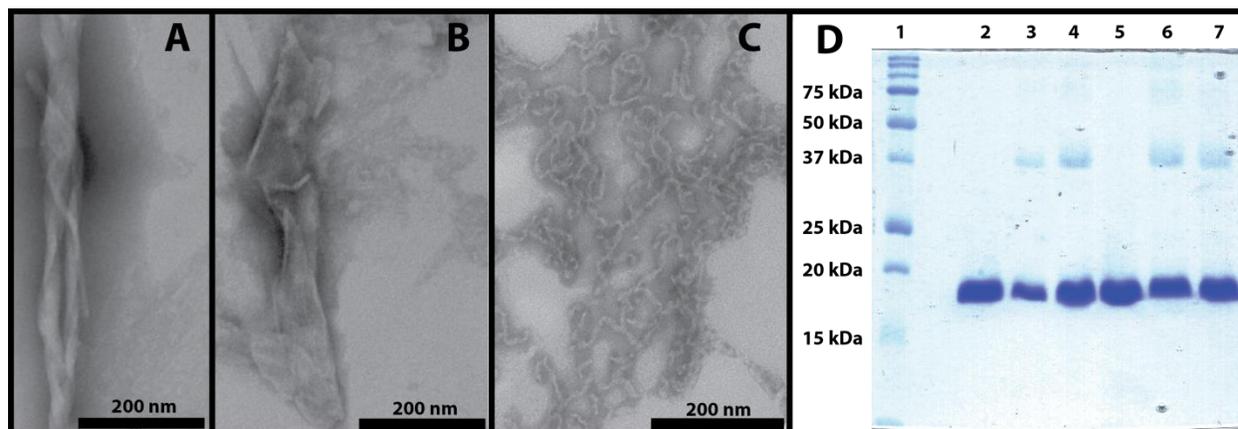


Figure 4.1 Characterization of thermally induced γ D-crystallin aggregates.

(A) TEM image of thermally induced aggregates with twisted sheet morphology. (B) TEM image of thermally induced aggregates with flat sheet morphology. (C) Acid-induced amyloid fibers. (D) SDS-PAGE analysis of thermally induced γ D-crystallin aggregates. Lanes from left: (1) Molecular weight marker; (2) native γ D-crystallin (S84C); (3) thermally induced aggregates of γ D-crystallin (S84C); (4) thermally induced aggregates of γ D-crystallin (S84C) with 2-mercaptoethanol; (5) native wild type γ D-crystallin; (6) thermally induced aggregates of wild type γ D-crystallin; (7) thermally induced aggregates of wild type γ D-crystallin with 2-mercaptoethanol. Image (C) is reproduced from data described in Ref. 11.

To characterize the secondary structure of the thermally induced aggregates, we turn to 2D IR spectroscopy. The secondary structure sensitivity of infrared spectroscopy is largely a result of the differences in vibrational coupling between peptide bonds in α -helices, β -sheets, and random coil structures.^{26,54,55} The relative orientations of the vibrational transition dipole moments and their respective coupling constants associated with each secondary structure cause delocalization

of vibrational motions that results in frequency shifts and intensity changes.²⁶ The 2D IR spectrum of unlabeled, thermally induced human γ D-crystallin (Figure 4.2A) shows features that are typical of amyloid fibers, including a narrow, low-frequency peak pair ($\omega_{\text{pump}} = 1614 \text{ cm}^{-1}$) assigned to β -sheets and a broader, high frequency peak pair ($\omega_{\text{pump}} = 1625\text{--}1700 \text{ cm}^{-1}$) assigned to disordered structures.¹² In 2D IR spectra, peaks come in pairs because the pulse sequence probes both the (0–1) and (1–2) vibrational transitions.²⁶ The characteristic amyloid β -sheet peak at $1614\text{--}1620 \text{ cm}^{-1}$ appears at low frequency because coupling constants between adjacent residues in β -sheets are strongly negative,^{39,54,55} resulting in a red shift of the amide band that scales with the size of the β -sheet.^{54,55} The β -sheet amide I band is also very narrow compared to that of the native protein and the disordered signal in the aggregates, reflecting a relatively small inhomogeneous contribution to the line width that reveals a highly uniform environment for these β -sheet modes.^{12,26} The ^{13}C -labeled aggregates (Figure 4.2B) contain similar sets of peak pairs at 1570 cm^{-1} and $1580\text{--}1660 \text{ cm}^{-1}$, respectively, consistent with the $\sim 40 \text{ cm}^{-1}$ shift in frequency that occurs upon replacement of ^{12}C with ^{13}C .¹²

For comparison, the 2D IR spectra of unlabeled (^{12}C) and labeled (^{13}C) γ D-crystallin fibers prepared by acid denaturation are reproduced in Figure 4.2 (C,D) from data reported previously.¹² Slices through the diagonals of the acid and thermally denatured 2D IR spectra (Figure 4.2E and F), which reflect the same modes that would be observed in FTIR spectra, aid in their comparison. The spectra are similar but the frequency of the β -sheet feature near 1620 cm^{-1} in the thermally induced aggregates is lower than that in the acid-induced fibers. This frequency difference, though small, reflects structural differences between the thermally induced and acid induced fibers. The frequency of this mode depends on the size and structural order of the β -sheets, which indicates

that the thermally induced fibers are larger and more ordered than those formed from acid denaturation.

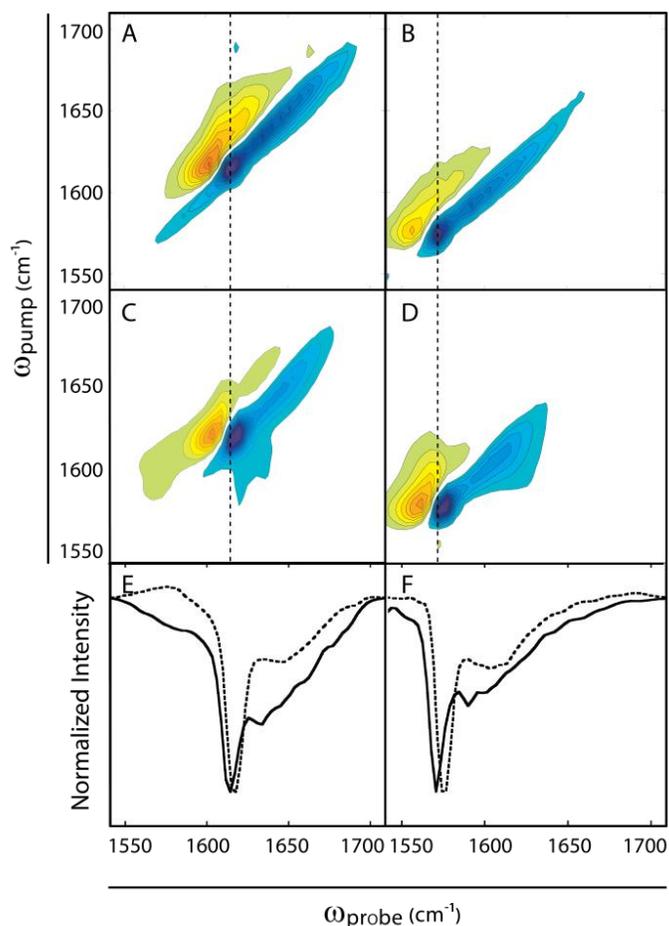


Figure 4.2 2D IR spectra of γ D-crystallin amyloid aggregates.

(A) Thermally induced ^{12}C γ D-crystallin aggregates. (B) Thermally induced ^{13}C γ D-crystallin aggregates. (C) Acid-induced ^{12}C γ D-crystallin amyloid fibers. (D) Acid-induced ^{13}C γ D-crystallin amyloid fibers. (E) Diagonal slices of spectra in (A) (solid lines) and (C) (dashed lines). (F) Diagonal slices of spectra in (B) (solid lines) and (D) (dashed lines). Dotted lines though the maxima of the β -sheet peaks reveal frequency differences between thermally induced and acid-induced β -sheets. Spectra of the acid-induced aggregates are reproduced from data described in Ref. 12.

Another difference between the thermally induced and acid-induced aggregates is apparent in the broad peak pairs that are observed between $\omega_{\text{pump}} = 1625\text{-}1700\text{ cm}^{-1}$ in unlabeled aggregates and $\omega_{\text{pump}} = 1585\text{-}1660\text{ cm}^{-1}$ in the labeled aggregates. These signals are due to disordered regions

of the proteins in the aggregates.^{11,12} It is clear from the diagonal slices in Figure 4.2(E,F) that the disordered structures in the thermally induced aggregates have greater relative intensities than their acid-induced counterparts, and also appear to contain multiple discrete peaks. Because the samples were cooled to room temperature for analysis, some refolding of denatured structures is possible, and may explain the presence of multiple features in the disordered region of the spectrum. Thus, the unlabeled spectra in Figure 4.2 provide an overview of the secondary structure of the proteins, but do not reveal which residues contribute to the β -sheets or the disordered regions. To locate these regions within the proteins, we turn to isotope labeling.

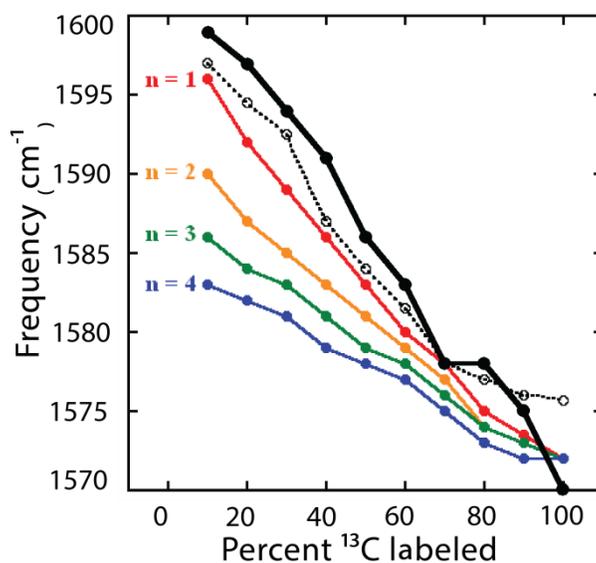


Figure 4.3 Isotope dilution of γ D-crystallin aggregates.

Isotope dilution study of thermally induced γ D-crystallin aggregates (solid black line, closed circles) compared to acid-induced amyloid fibers (dashed black line, open circles) and transition dipole coupling simulations (colored lines, closed circles) for aggregate models with $n = 1-4$ strands per β -sheet per protein. Acid-induced fiber data and simulation results are reproduced from Ref. 12.

To obtain additional information about the architecture of the aggregates, we use isotope dilution. The frequency of the amide I β -sheet signal is related to the extent of vibrational

delocalization in the aggregates.^{12,55} The vibrational coupling in amyloid β -sheets is very strong and so if all strands are composed of the same isotopes, the vibrational modes will extend over many β -strands in the β -sheets. The larger the delocalization of the vibrational mode, the lower the β -sheet frequency.^{26,54,55} If the strands are not all composed of the same isotopes, then the couplings do not delocalize vibrational modes as effectively, giving rise to a higher frequency.^{12,28,55} We can use this fact to provide an estimate of the number of β -strands each protein molecule contributes to each β -sheet.¹² By mixing ^{12}C - and ^{13}C -labeled proteins in various proportions prior to aggregation, we observe a ^{13}C β -sheet frequency trend in the aggregates that reflects the coupling between ^{13}C -labeled β -strands (Figure 4.3). Thus, the frequency of the ^{13}C amide I band provides a measure of the number of β -strands each protein contributes to the β -sheets. We quantify the frequency as a function of the number of contributed β -strands using transition dipole coupling simulations,⁵⁵ described in detail in a previous publication.¹² Shown in Figure 4.3 are the frequency trends for ^{13}C amide I maxima for amyloid β -sheets in which each individual protein contributes between one and four β -strands to each β -sheet. We find that the frequencies are consistent with approximately one β -strand per β -sheet for each protein molecule in the aggregates. We previously found a similar result for the acid-induced fibers.¹²

We have also learned that each of the two domains contribute β -strands to the amyloid β -sheets of the thermally induced aggregates. Using expressed protein ligation, we generated a variant of γD -crystallin (S84C) in which the N-terminal domain is ^{13}C -labeled and the C-terminal domain is unlabeled (^{12}C).^{7,11,12} The spectrum of segmentally labeled, thermally induced aggregates of γD -crystallin is shown in Figure 4.4A. First, we consider the signal from the unlabeled C-terminal domain. In the thermally induced aggregates, this signal appears at $\omega_{\text{pump}} = 1624 \text{ cm}^{-1}$, a much lower frequency than the native ^{12}C C-terminal domain (which appears at ω_{pump}

= 1640 cm^{-1} , Figure 4.4B) and within the typical frequency range for amyloid fibers.⁷ Thus, we conclude that the C-terminal domain still forms amyloid β -sheets. Turning to the ^{13}C -labeled N-terminal domain, we see a peak with a maximum at $\omega_{\text{pump}} = 1600 \text{ cm}^{-1}$, similar to the native frequency.¹² However, we also see a low frequency shoulder at $\omega_{\text{pump}} = 1591 \text{ cm}^{-1}$. To better visualize the appearance of this shoulder, we subtracted the native protein spectrum (Figure 4.4B) from the spectrum of the thermally induced aggregates (Figure 4.4A), each of which was normalized to the most intense diagonal amide I signal. This difference spectrum (Figure 4.4C), which is useful as a qualitative indicator of frequency shifts, contains three difference peak pairs along the diagonal. The lowest frequency pair, centered at $\omega_{\text{pump}} = 1591 \text{ cm}^{-1}$, results from the structural transition of the N-terminal domain. The two other peak pairs, at $\omega_{\text{pump}} = 1625 \text{ cm}^{-1}$ and with opposite sign at $\omega_{\text{pump}} = 1638 \text{ cm}^{-1}$, are the signal from the amyloid-like C-terminal β -sheets and the loss of the native β -sandwich fold of the C-terminal domain, respectively. Thus, it is clear that both the N- and C-terminal domains form β -sheets upon thermal denaturation, and that both domains undergo conformational changes from the native state to the aggregated state.

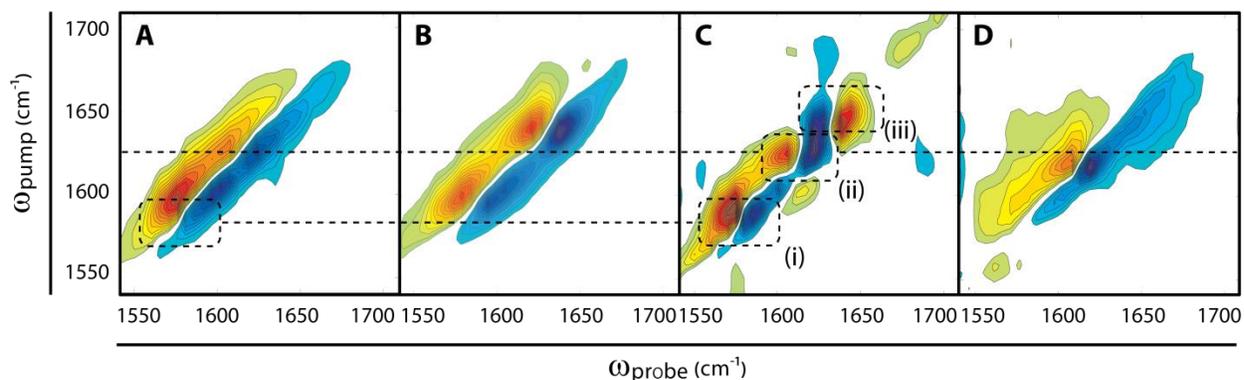


Figure 4.4 2D IR spectra of segmentally labeled γ D-crystallin aggregates.

(A) Thermally induced amyloid aggregates of N-terminally ^{13}C -labeled γ D-crystallin. (B) Native N-terminally ^{13}C -labeled γ D-crystallin. (C) Difference spectrum of thermally induced aggregates (A) minus native (B) N-terminally ^{13}C -labeled γ D-crystallin. Features in the difference spectrum clarify the amide I peak shifts associated with thermally induced aggregation: (i) additional low frequency intensity in the ^{13}C -labeled (N-terminal domain) region; (ii) additional low-frequency intensity in the unlabeled (C-terminal domain) region; (iii) loss of intensity in the region of the native, unlabeled C-terminal domain indicated by a difference peak with opposite sign. (D) Acid-induced amyloid fibers of ^{13}C N-terminally labeled γ D-crystallin. Spectra (B) and (D) are reproduced from data described in Ref. 12.

Also shown in Figure 4.4 is the spectrum of acid induced amyloid fibers with the ^{13}C N-terminally labeled (Figure 4.4D).¹² The acid-induced spectrum indicates that only the C-terminal domain is involved in amyloid formation, since sharp peaks only appear at $\omega_{\text{pump}} = 1617 \text{ cm}^{-1}$ and are identical to that in the unlabeled acid-induced amyloid fibers (Figure 4.2C), as previously reported and confirmed with mass spectrometry.^{11,12} Thus, it is immediately apparent that both domains contribute to the amyloid β -sheets of the thermally induced aggregates whereas in acid denaturation only the C-terminal domain forms amyloid β -sheets. Moreover, in the thermally induced aggregates, there is a $\sim 10 \text{ cm}^{-1}$ difference between the unlabeled amyloid β -sheet peak in Figure 4.2A and the amyloid signal from the unlabeled C-terminal domain in Figure 4.4A. Because the frequencies do not match but the protein structures are identical, we must conclude that it is the ^{13}C labeling of the N-terminal domain that causes the frequency shift in the ^{12}C C-terminal

domain. In other words, the ^{13}C labels of the N-terminal domain disrupt the vibrational coupling between the ^{12}C residues of the C-terminal domain. In order for one domain to alter the frequency of the other domain, both domains must contribute to the same amyloid β -sheets, as we describe below. Three hypothetical structural models for thermally-induced aggregates of full-length γD -crystallin (and its cross-linked products) are postulated in Figure 4.5 assuming parallel β -sheets, which are common in amyloid aggregates.⁵⁶ Corresponding models with antiparallel strands could also be constructed, but because β -sheets containing parallel and antiparallel β -strands have similar amide I frequencies, our conclusions (described below) would be the same. We consider simple models in which the N-terminal domain and C-terminal domain form interacting amyloid β -sheets, consistent with our data above. These models, shown in Figure 4.5, are: (A) amyloid aggregates with the N-terminal and C-terminal domains forming non-interspersed β -sheets; (B) amyloid aggregates in which the N-terminal and C-terminal domains of the same protein are hydrogen-bonded to one another in the same β -sheet; and (C) amyloid aggregates in which the N-terminal and C-terminal domains of different proteins are hydrogen-bonded to one another in the same β -sheet. Models like these have been proposed for other amyloid-forming proteins, and so might be thought of as representative of the canonical set of plausible amyloid aggregate structures.^{52,56–58} Based on the frequencies of both the N- and C-terminal domains in Figure 4.4A, we exclude model (A), because although β -strands may interact at the boundary of labeled and unlabeled regions the intra-strand coupling constants are too weak to cause a significant shift in β -sheet frequency.^{39,54,55} Additionally, the frequency of the shoulder on the labeled N-terminal domain signal (Figure 4.4A and C) is too high for extensively coupled N-terminal domains. We can also exclude model (B) because the isotope dilution data obtained from uniformly labeled proteins shows a single strand per β -sheet per protein, whereas a lower frequency would be expected in dilute ^{13}C aggregates

with more than one consecutive β -strand from each protein.¹² Based on the exclusion of models (A) and (B), we are thus left with model (C) as the most likely explanation for our data. Model (C) is the only model that is consistent with the low frequency β -sheet peaks in the 2D IR spectra of uniformly labeled samples, the β -sheet frequency trend observed upon isotope dilution of uniformly labeled aggregates, and the frequency shifts of β -sheet peaks that occur when the N-terminal domain of the aggregates protein is ^{13}C labeled.

It is known that both the N-terminal domain and C-terminal domain can individually form amyloid structures after denaturation with acid,¹³ which may not be surprising given the sequence and native structure similarity of the N- and C-terminal domains.¹⁷ However, in our previous studies of full-length γD -crystallin, only the C-terminal domain formed amyloid β -sheets.^{11,12} In fact, even when the polypeptide backbone is cleaved by UV-B radiation, thereby separating N- and C-terminal sequences, no amyloid structures containing N-terminal domain residues were observed.⁷ Moreover, it is clear from our SDS-PAGE results that polypeptide cleavage does not occur upon heating, so domain separation cannot explain the inclusion of both domains in the aggregate β -sheets. Additionally, because the N-terminal domain is less stable than the C-terminal domain under our experimental conditions,¹⁹ the ratio of N-terminal and C-terminal strands cannot be explained by the relative proportions of unfolded domains at 80°C. Thus, we hypothesize that the observed cross-linking drives the formation of mixed-domain amyloid structures. Similar recruitment of peripheral β -strands into the amyloid fiber core, driven by cross-linking, has been observed before.⁵² Also, amyloid cross-seeding and copolymerization experiments have shown that amyloid proteins with different sequences or modifications can co-localize in the same fibers.^{59–62} Thus, it is possible that the N- and C-terminal domains can both contribute to the same β -sheets, even though they have different amino acid sequences. Our data supports that conclusion.

Based on the relative shifts of the signals from the N-terminal and C-terminal domains, we conclude that the amide I vibrations of the unlabeled C-terminal domain are more delocalized in the aggregates (14 cm^{-1} red shift vs. native Greek key) than those of the N-terminal domain (7 cm^{-1} red shift vs. native Greek key). This observation indicates that the proportion of C-terminal domains in the aggregate β -sheets is larger than the proportion of N-terminal domains (Figure 4.3). Because vibrational coupling constants are invariant with ^{13}C labeling,^{12,63} it is reasonable to assume that the ^{13}C β -sheet frequency trend observed upon isotope dilution is identical to that of ^{12}C β -sheets. Referring to this trend in Figure 4.3, we estimate that the β -sheets contain approximately 30–40% N-terminal strands and 60–70% C-terminal strands. This result is similar to the proportion of putatively cross-linked products observed in our SDS-PAGE results (Figure 4.1D), and thus corroborates our cross-linking hypothesis.

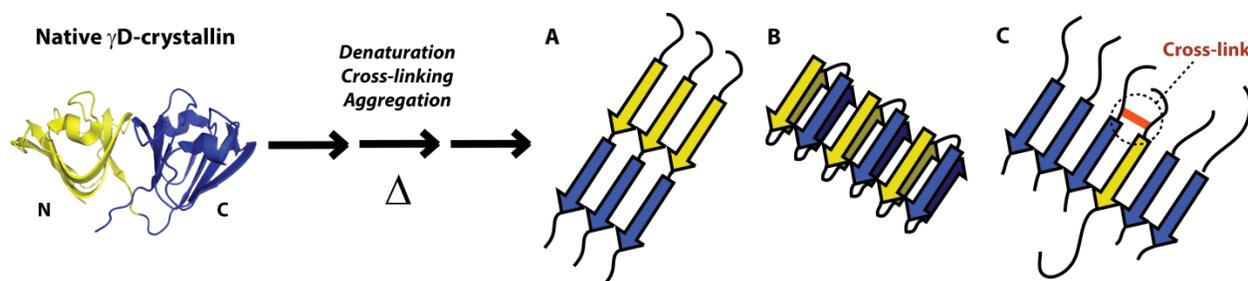


Figure 4.5 Structural models for the thermally induced aggregation of γ D-crystallin.

Native γ D-crystallin (left) is denatured by heat, and undergoes both cross-linking and aggregation into amyloid-like β -sheets. Three general aggregate structures are possible. (A) N-terminal (yellow) and C-terminal (blue) domains form consecutive, non-hydrogen-bonded β -strands. (B) Both the N-terminal and C-terminal domains of a single protein form β -strands in the same β -sheet. (C) N-terminal domains of a minority of proteins are recruited into a C-terminal domain β -sheet, possibly as a result of intermolecular cross-linking (red).

Although this difference in domain content may seem small, we note that the frequency shift and intensity increase due to vibrational coupling is dependent on the proximity of multiple β -strands of the same carbon isotopes. This may lead to large differences in the 2D IR spectra

based on small differences in the number of labeled N-terminal and unlabeled C-terminal strands. For example, in a randomly formed aggregate containing 65% unlabeled (^{12}C) β -strands, the probability of producing three consecutive ^{12}C strands is $(0.65)^3 = 0.275$ while the probability of producing three consecutive ^{13}C strands is $(0.35)^3 = 0.043$. Of course, cross-linked samples are not random; in our model (Figure 4.5C) labeled and unlabeled strands come in pairs. However, because we observe a greater shift in unlabeled C-terminal amide I frequency, we know that the aggregate β -sheets contain more C-terminal strands than N-terminal strands. This fact may be explained by the exclusion of some cross-linked N-terminal β -strands from the aggregate β -sheets, or from the incorporation of C-terminal strands from the non-cross linked portion of the sample. Both of these possibilities can be explained by the greater intrinsic amyloid propensity of the C-terminal domain, which we have observed in our previous studies.^{7,11,12} Thus, in model (C) of Figure 4.5, we draw a higher proportion of C-terminal domain than N-terminal domain in the β -sheets.

4.5 Conclusions

In this paper, we demonstrate the formation of sheet-like aggregates from thermally denatured human γD -crystallin with extended, ordered β -sheets that are similar to those that occur in amyloid fibers.⁵⁶ Although the morphologies of these aggregates are very different from amyloid fibers formed by acid^{11,13} and UV-B⁷-induced denaturation of the protein, one might conclude that the aggregate structures are the same based on the similarity of their FTIR (not shown) or 2D IR spectra. Aside from small differences in amide I frequency and lineshape, 2D IR spectra of uniformly labeled aggregates produced by thermal and acid denaturation are indistinguishable. However, that is far from correct. Segmental isotope labeling of the N-terminal domain shows that that the amyloid-like β -sheets of the thermally induced aggregates are actually

very different from those in the acid-induced fibers. In the thermally induced aggregates, both the N- and C-terminal domains contribute to the same β -sheets.

The C-terminal domain constitutes the majority of the β -strands in these aggregates, consistent with it being the more amyloidogenic of the two domains.^{7,11,12} The identification of this alternative aggregate structure highlights the importance of isotope labeling in structural studies of protein aggregates; without the resolution of domains provided by segmental ¹³C labeling, the participation of the N-terminal domain in the aggregate β -sheets could not have been observed. Our results also prove that ordered amyloid fibers can accommodate β -strands with sequences that are not identical, which is an important observation because sequence mismatches between β -strands are relevant to amyloid polymorphism,⁶⁴ amyloid cross-seeding,⁶⁵ co-aggregation⁶² of amyloid proteins, and some proposed amyloid structures of proteins with repeat sequences.⁶⁶ The 2D IR and isotope labeling methods presented here provide a way to understand β -strand organization in these and other complex amyloid aggregates.

These results indicate that amyloid-like aggregates of γ D-crystallin can occur in a variety of different isoforms depending on the mode of denaturation. In previous studies, we showed that only the C-terminal domain forms amyloid β -sheets after acid or UV-B-induced denaturation. Here, we attribute the recruitment of N-terminal domain sequences into the β -sheets to the formation of highly stable associations between protein molecules we observed in the SDS-PAGE analysis of the thermally induced aggregates, a phenomenon that is not observed in acid-induced aggregates¹² but has been identified in other systems.⁵² We tentatively attribute these associations to covalent cross-linking, which can result from the incubation of proteins at elevated temperatures.⁶⁷ However, some amyloid aggregates, such as those formed by A β , also form non-covalent SDS-stable dimers upon disruption,⁵³ suggesting that strong non-covalent interactions

between γ D-crystallin molecules may account for the observed dimers and oligomers. Although some cross-linking was also observed in UV-B-induced fibers, the presence of other modifications such as side-chain oxidation and peptide bond cleavage may prevent the recruitment of the N-terminal domain into the UV-B-induced fiber β -sheets.

The differences between amyloid fibers formed from thermal and acid denaturation highlight the complexity of amyloid aggregation processes. The final structures of amyloid aggregates likely depend on the details of the misfolding pathway, including the structural characteristics of the initial denatured state, barrier heights between folding intermediates, and the stability of the final aggregates. In γ D-crystallin, the N-terminal domain is less stable to chemical and thermal denaturation than the C-terminal domain,^{68,69} but it is the C-terminal domain that forms the majority of the amyloid β -sheet core in aggregates prepared by a variety of denaturation methods.^{11,70,71} Additionally, Brubaker et al.⁷² showed that the thermal stabilities of mutants of the closely related protein γ S-crystallin do not predict their aggregation propensities. Because the final γ D-crystallin amyloid aggregates prepared by thermal and acid denaturation are different at both the morphological and molecular structure level, it is clear that a detailed analysis of the entire aggregation pathway is required to understand how the different structural isoforms are favored under different conditions. Rapid-scan 2D IR spectroscopy³⁸ provides a possible means of deciphering the structural evolution of lens crystallin proteins during amyloid aggregate formation.

The results presented here also have important implications for the study of γ -crystallin aggregation in cataracts. To date, amyloid fibers have not been observed in protein aggregates taken from cataractous lenses, despite the apparent high propensity for crystallins to form amyloid fibers *in vitro*.^{1,7,11-16,20,25,73} However, in a few cases where the infrared spectra of such aggregates have been measured, low-frequency absorption near 1620 cm^{-1} , which is characteristic of extended

amyloid β -sheets, have been reported.⁷⁴ The fact that such spectral features are observed for aggregates that form sheets instead of fibers indicates that amyloid-like aggregation may occur even where fibers are not observed. In fact, amyloid-like infrared spectra of γ -crystallins, without the presence of fiber morphology in TEM images, has been observed *in vitro*.⁶ In addition, the presence of intermolecular cross-links in thermally induced γ D-crystallin aggregates mirrors the crosslinking observed in heavily damaged, aged lenses.^{2,9,75,76} Although the nature of our *in vitro* cross-links remains to be determined, our results suggest that the presence of intermolecular crosslinks may influence morphologies and molecular structures of lens crystallin aggregates.

4.6 Acknowledgements

The authors would like to thank Randall Massey of the University of Wisconsin Medical Sciences Electron Microscopy Facility for his help in the collection of TEM images, and would also like to thank Sean M. Decatur for useful discussion. This work was supported by the NIH NIDDK DK79895.

4.7 References

- (1) Wang Y, King JA (2010) Cataract as a protein aggregation disease. In: *Protein misfolding diseases*. Marina Ramirez-Alvarado, Jeffery W. Kelly, and Christopher M. Dobson, editor. Hoboken: John Wiley & Sons, pp 487–515.
- (2) Bloemendal H, de Jong W, Jaenicke R, Lubsen NH, Slingsby C, Tardieu A (2004) Ageing and vision: structure, stability and function of lens crystallins. *Prog Biophys Mol Biol* **86**:407–485.
- (3) Graw J (2004) Congenital hereditary cataracts. *Int J Dev Biol* **48**:1031–1044.
- (4) Taylor HR (1989) Ultraviolet radiation and the eye: an epidemiologic study. *Trans Am Ophthalmol Soc* **87**:802–853.
- (5) Davies MJ, Truscott RJW (2001) Photo-oxidation of proteins and its role in cataractogenesis. *J Photochem Photobiol B* **63**:114–125.
- (6) Fatima U, Sharma S, Guptasarma P (2010) Structures of differently aggregated and precipitated forms of γ B crystallin: an FTIR spectroscopic and EM study. *Protein ept Lett* **17**:1155–1162.
- (7) Moran SD, Zhang TO, Decatur SM, Zanni MT (2013) Amyloid fiber formation in human γ D-crystallin induced by UV-B photodamage. *Biochemistry*. doi: 10.1021/bi4008353.
- (8) Schafheimer N, King J (2013) Tryptophan cluster protects human γ D-crystallin from ultraviolet radiation-induced photoaggregation in vitro. *Photochem Photobiol* **89**:1106–1115.
- (9) Truscott RJW (2005) Age-related nuclear cataract—oxidation is the key. *Exp Eye Res* **80**:709–725.
- (10) Hains PG, Truscott RJ (2007) Post-translational modifications in the nuclear region of young, aged, and cataract human lenses. *J Proteome Res* **6**:3935–3943.
- (11) Moran SD, Decatur SM, Zanni MT (2012) Structural and sequence analysis of the human γ D-crystallin amyloid fibril core using 2D IR spectroscopy, segmental ^{13}C labeling, and mass spectrometry. *J Am Chem Soc* **134**:18410–18416.
- (12) Moran SD, Woys AM, Buchanan LE, Bixby E, Decatur SM, Zanni MT (2012) Two-dimensional IR spectroscopy and segmental ^{13}C labeling reveals the domain structure of human γ D-crystallin amyloid fibrils. *Proc Natl Acad Sci USA* **109**:3329–3334.
- (13) Papanikolopoulou K, Mills-Henry I, Tho SL, Wang YT, Gross AAR, Kirschner DA, Decatur SM, King J (2008) Formation of amyloid fibrils in vitro by human γ D-crystallin and its isolated domains. *Mol Vis* **14**:81–89.
- (14) Ecroyd H, Carver JA (2009) Crystallin proteins and amyloid fibrils. *Cell Mol Life Sci* **66**:62–81.
- (15) Meehan S, Knowles TPJ, Baldwin AJ, Smith JF, Squires AM, Clements P, Treweek TM, Ecroyd H, Tartaglia GG, Vendruscolo M, MacPhee CE, Dobson CM, Carver JA (2007) Characterisation of amyloid fibril formation by small heat-shock chaperone proteins human α A-, α B- and R120G α B-crystallins. *J Mol Biol* **372**:470–484.

- (16) Meehan S, Berry Y, Luisi B, Dobson CM, Carver JA, MacPhee CE (2004) Amyloid fibril formation by lens crystallin proteins and its implications for cataract formation. *J Biol Chem* **279**:3413–3419.
- (17) Basak A, Bateman O, Slingsby C, Pande A, Asherie N, Ogun O, Benedek GB, Pande J (2003) High-resolution X-ray crystal structures of human γ D crystallin (1.25 Å) and the R58H mutant (1.15 Å) associated with aculeiform cataract. *J Mol Biol* **328**:1137–1147.
- (18) Mills IA, Flaugh SL, Kosinski-Collins MS, King JA (2007) Folding and stability of the isolated Greek key domains of the long-lived human lens proteins γ D-crystallin and gS-crystallin. *Protein Sci* **16**:2427–2444.
- (19) Kosinski-Collins MS, Flaugh SL, King J (2004) Probing folding and fluorescence quenching in human γ D-crystallin Greek key domains using triple tryptophan mutant proteins. *Protein Sci* **13**:2223–2235.
- (20) Kosinski-Collins MS, King J (2003) In vitro unfolding, refolding, and polymerization of human γ D crystallin, a protein involved in cataract formation. *Protein Sci* **12**:480–490.
- (21) Flaugh SL, Kosinski-Collins MS, King J (2005) Interdomain side-chain interactions in human γ D crystallin influencing folding and stability. *Protein Sci* **14**:2030–2043.
- (22) Das P, King JA, Zhou R (2010) β -strand interactions at the domain interface critical for the stability of human lens γ D-crystallin. *Protein Sci* **19**:131–140.
- (23) Das P, King JA, Zhou R (2011) Aggregation of gamma-crystallins associated with human cataracts via domain swapping at the C-terminal β -strands. *Proc Natl Acad Sci USA* **108**:10514–10519.
- (24) Moreau KL, King JA (2012) Cataract-causing defect of a mutant g-crystallin proceeds through an aggregation pathway which bypasses recognition by the α -crystallin chaperone. *PLoS One* **7**:e37256.
- (25) Wang Y, Petty S, Trojanowski A, Knee K, Goulet D, Mukerji I, King J (2010) Formation of amyloid fibrils in vitro from partially unfolded intermediates of human gC-crystallin. *Invest Ophthalmol Vis Sci* **51**:672–678.
- (26) Hamm P, Zanni MT (2011) *Concepts and methods of 2D infrared spectroscopy*. New York: Cambridge University Press.
- (27) Woys AM, Mukherjee SS, Skoff DR, Moran SD, Zanni MT (2013) A strongly absorbing class of non-natural labels for probing protein electrostatics and solvation with FTIR and 2D IR spectroscopies. *J Phys Chem B* **117**:5009–5018.
- (28) Kim YS, Liu L, Axelsen PH, Hochstrasser RM (2008) Two-dimensional infrared spectra of isotopically diluted amyloid fibrils from Ab40. *Proc Natl Acad Sci USA* **105**:7720–7725.
- (29) Mukherjee P, Kass I, Arkin IT, Zanni MT (2006) Picosecond dynamics of a membrane protein revealed by 2D IR. *Proc Natl Acad Sci USA* **103**:3528–3533.

- (30) Middleton CT, Marek P, Cao P, Chiu CC, Singh S, Woys AM, de Pablo JJ, Raleigh DP, Zanni MT (2012) Two-dimensional infrared spectroscopy reveals the complex behaviour of an amyloid fibril inhibitor. *Nat Chem* **4**:355–360.
- (31) Dunkelberger EB, Woys AM, Zanni MT (2013) 2D IR cross peaks reveal hydrogen–deuterium exchange with single residue specificity. *J Phys Chem B*. doi: 10.1021/jp402942s.
- (32) Woys AM, Lin Y-S, Reddy AS, Xiong W, de Pablo JJ, Skinner JL, Zanni MT (2010) 2D IR line shapes probe ovispirin peptide conformation and depth in lipid bilayers. *J Am Chem Soc* **132**:2832–2838.
- (33) Remorino A, Korendovych IV, Wu Y, DeGrado WF, Hochstrasser RM (2011) Residue-specific vibrational echoes yield 3D structures of a transmembrane helix dimer. *Science* **332**:1206–1209.
- (34) Wang L, Middleton CT, Singh S, Reddy AS, Woys AM, Strasfeld DB, Marek P, Raleigh DP, de Pablo JJ, Zanni MT, Skinner JL (2011) 2D IR spectroscopy of human amylin fibrils reflects stable β -sheet structure. *J Am Chem Soc* **133**:16062–16071.
- (35) Dunkelberger EB, Buchanan LE, Marek P, Cao P, Raleigh DP, Zanni MT (2012) Deamidation accelerates amyloid formation and alters amylin fiber structure. *J Am Chem Soc* **134**:12658–12667.
- (36) Shim S-H, Gupta R, Ling YL, Strasfeld DB, Raleigh DP, Zanni MT (2009) Two-dimensional IR spectroscopy and isotope labeling defines the pathway of amyloid formation with residue-specific resolution. *Proc Natl Acad Sci USA* **106**:6614–6619.
- (37) Smith AW, Lessing J, Ganim Z, Peng CS, Tokmakoff A, Roy S, Jansen TLC, Knoester J (2010) Melting of a β -hairpin peptide using isotope-edited 2D IR spectroscopy and simulations. *J Phys Chem B* **114**:10913–10924.
- (38) Middleton CT, Woys AM, Mukherjee SS, Zanni MT (2010) Residue-specific structural kinetics of proteins through the union of isotope labeling, mid-IR pulse shaping, and coherent 2D IR spectroscopy. *Methods* **52**:12–22.
- (39) Woys AM, Almeida AM, Wang L, Chiu C-C, McGovern M, de Pablo JJ, Skinner JL, Gellman SH, Zanni MT (2012) Parallel β -sheet vibrational couplings revealed by 2D IR spectroscopy of an isotopically labeled macrocycle: quantitative benchmark for the interpretation of amyloid and protein infrared spectra. *J Am Chem Soc* **134**:19118–19128.
- (40) Backus EHG, Bloem R, Donaldson PM, Ihalainen JA, Pfister R, Paoli B, Caflisch A, Hamm P (2010) 2D IR study of a photoswitchable isotope-labeled α -helix. *J Phys Chem B* **114**:3735–3740.
- (41) Liang C, Louhivuori M, Marrink SJ, Jansen TLC, Knoester J (2013) Vibrational spectra of a mechanosensitive channel. *J Phys Chem Lett* **4**:448–452.
- (42) King JT, Kubarych KJ (2012) Site-specific coupling of hydration water and protein flexibility studied in solution with ultrafast 2D IR spectroscopy. *J Am Chem Soc* **134**:18705–18712.

- (43) Thielges MC, Axup JY, Wong D, Lee HS, Chung JK, Schultz PG, Fayer MD (2011) Two-dimensional IR spectroscopy of protein dynamics using two vibrational labels: a site-specific genetically encoded unnatural amino acid and an active site ligand. *J Phys Chem B* **115**:11294–11304.
- (44) Choi J-H, Raleigh D, Cho M (2011) Azido homoalanine is a useful infrared probe for monitoring local electrostatics and side-chain solvation in proteins. *J Phys Chem Lett* **2**:2158–2162.
- (45) Dutta S, Li Y-L, Rock W, Houtman JCD, Kohen A, Cheatum CM (2011) 3-Picolyl azide adenine dinucleotide as a probe of femtosecond to picosecond enzyme dynamics. *J Phys Chem B* **116**:542–548.
- (46) Jo H, Culik RM, Korendovych IV, Degrado WF, Gai F (2010) Selective incorporation of nitrile-based infrared probes into proteins via cysteine alkylation. *Biochemistry* **49**:10354–10356.
- (47) Rubtsov IV (2009) Relaxation-assisted two-dimensional infrared (RA 2DIR) method: accessing distances over 10 Å and measuring bond connectivity patterns. *Acc Chem Res* **42**:1385–1394.
- (48) Maekawa H, Poli MD, Moretto A, Toniolo C, Ge N-H (2009) Toward detecting the formation of a single helical turn by 2D IR cross peaks between the amide-I and -II modes. *J Phys Chem B* **113**:11775–11786.
- (49) Kwak K, Park S, Finkelstein IJ, Fayer MD (2007) Frequency-frequency correlation functions and apodization in two-dimensional infrared vibrational echo spectroscopy: a new approach. *J Chem Phys* **127**:124503-124517.
- (50) Roberts ST, Loparo JJ, Tokmakoff A (2006) Characterization of spectral diffusion from two-dimensional line shapes. *J Chem Phys* **125**:084502.
- (51) Skoff DR, Laaser JE, Mukherjee SS, Middleton C, Zanni MT (2012) Simplified and economical 2D IR spectrometer design using a dual acousto-optic modulator. *Chem Phys* **422**:8-15.
- (52) Louis JM, Byeon I-JL, Baxa U, Gronenborn AM (2005) The GB1 amyloid fibril: Recruitment of the peripheral β -strands of the domain swapped dimer into the polymeric interface. *J Mol Biol* **348**:687–698.
- (53) O’Nuallain B, Freir DB, Nicoll AJ, Risse E, Ferguson N, Herron CE, Collinge J, Walsh DM (2010) Amyloid β -protein dimers rapidly form stable synaptotoxic protofibrils. *J Neurosci* **30**:14411–14419.
- (54) Hahn S, Kim S-S, Lee C, Cho M (2005) Characteristic two-dimensional IR spectroscopic features of antiparallel and parallel β -sheet polypeptides: simulation studies. *J Chem Phys* **123**:084905.
- (55) Strasfeld DB, Ling YL, Gupta R, Raleigh DP, Zanni MT (2009) Strategies for extracting structural information from 2D IR spectroscopy of amyloid: application to islet amyloid polypeptide. *J Phys Chem B* **113**:15679-15691.
- (56) Jahn TR, Makin OS, Morris KL, Marshall KE, Tian P, Sikorski P, Serpell LC (2010) The common architecture of cross- β amyloid. *J Mol Biol* **395**:717–727.

- (57) Tycko R (2006) Molecular structure of amyloid fibrils: insights from solid-state NMR. *Q Rev Biophys* **39**:1–55.
- (58) Teoh CL, Pham CLL, Todorova N, Hung A, Lincoln CN, Lees E, Lam YH, Binger KJ, Thomson NH, Radford SE, Smith TA, McEuler SA, Engel A, Griffin MDW, Yarovsky I, Gooley PR, Howlett GJ (2011) A structural model for apolipoprotein C-II amyloid fibrils: 330 PROTEINSCIENCE.ORG An alternative γ D-crystallin aggregate structure experimental characterization and molecular dynamics simulations. *J Mol Biol* **405**:1246–1266.
- (59) Gal N, Morag A, Kolusheva S, Winter R, Landau M, Jelinek R (2013) Lipid bilayers significantly modulate cross-fibrillation of two distinct amyloidogenic peptides. *J Am Chem Soc* **135**:13582–13589.
- (60) Seeliger J, Evers F, Jeworrek C, Kapoor S, Weise K, Andreetto E, Tolan M, Kapurniotu A, Winter R (2012) Cross-amyloid interaction of A β and IAPP at lipid membranes. *Angew Chem Intl Ed* **51**:679–683.
- (61) Sarell CJ, Stockley PG, Radford SE (2013) Assessing the causes and consequences of copolymerization in amyloid formation. *Prion* **7**:0–1.
- (62) Sarell CJ, Woods LA, Su Y, Debelouchina GT, Ashcroft AE, Griffin RG, Stockley PG, Radford SE (2013) Expanding the repertoire of amyloid polymorphs by copolymerization of related protein precursors. *J Biol Chem* **288**:7327–7337.
- (63) Lam AR, Moran SD, Preketes NK, Zhang TO, Zanni MT, Mukamel S (2013) Study of the γ D-crystallin protein using two-dimensional infrared (2DIR) spectroscopy: experiment and simulation. *J Phys Chem B* **117**: 15436–15443.
- (64) Eisenberg D, Jucker M (2012) The amyloid state of proteins in human diseases. *Cell* **148**:1188–1203.
- (65) Ma B, Nussinov R (2012) Selective molecular recognition in amyloid growth and transmission and cross species barriers. *J Mol Biol* **421**:172–184.
- (66) McGlinchey RP, Yap TL, Lee JC (2011) The yin and yang of amyloid: insights from [small alpha]-synuclein and repeat domain of Pmel17. *Phys Chem Chem Phys* **13**:20066–20075.
- (67) Luppi B, Bigucci F, Corace G, Delucca A, Cerchiara T, Sorrenti M, Catenacci L, Di Pietra AM, Zecchi V (2011) Albumin nanoparticles carrying cyclodextrins for nasal delivery of the anti-Alzheimer drug tacrine. *Eur J Pharm Sci* **44**:559–565.
- (68) Kosinski-Collins MS, Flaugh SL, King J (2004) Probing folding and fluorescence quenching in human γ D crystallin Greek key domains using triple tryptophan mutant proteins. *Protein Sci* **13**:2223–2235.
- (69) Mills IA, Flaugh SL, Kosinski-Collins MS, King JA (2007) Folding and stability of the isolated Greek key domains of the long-lived human lens proteins γ D-crystallin and gS-crystallin. *Protein Sci* **16**:2427–2444.

- (70) Moran SD, Decatur SM, Zanni MT (2012) Structural and sequence analysis of the human γ D-crystallin amyloid fibril core using 2D IR spectroscopy, segmental (13)C labeling, and mass spectrometry. *J Am Chem Soc* **134**:18410–18416.
- (71) Moran SD, Zhang TO, Decatur SM, Zanni MT (2013) Amyloid fiber formation in human γ D-crystallin induced by UV-B photodamage. *Biochemistry* **52**:6169-6181.
- (72) Brubaker WD, Freites JA, Golchert KJ, Shapiro RA, Morikis V, Tobias DJ, Martin RW (2011) Separating instability from aggregation propensity in γ S-crystallin variants. *Biophys J* **100**:498–506.
- (73) Zhang W, Cai H-C, Li F-F, Xi Y-B, Ma X, Yan Y-B. (2011) The congenital cataract-linked G61C mutation destabilizes γ D-crystallin and promotes non-native aggregation. *PLoS One* **6**:e20564.
- (74) Lee S, Lin S, Li M, Liang R (1997) Possible mechanism of exacerbating cataract formation in cataractous human lens capsules induced by systemic hypertension or glaucoma. *Ophthalmic Res* **29**:83–90.
- (75) Hanson SRA, Hasan A, Smith DL, Smith JB (2000) The major in vivo modifications of the human water insoluble lens crystallins are disulfide bonds, deamidation, methionine oxidation and backbone cleavage. *Exp Eye Res* **71**:195–207.
- (76) Takemoto L, Sorensen CM (2008) Protein–protein interactions and lens transparency. *Exp Eye Res* **87**: 496–501.

5 Chapter 5

Amyloid Identified in UV-Induced Cataracts of Porcine Lenses using 2D IR Spectroscopy*

5.1 Abstract

Cataracts are formed by the aggregation of crystallin proteins in the lens of the eye. Many *in vitro* studies have established that crystallin proteins precipitate into aggregates that contain amyloid fibers when denatured with acid or irradiated with UV light, but there is little evidence that *ex vivo* cataracts contain amyloid. In this study, we collect two dimensional infrared (2D IR) spectra on tissue slices dissected from the lenses of porcine eyes. 2D IR spectroscopy can identify amyloid fibers too short to be resolved with TEM, as shown in control experiments on *in vitro* α B- and γ D-crystallin. After acid treatment of tissue slices, amyloid is identified by both 2D IR spectroscopic features and fiber morphologies in TEM images. In UV-irradiated tissues, fibers are not observed with TEM, but amyloid is identified from the 2D IR spectra which exhibit the characteristic diagonal and cross peaks of amyloid β -sheets. The amount of amyloid increases linearly with UV-irradiation time until 20 hours (equivalent to 15 years of cumulative sunlight exposure), and then plateaus. We discuss these findings in the context of the chaperone system of the lens, which we hypothesize sequester small aggregates of amyloid β -sheets, preventing long fibers from forming. This study expands the scope of heterodyned 2D IR spectroscopy to tissues.

* The material of this chapter has been submitted for publication in fall 2016 and is currently under review. This work was done in collaboration with Ariel Alperstein and Martin Zanni.

Our results provide a link between *in vitro* and *ex vivo* studies, and support the hypothesis that cataract formation is an amyloid disease.

5.2 Introduction

Senile cataracts are an aggregation disease caused by the precipitation of lens proteins¹. Healthy eye lenses are transparent and focus light onto the retina, enabling vision². The focus is dictated by the refractive index, which is maintained by high protein concentration within the capsule. Lens proteins have long lifetimes (up to 90 years), but eventually precipitate to form aggregates due to aging¹. These protein aggregates create opaque cataract deposits, resulting initially in blurred vision, and ultimately in blindness if not corrected with surgery³. Senile cataracts are the leading cause of blindness worldwide. Because of their prevalence, there is great interest in understanding the composition, structure and mechanism of cataract deposits. It is difficult, however, to study molecular structures directly in tissues (*ex vivo*), and so most structural and biophysical experiments are performed *in vitro* on individual families of crystallin proteins.

One observation from *in vitro* experiments is that crystallin proteins readily form amyloid fibers under common denaturing conditions. More than 90% of the total protein mass in the ocular lens is composed of crystallin proteins⁴. There are three families of crystallin proteins: α -, β -, and γ -crystallin, and each has several different members in its family. *In vitro*, all three of these families form amyloid fibers under denaturing conditions^{5; 6}. Extensive structural and mechanistic studies have been performed on γ D- and α B-crystallins, making them models for their families^{7; 8}. γ D-crystallin exclusively forms amyloid fibers when denatured by acid and precipitates as both disordered aggregates and amyloid fibers upon UV-irradiation^{9; 10; 11; 12}. α B-crystallin also forms amyloid when incubated at 60°C with 10% TFE or under other conditions⁶. Thus, *in vitro*, the

crystallin proteins have multiple pathways to aggregation with different propensities and structures, but amyloid fiber formation is very common.

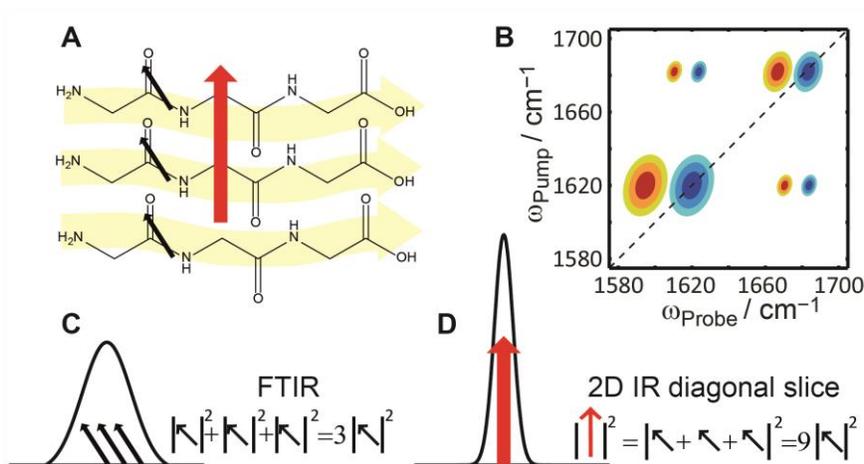


Figure 5.1 Cartoon illustration of amide I modes of (amyloid) β -sheets.

(A) Amide I vibrational modes in β -sheets (parallel). Local modes are shown with black arrow. Vibrational coupling results in the redistribution of the transition dipole creating a normal mode along the fiber axis (red arrow). (B) A cartoon representation of a typical 2D IR spectrum with two pairs of diagonal peaks that are coupled to each other, creating two pairs of cross peaks below and above the diagonal. (C) A cartoon illustration of the different peak intensity and resolution of FTIR and 2D IR diagonal slices.

Even though there is a preponderance of evidence for amyloid fiber formation *in vitro*, there is little evidence for amyloid fibers *ex vivo* in senile cataracts. Amyloid-like structures were reported in mouse lenses with a point mutation on γ S-crystallin^{13; 14} or γ B-crystallin¹⁵, using TEM or Congo red staining. However, the aggregation mechanisms of mutated proteins might be different from the senile cataracts depending on the mutation. In the study of senile cataracts, early biochemical measurements suggested that protein unfolding occurred in cataractous lenses¹, but no amyloid was observed. FTIR spectra of lenses from patients with familial congenital cataracts, myopia, and glaucoma might contain amyloid features¹⁶, but the spectra are spectrally congested and the fits are not unique. Fibers were observed in the nucleus region in the lens of an OXYS

strain of rat¹⁷, but those fibers do not stain for amyloid. Considering the paucity of evidence for amyloid *ex vivo*, it is not clear how to interpret the many structural and mechanistic studies of *in vitro* amyloid formation.

In this article, we use 2D IR spectroscopy to probe the secondary structures of *ex vivo* cataracts induced in slices of porcine lens tissues to mimic age-related cataracts. 2D IR spectroscopy is sensitive to protein structure because of coupling along the protein backbone, such as between carbonyl groups (amides, shown as black arrows in Figure 5.1A). The couplings delocalize the vibrations, so that multiple amides vibrate in unison. Delocalization, in turn, changes the absorption frequencies, absorptivity coefficients, and anharmonicities, creating diagonal peaks and off-diagonal cross peaks (Figure 5.1B), all of which are sensitive to structures¹⁸. The diagonal peaks in 2D IR spectra are analogous to the absorption features in FTIR spectroscopy, but are more sensitive to secondary structures because they scale non-linearly with the absorptivity coefficients¹⁹ (Figure 5.1C). The absorption coefficient for amyloid β -sheets is very large, and so amyloid features are enhanced in 2D IR relative to FTIR spectra²⁰. Cross peaks provide another diagnostic for amyloid not available from FTIR. They are a very direct measure of coupling, and for amyloids, appear in a region of the 2D spectrum with little contribution from other structures. In this article, we utilize these observables unique to 2D IR spectroscopy to discover that cataracts can be induced in the porcine lenses by UV-irradiation and acid denaturation. The spectra provide a link between *ex vivo* tissues and *in vitro* assays. The data implicates amyloid fibers in the disease mechanism of age-related cataract formation.

5.3 Material and Methods

Full-length γ D-crystallin is expressed in *E. coli* and purified according the method described previously¹². Human α B-crystallin pAED4 plasmid was a kind gift from Jonathan King

at MIT University. The plasmid was electroporated into BL21 cells for expression and purified using an established protocol²⁴. 2D IR spectra were collected as described previously¹². The contours on 2D IR spectra were plotted exponentially to the third power to emphasize the lower intensity peaks. All buffers solutions used in 2D IR and FTIR studies are 20 mM NaH₂PO₄ and 100 mM NaCl unless stated otherwise. Porcine eyeballs are supplied by the Meat Science and Muscle Biology Laboratory at UW-Madison. The lenses are extracted from the eyeballs within approximately 2 hours of TOD of the animals. The lenses were stored at -20 °C in embedding medium (Tissue-Tek) until they were sliced using a microtome at -26 °C to a thickness of 25 microns (or 60 microns for TEM samples). The slices were deposited directly onto a CaF₂ window (or a coverslip for TEM samples) and deuterated with D₂O. For native and UV-irradiation samples, pD 7.4 buffer was added to the samples. For acid treated tissue samples, pD 2.0 buffer were used (instead of pD 3.0 for the protein solutions as previous published¹²) because the high protein concentrations in lens tissues served as additional buffering system. After the addition of buffer, the sample cell was assembled with a 56-micron Teflon spacer. For UV-irradiation, the samples cells were exposed to four 8 W medium pressure mercury lamps at a distance of 30 cm, which corresponds to a power intensity of 35 W/m² measured by a power meter. For TEM, the 60-micron lens tissues were treated under the same conditions as for the 2D IR samples. The tissues were then fixed and sectioned to 100 nm for imaging. Lens slices were deposited onto coverslips with a hybridization chamber. Samples were irradiated with UV light before buffer containing 25 μM ThT was added. A 405 nm laser and a 482 nm filter were used for the measurement. Additional details are given in SI Materials and Methods.

5.4 Results

5.4.1 Signature of amyloid formation in 2D IR spectra of *in vitro* proteins.

The crystallin proteins are the major protein component found in senile cataracts²¹. To establish the spectroscopic features expected for amyloid fibers formed from crystallin proteins, we collected 2D IR spectra of acid and UV treated γ D-crystallin (with an S84C mutation) and α B-crystallin in solutions (structures shown in Figure 5.2). It was established in previous studies that the S84C mutation on γ D-crystallin does not alter its native or its aggregated amyloid structure¹². Shown in Figure 5.2A is the 2D IR spectrum of γ D-crystallin in its native fold, prior to acid or UV treatments. 2D IR spectra of γ D-crystallin have been previously reported and computational modeling has helped to interpret these spectra²².

The primary features in γ D-crystallin 2D IR spectra are a pair of out-of-phase peaks (red and blue) at pump axis (ω_{pump} or y-axis) = 1636 cm^{-1} . The 2D IR pulse sequences always creates pairs of out-of-phase peaks, separated in frequency along probe axis (ω_{probe} or x-axis). 1636 cm^{-1} is a typical frequency for short β -sheets found in many soluble proteins. In the case of γ D-crystallin, the 1636 cm^{-1} frequency is formed by 2 Greek Key folds (4 strands of antiparallel β -sheets) found in each of the two domains (Figure 5.2)²². The 2D peaks have elongated shapes that extend up to about 1660 cm^{-1} because of loops and disordered segments, as well as weaker modes of the β -sheets^{19; 23}. Shown in Figure 5.2D is a 2D IR spectrum of native α B-crystallin, which also absorbs near 1636 cm^{-1} . Although the complete structure of α B-crystallin is not known, we assign the 1636 cm^{-1} absorption to the core domain because it consists of short stands of antiparallel β -sheets (Figure 5.2)^{7; 24}. The other two α B-crystallin domains are disordered. Thus, the 2D IR spectra of γ D- and α B-crystallin are quite similar. The diagonal slices ($\omega_{\text{pump}} = \omega_{\text{probe}}$) of the 2D IR

spectra, which are somewhat analogous to a FTIR spectrum, are shown in Figure 5.2G for native γ D and α B-crystallin to help illustrate the similarity between the spectra.

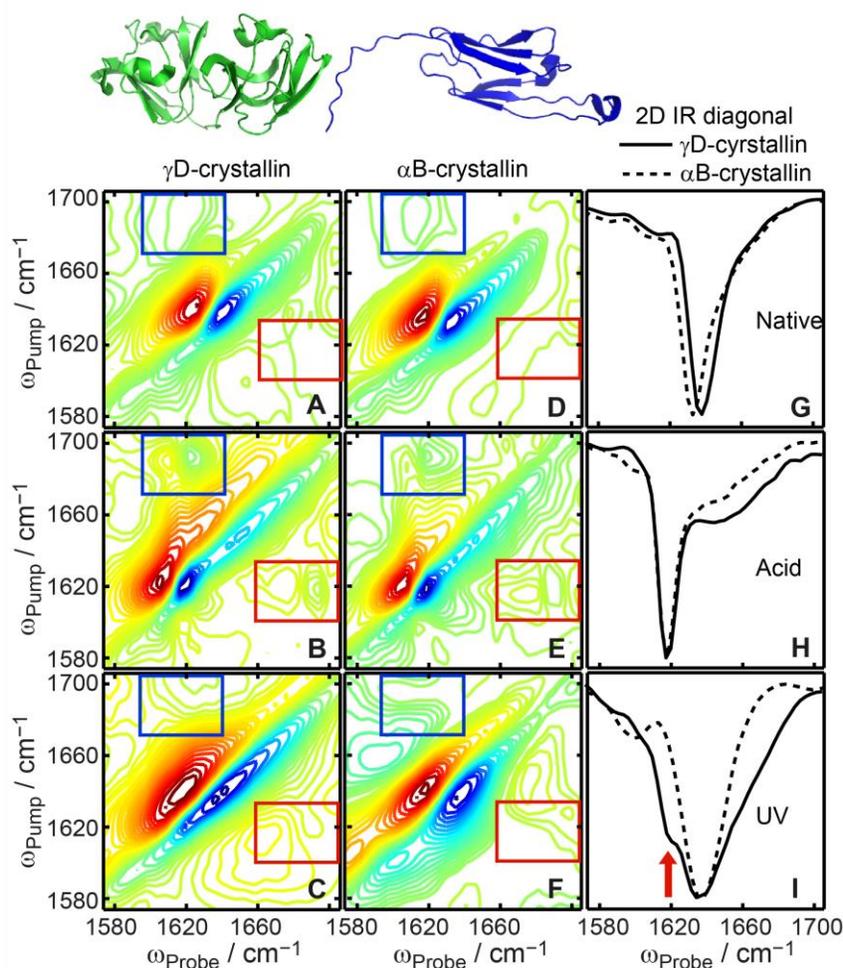


Figure 5.2 2D IR spectra of γ D- and α B-crystallin proteins in deuterated buffer.

(A) Native, (B) acid denatured and (C) UV-irradiated (18 hrs) γ D-crystallin; and (D) Native, (E) acid denatured and (F) UV-irradiated (18 hrs) α B-crystallin 2D IR spectra. The structure illustration of α B shows only the core structure and the N-terminal domain. (G-I) The corresponding diagonal slices of the 2D IR spectra.

Shown in Figure 5.2B is γ D-crystallin denatured in pD 3.0 buffer overnight (TEM images have been previously published²⁵). After the incubation, we observe a well-defined and narrow set

of peaks around $\omega_{\text{pump}}=1621 \text{ cm}^{-1}$. These features are created by highly-ordered amyloid β -sheets and have been observed in other peptide samples such as human amylin²⁶ and polyQ²⁷. We know from segmentally labeled γ D-crystallin studies that the amyloid peak at 1621 cm^{-1} comes from the amyloid fiber formed by the C-terminal domain and that the broad peak around 1650 cm^{-1} corresponds to the random coil structure in the N-terminal domain, which is unfolded in the acid induced structure¹².

We also observe two sets of cross peaks in the spectrum for acid denatured γ D-crystallin (red and blue boxes in Figure 5.2B), with frequencies $\omega_{\text{pump}}= 1621 \text{ cm}^{-1}$ and 1687 cm^{-1} , respectively. These cross peaks are often observed in the 2D IR spectra of amyloid fibers, and are a second signature for amyloid fiber formation.

According to 2D IR and previous study⁶, α B-crystallin also forms highly ordered β -sheets when incubated in acid. Shown in Figure 5.2E is a 2D IR spectrum of α B-crystallin incubated in pD 3.0 buffer overnight (same as for γ D-crystallin). The pair of diagonal peaks at $\omega_{\text{pump}}=1621 \text{ cm}^{-1}$ is observed as are cross peaks on either side of the diagonal correlating the 1621 cm^{-1} band to about 1687 cm^{-1} (red and blue boxes, respectively, in Figure 5.2E). We note that not all amyloid forming proteins exhibit the exact frequencies or cross peaks, because these features depend on the size of the β -sheet and the amount of structural disorder. The absorption wavelength of amyloid β -sheets scales inversely with the number of coupled amino acids in the β -sheet²³. The cross peaks are caused either by coupling between the 1621 cm^{-1} β -sheet mode to a higher mode or to a turn or disordered loop^{19;28}. In principle, the pairs of cross peaks above (1687 cm^{-1} , blue box) and below (1621 cm^{-1} , red box) the diagonal should have the same intensity, but in practice the cross peaks below the diagonal are often better pronounced because they have less overlap with the diagonal peaks. Cross peaks below the diagonal give the amyloid signature frequency by their positions

along ω_{pump} of the 2D IR spectra. Thus, they are strong indicators of amyloid formation, even when the diagonal peaks are congested.

We also used UV-irradiation to induce aggregation. Shown in Figure 5.2C is a 2D IR spectrum of γ D-crystallin irradiated with UV light for 18 hours. Two changes are observed as compared to the spectrum of native γ D-crystallin. First, a shoulder appears at $\omega_{\text{pump}}=1621 \text{ cm}^{-1}$ along with a set of cross peaks below the diagonal (red box), indicating amyloid formation. The diagonal slice of the spectrum (Figure 5.2I) shows this the shoulder more clearly (red arrow). Second, there is the set of cross peaks expected for amyloid (Figure 5.2C, red box). Third, the peaks along the diagonal have intensities above $\omega_{\text{pump}} = 1640 \text{ cm}^{-1}$ and are very broad, consistent with unfolding. This 2D IR spectrum is representative of a situation in which multiple products are formed. Previously, we determined that only the C-terminal domain is responsible for amyloid formation and mass spectrometry experiments found that UV radiation causes fragmentation of γ D-crystallin into shorter peptides⁹. As a result, the diagonal peaks have overlapping contributions. Nonetheless, the diagonal peak at $\omega_{\text{pump}} = 1621 \text{ cm}^{-1}$ and the expected cross peaks indicates that amyloid is produced. The cross peaks are especially useful for identifying amyloid because they lie far from the congested diagonal features.

In contrast to our observations for γ D-crystallin, we do not observe amyloid formation upon UV-irradiation of α B-crystallin. Shown in Figure 5.2F is a 2D IR spectrum of α B-crystallin irradiated under the same conditions as γ D-crystallin for 18 hours. The diagonal peaks broaden but the maximum absorbance is still at about 1636 cm^{-1} after 24 hours (Figure A2. 1) and there is little to no intensity in the lower cross peak region (red box), even after 48 hours of UV-irradiation (Figure A2. 1). Thus, UV-irradiation causes α B-crystallin to unfolding but it does not form amyloid, in contrast to acid denaturation which clearly leads to amyloid formation.

5.4.2 2D IR spectroscopy identifies highly-ordered β -sheets in α B-crystallin amyloid structures that are not easily classified with TEM.

Shown in Figure 5.3 are 2D IR spectra of α B-crystallin treated under conditions in which they do form amyloid. Figure 5.3A, C and E presents data on α B-crystallin solvated in 10% TFE buffer and heated at 60 C° for 35 hours. Large and long amyloid fibers are present in the TEM. The corresponding 2D IR spectrum and diagonal slice (Figure 5.3A and C) exhibit the sharp 1621 cm^{-1} diagonal peak and cross peaks between 1621 cm^{-1} and 1687 cm^{-1} (red box). Thus, according to either the TEM or 2D IR spectroscopy, under this condition α B-crystallin forms amyloid fiber. Shown in Figure 5.3B, D and F, are 2D IR spectra and a TEM image of α B-crystallin when incubated in pH 3.0 buffer. The TEM image (Figure 5.3F) reveals aggregates that are about 50 nm in length. The 2D IR spectra exhibit the 1621 cm^{-1} peak and the corresponding cross peaks of amyloid. Short amyloid morphologies are not uncommon and are observed in TEM images other proteins as fibrillary^{29;30}, prefibrillar^{31;32}, oligomeric³³, or amorphous species³⁴. These images and spectra are shown to illustrate that diversity in length that amyloids can adopt, as measured by TEM, but that 2D IR spectra exhibits characteristic amyloid features regardless of morphology. 2D IR spectroscopy is more sensitive to amyloid formation than TEM because only 3 or 4 strands are needed to create the spectroscopic signatures for amyloid, as explained more below.

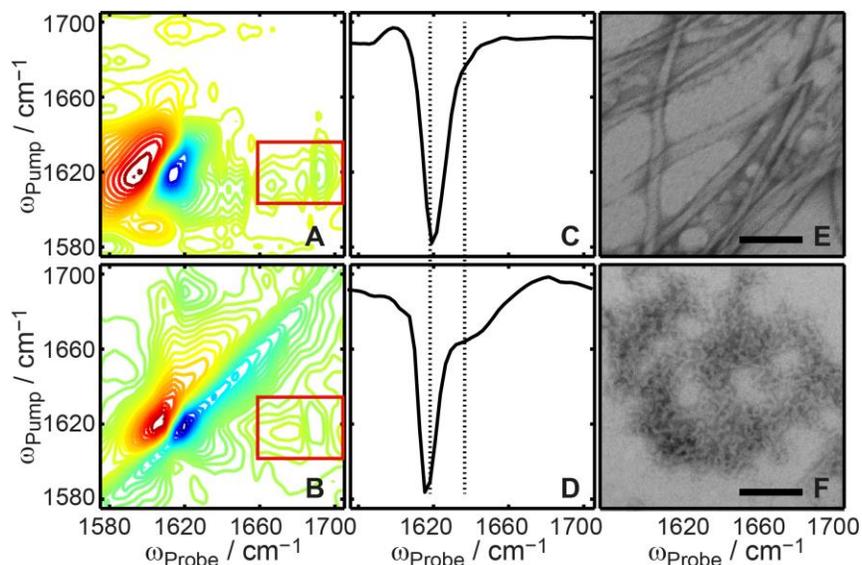


Figure 5.3 2D IR spectra and TEM images of aggregated α B-crystallin.

2D IR spectra of (A) α B-crystallin in 10% TFE phosphate buffer, incubated at 60 °C for 35 hrs. (B) α B-crystallin in phosphate buffer, pH 3.0. (C-D) 2D IR diagonal slices of corresponding 2D IR spectrum on the left. (E-F) TEM images of corresponding 2D IR samples on the left. Scale bars are 100 nm. The concentration of α B-crystallin is 30 μ M in all the solutions.

5.4.3 TEM images of porcine lens tissue treated with acid and UV-irradiation.

TEM of amyloid fiber deposits in tissues have been reported for various parts of the body, including the brain³⁵, the heart³⁶, and the liver³⁷. Those examples of amyloid fibers all show elongated filaments that are easily distinguished from the tissue background. Shown in Figure 5.4 are TEM images for slices of porcine lens tissues. Porcine eye lenses are initially sliced, treated with acid or UV-irradiation, chemically fixed, and then sectioned to 100 nm for TEM imaging (see Materials and Methods). The native tissue (Figure 5.4A) has amorphous features no larger than 5 nm in size that are somewhat evenly distributed throughout the image. While whiter features correspond to higher protein content, the lens tissue has protein everywhere. So the “texture” of the image is essentially a relative contrast between regions of higher and lower protein content.

Nevertheless, the lens is essentially packed with sacks full of proteins that are all subject to negatively staining for TEM. Such high protein concentration makes the analysis of aggregation morphology intrinsically harder than that for other tissues.

Figure 5.4B shows TEM of acid treated tissue. There are much larger sized features, indicating that the spatial variation of protein content is much less uniform than for the native tissue. Also apparent are numerous sharper features that appear to be long fibers. Figure 5.4D shows a copy of Figure 5.4B, in which we have drawn pink lines along linear features that might be fibers. *In vitro*, TEM images of fibers often have periodic twists, which can be visualized because *in vitro* TEM has low background³⁸. In these tissues, there are no regions without protein and so their TEM contrast is very low, because potential fibers are enveloped by tissue. Shown in Figure 5.4C are TEM images for tissue exposed to 90 hours of UV-irradiation. Large amorphous features are observed, but there are few, if any, linear features. These images illustrate the difficulty in identifying amyloid with TEM. While there may be amyloid fibers in acid treated lenses, UV-irradiation does not produce features in TEM images that one could definitively assign to amyloid. In what follows, we prove using 2D IR spectroscopy that amyloids are indeed produced in lens by both acid treatment and UV-irradiation.

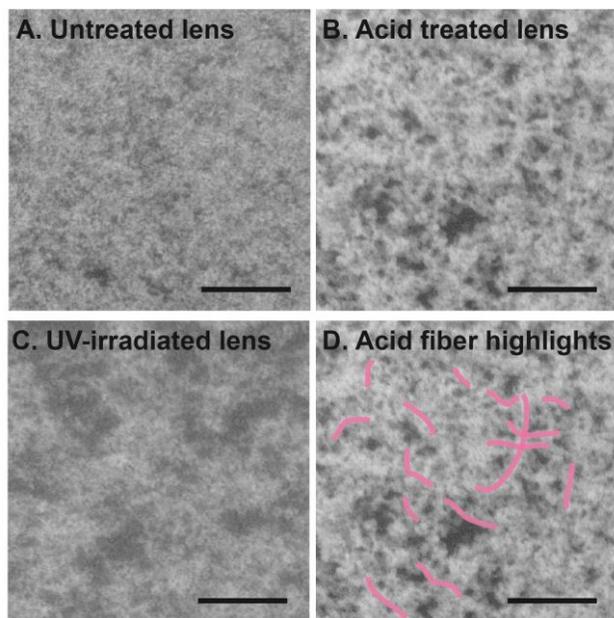


Figure 5.4 TEM images of porcine lens tissues upon acid and UV treatment.

(A) Untreated lens tissue. (B) Lens tissue treated with acid overnight. (C) Len tissue treated with UV-irradiation. (D) Duplicate image of (B) with pink highlights of amyloid fibers. Scale bars are 200 nm.

5.4.4 2D IR spectra reveal amyloid fibers in sliced porcine lens tissues upon acid denaturation and UV-irradiation.

2D IR measurements were made for tissue slices taken from the same lens and prepared in the same way as for the TEM studies above. The 2D IR spectra of untreated lens tissue (Figure 5.5A and F) are remarkably similar to those of native γ D- and α B-crystallin i.e. there are no signatures of amyloid. The diagonal peak has a maximum absorbance at about 1636 cm^{-1} and extends up to 1660 cm^{-1} . The similarity between the spectra of the recombinant protein samples and tissues reflects the fact that crystallin proteins account for 90% of the proteins in mammalian lenses². The exact composition of porcine lenses has not been determined, but is probably similar to humans, according to sequence comparisons³⁹. Thus, we interpret the 2D IR spectra of the tissue using the *in vitro* spectra of γ D- and α B-crystallin as references.

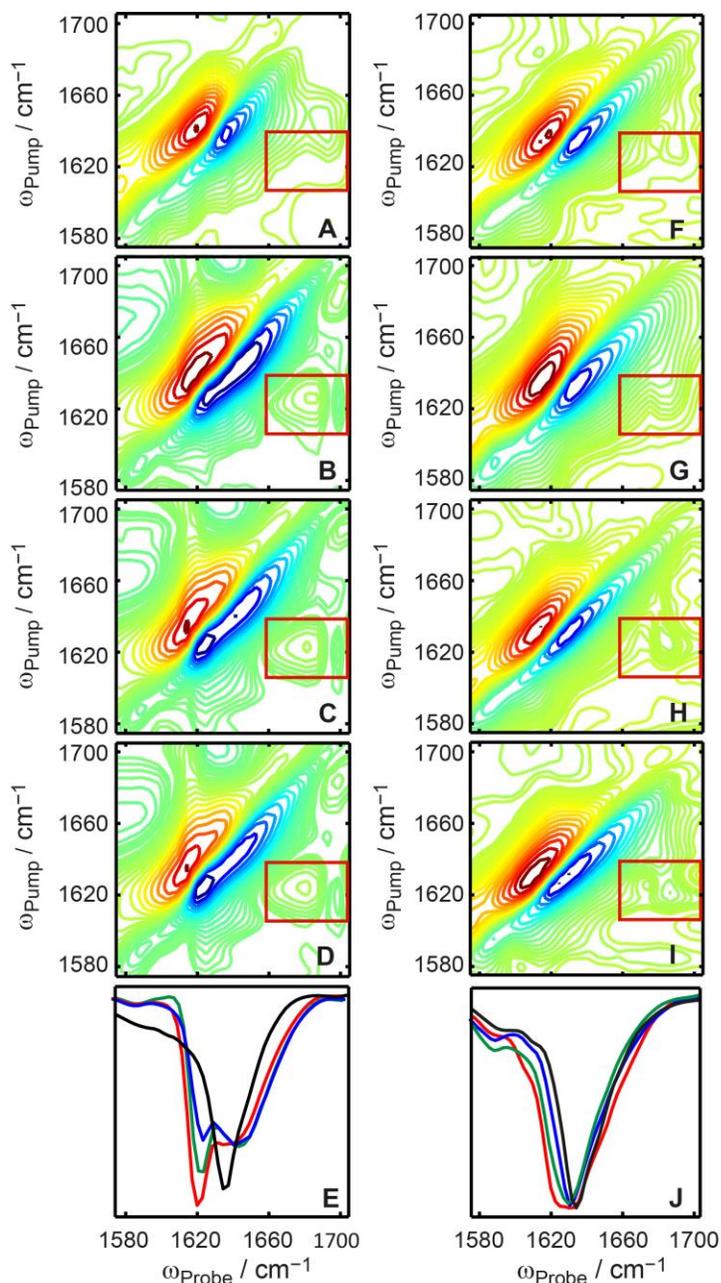


Figure 5.5 2D IR spectra of porcine lens tissues upon acid and UV treatment.

(A and F) Untreated porcine lens tissue in pD 7.4 D₂O buffer, (B) 5 min, (C) 45 min, and (D) 145 min after addition of pD 2.0 buffer. (E) Overlay of diagonal slices of the 2D IR spectra above for (A) in black, (B) in blue, (C) in green and (D) in red. (G) 8 hrs, (H) 16 hrs, and (I) 24 hrs of UV-irradiation in pD 7.4 buffer. (J) Overlay of diagonal slices of the 2D IR spectra above for (F) in black, (G) in blue, (H) in green and (I) in red. Pump slices of these spectra through $\omega_{\text{pump}} = 1621 \text{ cm}^{-1}$ are in Figure A2. 5 further visualize amyloid cross peaks.

Shown in Figure 5.5B-D are 2D IR spectra of lens slices at different time points after acid denaturation. There was a dead time of 5 min between the addition of the acid and when the first 2D IR spectrum was measured (Figure 5.5B). Prior to acidification, the spectra of the tissue exhibits a cross peak at $\omega_{\text{pump}} = 1636 \text{ cm}^{-1}$, which is the frequency for native crystallin β -sheets, but there is no cross peak within the red box where we expect to see amyloid cross peaks (Figure 5.5A, red box). Five min after acidification, cross peaks within the red boxes are clearly present, indicating of amyloid formation, along with a diagonal peak initially at about 1625 cm^{-1} (Figure 5.5B, red box). As time progresses, the lower frequency diagonal peak grows to become the largest features in the spectrum and drops in frequency to about 1621 cm^{-1} . Both the lower diagonal peak and the amyloid cross peaks grow in with about a 20 minute time constant (Figure A2. 2). Figure 5.5E shows the change of 2D IR diagonal slices. Thus, 2D IR spectroscopy provides clear evidence that acidification of the lens tissue causes amyloid β -sheets to form.

Shown in Figure 5.5G-I are 2D IR spectra of lens tissues irradiated with a UV lamp after 0, 8, 16 and 24 hours of irradiation. The temperature of the cell is monitored and kept between 25 to $30 \text{ }^{\circ}\text{C}$ throughout irradiation. UV-irradiation was halted every hour to collect 2D IR spectra. After 8 hrs, the tissue was no longer clear, but was cloudy with a white color. The peak at 1636 cm^{-1} has broadened compared to the peak width in the native lens spectrum (Figure 5.5F), indicating unfolding. A slight shoulder around 1621 cm^{-1} might be present, but there is distinct pair of cross peaks observed (red box), which is strong evidence of amyloid fiber formation¹². After 16 hrs of UV-irradiation (Figure 5.5H), the 1636 cm^{-1} peak of the lens slice further broadens and the pair of cross peaks persists. After 24 hours of UV-irradiation, a diagonal peak near 1621 cm^{-1} becomes prominent enough that it is resolvable from the other diagonal features (Figure 5.5I). The 2D IR diagonal slices better illustrate the formation of a diagonal peak (Figure 5.5).

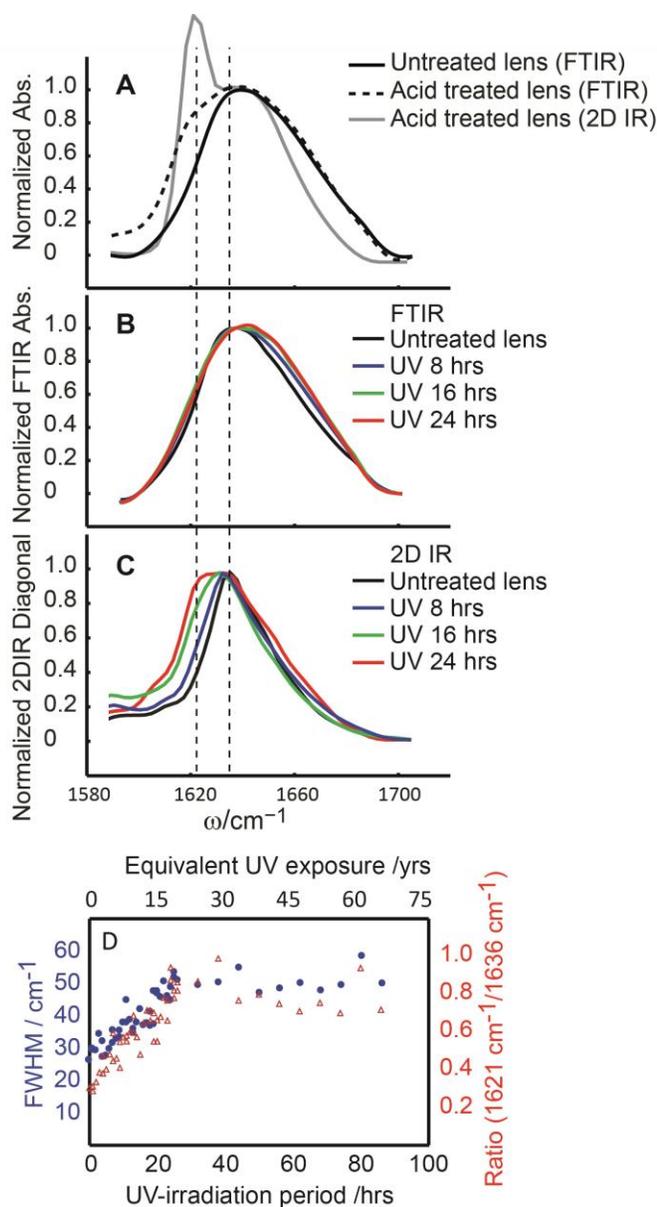


Figure 5.6 FTIR and 2D IR diagonal slices of porcine lens tissues.

(A) Overlay of FTIR spectrum of untreated (black solid) and acid denatured (black dashed) lens tissue, and 2D IR diagonal slice (gray) of acid denatured lens tissues. (B) FTIR spectra of native and UV-irradiated lens tissues. (C) 2D IR diagonal slices of UV-irradiated lens tissues collected on the same samples in (B). (D) Kinetics of amyloid formation plotted as the FWHM (blue circles, left y-axis) and as the ratio between the 1621 cm^{-1} and the 1636 cm^{-1} intensity of the 2D IR diagonal slices (red triangles, right y-axis).

5.4.5 2D IR is more sensitive to amyloid formation than FTIR spectroscopy.

The kinetics of amyloid fiber formation is plotted in Figure 5.6D as measured by the FWHM of the 1636 cm^{-1} peak and by the ratio of intensities at 1621 cm^{-1} and 1636 cm^{-1} . The former metric follows protein denaturation while the latter is a measure of amyloid formation. These ratios increase linearly with exposure time up to 20 hours. Interestingly, after 20 hours the kinetics of both plots plateaus, indicating little additional changes in the secondary structure. If UV irradiation is halted at any point, the sample stops evolving until UV is reapplied (measured up to 48 hours later, see Figure A2. 1), indicating that the protein structures are arrested until further damage occurs. *In vitro* UV-irradiation⁹ of solutions of γ D- and α B-crystallin do not exhibit this plateau (Figure A2. 3).

Amyloid formation *in vitro* is often studied with FTIR spectroscopy. Figure 5.6 shows a comparison between FTIR and 2D IR spectra, collected on the same tissue samples as reported above for the 2D IR studies. Note that the amyloid peak at 1621 cm^{-1} is much more apparent in the 2D IR spectra. It is enhanced by a factor of 5-10 over the FTIR absorbance. For example, from the FTIR spectra alone, one could not conclude that UV-irradiation causes amyloid. In contrast, the diagonal slices from the 2D IR spectra shows a pronounced intensity near 1621 cm^{-1} . This comparison between FTIR and 2D IR spectra is presented to illustrate the improvement that 2D IR spectroscopy makes for diagnosing amyloid formation in complex mixtures. The physics behind this enhancement is explained below.

5.4.6 Thioflavin T (ThT) fluorescence essay of lens tissue is consistent with the formation of amyloid fiber after UV-irradiation.

ThT is a dye that binds preferentially to amyloid fiber structures and exhibits an enhanced fluorescence upon binding. The ThT fluorescence essay of native and UV-irradiated lens is shown

in Figure 5.7. The integrated fluorescence of the UV-irradiated lens samples is at least five-fold higher than that of the native lens samples (Figure 5.7). The higher fluorescence is suggestive of the presence of amyloid fibers in the UV-irradiated samples. Similar increase of ThT fluorescence was observed in UV-irradiated γ D-crystallin solution. While ThT fluorescence alone is not conclusive, because it can bind non-specifically to β -sheet structures⁴⁰, the increase in fluorescence is consistent with the 2D IR results.

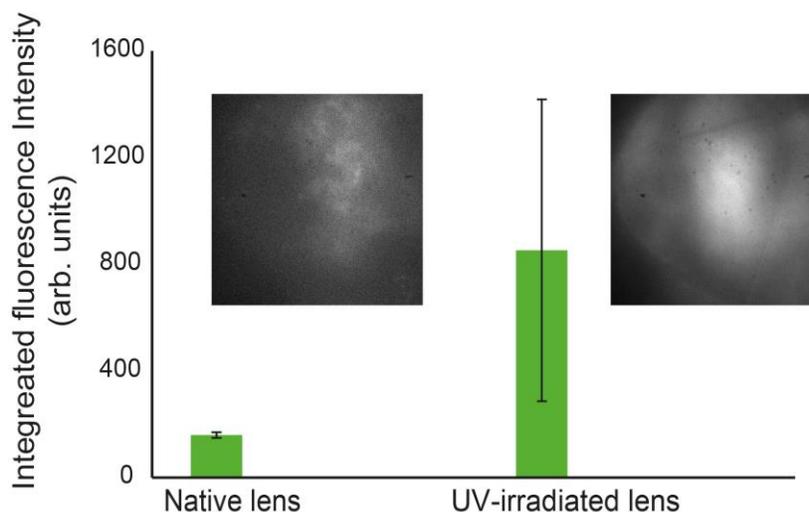


Figure 5.7 ThT fluorescence histogram.

Histogram of ThT fluorescence essay of native and UV-irradiated lens tissue. Phosphate buffer with ThT (25 μ M) is added to 25 μ m lens tissues. The fluorescence is measured with a 405 nm excitation laser and a 482 nm filter. Histogram shows the integrated intensity over the camera spots. The fluorescence images are examples of the ThT fluorescence essay.

5.5 Discussion

FTIR spectroscopy is often used to study amyloid fiber formation because of the characteristic 1621 cm^{-1} absorption created by amyloid β -sheets. While a 1621 cm^{-1} band is a clear identification of amyloid, FTIR spectra are often congested, especially for *ex vivo* tissues. To identify the amyloid mode, deconvolutions are used or spectra are sometimes fit to a sum of Gaussians. Deconvolutions are susceptible to noise and Gaussian fits are almost never unique (nor

correct representations of secondary structures). Thus, assigning amyloid formation based on FTIR spectroscopy alone can be circumspect unless the population of amyloid is the dominant protein structure.

2D IR spectroscopy provides a much more rigorous and sensitive method for identifying amyloid fiber, for two reasons. First, 2D IR spectra scale non-linearly with absorptivity. Beer's Law for FTIR absorption, A , is given by $A^{FTIR} = \epsilon \cdot c \cdot l$, where ϵ is the molar absorptivity, c is the concentration, and l is the pathlength. Each normal mode of the protein, such as the β -sheet modes, has a molar absorptivity. The molar absorptivities are related to the transition dipoles, μ , by $\epsilon \propto |\mu|^2$, which are created by linear combinations of the individual (local mode) transition dipoles for each carbonyl group that contributes to that normal mode. For a given secondary structure, the couplings dictate the precise linear combination, which sets the frequency and magnitude of ϵ for that normal mode. The linear combination of transition dipoles that creates the 1621 cm^{-1} amyloid mode is illustrated in Figure 5.1. This amyloid mode is due mostly to a linear combination of carbonyl groups across β -strands, creating a normal mode that oscillates along the length of the fiber. Up to 12 carbonyl groups vibrate in-phase²⁰, which produces a very large ϵ . There are many other β -sheet modes, which are not illustrated here, that have smaller ϵ . An interesting consequence of Beer's Law is that the integrated area of an FTIR spectrum is conserved for any given protein. That means that if a protein transitions from a random coil to a β -sheet, for example, the spectrum will change shape but the area under the curve is the same. There exist more comprehensive discussions for the physics of these modes elsewhere.

For 2D IR spectroscopy, the analogous equation to Beer's Law is $A^{2DIR} = \epsilon^2 \cdot c \cdot l$. Thus, for a given normal mode, the frequency is the same in a 2D IR spectrum as it is in an FTIR, but the intensities will be different due to the non-linear dependence on ϵ . An interesting consequence of

the ϵ^2 scaling is that the integrated area of a 2D IR spectrum is not conserved^{18; 23}. For example, for a protein with equal amounts of β -sheet and random coil structures, the area under the peaks in FTIR are the same. For the same protein in 2D IR, the β -sheet feature will dominate the spectrum because ϵ for β -sheets is much larger than for random coils (illustrated in Figure 5.1 for a normal mode made from three local modes). Thus, the diagonal peaks of 2D IR spectra are much more sensitive to amyloid fiber formation than is FTIR spectroscopy.

The second reason that 2D IR is more sensitive than FTIR spectroscopy for identifying amyloids is because of the cross peaks. Couplings create the linear combinations that give the frequencies of the normal modes, and in doing so, also create cross peaks at the frequencies between the normal modes, like schematically shown in Figure 5.1. The intensities of cross peaks are proportional to $A^{2DIR}(\text{cross peak}) \propto e_i \cdot e_j \cdot c \cdot l$, where e_i and e_j are the molar absorptivities for two different normal modes. The strongest cross peak in the crystallin amyloids is between the strong 1621 cm^{-1} β -sheet mode, and a weaker normal mode at around 1687 cm^{-1} . Cross peaks appear off of the diagonal and so are better resolved than the diagonal peaks in 2D IR spectra or the corresponding modes in FTIR spectra. Among coupled modes, the amyloid β -sheet cross peaks are particularly well-resolved, because they lie far from the diagonal peaks in a region of the 2D IR spectrum that is largely background free. Thus, a cross peak can be used to determine the frequency of a given mode by its ω_{pump} even if it cannot be identified by the diagonal peaks in a 2D IR spectrum or a congested FTIR spectrum.

The *in vitro* control experiments presented here on γ D- and α B-crystallins clearly shows that they are capable of forming amyloid fibers (Figure 5.2). Acid denaturation of γ D- and α B-crystallin is known to create long amyloid fibers with few other structural products^{5; 6; 38}, which is consistent with the 2D IR spectroscopic features of amyloid: a diagonal peak at 1621 cm^{-1} and the

corresponding cross peaks. For UV irradiation *in vitro*, a shoulder along the diagonal (Figure 5.2I) at 1621 cm^{-1} and a weak cross peak (Figure 5.2C, red box), indicates that γ D forms amyloid although it is not the predominant structure. In contrast, we see no evidence that *in vitro* α B forms amyloid upon UV-irradiation, because there is no resolvable diagonal feature nor cross peaks, even when α B is irradiated more than three times as long as γ D (48 versus 18 hours, see Figure A2. 1). Therefore, at least some subset of the crystallin proteins have the capacity to form amyloid fibers upon UV-irradiation, but not all. These 2D IR spectra of *in vitro* γ D- and α B-crystallin serve as benchmarks against which more complicated spectra can be compared.

TEM imaging is the most common imaging technique for studying amyloid fibers *in vitro*. *In vitro*, samples are prepared by evaporating solution of amyloid proteins onto electron microscopy grids. The background is very low, because there is little protein present in regions of the grid with no fibers. The TEM images of the *in vitro* control experiments are consistent with the conclusions drawn from 2D IR spectroscopy. But much shorter and less linear amyloid fibers can be identified by 2D IR spectroscopy than by TEM, as illustrated by the data in Figure 5.3. For TEM, fibers needs to be at least 25 nm in length to be resolved. In contrast, 2D IR spectroscopy is sensitive to fibers made from only 4 or 5 strands of beta-sheets, which is enough to form the delocalized vibrational modes that create the 2D IR spectroscopic signatures. 4 or 5 strands corresponds to a fiber length of $<2\text{ nm}$. Thus, fibers can be short or twisted, but the amyloidogenic structure still identified by 2D IR.

Another difficulty with TEM is contrast, which is especially problematic for tissues. Tissue imaging with TEM requires samples sliced 100 nm thick, so that there is protein everywhere (especially for lens tissues), and thus, a high background. As a result, *ex vivo* imaging has much lower contrast than *in vitro* imaging. While TEM of cataract lens tissues (that is not due to a

mutation on the crystallin proteins) have been reported, few examples of fibers exist, and their assignment as amyloid is not certain¹⁷. In our TEM results, we do observe a difference between the native and UV-irradiated lens tissues. There may be amyloid fiber structures present in the acid treated lens tissues as we highlighted in Figure 5.4, but none can be identified in the TEM images of the UV irradiated tissues. FTIR spectra of these tissues is also inconclusive (Figure 5.4).

While assigning amyloid in these tissues based on TEM images and FTIR spectroscopy are both problematic, as is illustrated by Figure 5.4 and Figure 5.6 (see also Figure A2. 4), amyloid is clearly observed with 2D IR spectroscopy. Acid treatment of the lens tissues creates 2D IR spectra with a prominent diagonal peak at 1621 cm^{-1} and a cross peak at $\omega_{\text{pump}}=1621\text{ cm}^{-1}$ (red box, Figure 5.2B and C), both of which are the indisputable features of large β -sheets from amyloid fibers. The rest of the spectrum broadens and shifts in intensity to 1645 cm^{-1} , consistent with denaturation into disordered or random coil-like structures. Using the intensities at 1621 versus 1645 cm^{-1} , as representative of amyloid sheet to native structure, we estimate that $\sim 14\%$ of the proteins in the sample are involved in amyloid sheets at 5 min and up to $\sim 21\%$ by the end of incubation at 140 minutes. Thus, while acidification is not physiological, this experiment establishes that the proteins in lens tissue can form amyloid like the individual crystallin proteins do *in vitro*. UV-irradiation of the lens tissues also results in amyloid formation, which we conclude from the signature cross peaks in the 2D IR spectra. The signature cross peak signifies amyloid formation. It correlates to intensity at 1621 cm^{-1} along the diagonal (Figure 5.5J) in which a shoulder appears, similar to the *in vitro* control experiment on αB (Figure 5.2). Thus, amyloid accounts for a fraction of the products. From the intensity of the cross peak relative to the 1636 cm^{-1} native protein absorption, we estimate that about 15% of the protein content is amyloid (See

SI for the detail). Amyloid fiber formation upon UV-irradiation is further confirmed with ThT fluorescence on the lens tissues (Figure 5.7).

Porcine lenses are often used as a model for humans. The soluble proteins in human lenses is estimated to be 40% α -crystallin and 50% β - plus γ -crystallin; a similar distribution is found in porcine lenses³⁹. The X-ray crystallography structures of porcine crystallin proteins have not been reported so we cannot compare the secondary or tertiary structure of porcine to human crystallin proteins, but the protein sequence of porcine and human α A- and α B-crystallin proteins are highly identical (over 91% for α A and 97% for α B); a uncharacterized protein in porcine (gene: LOC100154892) is 82% identical with human γ D-crystallin. Furthermore, the 2D IR spectra of the native lens tissue produces a spectrum that resembles *in vitro* γ D- and α B-crystallins. The spectra of the tissues are broader along the diagonal than the *in vitro* spectra (see slices in Figure 5.2 and Figure 5.5), indicating that there is a larger diversity of structures, but they have the same absorbance maximum $\sim 1636\text{ cm}^{-1}$ and the same general features. Considering these facts, it follows that what we have observed for pig lenses may also apply to human lenses, suggesting that human cataracts might also contain amyloid deposits.

Cataract formation in human is linked to sunlight exposure. Oxidative stress, diabetic conditions, alcohol consumption, and other factors also contribute, but studies have established a strong correlation between cataract prevalence and the exposure time to UV-irradiation. A study done in Nepal by Brilliant *et al.* established that people living in an area with 12 hrs of sun exposure are 4 times more likely to develop cataracts than those with 7 hrs of sun exposure⁴¹.

Estimates put the UV exposure for an average American at 30000 minimal erythema doses (MED) for a person's lifetime, which is equivalent to 12 million J/m^2 . As a result, UV light is often used to simulate cataract formation. Under our experimental conditions, 1 year of natural UV

exposure corresponds to 80 minutes of irradiation with our UV lamp (For detailed calculation, see SI). In the plot of UV induced kinetics (Figure 5.6D) both the intensity at 1621 cm^{-1} and the HWFM of the entire peak grow as the time period of UV-irradiation increases. At early times, denaturation and amyloid growth occur proportionally with UV exposure time, suggesting that natural damage and aggregation may occur in human lenses even with small amounts of UV light.

Cataract formation in humans typically appears at around 50-60 years old. At least 50% of all people will have cataracts by the age of 75³. In order for cataracts to impair vision, the aggregates must be large enough to scatter light. 2D IR spectra require only a few beta-strands to create the 1621 cm^{-1} features, and are thus sensitive to aggregates far smaller than those that scatter light or can be detected by a light microscope. Our TEM images find no fibers in the UV-irradiated lens tissues, putting an upper limit on their size of about 50 nm. Thus, our results suggest that damage and amyloid aggregation occurs continuously throughout a lifetime, but do not form deposits large enough to scatter light and impair vision until major amounts of accumulation occur.

If amyloid fibers do indeed form *ex vivo* as our results support, then why is there not more evidence for amyloid from images of cataractous lens tissues? Or simply, why are amyloid fibers formed from UV-irradiation so short? As mentioned in the Introduction, we find only one report of an image with a structure resembling an amyloid fiber, reported in OXYS strain of rat. One answer to this question might be the difficulty associated with obtaining high quality images with sufficient spatial resolution. Tissue imaging with TEM usually requires samples sliced $< 150\text{ nm}$ thick. For lens tissues, the background from high protein concentration is relatively high compared to that of other tissues.

There may also be a physiological reason for a lack of images of fibers in tissues. The α B-crystallins are small heat shock proteins that are molecular chaperones and are found throughout

the body⁴²; α B-crystallin is found in the heart, brain, and muscle tissues. α B-crystallin increases the stability of proteins and adheres to damaged proteins⁸. It has also been shown that α B-crystallin slows amyloid formation and adheres to amyloid fibers⁴³. The lens contains ~40% of α B-crystallins. Having a large percentage of chaperones is thought necessary to enable the long-term stability of the eye proteins, since the lens proteins are not regenerated and so need to be stable for the entire human lifetime². Thus, we hypothesize that the chaperone proteins in the lens adhere to amyloid aggregates, but do so while they are very small, inhibiting them from forming microscopic fibers visible in EM microscopy. If UV-irradiation is discontinued at any time, the 2D IR spectra stop evolving, indicating that protein aggregation is arrested, which also suggests the work of chaperones. The fibers may only be 4 or 5 strands in lens, since that is all that is needed to create the characteristic 2D IR feature of amyloid.

Another interesting implication of the 2D IR data has to do with the UV-induced kinetics. According to the intensity of the 1621 cm^{-1} peaks (Figure 5.6D), the amount of amyloid beta-sheets forming in the lens increases proportionally with UV exposure time until about 20 hours (equivalent to 15 years). After 20 hours, the signal is roughly constant, indicating that no additional amyloid is formed. What causes the plateau? Our *in vitro* experiments on γ D and α B teach us that γ D is much more susceptible to UV induced degradation than is α B-crystallin. Thus, we hypothesize that the plateau occurs when a subset of proteins, maybe the γ -crystallins, have largely photodegraded so that they no longer strongly absorb UV light. An alternative hypothesis is that the chaperone proteins are more effective after 20 hours exposure than before. Studies have shown that partial unfolding of α B-crystallin can enhance its chaperone activity⁴⁴. We cannot rule out either hypothesis with our current data, but the kinetics reveal that competing factors are at play in the formation of UV induced cataracts.

5.6 Conclusion

There exist many *in vitro* experiments on the amyloid formation of the crystallin proteins, but with little evidence that cataracts contain amyloid. Using 2D IR spectroscopy and *in vitro* control experiments of γ D- and α B-crystallin, we have shown that tissues of porcine eye lenses, when irradiated with UV light to simulate cataract formation, contain aggregates with about 15% amyloid fibers. Acidification of lens tissue also creates amyloid. The population of amyloid beta-sheets increases proportionally with UV exposure until about 20 hours (15 years), but then stops. Amyloid formation is also halted if UV-irradiation is stopped. We attribute these observations to the differences in susceptibility to UV induced aggregation between different families of crystallin proteins and to the role of chaperones. Chaperones may be responsible for inhibiting growth of the fibers, which could explain the lack of amyloid images in cataract tissues. Many of the observations in this report were made possible by the non-linear scaling of 2D IR intensities and cross peaks, which enhances the technique's sensitivity to amyloids. Similar applications could be applied to identify and image tissues in other amyloid diseases. Our results support many of the conclusions drawn from *in vitro* experiments on the crystalline proteins and suggests that natural cataract formation in humans is an amyloid disease.

5.7 Acknowledgements

The Support for this research was provided by the National Institutes of Health (NIH) Grant 1R01GM102387. The authors thank Robby Weyker at UW Meat Science and Muscle Biology Laboratory for his kind supply of the porcine eye balls, Dr. Karla M. Knobel at the Cellular and Molecular Neuroscience Core at the Waisman Center for use of the cryo-microtome, Dr. Leandro B. C. Teixeira at Comparative Ocular Pathology Laboratory of Wisconsin for helpful discussion and dissection instruction

5.8 References

- (1) Harding, J. J. (1972). Conformational Changes in Human Lens Proteins in Cataract. *Biochem. J.* **129**, 97-100.
- (2) Hejtmancik, J. F., Riazuddin, S. A., McGreal, R., Liu, W., Cvekl, A. & Shiels, A. (2015). Lens Biology and Biochemistry.
- (3) Dawson, C. R. & Schwab, I. R. (1981). Epidemiology of cataract—a major cause of preventable blindness. *Bulletin of the World Health Organization* **59**, 493-501.
- (4) Mason, C. V. & Hines, M. C. (1966). Alpha, beta, and gamma crystallins in the ocular lens of rabbits: Preparation and partial characterization. *Investigative Ophthalmology* **5**, 601-609.
- (5) Wang, Y., Petty, S., Trojanowski, A., Knee, K., Goulet, D., Mukerji, I. & King, J. (2010). Formation of amyloid fibrils in vitro from partially unfolded intermediates of human gammaC-crystallin. *Invest Ophthalmol Vis Sci* **51**, 672-8.
- (6) Meehan, S., Berry, Y., Luisi, B., Dobson, C. M., Carver, J. A. & MacPhee, C. E. (2004). Amyloid fibril formation by lens crystallin proteins and its implications for cataract formation. *J Biol Chem* **279**, 3413-9.
- (7) Laganowsky, A., Benesch, J. L., Landau, M., Ding, L., Sawaya, M. R., Cascio, D., Huang, Q., Robinson, C. V., Horwitz, J. & Eisenberg, D. (2010). Crystal structures of truncated alphaA and alphaB crystallins reveal structural mechanisms of polydispersity important for eye lens function. *Protein Sci* **19**, 1031-43.
- (8) Acosta-Sampson, L. & King, J. (2010). Partially folded aggregation intermediates of human gammaD-, gammaC-, and gammaS-crystallin are recognized and bound by human alphaB-crystallin chaperone. *J Mol Biol* **401**, 134-52.
- (9) Moran, S. D., Zhang, T. O., Decatur, S. M. & Zanni, M. T. (2013). Amyloid fiber formation in human gammaD-Crystallin induced by UV-B photodamage. *Biochemistry* **52**, 6169-81.
- (10) Xia, Z., Yang, Z., Huynh, T., King, J. A. & Zhou, R. (2013). UV-radiation induced disruption of dry-cavities in human gammaD-crystallin results in decreased stability and faster unfolding. *Sci Rep* **3**, 1560.
- (11) Schafheimer, N. & King, J. (2013). Tryptophan cluster protects human gammaD-crystallin from ultraviolet radiation-induced photoaggregation in vitro. *Photochem Photobiol* **89**, 1106-15.
- (12) Moran, S. D., Woys, A. M., Buchanan, L. E., Bixby, E., Decatur, S. M. & Zanni, M. T. (2012). Two-dimensional IR spectroscopy and segmental ¹³C labeling reveals the domain structure of human gammaD-crystallin amyloid fibrils. *PNAS*.
- (13) Lee, S., Mahler, B., Toward, J., Jones, B., Wyatt, K., Dong, L., Wistow, G. & Wu, Z. (2010). A single destabilizing mutation (F9S) promotes concerted unfolding of an entire globular domain in gammaS-crystallin. *J Mol Biol* **399**, 320-30.

- (14) Sinha, D., Wyatt, M. K., Sarra, R., Jaworski, C., Slingsby, C., Thaug, C., Pannell, L., Robison, W. G., Favor, J., Lyon, M. & Wistow, G. (2001). A temperature-sensitive mutation of Crygs in the murine Opj cataract. *J Biol Chem* **276**, 9308-15.
- (15) Sandilands, A., Hutcheson, A. M., Long, H. A., Prescott, A. R., Vrensen, G., Löster, J., Klopp, N., Lutz, R. B., Graw, J., Masaki, S., Dobson, C. M., MacPhee, C. E. & Quinlan, R. A. (2002). Altered aggregation properties of mutant γ -crystallins cause inherited cataract. *The EMBO Journal* **21**, 6005-6014.
- (16) Lee, S. M., Lin, S. Y., Cheng, C. L. & Liang, R. C. (1996). Possible changes in secondary structure and composition of human lens capsules in hereditary congenital cataract. *Graefe's Arch Clin Exp Ophthalmol* **234**, 342-348.
- (17) Marsili, S., Salganik, R. I., Albright, C. D., Freel, C. D., Johnsen, S., Peiffer, R. L. & Costello, M. J. (2004). Cataract formation in a strain of rats selected for high oxidative stress. *Exp Eye Res* **79**, 595-612.
- (18) Baiz, C. R., Reppert, M. & Tokmakoff, A. (2013). Amide I Two-Dimensional Infrared Spectroscopy: Methods for Visualizing the Vibrational Structure of Large Proteins. *The Journal of Physical Chemistry A* **117**, 5955-5961.
- (19) Chung, H. S., Khalil, M. & Tokmakoff, A. (2004). Nonlinear Infrared Spectroscopy of Protein Conformational Change during Thermal Unfolding. *The Journal of Physical Chemistry B* **108**, 15332-15342.
- (20) Dunkelberger, E. B., Grechko, M. & Zanni, M. T. (2015). Transition Dipoles from 1D and 2D Infrared Spectroscopy Help Reveal the Secondary Structures of Proteins: Application to Amyloids. *J Phys Chem B* **119**, 14065-75.
- (21) Hanson, S. R., Hasan, A., Smith, D. L. & Smith, J. B. (2000). The major in vivo modifications of the human water-insoluble lens crystallins are disulfide bonds, deamidation, methionine oxidation and backbone cleavage. *Exp Eye Res* **71**, 195-207.
- (22) Lam, A. R., Moran, S. D., Preketes, N. K., Zhang, T. O., Zanni, M. T. & Mukamel, S. (2013). Study of the gammaD-crystallin protein using two-dimensional infrared (2DIR) spectroscopy: experiment and simulation. *J Phys Chem B* **117**, 15436-43.
- (23) Hahn, S., Kim, S. S., Lee, C. & Cho, M. (2005). Characteristic two-dimensional IR spectroscopic features of antiparallel and parallel beta-sheet polypeptides: simulation studies. *J Chem Phys* **123**, 084905.
- (24) Hochberg, G. K., Ecroyd, H., Liu, C., Cox, D., Cascio, D., Sawaya, M. R., Collier, M. P., Stroud, J., Carver, J. A., Baldwin, A. J., Robinson, C. V., Eisenberg, D. S., Benesch, J. L. & Laganowsky, A. (2014). The structured core domain of alphaB-crystallin can prevent amyloid fibrillation and associated toxicity. *Proc Natl Acad Sci U S A* **111**, E1562-70.
- (25) Moran, S. D., Decatur, S. M. & Zanni, M. T. (2012). Structural and sequence analysis of the human gammaD-crystallin amyloid fibril core using 2D IR spectroscopy, segmental ¹³C labeling, and mass spectrometry. *J Am Chem Soc* **134**, 18410-6.

- (26) Shim, S. H., Gupta, R., Ling, Y. L., Strasfeld, D. B., Raleigh, D. P. & Zanni, M. T. (2009). Two-dimensional IR spectroscopy and isotope labeling defines the pathway of amyloid formation with residue-specific resolution. *Proc Natl Acad Sci U S A* **106**, 6614-9.
- (27) Buchanan, L. E., Carr, J. K., Fluitt, A. M., Hoganson, A. J., Moran, S. D., de Pablo, J. J., Skinner, J. L. & Zanni, M. T. (2014). Structural motif of polyglutamine amyloid fibrils discerned with mixed-isotope infrared spectroscopy. *Proc Natl Acad Sci U S A* **111**, 5796-801.
- (28) Strasfeld, D. B., Ling, Y. L., Gupta, R., Raleigh, D. P. & Zanni, M. T. (2009). Strategies for extracting structural information from 2D IR spectroscopy of amyloid: application to islet amyloid polypeptide. *J Phys Chem B* **113**, 15679-91.
- (29) Khan, J. M., Qadeer, A., Chaturvedi, S. K., Ahmad, E., Rehman, S. A., Gourinath, S. & Khan, R. H. (2012). SDS can be utilized as an amyloid inducer: a case study on diverse proteins. *PLoS One* **7**, e29694.
- (30) Jurado, R., Castello, F., Bondia, P., Casado, S., Flors, C., Cuesta, R., Dominguez-Vera, J. M., Orte, A. & Galvez, N. (2016). Apoferritin fibers: a new template for 1D fluorescent hybrid nanostructures. *Nanoscale* **8**, 9648-56.
- (31) Tycko, R. (2015). Amyloid polymorphism: structural basis and neurobiological relevance. *Neuron* **86**, 632-45.
- (32) Starck, C. S. & Sutherland-Smith, A. J. (2010). Cytotoxic aggregation and amyloid formation by the myostatin precursor protein. *PLoS One* **5**, e9170.
- (33) Matheou, C. J., Younan, N. D. & Viles, J. H. (2015). Cu(2)(+) accentuates distinct misfolding of Abeta(1)(-)(4)(0) and Abeta(1)(-)(4)(2) peptides, and potentiates membrane disruption. *Biochem J* **466**, 233-42.
- (34) Romero, D., Aguilar, C., Losick, R. & Kolter, R. (2010). Amyloid fibers provide structural integrity to *Bacillus subtilis* biofilms. *Proc Natl Acad Sci U S A* **107**, 2230-4.
- (35) Eisenberg, D. & Jucker, M. (2012). The amyloid state of proteins in human diseases. *Cell* **148**, 1188-20.
- (36) Christoph, D. C., Boese, D., Johnson, K. T., Schlosser, T. W., Hunold, P., Baba, H. A., Erbel, R. & Philipp, S. (2009). Heart failure and cardiac involvement as isolated manifestation of familial form of transthyretin amyloidosis resulting from Val30Met mutation with no clinical signs of polyneuropathy. *Circ Heart Fail* **2**, 512-5.
- (37) Tosoni, A., Barbiano di Belgiojoso, G. & Nebuloni, M. (2011). Electron Microscopy in the Diagnosis of Amyloidosis. In *Amyloidosis - Mechanisms and Prospects for Therapy* (Sarantseva, S., ed.). InTech.
- (38) Papanikolopoulou, K., Mills-Henry, I., Thol, S. L., Wang, Y., Gross, A. A. R., Kirschner, D. A., Decatur, S. M. & King, J. (2008). Formation of amyloid fibrils in vitro by human γ D-crystallin and its isolated domains. *Molecular Vision* **14**, 81-89.
- (39) Keenan, J., Orr, D. F. & Pierscionek, B. K. (2008). Patterns of crystallin distribution in porcine eye lenses. *Mol Vis* **14**, 1245-53.

- (40) Eisert, R., Felau, L. & Brown, L. R. (2006). Methods for enhancing the accuracy and reproducibility of Congo red and thioflavin T assays. *Anal Biochem* **353**, 144-6.
- (41) West, S. K. & Valmadrid, C. T. (1995). Epidemiology of Risk Factors for Age-Related Cataract. *Survey of Ophthalmology* **39**, 323-334.
- (42) Iwaki, T., Kume-Iwaki, A. & Goldman, J. E. (1990). Cellular distribution of alpha B-crystallin in non-lenticular tissues. *Journal of Histochemistry & Cytochemistry* **38**, 31-39.
- (43) Raman, B., Ban, T., Sakai, M., Pasta, S. Y., Ramakrishna, T., Naiki, H., Goto, Y. & Rao Ch, M. (2005). AlphaB-crystallin, a small heat-shock protein, prevents the amyloid fibril growth of an amyloid beta-peptide and beta2-microglobulin. *Biochem J* **392**, 573-81.
- (44) Foit, L., George, J. S., Zhang, B. W., Brooks, C. L., 3rd & Bardwell, J. C. (2013). Chaperone activation by unfolding. *Proc Natl Acad Sci U S A* **110**, E1254-62.

6 Chapter 6

GXXXG-mediated parallel and antiparallel dimerization of transmembrane helices and its inhibition by cholesterol: Single-molecule FRET and 2D IR studies*

6.1 Abstract

Small-residue-mediated interhelical packings are ubiquitously found in helical membrane proteins, although their interaction dynamics and lipid dependence remain mostly uncharacterized. We used a single-molecule FRET technique to examine the effect of a GXXXG motif on the association of de-novo designed (AALALAA)₃ helices in liposomes. Dimerization occurred with sub-second lifetimes, which was abolished by cholesterol. Utilizing the nearly instantaneous time-resolution of 2D IR spectroscopy, parallel and antiparallel helix associations were identified by vibrational couplings across helices at their interface. Taken together, the data illustrate that the GXXXG motif controls helix packing but still allows for a dynamic and lipid-regulated oligomeric state.

* The material of this chapter was prepared as a manuscript submitted to *Angewandte Chemie* in fall 2016 as: Yoshiaki Yano, Kotaro Kondo, Yuta Watanabe, Tianqi O. Zhang, Jia-Jung Ho, Shinya Oishi, Nobutaka Fujii, Martin T. Zanni, and Katsumi Matsuzaki. “GXXXG-mediated parallel and antiparallel dimerization of transmembrane helices and its inhibition by cholesterol: single-molecule FRET and 2D IR studies”. I contributed to this work in the 2D IR spectroscopy experimental design, data collection and data analysis.

6.2 Introduction

Small amino acid residues such as Gly, Ala, and Ser are ubiquitously found at contact interfaces of transmembrane helices in clusters of small residues with 2 or 3 interval residues (such as GXXXG and GXXXGXXXG) with various topologies and orientations.¹⁻³ Because these small residues do not have sufficient side chain hydrophobicities to compensate for the partitioning of the polar main chain of the helix into membranes,⁴ the polar surfaces can preferentially pack with each other to avoid lipid exposure. The energetic stability of small-residue packing markedly depends on the protein sequence and membrane lipids,⁵ although the regulation rules are poorly understood.

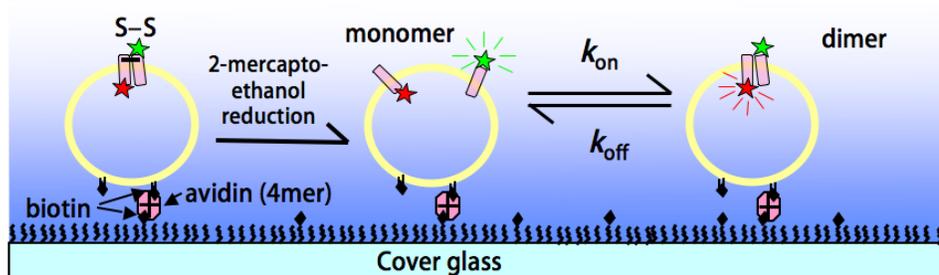
While it is clear from the sequence homology that small-residue packing is important for stabilizing interhelical interactions in membranes, the role that these sequences play in dynamics is less well understood. Rearrangements of transmembrane helices due to transient helix–helix interaction are known to occur in membrane proteins.⁶ For example, G-protein-coupled receptors and growth factor receptors fluctuate among distinct conformations and oligomeric states in millisecond and subsecond timescales, respectively.^{7,8} Single-molecule and ensemble methods, such as fluorescence, atomic force microscopy, and NMR, have enabled the observation of membrane protein dynamics in model membranes and living cells.⁹⁻¹¹ However, while the existence of dynamics has been established, obtaining specific structural information is difficult in membrane-bound systems, obscuring the link between structure and dynamics.

6.3 Result and Discussion

To investigate the effects of sequence motifs and lipid compositions on helix associations, we utilize a de-novo designed hydrophobic peptide (AALALAA)₃. The peptide assumes a stable transmembrane helix with a tendency to form antiparallel dimers due to electrostatic attraction

between the helix macrodipoles.¹² Recently, single-molecule fluorescence resonance energy transfer (sm-FRET) experiments in liposomes revealed that cholesterol stabilizes an hourglass-shaped antiparallel dimer by ‘lipophobic’ forces.¹³ Using this transmembrane helix as a host, we investigated the effect of the guest GXXXG motif at the center of the sequence (AALALAA-AGLALGA-AALALAA).

a)



b) Host in POPC/cholesterol

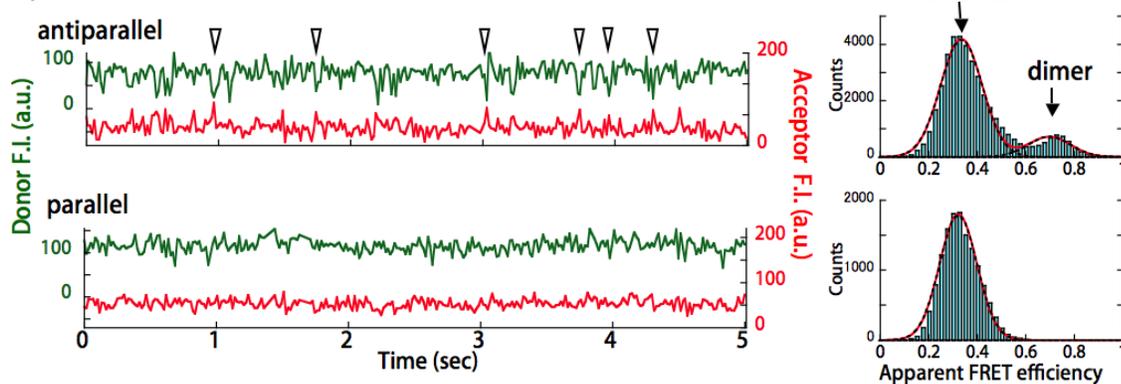


Figure 6.1 Cartoon illustration of dimer formation.

The concept (a) and proof-of-concept (b) of controlling the helix topology in liposomal sm-FRET experiments. (a) Liposomes containing disulfide-bridged dimers are attached to a cover glass via biotin–avidin interaction. Reduction of the disulfide with 2-mercaptoethanol results in two helix monomers that can noncovalently dimerize with a defined topology. Note that the flip-flop of the helix does not occur.²⁶ (b) Fluorescence time traces for host helix dimers in POPC/cholesterol (7/3) vesicles (25°C). Arrowheads indicate transient FRET signals due to helix association. In antiparallel and parallel topologies, 86 and 0% of the bright spots including donor and acceptor pair exhibited FRET fluctuations, respectively ($n = 28$). Right panels show histograms of the apparent FRET efficiency ($F_{\text{acceptor}}/(F_{\text{donor}}+F_{\text{acceptor}})$).

To monitor the relative helix topology in the liposomal sm-FRET, two helices labeled with the FRET donor Cy3B and the acceptor Cy5 were disulfide-linked between the N and N, or N and C termini (Supporting Information, Figure A3. 1). After dilution of the dimer helices into vesicles (1 dimer per 10 vesicles on average), reduction of the disulfide resulted in two transmembrane helices with controlled topology (Figure 6.1a). As a proof-of-concept, antiparallel and parallel host helix dimers were prepared in palmitoyl oleoyl phosphocholine (POPC)/cholesterol (7/3) vesicles, and observed with total internal reflection fluorescence microscopy (Figure 6.1b). As expected, most of the bright spots including a donor and acceptor pair exhibited FRET fluctuation due to transient associations in antiparallel topology (86%, $n = 28$),¹³ whereas no FRET was observed for parallel topology (0%, $n = 28$) in the time resolution (17 msec).

In POPC, no association was detected for the host helices, whereas the introduction of the GXXXG motif evoked transient dimerizations in both topologies (Figure 6.2a,b). The addition of cholesterol completely inhibited these Gly-mediated associations (Figure 6.2a). The association of glycoporphin A with a GXXXG motif was also weakened by cholesterol and other membrane components.^{14,15} A possible reason is preferable interaction between GXXXG and cholesterol to suppress helix–helix contact. However, the degree of FRET from a fluorescent sterol, dehydroergosterol, to NBD-labeled helices suggested no association between them (Supporting Information, Figure A3. 2).

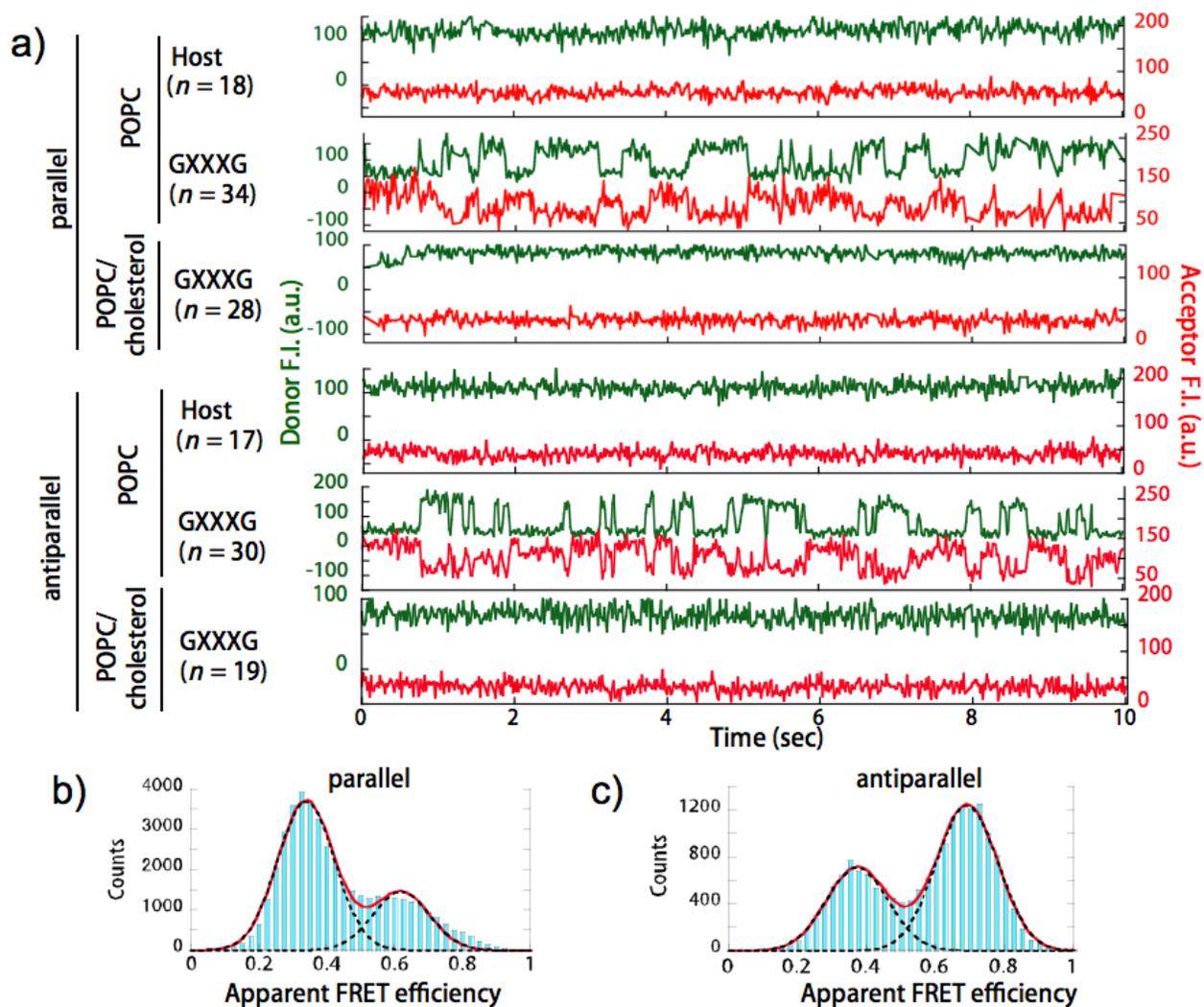


Figure 6.2 GXXXG-mediated association of transmembrane helices (25°C).

(a) Representative sm-FRET trajectories under different association topologies and lipid compositions. (b)(c) Histograms of the apparent FRET efficiency for parallel (b) and antiparallel (c) associations of GXXXG helices in POPC.

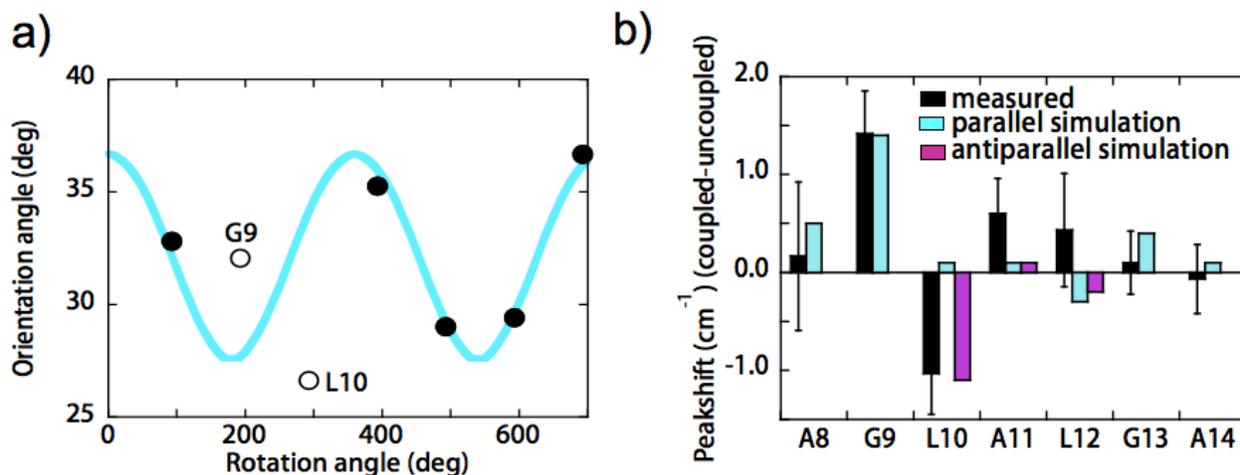


Figure 6.3 Angle orientation and peakshift of isotope labeled residues.

Helix orientation and association interface measured with $^{13}\text{C}=^{18}\text{O}$ isotope-labeled helices. (a) Position-dependent orientation angle of amide I bands (circles), determined from dichroic ratios of amide I' bands for PATR FTIR spectra. The curve fitting did not consider the data for G9 and L10. The best fit (cyan) gave a helix orientation angle of 4.5° . (b) Frequency shift of 100% isotope samples from isotope-diluted samples (25% isotopes) determined from the diagonal cut of 2D IR spectra (black bars). Cyan and purple bars indicate simulated shifts for the parallel and antiparallel dimers, respectively.

To obtain more detailed structural insights, we performed 2D IR and polarized-attenuated total internal reflection (PATR) FTIR in POPC for uncrosslinked GXXXG helices that were $^{13}\text{C}=^{18}\text{O}$ labeled at each residue in the central heptad (A8, G9, L10, A11, L12, G13, and A14). The isotope amide I' peaks were $60\text{--}70\text{ cm}^{-1}$ red shifted ($\sim 1,590\text{ cm}^{-1}$) compared with unlabeled peaks (Supporting Information, Figure A3. 3), as reported.¹⁶ The position-dependent tilt of amide bonds, calculated from amide I' PATR dichroic ratios, was best described assuming a helix orientation angle of $\sim 5^\circ$, when the residues at interhelical contacts (G9 and L10, see below) were excluded from the fitting (Figure 6.3a). Using 2D IR spectroscopy, we also measured the vibrational coupling between residues, as determined by the shift in frequency as compared with isotope-diluted samples (Figure 6.3b, black bars). 2D IR spectroscopy is more advantageous than FTIR for vibrational coupling measurements because the isotope bands appear more prominently relative to weaker background bands.¹⁷ A positive shift at G9 is consistent with a parallel contact,¹⁷

whereas a negative shift at L10 suggests an antiparallel approach with head-to-tail dipole geometry.¹⁷ Indeed, spectral simulations with crossing angles of $+10^\circ$ and -170° gave the G9 and L10 shifts, respectively (Figure 6.3b). Note that the parallel and antiparallel contacts did not affect L10 and G9 shifts, respectively. We therefore conclude that interhelical contacts with both topologies are populated with approximately equal weighting, supporting the findings from the smFRET experiments. A similar parallel interface was independently predicted by a surface-based packing algorithm (Supporting Information, Supporting Discussion and Figure A3. 4).

	POPC		POPC/cholesterol (7/3)	
	Parallel	Antiparallel	Parallel	Antiparallel
Host ΔG_a (kJ mol ⁻¹)	ca. -2 ± 1 ^[a]	-13.2 ± 0.2 ^[b]	ca. -7 ± 1 ^[a]	-19.1 ± 0.2
GXXXG ΔG_a (kJ mol ⁻¹)	-21.4 ± 1.2	-28.3 ± 1.4	-15 or weaker ^[c]	-15 or weaker ^[c]
$\Delta\Delta G_a$ (kJ mol ⁻¹)	-20 ($-8 kT$)	-15 ± 1.4 ($-6 kT$)	-8 or weaker	+4 or more positive

Table 6.1 Energetic contribution of GXXXG motif at 25°C

[a] Estimated from self-quenching of NBD-labeled helices. [b] Determined by ensemble FRET. [c] Estimated from the detection limit of $k_{\text{off}} = 1/(17 \times 10^{-3}) = 59 \text{ s}^{-1}$ and assuming k_{on} of GXXXG parallel dimer (2.3 s^{-1}).

The lifetimes of monomers and dimers were estimated by HaMMY fitting¹⁸ to the apparent FRET efficiency (Figure 6.4). Helix macrodipole attraction in antiparallel orientation resulted in a faster association and slower dissociation of the helices. Energetic contributions of the GXXXG motif are summarized in

Table 6.1, obtained from smFRET, ensemble FRET, or self-quenching (Supporting Information, Figure A3. 5) experiments. The introduction of terminal labels and linker residues was assumed to have no effect on the association, although it can slightly alter the association energy ($\sim 3 \text{ kJ mol}^{-1}$ for host antiparallel dimer in POPC/cholesterol).¹³ In POPC, the motif contributed to $\Delta\Delta G_a$ of -20 and -15 kJ mol^{-1} (in doubled mole fraction units) for parallel and antiparallel associations, respectively, comparable with the effect of a Gly mutation on

glycophorin A dimerization.¹⁹ However, in the presence of cholesterol, the contribution became weaker in both topologies.

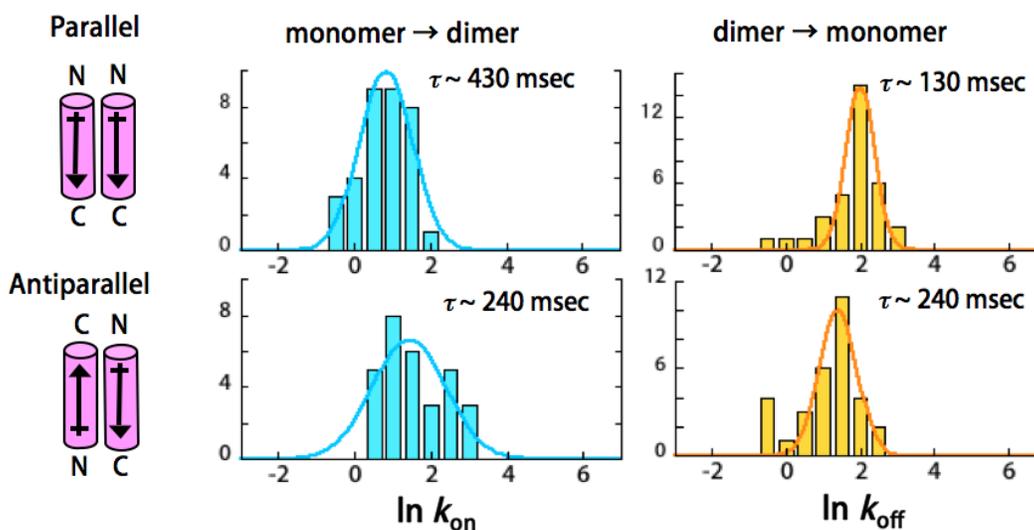


Figure 6.4 Kinetics of the monomer–dimer transition of GXXXG helices in POPC.

Distributions of rate constants for association and dissociation processes were estimated by the HaMMY program, and the histograms were fitted with Gaussians to estimate the average lifetimes.

It is interesting to note that the dissociation rates of the GXXXG dimers studied here in POPC ($k_{off} = 3\text{--}7\text{ s}^{-1}$) are similar or even higher than those of aqueous coiled-coils,^{20,21} despite a comparable bulk K_d of 200–10 μM , estimated from the ΔG_a , assuming that the helices are confined into the bilayers with an average area and thickness of 60 \AA^2 and 15 \AA per lipid, respectively. Frictional forces from both solvents and protein can affect protein dissociation dynamics.²² The helix termini are anchored at the water–membrane interface with high viscosity (100–200 cP)^{23,24}; therefore, the dissociation should be slowed down compared with that of an aqueous coiled-coil. On the other hand, helix association–dissociation in membranes does not require dihedral isomerization of the backbone as helix–coil transition for aqueous coiled-coils. As a result of these

opposing effects, interhelical interactions in membranes appear to be dynamic as those typically seen in the aqueous phase.

6.4 Conclusion

In conclusion, two emerging techniques, the liposomal smFRET with controlled association topology and isotope-edited 2D IR, shed light on the dynamic nature of small-residue-mediated helix associations with a defined interface, which is useful for the rapid rearrangement and transient association of membrane proteins. The motif drove both parallel and antiparallel associations, whereas cholesterol impaired the motif-mediated attraction. In animal cells, cholesterol levels are tightly controlled among organelles,²⁵ suggesting a mechanism for membrane proteins in which the strength of small-residue interaction is regulated by cholesterol in organelles.

6.5 References

- (1) Zhang, S. Q.; Kulp, D. W.; Schramm, C. A.; Mravic, M.; Samish, I.; DeGrado, W. F.: The membrane- and soluble-protein helix-helix interactome: similar geometry via different interactions. *Structure* **2015**, *23*, 527–541.
- (2) Walters, R. F.; DeGrado, W. F.: Helix-packing motifs in membrane proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 13658–13663.
- (3) Russ, W. P.; Engelman, D. M.: The GxxxG motif: a framework for transmembrane helix-helix association. *J. Mol. Biol.* **2000**, *296*, 911–919.
- (4) White, S. H.; Wimley, W. C.: Membrane protein folding and stability: physical principles. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 319–365.
- (5) Teese, M. G.; Langosch, D.: Role of GxxxG Motifs in Transmembrane Domain Interactions. *Biochemistry* **2015**, *54*, 5125–5135.
- (6) Langosch, D.; Arkin, I. T.: Interaction and conformational dynamics of membrane-spanning protein helices. *Protein Sci.* **2009**, *18*, 1343–1358.
- (7) Kofuku, Y.; Ueda, T.; Okude, J.; Shiraishi, Y.; Kondo, K.; Mizumura, T.; Suzuki, S.; Shimada, I.: Functional dynamics of deuterated beta2 -adrenergic receptor in lipid bilayers revealed by NMR spectroscopy. *Angew. Chem. Int. Ed. Engl.* **2014**, *53*, 13376–13379.

- (8) Low-Nam, S. T.; Lidke, K. A.; Cutler, P. J.; Roovers, R. C.; van Bergen en Henegouwen, P. M.; Wilson, B. S.; Lidke, D. S.: ErbB1 dimerization is promoted by domain co-confinement and stabilized by ligand binding. *Nat. Struct. Mol. Biol.* **2011**, *18*, 1244–1249.
- (9) Kasai, R. S.; Kusumi, A.: Single-molecule imaging revealed dynamic GPCR dimerization. *Curr. Opin. Cell Biol.* **2014**, *27*, 78–86.
- (10) Jefferson, R. E.; Blois, T. M.; Bowie, J. U.: Membrane proteins can have high kinetic stability. *J. Am. Chem. Soc.* **2013**, *135*, 15183–15190.
- (11) Zocher, M.; Zhang, C.; Rasmussen, S. G.; Kobilka, B. K.; Muller, D. J.: Cholesterol increases kinetic, energetic, and mechanical stability of the human beta2-adrenergic receptor. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, E3463–3472.
- (12) Yano, Y.; Matsuzaki, K.: Measurement of thermodynamic parameters for hydrophobic mismatch 1: self-association of a transmembrane helix. *Biochemistry* **2006**, *45*, 3370–3378.
- (13) Yano, Y.; Kondo, K.; Kitani, R.; Yamamoto, A.; Matsuzaki, K.: Cholesterol-induced lipophobic interaction between transmembrane helices using ensemble and single-molecule fluorescence resonance energy transfer. *Biochemistry* **2015**, *54*, 1371–1379.
- (14) Hong, H.; Bowie, J. U.: Dramatic destabilization of transmembrane helix interactions by features of natural membrane environments. *J. Am. Chem. Soc.* **2011**, *133*, 11389–11398.
- (15) Yano, Y.; Furukawa, N.; Ono, S.; Takeda, Y.; Matsuzaki, K.: Selective amine labeling of cell surface proteins guided by coiled-coil assembly. *Biopolymers* **2016**, *106*, 484–490.
- (16) Manor, J.; Mukherjee, P.; Lin, Y. S.; Leonov, H.; Skinner, J. L.; Zanni, M. T.; Arkin, I. T.: Gating mechanism of the influenza A M2 channel revealed by 1D and 2D IR spectroscopies. *Structure* **2009**, *17*, 247–254.
- (17) Hamm, P.; Zanni, M.: *Concepts and methods of 2D infrared spectroscopy*; 1 ed.; Cambridge University Press, 2011.
- (18) McKinney, S. A.; Joo, C.; Ha, T.: Analysis of single-molecule FRET trajectories using hidden Markov modeling. *Biophys. J.* **2006**, *91*, 1941–1951.
- (19) Hong, H.; Blois, T. M.; Cao, Z.; Bowie, J. U.: Method to measure strong protein-protein interactions in lipid bilayers using a steric trap. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 19802–19807.
- (20) Wendt, H.; Berger, C.; Baici, A.; Thomas, R. M.; Bosshard, H. R.: Kinetics of folding of leucine zipper domains. *Biochemistry* **1995**, *34*, 4097–4107.
- (21) De Crescenzo, G.; Litowski, J. R.; Hodges, R. S.; O'Connor-McCourt, M. D.: Real-time monitoring of the interactions of two-stranded de novo designed coiled-coils: effect of chain length on the kinetic and thermodynamic constants of binding. *Biochemistry* **2003**, *42*, 1754–1763.
- (22) Hagen, S. J.: Solvent viscosity and friction in protein folding dynamics. *Curr. Protein Pept. Sci.* **2010**, *11*, 385–395.

- (23) Lenaz, G.: Lipid fluidity and membrane protein dynamics. *Biosci. Rep.* **1987**, 7, 823–837.
- (24) Venable, R. M.; Zhang, Y.; Hardy, B. J.; Pastor, R. W.: Molecular dynamics simulations of a lipid bilayer and of hexadecane: an investigation of membrane fluidity. *Science* **1993**, 262, 223–236.
- (25) Ikonen, E.: Cellular cholesterol trafficking and compartmentalization. *Nat. Rev. Mol. Cell. Biol.* **2008**, 9, 125–138.
- (26) Yano, Y.; Takemoto, T.; Kobayashi, S.; Yasui, H.; Sakurai, H.; Ohashi, W.; Niwa, M.; Futaki, S.; Sugiura, Y.; Matsuzaki, K.: Topological stability and self-association of a completely hydrophobic model transmembrane helix in lipid bilayers. *Biochemistry* **2002**, 41, 3073–3080.

7 Chapter 7

Conclusions and Future Work

7.1 Introduction

2D IR spectroscopy is a powerful tool for studying the biophysics of a large variety of biomolecules. During my time in the Zanni group, I worked with many different peptides and proteins with unique properties. However, besides what we have already explored, there are other aspects of these biomolecules that we can examine with 2D IR spectroscopy. For example, we have studied the unfolding and aggregation of α B-crystallin with regard to cataract formation. Meanwhile, α B-crystallin is also known as a molecular chaperone that binds to misfolded proteins and amyloidogenic peptides. With isotope-labeled proteins and peptides, 2D IR can be used to probe the chaperone activity of α B-crystallin. In this chapter, I show the preliminary data on some of the ongoing projects and some interesting observations that may potentially lead to new projects. Different biomolecules are separated by sections. In Section 7.2, I will talk about α B-crystallin as molecular chaperone that binding to different proteins. In Section 7.3, I will discuss the aggregation of hIAPP fragments. Variations in a few amino acids can change the aggregation of these fragments dramatically. In Section 7.4, I will briefly talk about a small peptide A₆K that we examined with 2D IR. Last in Section 7.5, I will talk about a protein ApoA-1 that we worked on in collaboration with Dr. Giorgio Cavigliolo.

7.2 α B-crystallin as a molecular chaperone

7.2.1 Chaperone activity of α B-crystallin towards γ D-crystallin

We have been very interested in the chaperone activity of α B-crystallin towards γ D-crystallin and the role it plays in cataract formation. The binding of α B-crystallin to either misfolded or unfolded γ D-crystallin has been reported previously using technique like gel filtration chromatography.¹ α B-crystallin forms passive molecular chaperones that bind to misfolded proteins but cannot refold them. The binding can, nevertheless, slow down aggregation or prevent the amyloid fibers from elongating. However, to my knowledge an IR study that directly probes the protein-protein interaction has not been reported yet. In this chapter I will discuss my efforts to study the interaction between aggregated γ D-crystallin (¹³C labeled) and native α B-crystallin (¹²C) with 2D IR spectroscopy. By isotope labeling full length γ D-crystallin, it can be spectrally resolved from α B-crystallin when the two proteins are mixed in the solution. The hypothesis is that we can visualize the interaction between these two proteins through the cross peaks or other features on a 2D IR spectrum.

Figure 7.1A shows the 2D IR spectrum of a mixture of aggregated γ D-crystallin (uniformly ¹³C labeled) and native α B-crystallin (unlabeled). The amyloid fibers of γ D-crystallin were formed by incubation in phosphate buffer at pH 3.0. α B-crystallin also forms amyloid fibers when incubated at low pH so it is important to change the pH of the γ D-crystallin solution before adding α B-crystallin. Buffer with low buffering capacity (1-2 mM phosphate) can be used here to avoid changing the pH of the γ D-crystallin solution. When α B-crystallin solution (pH 7.4, 20 mM phosphate) is added to a smaller amount of 1 mM phosphate buffer at pH 3.0, the pH of the final solution should stay around 7.4. In Figure 7.1A, we observe a pair of diagonal peaks at $\omega_{\text{pump}} = 1580 \text{ cm}^{-1}$ corresponding to γ D-crystallin amyloid fibers and another pair at $\omega_{\text{pump}} = 1633 \text{ cm}^{-1}$

corresponding to unlabeled native α B-crystallin. Two pairs of off-diagonal peaks are the signatures of amyloid fibers, resulting from the vibrational mode at $\sim 1580\text{ cm}^{-1}$ coupling to a vibrational mode at $\sim 1643\text{ cm}^{-1}$. The diagonal peak from the 1643 cm^{-1} mode is too weak to see on the 2D IR spectrum.

Figure 7.1B shows a 2D IR spectrum of γ D-crystallin amyloid fiber without α B-crystallin. If there is no interaction between aggregated γ D-crystallin and native α B-crystallin, we should get a spectrum of α B-crystallin when subtracting spectrum (B) from (A). Figure 7.1C shows the subtraction result. Note that the subtraction is scaled so that there is still residual γ D-crystallin amyloid fiber intensity at $\sim 1580\text{ cm}^{-1}$. We also see clearly a pair of diagonal peaks at $\omega_{\text{pump}} = 1633\text{ cm}^{-1}$ that are from native α B-crystallin.

There are two things in this spectrum that may indicate protein-protein interaction. First, the cross peaks at $\omega_{\text{pump}} = 1643\text{ cm}^{-1}$ and $\omega_{\text{probe}} = 1580\text{ cm}^{-1}$ have opposite signs from the ones in Figure 7.1A and B, but the diagonal peaks at $\omega_{\text{pump}} = 1580\text{ cm}^{-1}$ have the same signs (colors of the peaks). This difference in signs indicates that the cross peaks in Figure 7.1A have lower intensity than those in Figure 7.1B relative to the diagonal peaks. One possible explanation for the lowered cross peak intensity is due to α B-crystallin binding to γ D-crystallin amyloid fiber, which decreases the amount of coupling between the two vibrational modes of γ D-crystallin.

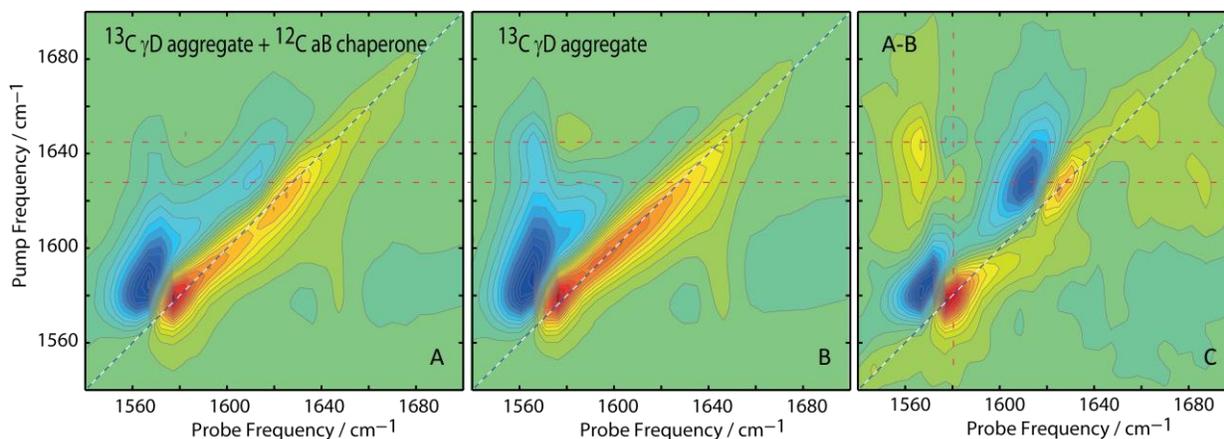


Figure 7.1 2D IR spectra of aggregated γ D-crystallin and native α B-crystallin.

(A) Aggregated γ D-crystallin with uniformly ^{13}C labeling and unlabeled native α B-crystallin. (B) Aggregated γ D-crystallin with ^{13}C labels. (C) Subtraction spectrum of not normalized (A)-(B).

The second interesting point to note in Figure 7.1C is the positive signal (orange) at $\omega_{\text{pump}} = 1633 \text{ cm}^{-1}$ and $\omega_{\text{probe}} = 1580 \text{ cm}^{-1}$. This signal is at the frequencies of γ D-crystallin amyloid fiber and native α B-crystallin, which makes it possibly a cross peak due to coupling between the two proteins. The sign (negative) of this cross peak is also correct because it should only be in (A) and not (B).

There is another way we can look for evidence of α B-crystallin binding to γ D-crystallin. Figure 7.2 shows the 2D IR spectra of aggregated γ D-crystallin and native α B-crystallin, similar to those shown in Figure 7.1. In Figure 7.2, the subtraction is scaled so that all the intensity of the γ D-crystallin amyloid fibers will be canceled. The result of this “complete” subtraction (Figure 7.2C) should be a spectrum of native α B-crystallin if there is no binding. Figure 7.2D is the 2D IR spectrum of native α B-crystallin that looks rather different from Figure 7.2C. In Figure 7.2C, the cross peaks with opposite signs from Figure 7.2B again indicate the weakened coupling between the two vibrational modes of γ D-crystallin amyloid fibers, similar to what we observe in Figure 7.1. This weakened coupling may be due to structural change of the amyloid fiber or due to the

interaction between the amyloid fibers with α B-crystallin. However, a structural change may also happen to α B-crystallin upon binding. Below Figure 7.2 C and D are the overlay of their 2D IR diagonal slices. The green solid curve corresponds to spectra C and the red dashed curve corresponds to spectra D. Note that the frequency of α B-crystallin redshifts to a lower frequency upon binding. Structurally, the redshift may be a result of the α B-crystallin core becoming more ordered when binding to other proteins. There is also a depletion of intensity at $\sim 1645\text{ cm}^{-1}$, which indicates a loss of random coil or α -helix structures from native α B-crystallin as a result of binding.

The preliminary data shown above is suggestive of α B-crystallin binding to aggregated γ D-crystallin in a possible chaperone mechanism. However, further experiments are necessary to confirm this assignment. Spectra subtraction is very sensitive to small errors in concentration and laser alignment, and so the spectra subtraction may not be perfect. Artifacts can also be generated during subtraction that influence our conclusions. These hypothesis should be validated and checked by repeating the experiment or complementary analytical techniques.

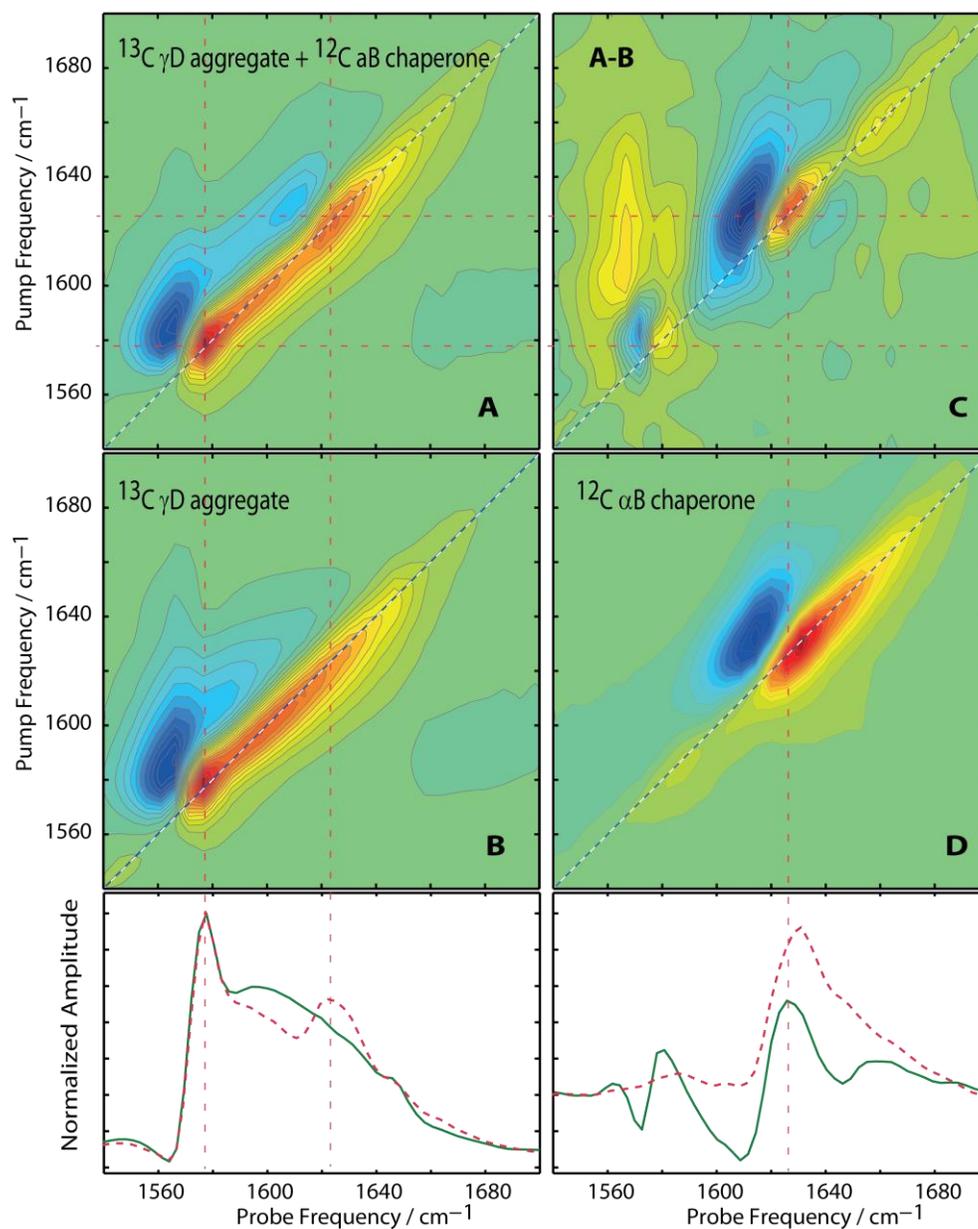


Figure 7.2 2D IR spectra of aggregated γ D-crystallin, native α B-crystallin, and their subtraction result.

2D IR spectrum of (A) 1:1 ratio mix of aggregated ^{13}C γ D-crystallin and native α B-crystallin, (B) aggregated ^{13}C γ D-crystallin with same concentration in (A), and (D) and native α B-crystallin. (C) is the subtraction spectrum of (B) from (A). The diagonal slices of spectra A and C are shown in green lines below the spectra. Slices of spectra B and D are shown in red dashed lines.

7.2.2 Inhibition of hIAPP fiber elongation by α B-crystallin chaperone

Inhibition of fiber elongation by α B-crystallin chaperone has been reported for vimentin, actin,² β -amyloid,³ α -synuclein,⁴ κ -Casein,⁵ and unfolded γ -crystallin.⁶ The research field is very excited about the therapeutic potential of α B-crystallin because α B-crystallin naturally exists in many parts of the body where amyloid fibers are found, for example in the brain and the heart.⁷ However, there are many unsolved questions about the chaperone activity of α B-crystallin. The binding specificity of α B-crystallin is still not well known. It is still under debate whether α B-crystallin binds to unfolded structures or aggregated structures, or both. There have been several studies trying to identify the binding interface of α B-crystallin or the origin of their chaperone activities, but the conclusions varies depending on the system.⁸⁻¹¹ It turns out that the chaperone activity of α B-crystallin might be more complicated than we thought. Different interfaces may be involved when the chaperone is binding to different structures.¹² I am interested in the inhibition of hIAPP fiber elongation by α B-crystallin, which has not been reported previously. This study will provide insights into the previously mentioned questions and allow us to gain more understanding of the chaperone activity of α B-crystallin.

Figure 7.3 shows the kinetics of hIAPP aggregation with and without α B-crystallin. With α B-crystallin, the aggregation of hIAPP is greatly suppressed. Using 2D IR spectroscopy, we can observe not only the kinetics of hIAPP aggregation but structural changes as well. Figure 7.4 shows the 2D IR spectra of aggregated hIAPP (uniformly ^{13}C labeled) and native α B-crystallin (unlabeled). The hypothesis is that we might be able to observe the interaction between ^{13}C hIAPP and ^{12}C α B-crystallin from the cross peaks on a 2D IR spectrum. A waiting time experiment allows us to see how the spectra evolves as we increase the waiting time between the pump and the probe pulses. As the waiting time increases (within a certain range), the intensity of the diagonal peaks

diminish and the cross peaks grow in (Figure 7.4). An ideal experimental design includes 1) a waiting time 2D IR measurement for a mixture of ^{13}C hIAPP and ^{12}C αB -crystallin, 2) ^{13}C hIAPP, and 3) ^{12}C αB -crystallin. We would then subtract 2) and 3) from 1) to see the changes of the spectra at each waiting time point. We would also run this experiment along the kinetics curve of hIAPP to examine what structure αB -crystallin binds to (monomer, oligomer, or amyloid fibers). A series of inhibition experiments can be performed with different concentrations of αB -crystallin while keeping the same concentration for hIAPP to test the binding capacity of αB -crystallin.

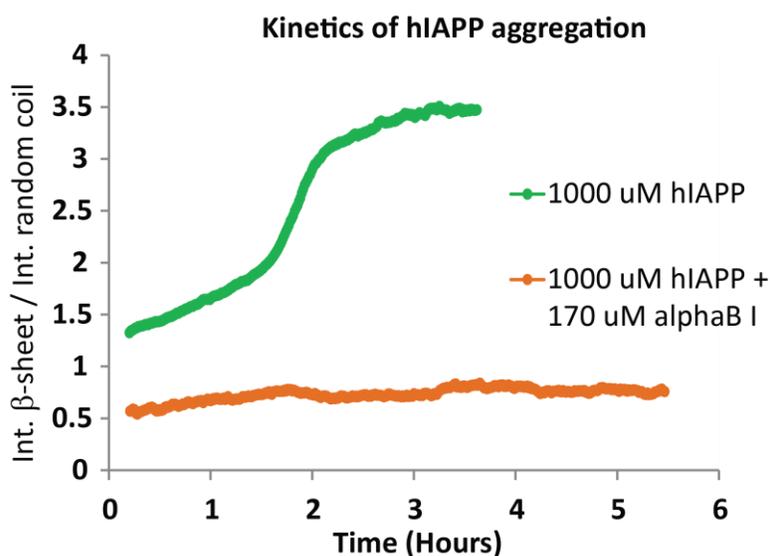


Figure 7.3 Kinetics of hIAPP aggregation with and without αB -crystallin.

The kinetics of 1 mM hIAPP is plotted with green curve, and that of 1 mM hIAPP with 170 μM native αB -crystallin is plotted in orange curve. The kinetics is calculated as the ratio of the amyloid fiber peak ($\sim 1620\text{ cm}^{-1}$) over the random coil peak ($\sim 1650\text{ cm}^{-1}$).

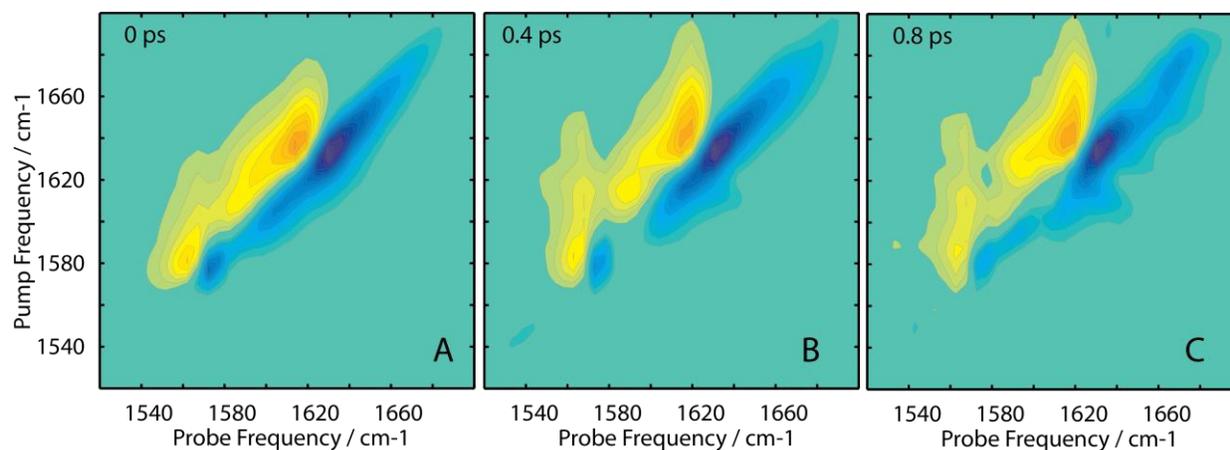


Figure 7.4 Waiting time 2D IR spectra of aggregated hIAPP and α B-crystallin. 2D IR spectra of aggregated hIAPP (uniformly ^{13}C labeled) and native α B-crystallin (unlabeled) at different waiting times: (A) 0 ps, (B) 0.4 ps, and (C) 0.8 ps.

7.2.3 α B-crystallin chaperone dynamics

α B-crystallin forms oligomers in solution. These oligomers are dynamic structures and can exchange subunits that affect the chaperone activity of α B-crystallin.^{13,14} 2D IR can be used to measure this dynamics because of its temporal resolution. The α B-crystallin proteins in a molecular chaperone will have to be spatially close to each other, so they might be vibrationally coupled. By mixing ^{13}C and ^{12}C α B-crystallin proteins in the same solution, we can observe their interaction with 2D IR spectroscopy. A temperature dependent waiting time experiment can be used to measure the coupling between α B-crystallin proteins. Three different samples should be measured at each waiting time: 1) a 1:1 mixture of ^{13}C and ^{12}C α B-crystallin, 2) ^{13}C α B-crystallin, and 3) ^{12}C α B-crystallin. Subtraction of 2) and 3) from 1) allows us to observe difference in the spectra. Cross peaks between the labeled and unlabeled crystallin proteins are more likely to be observed at high temperature (above 25 °C and below 50°C) and longer waiting times.

7.3 Fragments of hIAPP

7.3.1 FGAIL fragment

Although the FGAIL residues are important in hIAPP amyloid fiber formation, I found that the amidated FGAIL fragment by itself either does not aggregate or aggregates very slowly in solution. Figure 7.5 shows the 2D IR spectrum and diagonal slice of amidated FGAIL fragment after overnight incubation in 20 mM Tris buffer, pH 7.4.

FGAIL fragments without C-terminal amidation (Free-FGAIL) do not aggregate in solution either. Figure 7.6 shows the 2D IR spectrum and diagonal slice of Free-FGAIL fragment in 20 mM Tris buffer, pH 7.4. The spectrum of Free-FGAIL fragments has an additional diagonal peak around 1595 cm^{-1} , resulting from the COO^- on the C-terminal domain. See Appendix 4 for the storage locations of these peptides.

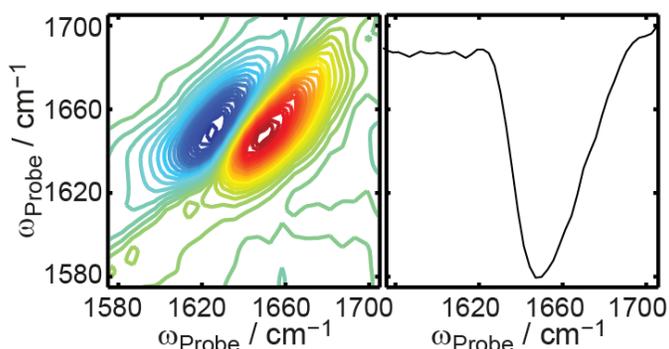


Figure 7.5 2D IR spectrum of amidated FGAIL fragment in D_2O buffer

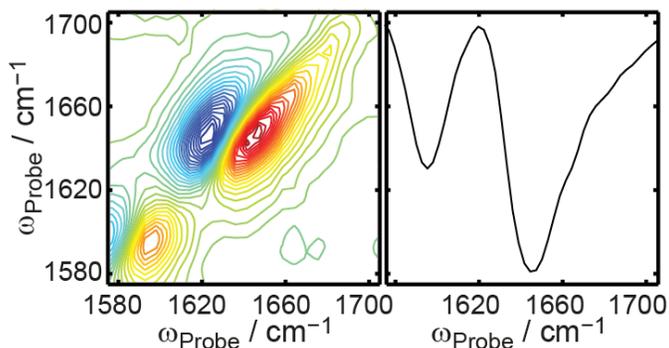


Figure 7.6 2D IR spectrum of FGAIL fragment in D_2O buffer

7.3.2 CFGAILSS fragment

Another fragment of hIAPP was made for 2D SFG studies. We added a cysteine on the N-terminus of the peptide so that it can be used to anchor the peptide on a gold surface. With the 3 extra amino acids, CFGAILSS behaves very differently from the FGAIL fragments. CFGAILSS forms amyloid fibers readily in solution at pH 7.4 (Figure 7.7). We can observe similar cross peaks in Figure 7.7 as the ones in full length hIAPP.

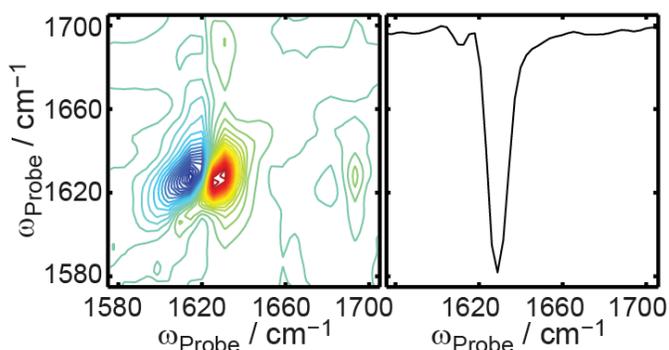


Figure 7.7 2D IR spectrum of CFGAILSS fragment in D₂O buffer

7.3.3 NNFGAILSS fragment

Another interesting fragment of hIAPP is amidated NNFGAILSS. Unlike in CFGAILSS, where we had to introduce a “mutation” (Cys) into the original hIAPP sequence, NNFGAILSS is the same as in the hIAPP sequence. However, this fragment forms aggregates that are very different from the aggregates formed by CGFAILSS or the full length hIAPP. Figure 7.8 shows the 2D IR spectra of NNFGAILSS in its random coil form (A) and its amyloid form (B). Note in Figure 7.8B, the cross peaks are formed at 1620 cm⁻¹ and another vibrational mode ~1665 cm⁻¹, which is much lower in frequency than the vibrational mode in full length hIAPP. We are not sure what the structure is of this aggregate but it is very interesting. It is worth investigating how

CFGAILSS and NNFGAILSS form aggregates that are so different in structure while they are only two amino acids different.

I did not observe the kinetics for CFGAILSS aggregation because the peptide was already all aggregated during the 5 min dead time between initiating the aggregation and measuring its 2D IR spectrum. On the other hand, the aggregation of NNFGAILSS is interesting because it follows a very similar kinetics as full length hIAPP (Figure 7.9). The kinetics of aggregation is plotted as the intensities at $\sim 1650\text{ cm}^{-1}$ and 1622 cm^{-1} .

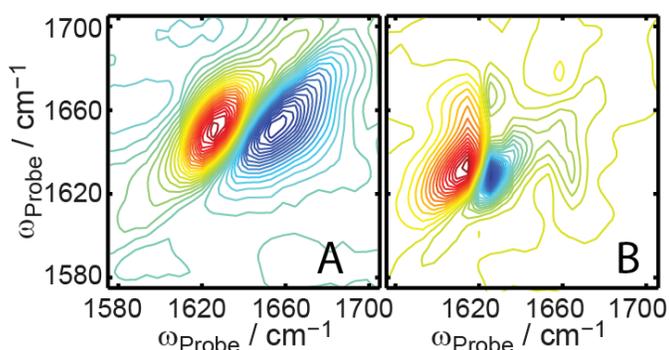


Figure 7.8 2D IR spectrum of NNFGAILSS fragment in D₂O buffer

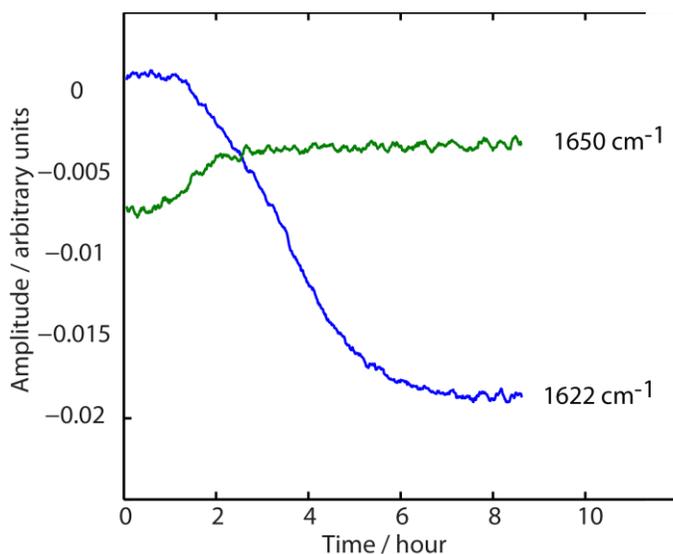


Figure 7.9 Kinetics of NNFGAILSS aggregation

The kinetics of NNFGAILSS aggregation is plotted as the intensity at $\sim 1622\text{ cm}^{-1}$ for amyloid fiber and $\sim 1650\text{ cm}^{-1}$ for random coil.

7.4 Small peptide A₆K

Small peptides constructed with several hydrophobic amino acids and a polar C-terminal amino acid can have properties that are similar to amphiphilic surfactants. These small peptides are also found to form self-assembled nanofibers that can have various applications. The commonly studied amphiphilic peptides include A₉K, V₆K, L₆K₂, and A₆K. We synthesized A₆K (Acetyl-A₆K-NH₂) and measured it in solution with 2D IR spectroscopy. What I found is that the formation of A₆K nanofibers is highly dependent on the pH of the aqueous buffer. Figure 7.10 shows the 2D IR spectra of A₆K peptide in D₂O buffer with different pD. There is a faint peak at around 1610 cm⁻¹ that should be investigated further.

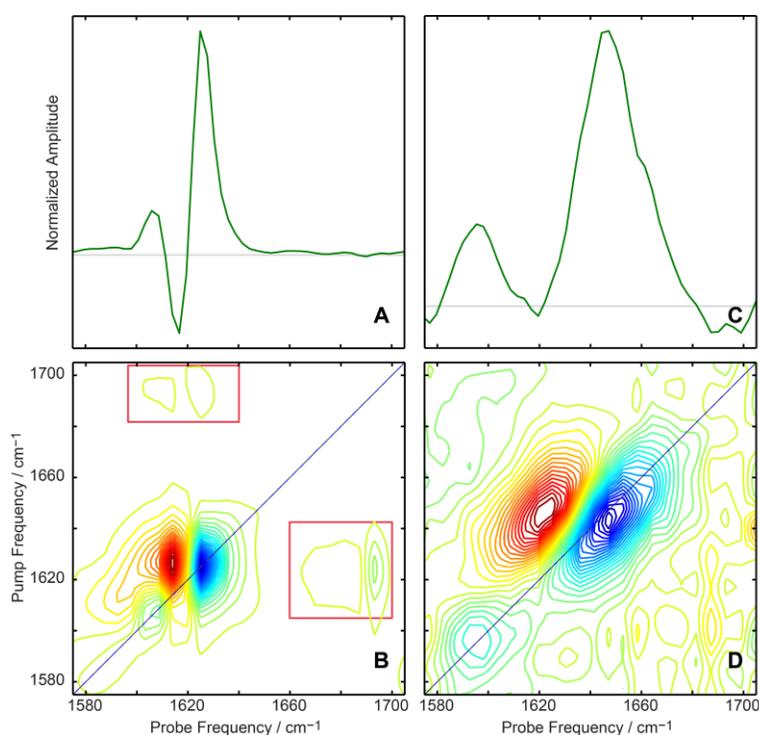


Figure 7.10 2D IR spectra of A₆K in aqueous buffers.

(A) 2D IR diagonal slice of A₆K in pD 7.4 phosphate buffer, (B) 2D IR spectrum of A₆K in pD 7.4 phosphate buffer, (C) 2D IR diagonal slice of A₆K in pD 10.0 phosphate buffer, and (D) 2D IR spectrum of A₆K in pD 10.0 phosphate buffer. (Sample concentration is estimated to be around 10 mM).

Note that the A₆K peptides do not contain any amino acid that can be used for quantifying peptide concentration by absorbance at 280 nm with Nanodrop. The concentration of the peptide samples for 2D IR measurement (Figure 7.10) is estimated using the mass of the peptide measured on a scale, which is not very accurate. I recommend measuring the peptide concentration by running a Bradford protein assay.

The A₆K peptide can be treated similar to hIAPP. It is most preferable to store the peptide as dried powder in a cold and dry environment. For short term storage, the peptide can be dissolved in HFIP-d as a stock solution. A small amount (1-2 μ L) of stock solution, with known concentration, should be dried under N₂ and dissolved in D₂O buffer to obtain the desired peptide concentration.

7.5 Apolipoprotein A1

I worked with Dr. Giorgio Cavigliolo on the amyloid formation of ApoA-1 (Apolipoprotein A1) due to oxidation caused by myeloperoxidase (an enzymes in the blood) (See Appendix 5. List of publication). We have a variety of the samples in stock that can be used for 2D IR measurements (See Appendix 4.2) and future collaborations.

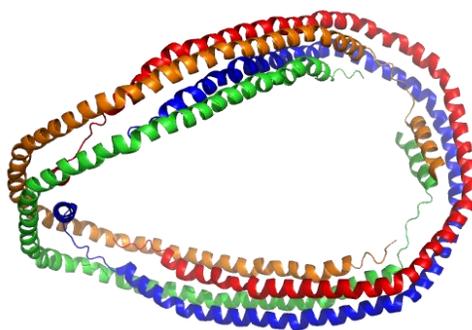


Figure 7.11 Crystal structure of Apolipoprotein A1 tetramer (PDB: 1AVA)

ApoA-1 is a large protein with 396 residues. The structure of an ApoA-1 tetramer is shown in

Figure 7.11 with colored monomers. Interestingly, in the native state, this protein is consisted of mostly all α -helices, which can be used as a standard protein for α -helix transition dipole calculations.¹⁵ A 2D IR spectrum of native ApoA-1 is shown in Figure 7.12. Amyloid fibers of apoA-1 are found in atherosclerotic plaques, which can be a potential subject for 2D IR tissue studies.

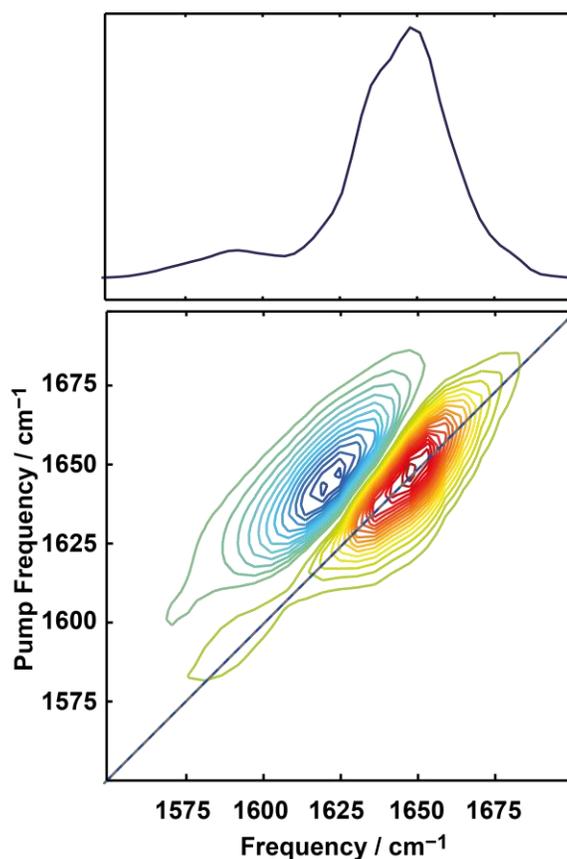


Figure 7.12 2D IR spectrum of native ApoA-1

7.6 References

- (1) Moreau, K. L., and King, J. A. (2012) Cataract-causing defect of a mutant gamma-crystallin proceeds through an aggregation pathway which bypasses recognition by the alpha-crystallin chaperone, *PLoS one* **7**, e37256.
- (2) Nagaraj, R. H., Nahomi, R. B., Mueller, N. H., Raghavan, C. T., Ammar, D. A., and Petrash, J. M. (2015) Therapeutic potential of alpha-crystallin, *Biochimica et biophysica acta*.
- (3) Raman, B., Ban, T., Sakai, M., Pasta, S. Y., Ramakrishna, T., Naiki, H., Goto, Y., and Rao Ch, M. (2005) AlphaB-crystallin, a small heat-shock protein, prevents the amyloid fibril growth of an amyloid beta-peptide and beta2-microglobulin, *The Biochemical journal* **392**, 573-581.
- (4) Waudby, C. A., Knowles, T. P., Devlin, G. L., Skepper, J. N., Ecroyd, H., Carver, J. A., Welland, M. E., Christodoulou, J., Dobson, C. M., and Meehan, S. (2010) The interaction of alphaB-crystallin with mature alpha-synuclein amyloid fibrils inhibits their elongation, *Biophysical journal* **98**, 843-851.
- (5) Thorn, D. C., Meehan, S., Sunde, M., Rekas, A., Gras, S. L., MacPhee, C. E., Dobson, C. M., Wilson, M. R., and Carver, J. A. (2005) Amyloid fibril formation by bovine milk kappa-casein and its inhibition by the molecular chaperones alphaS- and beta-casein, *Biochemistry* **44**, 17027-17036.
- (6) Acosta-Sampson, L., and King, J. (2010) Partially folded aggregation intermediates of human gammaD-, gammaC-, and gammaS-crystallin are recognized and bound by human alphaB-crystallin chaperone, *Journal of molecular biology* **401**, 134-152.
- (7) Reddy, V. S., and Reddy, G. B. (2015) Emerging Role for α B-Crystallin as a Therapeutic Agent: Pros and Cons, *Current Molecular Medicine 2015* **15**, 47-61.
- (8) Kannan, R., Santhoshkumar, P., Mooney, B. P., and Sharma, K. K. (2013) The alphaA66-80 peptide interacts with soluble alpha-crystallin and induces its aggregation and precipitation: a contribution to age-related cataract formation, *Biochemistry* **52**, 3638-3650.
- (9) Delbecq, S. P., Jehle, S., and Klevit, R. (2012) Binding determinants of the small heat shock protein, alphaB-crystallin: recognition of the 'IxI' motif, *The EMBO journal* **31**, 4587-4594.
- (10) Ghosh, J. G., Estrada, M. R., and Clark, J. I. (2005) Interactive domains for chaperone activity in the small heat shock protein, human alphaB crystallin, *Biochemistry* **44**, 14854-14869.
- (11) Bhattacharyya, J., Padmanabha Udupa, E. G., Wang, J., and Sharma, K. K. (2006) Mini-alphaB-crystallin: a functional element of alphaB-crystallin with chaperone-like activity, *Biochemistry* **45**, 3069-3076.
- (12) Mainz, A., Peschek, J., Stavropoulou, M., Back, K. C., Bardiaux, B., Asami, S., Prade, E., Peters, C., Weinkauff, S., Buchner, J., and Reif, B. (2015) The chaperone alphaB-crystallin uses different interfaces to capture an amorphous and an amyloid client, *Nature structural & molecular biology* **22**, 898-905.

- (13) Koteiche, H. A., and Mchaourab, H. S. (2002) The determinants of the oligomeric structure in Hsp16.5 are encoded in the alpha-crystallin domain, *FEBS letters* **519**, 16-22.
- (14) Sreelakshmi, Y., and Sharma, K. K. (2005) Recognition sequence 2 (residues 60-71) plays a role in oligomerization and exchange dynamics of alphaB-crystallin, *Biochemistry* **44**, 12245-12252.
- (15) Grechko, M., and Zanni, M. T. (2012) Quantification of transition dipole strengths using 1D and 2D spectroscopy for the identification of molecular structures via exciton delocalization: application to alpha-helices, *The Journal of chemical physics* **137**, 184202.

A1. Appendix 1

Supporting Information for Chapter 3: Amyloid fiber formation in human γ D-crystallin induced by UV-B photodamage

		++	*	*++	0	0++	eq.		++	*	*++	0	0++	2
	48.0757	4.5415												3
	62.1186	31.5629	45.0921	23.0497				635.315	318.161	618.289	309.648	617.305	309.156	2
	91.1612	96.0842	74.1347	87.571	73.1506	87.079		521.272	261.14	504.246	252.626	503.262	252.134	1
	04.2453	52.6263	87.2187	44.113	86.2347	43.621		392.23	196.618	375.203	188.105	374.219	187.613	0
	41.3042	21.1557	24.2776	12.6425	23.2936	12.1504		279.146	140.076	262.119	131.563	261.135	131.071	9
	28.3362	64.6717	11.3097	56.1585	10.3256	55.6665		142.087	071.547	125.06	063.034	124.076	062.542	8
	41.4203	21.2138	24.3937	12.7005	23.4097	12.2085		055.055	028.031	038.028	019.518	037.044	019.026	7
	55.4632	78.2352	38.4367	69.722	37.4526	69.23		941.971	71.4889	924.944	62.9756	923.96	62.4836	6
	054.532	27.7694	037.505	19.2562	036.521	18.7642		827.928	14.4674	810.901	05.9542	809.917	05.4621	5
0	167.616	84.3115	150.589	75.7982	149.605	75.3062		728.859	64.9332	711.833	56.4199	710.849	55.9279	4
1	296.658	48.8328	279.632	40.3195	278.648	39.8275		615.775	08.3912	598.749	99.8779	597.765	99.3859	3
2	353.68	77.3435	336.653	68.8302	335.669	68.3382		486.733	43.8699	469.706	35.3566	468.722	34.8646	2
3	440.712	20.8595	423.685	12.3462	422.701	11.8542		429.711	15.3592	412.685	06.8459	411.701	06.3539	1
4	626.791	13.8992	609.765	05.3859	608.781	04.8939		342.679	71.8431	325.653	63.3299	324.668	62.8379	0
5	725.86	63.4334	708.833	54.9201	707.849	54.4281		156.6	78.8035	139.573	70.2902	138.589	69.7982	
6	838.944	19.9754	821.917	11.4621	820.933	10.9701		057.531	29.2693	040.505	20.756	039.521	20.264	
7	002.007	001.507	984.98	92.9938	983.996	92.5018		44.4472	72.7272	27.4207	64.214	26.4367	63.722	
8	131.05	066.028	114.023	057.515	113.039	057.023		81.3839	91.1956	64.3573	82.6823	63.3733	82.1903	
9	244.134	122.57	227.107	114.057	226.123	113.565		52.3413	26.6743	35.3148	18.161	34.3307	17.669	
0	331.166	166.086	314.139	157.573	313.155	157.081		39.2572	70.1323	22.2307	61.619	21.2467	61.127	
1	445.209	223.108	428.182	214.595	427.198	214.103		52.2252	26.6162	35.1987	18.103			
2	608.272	304.64	591.245	296.126	590.261	295.634		38.1823	69.5948	21.1557	61.0815			
3								75.119	8.0631	58.0924	9.5498			

Table A1. 1

Fragment ions in the CID spectrum of parent peptide FNEIHSLSNVLEGSWVLYELSNYR (2781.39, +2), corresponding to Figure 3.2A. Assigned fragments used in sequence analysis are shown in red.

		++	*	*++	0	0++	eq.		++	*	*++	0	0++	2
	48.0757	4.5415												3
	62.1186	31.5629	45.0921	23.0497				667.305	334.156	650.278	325.643	649.294	325.151	2
	91.1612	96.0842	74.1347	87.571	73.1506	87.079		553.262	277.135	536.236	268.621	535.251	268.129	1
	04.2453	52.6263	87.2187	44.113	86.2347	43.621		424.219	212.613	407.193	204.1	406.209	203.608	0
	57.2991	29.1532	40.2726	20.6399	39.2885	20.1479		311.135	156.071	294.109	147.558	293.125	147.066	9
	44.3311	72.6692	27.3046	64.1559	26.3206	63.6639		158.082	079.544	141.055	071.031	140.071	070.539	8
	57.4152	29.2112	40.3886	20.698	39.4046	20.206		071.05	036.028	054.023	027.515	053.039	027.023	7
	71.4581	86.2327	54.4316	77.7194	53.4476	77.2274		957.965	79.4863	940.939	70.9731	939.955	70.4811	6
	070.527	35.7669	053.5	27.2536	052.516	26.7616		843.923	22.4649	826.896	13.9516	825.912	13.4596	5
0	183.611	92.3089	166.584	83.7957	165.6	83.3037		744.854	72.9307	727.828	64.4174	726.844	63.9254	4
1	312.653	56.8302	295.627	48.317	294.643	47.825		631.77	16.3886	614.744	07.8754	613.76	07.3834	3
2	369.675	85.341	352.648	76.8277	351.664	76.3357		502.727	51.8674	485.701	43.3541	484.717	42.8621	2
3	456.707	28.857	439.68	20.3437	438.696	19.8517		445.706	23.3566	428.679	14.8433	427.695	14.3513	1
4	658.781	29.8941	641.754	21.3808	640.77	20.8888		358.674	79.8406	341.647	71.3273	340.663	70.8353	0
5	757.849	79.4283	740.823	70.915	739.839	70.423		156.6	78.8035	139.573	70.2902	138.589	69.7982	
6	870.933	35.9703	853.907	27.4571	852.923	26.9651		057.531	29.2693	040.505	20.756	039.521	20.264	
7	033.997	017.502	016.97	008.989	015.986	008.497		44.4472	72.7272	27.4207	64.214	26.4367	63.722	
8	163.039	082.023	146.013	073.51	145.029	073.018		81.3839	91.1956	64.3573	82.6823	63.3733	82.1903	
9	276.123	138.565	259.097	130.052	258.113	129.56		52.3413	26.6743	35.3148	18.161	34.3307	17.669	
0	363.155	182.081	346.129	173.568	345.145	173.076		39.2572	70.1323	22.2307	61.619	21.2467	61.127	
1	477.198	239.103	460.172	230.59	459.188	230.098		52.2252	26.6162	35.1987	18.103			
2	640.262	320.635	623.235	312.121	622.251	311.629		38.1823	69.5948	21.1557	61.0815			
3								75.119	8.0631	58.0924	9.5498			

Table A1. 2

Table S2. Fragment ions in the CID spectrum of parent peptide FNEIHSLNVLEGSWVLYELSNYR (2813.39, 2 Ox., +3), corresponding to Figure 3.2B. Assigned fragments used in sequence analysis are shown in red.

		++	*	*++	0	0++	eq.		++	*	*++	0	0++	2
	48.0757	4.5415												4
	62.1186	31.5629	45.0921	23.0497				497.733	49.3703	480.707	40.857	479.723	40.365	3
	91.1612	96.0842	74.1347	87.571	73.1506	87.079		383.69	92.3488	366.664	83.8355	365.68	83.3435	2
	04.2453	52.6263	87.2187	44.113	86.2347	43.621		254.648	27.8275	237.621	19.3142	236.637	18.8222	1
	41.3042	21.1557	24.2776	12.6425	23.2936	12.1504		141.564	71.2855	124.537	62.7722	123.553	62.2802	0
	28.3362	64.6717	11.3097	56.1585	10.3256	55.6665		004.505	02.756	87.4782	94.2427	86.4942	93.7507	
	41.4203	21.2138	24.3937	12.7005	23.4097	12.2085		17.4727	59.24	00.4462	50.7267	99.4621	50.2347	
	55.4632	78.2352	38.4367	69.722	37.4526	69.23		04.3886	02.698	87.3621	94.1847	86.3781	93.6927	
	054.532	27.7694	037.505	19.2562	036.521	18.7642		90.3457	45.6765			72.3352	36.6712	
0	167.616	84.3115	150.589	75.7982	149.605	75.3062		91.2773	96.1423			73.2667	87.137	
1	296.658	48.8328	279.632	40.3195	278.648	39.8275		78.1932	39.6003			60.1827	30.595	
2	353.68	77.3435	336.653	68.8302	335.669	68.3382		49.1506	75.079			31.1401	66.0737	
3	440.712	20.8595	423.685	12.3462	422.701	11.8542		92.1292	46.5682			74.1186	37.5629	
4								05.0972	03.0522					

Table A1. 3

Fragment ions in the CID spectrum of parent peptide FNEIHSLNVLEGSW (1643.81, +2), corresponding to Figure 3.2C. Assigned fragments used in sequence analysis are shown in red.

		++	*	*++	0	0++	eq.		++	*	*++	0	0++	2
	00.0757	0.5415												
	13.1598	07.0835						057.531	29.2693	040.505	20.756	039.521	20.264	
	76.2231	88.6152						44.4472	72.7272	27.4207	64.214	26.4367	63.722	
	05.2657	53.1365			87.2551	44.1312		81.3839	91.1956	64.3573	82.6823	63.3733	82.1903	
	18.3497	09.6785			00.3392	00.6732		52.3413	26.6743	35.3148	18.161	34.3307	17.669	
	05.3818	53.1945			87.3712	44.1892		39.2572	70.1323	22.2307	61.619	21.2467	61.127	
	19.4247	10.216	02.3981	01.7027	01.4141	01.2107		52.2252	26.6162	35.1987	18.103			
	82.488	91.7477	65.4615	83.2344	64.4775	82.7424		38.1823	69.5948	21.1557	61.0815			
								75.119	8.0631	58.0924	9.5498			

Table A1. 4

Fragment ions in the CID spectrum of parent peptide VLYELSNYR (1155.60, +2), corresponding to Figure 3.2D. Assigned fragments used in sequence analysis are shown in red.

		++	*	*++	0	0++	eq.		++	*	*++	0	0++	2
	87.0866	4.0469												0
	86.1559	43.5811						156.6	78.8035	139.573	70.2902	138.589	69.7982	
	99.2391	00.1232						057.531	29.2693	040.505	20.756	039.521	20.264	
	62.3024	81.6548						44.4472	72.7272	27.4207	64.214	26.4367	63.722	
	91.345	46.1761			73.3344	37.1709		81.3839	91.1956	64.3573	82.6823	63.3733	82.1903	
	04.4291	02.7182			86.4185	93.7129		52.3413	26.6743	35.3148	18.161	34.3307	17.669	
	91.4611	46.2342			73.4505	37.2289		39.2572	70.1323	22.2307	61.619	21.2467	61.127	
	005.504	03.2556	88.4775	94.7424	87.4934	94.2504		52.2252	26.6162	35.1987	18.103			
	168.567	84.7873	151.541	76.274	150.557	75.782		38.1823	69.5948	21.1557	61.0815			
0								75.119	8.0631	58.0924	9.5498			

Table A1. 5

Fragment ions in the CID spectrum of parent peptide WVLVELSNYR (1341.68, +2), corresponding to Figure 3.2E. Assigned fragments used in sequence analysis are shown in red.

		++	*	*++	0	0++	eq.		++	*	*++	0	0++	2
	03.0815	02.0444												0
	02.1499	51.5786						156.6	78.8035	139.573	70.2902	138.589	69.7982	
	15.234	08.1206						057.531	29.2693	040.505	20.756	039.521	20.264	
	78.2973	89.6523						44.4472	72.7272	27.4207	64.214	26.4367	63.722	
	07.3399	54.1736			89.3293	45.1683		81.3839	91.1956	64.3573	82.6823	63.3733	82.1903	
	20.424	10.7156			02.4134	01.7103		52.3413	26.6743	35.3148	18.161	34.3307	17.669	
	07.456	54.2316			89.4454	45.2264		39.2572	70.1323	22.2307	61.619	21.2467	61.127	
	021.499	11.2531	004.472	02.7398	003.488	02.2478		52.2252	26.6162	35.1987	18.103			
	184.562	92.7848	167.536	84.2715	166.552	83.7795		38.1823	69.5948	21.1557	61.0815			
0								75.119	8.0631	58.0924	9.5498			

Table A1. 6

Fragment ions in the CID spectrum of parent peptide WVLVELSNYR (1357.67, 1 Ox., +2), corresponding to Figure 3.2E. Assigned fragments used in sequence analysis are shown in red.

A2. Appendix 2

Supporting Information for Chapter 5: Amyloid Identified in UV-Induced Cataracts of Porcine Lenses using 2D IR Spectroscopy

A2.1 Supporting Material and Methods

A2.1.1 Purification of recombinant γ D-Crystallin

γ D-crystallin proteins were expressed in BL21 (DE3) PlysS *E. coli* with a His-tag on the N-terminus in a Pet32 vector. A 10 mL starter was grown from a single colony and then used to grow 1 L of LB at 37 °C. Protein expression was induced with 2 mM of IPTG when OD₆₀₀ reached 0.6. Cells were harvested by centrifugation at 5,000 x G after 2.5 hours of incubation at 37 °C and then lysed using a French press. Lysate was obtained by spinning at 40,000 x G for 30 min and purified using a Ni Sepharose HisTrap HP column (GE). Proteins were then dialyzed into the correct buffer, lyophilized, and then stored at -20 °C.

A2.1.2 Purification of recombinant α B-Crystallin

Human α B-crystallin pAED4 plasmid was a gift from Jonathan King at MIT University. The plasmid was incorporated into BL21 (DE3) PlysS *E. coli* cells for expression. An overnight starter culture of 10 mL was grown from a single colony and was used to inoculate 1L of LB at 37 °C. Overexpression of protein was then induced by addition of IPTG (final concentration 1 mM) when OD₆₀₀ reached 0.6. The cells were allowed to growth overnight at 18 °C and were harvested the second day by 10 min centrifugation at 5,000 G. The Cell pellets were then resuspended in 40

mM Tris buffer, pH 8.0, and immediately crushed with French Pressure cell press at 4 °C, followed by centrifugation at 50,000 x G for one hour. The lysate was then loaded onto GE healthcare HiPrep Q HP ion exchange columns. Protein was washed with buffer A (20 mM Tris, pH 8.5), and then eluted with 10-20% Buffer B (20 mM Tris and 1.0 M NaCl, pH 8.5). The eluted protein was then concentrated with spin columns and purified further with size-exclusion columns. Purified protein was dialyzed into phosphate buffer before lyophilized and stored at -20 °C.

A2.1.3 Slices of porcine lenses

Porcine eyeballs are supplied by the Meat Science and Muscle Biology Laboratory, Department of Animal Science at University of Wisconsin-Madison. The lenses are extracted from the eyeballs within 2 hours of TOD of the animals. The lenses were stored at -20 °C in embedding medium for frozen tissue (Tissue-Tek, content: 10% polyvinyl alcohol, 4% polyethylene glycol) until they were sliced using a microtome at -26 °C to an optimal thickness of 25 μm (for 2D IR and ThT fluorescence assay) and 60 μm (for TEM samples). The slices were deposited directly onto a CaF₂ window.

A2.1.4 2D IR spectra of protein solutions

The proteins were deuterated by D₂O and lyophilized before being dissolved in pD 7.4 buffer. Protein concentration was measured by a NanoDrop spectrometer and diluted to the desired concentrations. 2 μL of protein solution was deposited between two CaF₂ windows with a 56-μm Teflon spacer. Each spectrum presented in this paper was averaged between 25 to 40 mins.

A2.1.5 2D IR spectra of lens slices

Lens slices are deuterated with D₂O and dried under gentle N₂ for three times before buffer was added. The sample cell was assembled with a 56-μm spacer and another CaF₂ window. Thus, the lens slice samples had the same path length as the protein solution samples.

A2.1.6 FTIR spectra of lens slices

The FTIR spectra were collected for the same samples using Thermo Scientific Nicolet iS10 model. Each spectrum was averaged for about one minute.

A2.1.7 UV-irradiation of protein solution and lens tissues

The samples were exposed to four 8 W medium pressure mercury UV-B lamps (spectrum shown in section 2). The distance between the samples and the lamps was approximately 30 cm, which corresponds to a power intensity of 35 W/m^2 measured by a power meter. The UV exposure condition was the same for protein solutions and lens slice samples. The exposure was interrupted when the 2D IR spectra were collected for these samples. For TEM, the 60 μm slice were UV irradiated for 90 hrs to account for the difference in the thicknesses.

A2.1.8 Acid treatment of protein solution and lens tissues.

In vitro protein aggregation was initiated by the addition of pD 3.0 buffer. For lens tissues, pD 2.0 buffer was used because the high protein concentration in lens slices functions as a buffer systems itself. Unlike UV-irradiation experiments where the samples cannot be measured during the UV treatment, acid induced samples can be monitored constantly.

A2.1.9 ThT fluorescence essay of lens slices

25 μm thick lens tissues were deposited onto a coverslip. A hybridization chamber (9.8 mm x 20 mm by Grace Bio-labs) is attached onto the coverslip so that the lens tissue is placed in the center of the chamber. Lens slices were irradiated same way as the 2D IR samples before the addition of 100 μL phosphate buffer containing 25 μM ThT at pH 7.4. ThT fluorescence was measured with a 405 nm laser for excitation and a 482 nm filter. For each sample, a phase contrast image is recorded with correct focus before switching to the fluorescence mode.

A2.1.10 TEM of lens tissues and protein solutions

Lens tissues (60 μm) on coverslips were fixed in a solution of 2.5% glutaraldehyde, 2% paraformaldehyde in 0.1 M sodium phosphate buffer, pH 7.1 for 1 hour at room temperature. The samples were then post fixed with 1% Osmium Tetroxide after intensive rinsing. Dehydration of samples was performed with ethanol series and then transitioned in propylene oxide. The dehydrated samples were then infiltrated and embedded with Durcupan ACM (Fluka AG, Switzerland) resin in open aluminum weighing dishes at 60°C overnight. The glass coverslip was then removed with concentrated hydrofluoric acid to reveal the embedded cell culture. 100 nm ultrathin samples were sectioned parallel to the former coverslip surface using a Leica EM UC6 ultramicrotome. The ultrathin sections were collected on pioloform coated 300 mesh Cu thin-bar and 2x1 Cu slot grids and contrasted with Reynolds lead citrate and 8% uranyl acetate in 50% ethanol. For solution TEM samples, protein solutions were evaporated on a grid before addition of uranyl acetate. Data were collected with a Philips CM120 TEM, and images were documented with an AMT BioSprint12 (Woburn, MA) side mounted digital camera.

A2.2 Calculation of UV dosage from UV lamp in lab and its equivalent to the average sun exposure per person.

In lab, a bank of four 8W medium pressure mercury UV-B lamps is used as the source of UV-irradiation. The output power measured with a power meter is 35 W/m^2 . In our calculation, we use the estimate made by Gregory Nole and Anthony Johnson¹ that on average the cumulative UV exposure for a person during his lifetime is about 30,000 minimal erythema doses (MEDs). MED varies from person to person due to the different skin types. In this calculation, we use a rough average MED based on a previous study² that shows

$$1 \text{ MED} \approx 400 \frac{\text{J}}{\text{m}^2} = 400 \frac{\text{W}\cdot\text{s}}{\text{m}^2}$$

Thus, a lifetime of UV exposure can be calculated as below, assuming a lifetime is 70 years.

$$UV\ exposure_{lifetime} = 30,000 * 400 \frac{W \cdot s}{m^2} = 12,000,000 \frac{W \cdot s}{m^2}$$

$$UV\ exposure_{per\ year} = \frac{UV\ exposure_{lifetime}}{70\ years} = 171428 \frac{W \cdot s}{m^2}$$

The incubation time that produces the same amount of UV-irradiation using a UV lamp is calculated as:

$$Time_{incubation} = \frac{UV\ exposure_{per\ year}}{35 \frac{W \cdot s}{m^2}} \approx 80\ min/year$$

The amount of UV exposure that one person receives per year is equivalent to 80 min of incubation with the UV lamp in our lab. Note that the estimation can vary quite much depends on the lifetime UV exposure amount and the MED values we choose. For a case with the lower bound we can use 20,000 MED for a lifetime UV exposure and 200 J/m² for the MED values. With these numbers $Time_{incubation} \approx 20\ min/year$, which is 4 times different from our previous estimate.

A2.3 Estimation of percent amyloid β -sheet in acid and UV treated lens tissues.

To estimate the percentage of protein that forms amyloid β -sheet when lens tissues are treated with acid or UV-irradiation, we fit the FTIR spectra of the acid denatured lens at 140 minutes to a sum of 3 Gaussians (Figure A2. 5), one of which had a frequency fixed at 1621 cm⁻¹. The 1621 cm⁻¹ band had an integrated area of 21%, which gives the percentage of β -sheets in the sample. For that same lens, the 2D IR diagonal slice is given in Figure 5.5, also fit to 3 Gaussians (Figure A2. 5). As discussed in the Discussion, the areas are not conserved in 2D IR spectra, but from the fitting of acid treated lens, we learn that 21% of protein in amyloid form creates a 1621

cm^{-1} peak with an intensity that is 1.4 times larger than that of the random coil and other structures in 2D IR. That ratio becomes a scaling factor to estimate β -sheet content from 2D IR spectra. We then calculate the β -sheet content from the 2D IR spectrum of UV-irradiated lens (Figure 5.5I). There are a number of assumptions with this approach, including that assumption that the transition dipole moments are the same magnitude for amyloid β -sheet in both acid and UV induced aggregation.

A2.4 Supporting Figures and legends

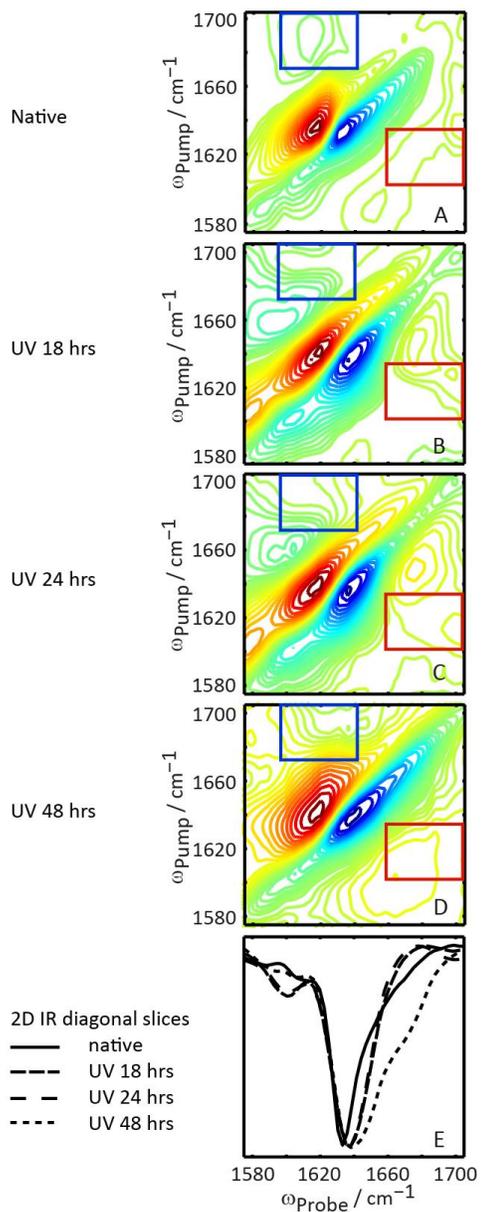


Figure A2. 1 2D IR spectra of UV-irradiated α B-crystallin proteins in D_2O buffer. (A) Native, (B) UV-irradiated (18 hrs), (C) UV-irradiated (24 hrs) and (D) UV-irradiated (48 hrs). (E) The overlay of the diagonal slices of the 2D IR spectra in (A-C).

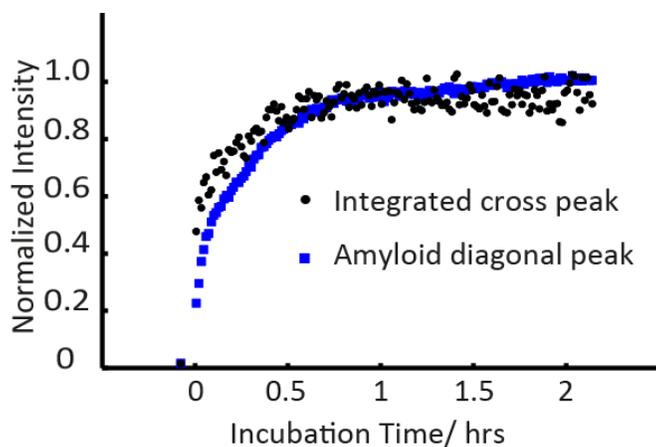


Figure A2. 2 Kinetics of amyloid formation in acid treated lens tissues.

Intensity of the 1621 cm^{-1} peak along diagonal was measured. The cross peak intensity was calculated by integrate the area of overtone because the fundamental cross peak interferes with the background. The initial data point at incubation time 0 min was obtained by measuring an untreated lens tissue sample.

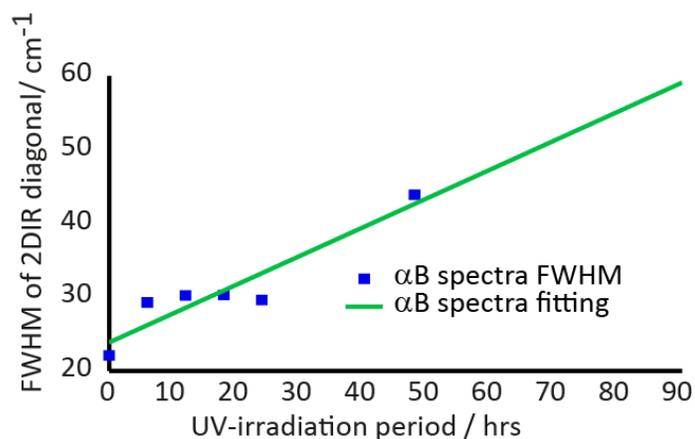


Figure A2. 3 FWHM of αB -crystallin protein 2D IR diagonal slices upon UV-irradiation. The data points follows a linear trend.

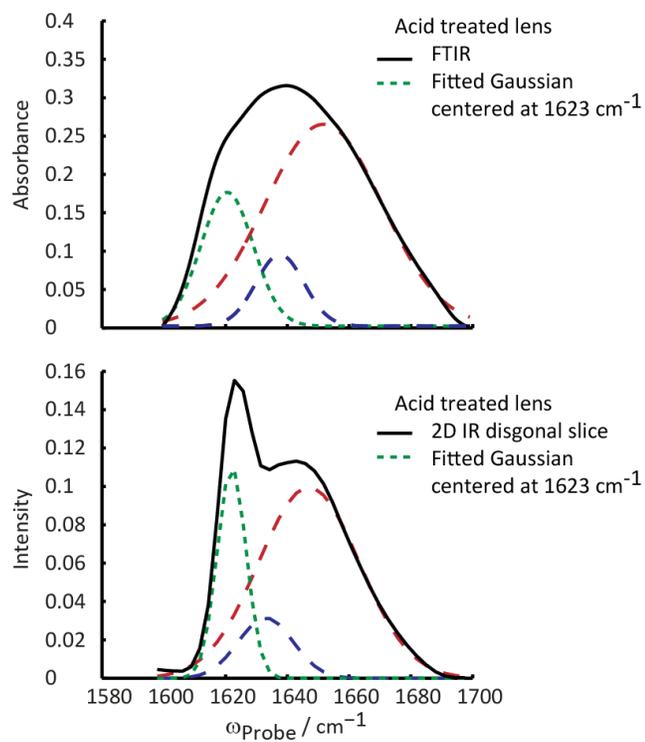


Figure A2. 4 Fitting of FTIR and 2D IR diagonal slice of acid treated lens tissue.

For 2D IR spectrum of the corresponding diagonal slice, see Figure 5.5 2D IR spectra of porcine lens tissues upon acid and UV treatment.. The FTIR and 2D IR diagonal slice are fitted to 3 Gaussians (with one centered around 1621 cm^{-1} , green dashed line). The original data were plotted as solid black lines.

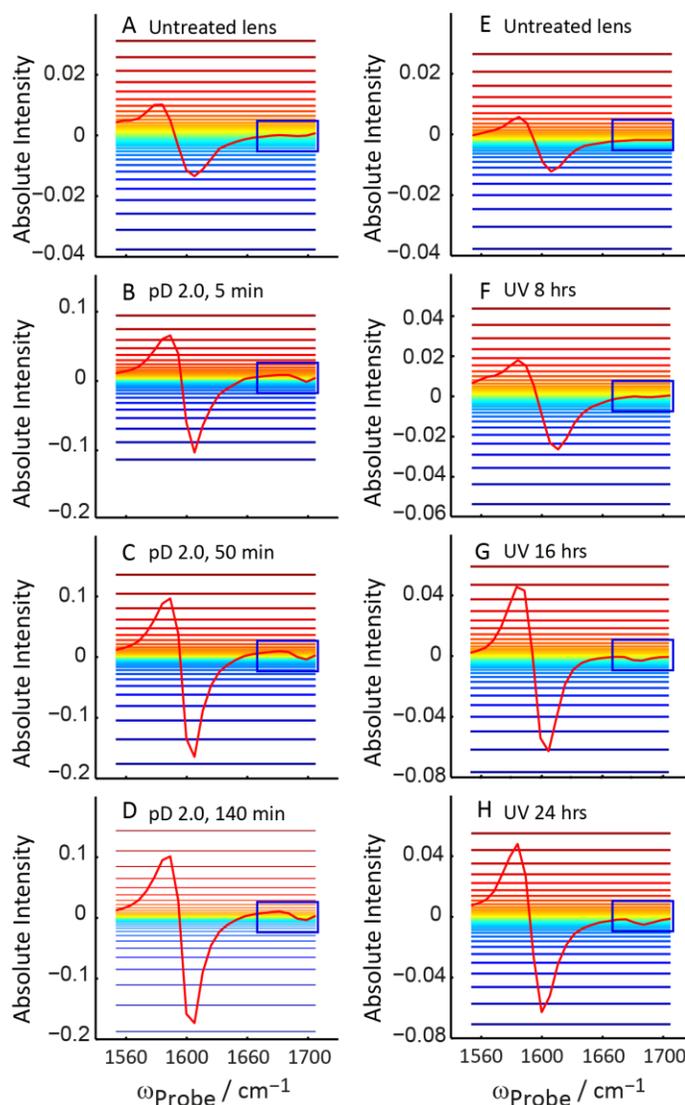


Figure A2. 5 Pump slices of 2D IR spectra.

Pump slices of 2D IR spectra at 1621 cm^{-1} of porcine lens tissue upon acid and UV treatment. Spectra (A-D, E-H) correspond to the 2D IR (A-D, F-I) spectra in Figure 5.5 2D IR spectra of porcine lens tissues upon acid and UV treatment. of the main text. The pair of negative and positive peaks in the blue box increases as incubation time extends. The colors of horizontal lines corresponds to contour plot color in 2D IR spectra.

A2.5 Supporting Reference

- (1) Nole G & Johnson AW (2004) An analysis of cumulative lifetime solar ultraviolet radiation exposure and the benefits of daily sun protection. *Dermatologic therapy* 17 Suppl 1:57-62.

- (2) Brenner AV (2002) Instrumental Measurements of Skin Color and Skin Ultraviolet Light Sensitivity and Risk of Cutaneous Malignant Melanoma: A Case-Control Study in an Italian Population. *American Journal of Epidemiology* 156(4):353-362.

A3. Appendix 3

Supporting Information for Chapter 6: GXXXG-mediated parallel and antiparallel dimerization of transmembrane helices and its inhibition by cholesterol: Single-molecule FRET and 2D IR studies

A3.1 Supporting Material and Methods

A3.1.1 Peptide Synthesis

Chromophore-labeled transmembrane peptides were manually synthesized by the standard 9-fluorenylmethoxycarbonyl (Fmoc)-based method on NovaSynTGR resin (Millipore, Billerica, MA, USA). Fmoc-amino acids were coupled for 2 hrs on the resin with 3 eq. amounts of amino acid, HOBt, and N,N'-diisopropylcarbodiimide in N,N-dimethylformamide (DMF). The reaction was monitored using the ninhydrin test. Fmoc was removed by treatment with 20% piperidine in DMF for 20 min. Chromophores (Cy3B, Cy5, and NBD) were labeled at the N-terminus of the peptides on resin by treatment with the succinimidyl esters (Cy3B and Cy5) (GE Healthcare, Little Chalfont, United Kingdom) or chloride derivative (NBD) (Nakalai Tesque, Kyoto, Japan) in DMF containing 5% N,N-diisopropylethylamine for 48 h. The peptide was cleaved from the resin with a deprotection cocktail of TFA/thioanisole/m-cresol/ethanedithiol/H₂O (16/1/1/1/1, v/v). For heterodimer formation, the cysteine thiols of Cy3B-labeled peptides were activated by the addition of 800 mM 2,2'-dithiodipyridine in the deprotection step.¹ The parallel dimer was synthesized by forming disulfide bonds between the Cy3B- β -C- β -HOST and Cy5- β -C- β -HOST, or Cy3B- β -C- β -

GXXXG and Cy5- β -C- β -GXXXG (β indicates β -alanine). The antiparallel dimer was synthesized by forming disulfide bonds between the Cy3B- β -C- β -HOST and Cy5- β -HOST- β -C, or Cy3B- β -C- β -GXXXG and Cy5- β -GXXXG- β -C. These disulfide-bridged dimer peptides were purified by a PLRP-S 300 Å 5 μ m reversed phase HPLC column (Agilent, Santa Clara, CA, USA) with a linear gradient from formic acid/H₂O (2/3, v/v) to formic acid/2-propanol (4/1, v/v) at 50°C. The eluted peptide solution was immediately neutralized with 28% NH₃ (aq) on ice, frozen with liquid N₂, and lyophilized. The peptide powder was dissolved in 1,1,1,3,3,3-hexafluoroisopropanol (HFIP) and identified by ion-spray or matrix-assisted laser desorption/ionization mass spectroscopy.

For isotope labeling, 1-¹³C amino acids (CIL, Andover, MA, USA) were ¹⁸O-exchanged with ¹⁸O water.² Fmoc-amino acid powder (Gly, Ala, or Leu) and a magnetic stir bar in a flask were vacuum-dried overnight. Immediately before starting the exchange reaction, magnesium sulfate was added to 4N-HCl/dioxane (~1 g per 10 mL) and mixed to dry the solvent, followed by paper-filtration to remove magnesium sulfate. The amino acid was dissolved in ¹⁸O water (> 40 eq. of amino acids) and 4N-HCl/dioxane (minimum amount to dissolve amino acids). The solution was refluxed at 100 °C for 3 hrs (Gly and Ala) or 7 hrs (Leu). The solvent was evaporated and vacuum-dried to obtain the amino acid powder. The ¹⁸O exchange efficiencies were ~95%, estimated from the ion-spray mass spectra. The amino acids (2 mmol) were dissolved in 7 mL of 10% Na₂CO₃ (w/v) and Fmoc-derivatized by incubating with Fmoc-OSu (3 mmol) in 7 mL of dimethoxyethane overnight. The solution was paper-filtrated, neutralized with HCl, and evaporated. Ethyl acetate (30 mL) was added and mixed, and the whole was washed with 0.1N HCl (twice) and saturated NaCl (twice) in a separatory funnel, and dried to obtain Fmoc-amino acids.

A3.1.2 Single-Molecule FRET

The biotin-PEG-coated slide chamber for fixing large unilamellar vesicles (LUVs) was prepared based on the protocol of Joo and Ha.³ Briefly, No. 1-S cover glass (24 × 60 mm, Matsunami Glass, Osaka, Japan) was washed by sonication in 1 M KOH and methanol, then amino-functionalized with aminopropylsilane, and coated with biotinylated PEG by treatment with PEG succinimidyl ester, with an average MW of 5,000 (mPEG-SVA, Laysan Bio, Arab, AL, USA), and its biotin derivative (biotin-PEG-SVA, Laysan Bio) at a ratio of 80/1 (w/w). A quartz slide (26 × 76 × 1 mm) with inlet/outlet holes (1-mm diameter) and an outlet tube was custom made by DAICO MFG (Kyoto, Japan). The chamber was assembled by putting a silicon spacer (0.2-mm thick) between the cover glass and slide to form a chamber space of ~60 μL. The samples were introduced into the chamber by suctioning a droplet on the inlet hole from the outlet side with a syringe pump (Nanojet, Chemix, Stafford, TX, USA).

The lipids 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphatidylcholine (POPC), 1,2-dipalmitoyl-*sn*-glycero-3-phosphoethanolamine-N-(cap biotinyl) (biotin-PE), and cholesterol were obtained from Avanti polar lipids (Alabaster, AL, USA). The disulfide-dimer-peptides were incorporated into membrane films by mixing with lipids in organic solvents. POPC/cholesterol/biotin-PE/peptide (630,000/270,000/9,000/1) or POPC/biotin-PE/peptide (900,000/9,000/1) was dissolved in ethanol or 2,2,2-trifluoroethanol, respectively, followed by evaporation. The membrane film was dried under a vacuum pump overnight. The film was hydrated with a fresh Tris buffer (10 mM Tris/150 mM NaCl/1 mM EDTA supplemented with 0.8% (w/v) D-glucose and 1 mM Trolox (pH 7.4)) at 50 °C for 20–30 min to obtain multilamellar vesicles (total lipid concentration of 5–15 mM). Large unilamellar vesicles (LUVs) were prepared

by an extrusion of the membrane suspension through a polycarbonate filter with 100-nm pores (21 times) at 50 °C.

The biotinylated LUVs (~ 1:100 dilution) were added into the biotin-PEG-coated chamber, and incubated for 10 min following a 10-min pretreatment with 0.2 mg/mL NeutrAvidin. After fixing LUVs, they were incubated with 100 mM 2-mercaptoethanol for 30 min to cleave the disulfide bonds.

The fluorescence images for Cy3B (575–635 nm) and Cy5 (645–745 nm) under Cy3B excitation at 561 nm were simultaneously acquired using a Nikon Ti-based total internal reflection microscope equipped with an Imagem EM-CCD camera and W-View optics (Hamamatsu Photonics, Hamamatsu, Japan) at a time resolution of 17 ms. To suppress photoblinking, the observation was performed in oxygen-depleted tris buffer containing 1 mM trolox, 1 mM methyl viologen, 0.8% (w/v) D-glucose, 0.25 mg/mL glucose oxidase, and 10.5 mg/mL catalase (pH 7.4). The apparent FRET efficiency, E_{app} , was calculated from fluorescence intensities for the donor (F_{Cy3B}) and acceptor (F_{Cy5}) as $E_{app} = F_{Cy5}/(F_{Cy3B} + F_{Cy5})$. The smFRET trajectories originating from the monomer–dimer transitions were analyzed with the HaMMY program (<http://bio.physics.illinois.edu/HaMMY.asp>) to deduce the rate constants between different states.

A3.1.3 FTIR

For IR measurements, the peptide main chains were deuterated in HFIP- d_2 (3 h, twice). Oriented films of lipids/peptides were prepared by uniformly spreading a 100 μ L TFE solution of lipids (5 μ mol) and peptides (2.5 nmol) on a germanium ATR plate (70 \times 10 \times 5 mm), followed by evaporation of the solvent with N₂ gas under a vacuum overnight. The films were hydrated with a D₂O-soaked piece of filter paper put over the plate for 3 h at 25°C. FTIR-PATR measurements were carried out on a BioRad FTS-3000MX spectrometer equipped with a Specac horizontal ATR

attachment with an AgBr polarizer and a temperature controller. The dichroic ratio, R , defined by $\Delta A_{\parallel}/\Delta A_{\perp}$, was calculated from the polarized spectra. The absorbance (ΔA) was obtained as the area for the $^{13}\text{C}=^{18}\text{O}$ isotope amide I' band. The subscripts \parallel and \perp refer to polarized light with the electric vector parallel and perpendicular to the plane of incidence, respectively. The orientation angle of the transition moment for amide I' to the bilayer normal, β , was calculated from R .⁴

$$\cos^2\beta = \frac{1}{3} \left(2 \cdot \frac{R - 2.00}{R + 1.45} + 1 \right) \quad (1)$$

To estimate the helix orientation angle to the bilayer normal, α , the position dependence of β was fitted

$$\cos\beta = \cos\theta \cdot \cos\alpha - \sin\theta \cdot \cos\phi \cdot \sin\alpha \quad (2)$$

where θ is the angle between the helix axis and transition moment. ϕ is the rotation angle that changes 100° per residue in the helix. In the fitting of β values, α , θ , and ϕ were treated as adjustable parameters.

A3.1.4 2D IR

For 2D IR measurements, films of lipids (2 μmol)/peptides (40 nmol) were hydrated with 10 μL of a deuterium buffer (10 mM Tris, 150 mM NaCl, 60% sucrose, pD = 7.4). Sucrose (deuterated in D_2O^5) was added to reduce scattering from the lipid vesicles. The suspension was thoroughly vortexed and sonicated to obtain a transparent vesicle solution. Small droplets of the vesicle samples ($\sim 0.5 \mu\text{L}$) were placed between CaF_2 windows (Spacer: 25 μm). To examine the vibrational coupling due to helix association, isotope-labeled (100%) and diluted (25%) samples

were put on the same window. The 2D IR measurements were performed at 23 °C according to reference 2.

A3.1.5 2D IR Simulation

The peak shift of the maximum intensity due to vibrational coupling between isotope-labeled amides in dimers was calculated with a simulation package based on Matlab script language (available from web: <http://zanni.chem.wisc.edu/content/software>). Two transmembrane helices were placed at a distance of 8 Å and a crossing angle of 10° (parallel) or -170° (antiparallel). Frequencies of the isotope label and line width were set at 1,590 and 10 cm⁻¹, respectively. Transition dipole-dipole coupling was used to calculate the frequency shifts associated with the dimer structure.

A3.2 Supporting Discussion

Although the helices may form higher-order oligomers in IR measurements due to a high peptide concentration (2 mol %), a similar parallel interface was expected in dilute conditions, because the same interface was independently predicted by a surface-based packing algorithm PREDDIMER (<http://model.nmr.ru/preddimer/>), which involved two Gly and four Ala residues (Figure A3. 4 Predicted dimer structure.). Two glycines were necessary for efficient packing in the prediction (data not shown). A similar heptad interface was reported for the EphA2 dimer.⁶ Although cholesterol⁷ and GXXXG⁸ possibly stabilize the hourglass-shaped helix dimer, it was not observed under the present conditions.

A3.3 Supporting Figures

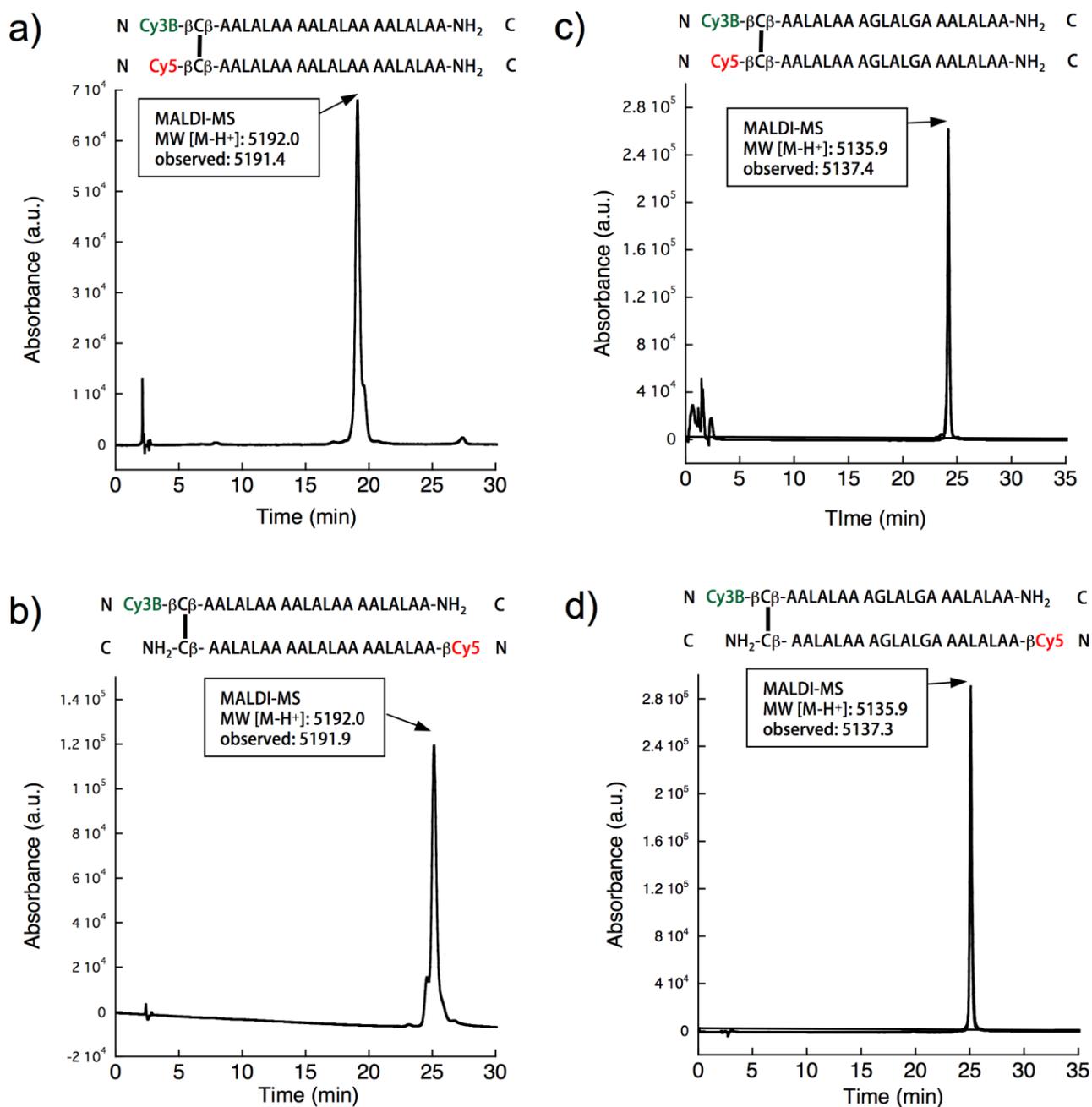


Figure A3. 1 Sequences and HPLC chromatograms of synthetic crosslinked transmembrane dimers.

The disulfide-bridged dimer peptides were purified by a PLRP-S column with a linear gradient from formic acid/H₂O (2/3, v/v) to formic acid/2-propanol (4/1, v/v) at 50°C. Host (a,b) and GXXXG (c,d) dimers were analyzed with gradients of 65–95 and 30–95%, respectively.

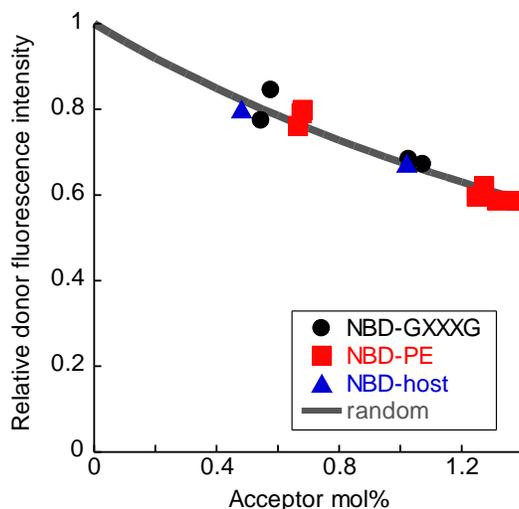


Figure A3. 2 Fluorescence quenching of dehydroergosterol.

Fluorescence quenching of dehydroergosterol with *N*-(7-nitrobenz-2-oxa-1,3-diazol-4-yl)-1,2-dihexadecanoyl-*sn*-glycero-3-phosphoethanolamine (NBD-PE, red squares), NBD-labeled host helices (blue triangles), or NBD-labeled GXXXG helices (black circles). The molar ratio of POPC/cholesterol/dehydroergosterol was fixed at 70/29/1. The fluorescence intensity from dehydroergosterol (excitation: 300 nm; emission: 374 nm) was measured in the presence of acceptor NBD molecules. The gray line indicates a theoretical quenching curve expected for random FRET assuming an R_0 of 30 Å and distances of the closest approach (R_c) of 25.8 Å.^{7,9}

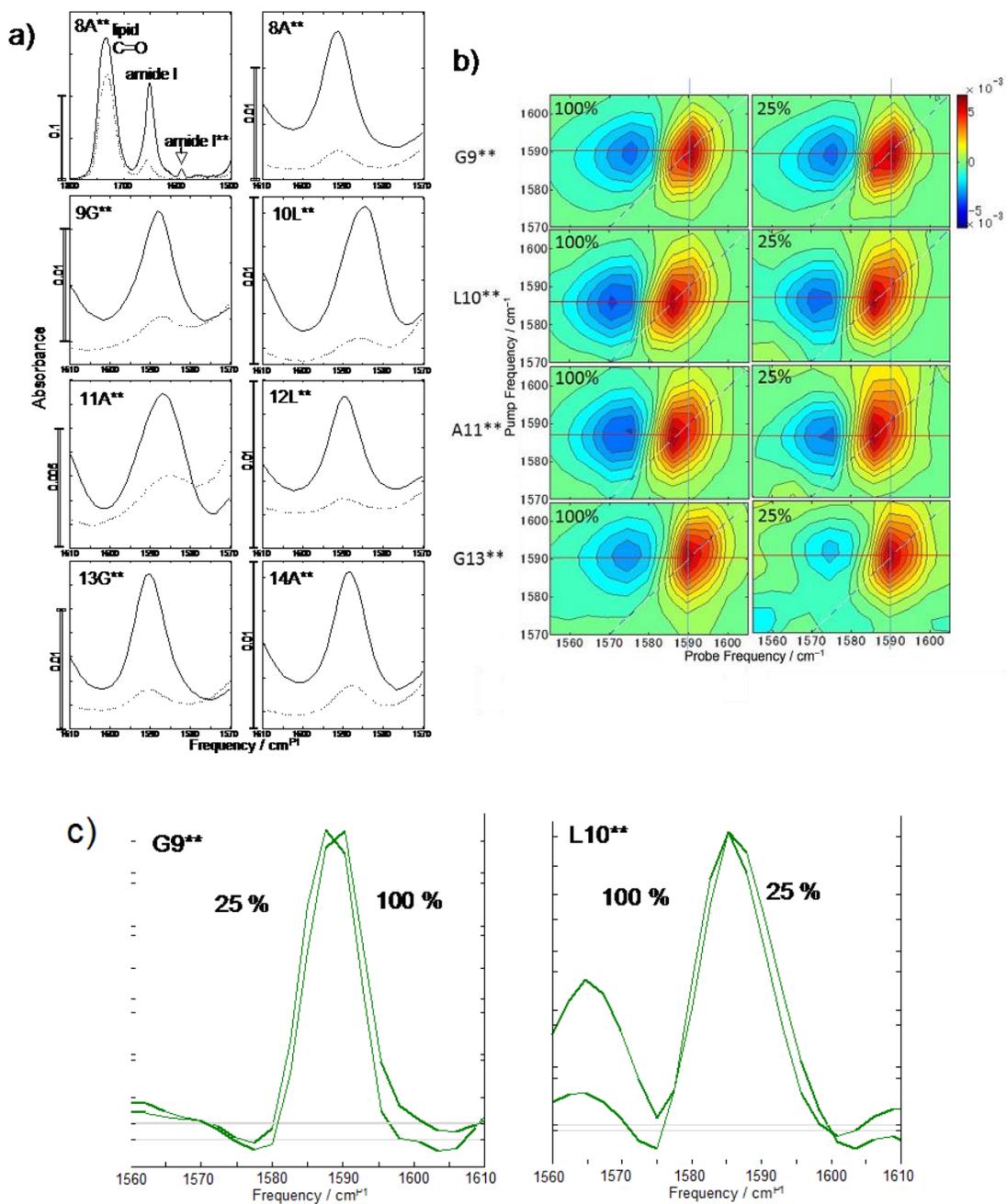


Figure A3.3 IR spectra of the isotope-edited GXXXG helix.

(a) PATR-FTIR spectra in the amide and lipid carbonyl region (8A**) and isotope amide I** region (all labels). (b) 2D-IR spectra for isotope-labeled samples (100%) and diluted samples (25%). (c) Diagonal cut of G9** and L10** spectra.

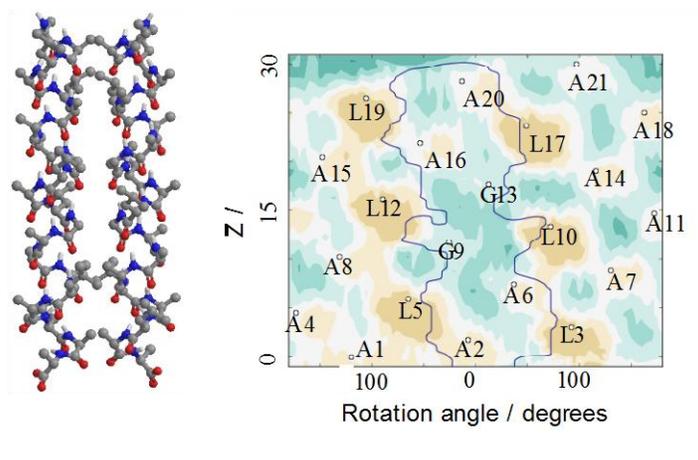


Figure A3. 4 Predicted dimer structure.

The dimer interface of the GXXXG helix was predicted by PREDDIMER (<http://model.nmr.ru/preddimer/>). Sand and green colors indicate convex and concave surfaces, respectively. The gray line shows the predicted helix–helix contact area, involving A2, A6, G9, G13, A16, and A20 residues.

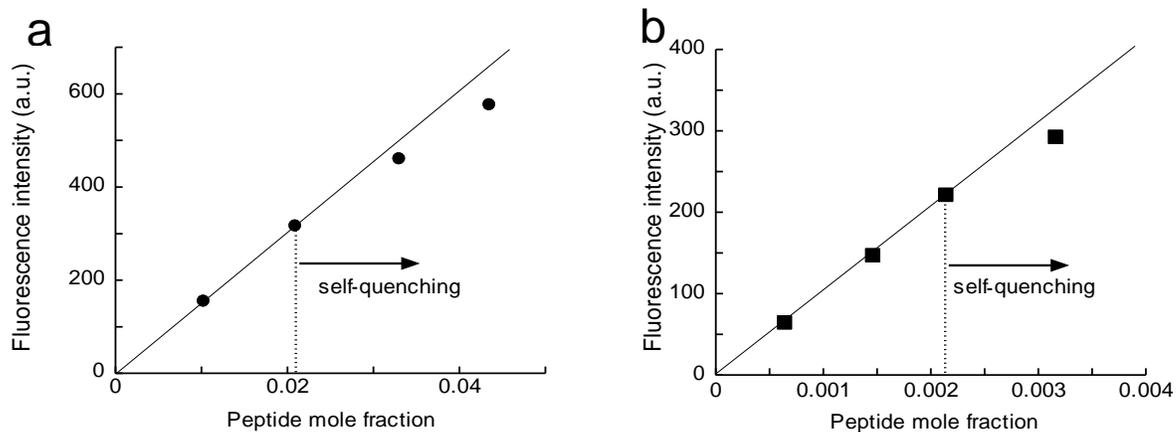


Figure A3. 5 Self-quenching of NBD-labeled helices.

Concentration dependences of NBD fluorescence of the host helix (25°C) indicate that parallel associations occur at peptide mole fractions of ~0.02 and ~0.002 in POPC (a) and POPC/cholesterol(7/3) (b) vesicles, respectively.

A3.4 Supporting References

- (1) C. M. Yamazaki, S. Asada, K. Kitagawa, T. Koide, Artificial collagen gels via self-assembly of de novo designed peptides, *Biopolymers* **2008**, *90*, 816.
- (2) C. T. Middleton, A. M. Woys, S. S. Mukherjee, M. T. Zanni, Residue-specific structural kinetics of proteins through the union of isotope labeling, mid-IR pulse shaping, and coherent 2D IR spectroscopy, *Methods* **2010**, *52*, 12.
- (3) C. Joo, T. Ha, Preparing sample chambers for single-molecule FRET, *Cold Spring Harbor Protocols* **2012**, 1104.
- (4) Y. Yano, T. Takemoto, S. Kobayashi, H. Yasui, H. Sakurai, W. Ohashi, M. Niwa, S. Futaki, Y. Sugiura, K. Matsuzaki, Topological stability and self-association of a completely hydrophobic model transmembrane helix in lipid bilayers, *Biochemistry* **2002**, *41*, 3073.
- (5) B. S. Kendrick, B. S. Chang, T. Arakawa, B. Peterson, T. W. Randolph, M. C. Manning, J. F. Carpenter, Preferential exclusion of sucrose from recombinant interleukin-1 receptor antagonist: Role in restricted conformational mobility and compaction of native state. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 11917.
- (6) E. V. Bocharov, M. L. Mayzel, P. E. Volynsky, K. S. Mineev, E. N. Tkach, Y. S. Ermolyuk, A. A. Schulga, R. G. Efremov, A. S. Arseniev, Left-handed dimer of EphA2 transmembrane domain: Helix packing diversity among receptor tyrosine kinases, *Biophys. J.* **2010**, *98*, 881.
- (7) Y. Yano, K. Kondo, R. Kitani, A. Yamamoto, K. Matsuzaki, Cholesterol-induced lipophobic interaction between transmembrane helices using ensemble and single-molecule fluorescence resonance energy transfer, *Biochemistry* **2015**, *54*, 1371.
- (8) B. K. Mueller, S. Subramaniam, A. Senes, A frequent, GxxxG-mediated, transmembrane association motif is optimized for the formation of interhelical C α -H hydrogen bonds, *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, E888.
- (9) P. K. Wolber, B. S. Hudson, An analytic solution to the Förster energy transfer problem in two dimensions, *Biophys. J.* **1979**, *28*, 197.

A4. Appendix 4

Plasmids, Genes and Peptides

In this appendix, I will list the plasmids and genes we own at best to my knowledge. In these files, the top line is DNA sequence and the bottom line is the corresponding amino acids. Most of these genes are sequenced with T7T and T7P primers (These are commercial primers that are embedded in most plasmids). Protein expression always starts from a methionine residue (M). The initial M is bolded in the sequences. A “-” in the protein sequence is a stop codon. An “X” in the protein sequence shows up when the base pairs are not recognizable. I am unsure about some of the vectors and these are indicated with a “*”.

There other genes we have that are incorporated in plasmids. For example, α -synuclein, ubiquitin and PolyQ. The cells containing these plasmids are stored in the -80 °C in our emergency box (check with Ariel for the inventory).

A4.1 Plasmids and genes

A4.1.1 γ D-crystallin

A4.1.1.1. γ D-crystallin WT

Vector: pet32*

Constructed by Sean Moran

```

agaaattccctcngaatatTTTTgtttaactTTaagaaggagatataccatgggcatcat
R N S L X I F C L T L R R R Y T M G H H
catcatcatcatcatcatcacagcagcggccatatcgaaggTCgtatgggcaaaatt
H H H H H H H H S S G H I E G R M G K I
actctgtatgaagatcgtggTTTccaggGCCgtcattatgagTgcagctccgatcaccCG
T L Y E D R G F Q G R H Y E C S S D H P
aacctgcagccgtacctgtctcgtTgcaactccgcacgcgttgactccggtTgttggatg
N L Q P Y L S R C N S A R V D S G C W M
ctgtatgaacagccgaactactctggTctgcagtattTcctgcgccgCGgactacgct
L Y E Q P N Y S G L Q Y F L R R G D Y A
gaccaccagcagTgggatgggtctgtccgactccgtgcgtTcctgccgtctgatcccgCAC
D H Q Q W M G L S D S V R S C R L I P H
tctggcagccaccgtatccgcctgtacgaacgtgaagattatcgtggTcaaTgattgag
S G S H R I R L Y E R E D Y R G Q M I E
ttcaccgaagactgctctTgcctgcaggatcgTTTccgTTTcaacgaaTccactctctg
F T E D C S C L Q D R F R F N E I H S L
aacgTtctggaaggctctTgggtcctgtatgaactgagcaactaccgtggccgTcagTAC
N V L E G S W V L Y E L S N Y R G R Q Y
ctgctgatgccgggCGactatcgtcGctaccaggactggggCGgaccaatgctcgtgTA
L L M P G D Y R R Y Q D W G A T N A R V
ggctctctgcgccgctgatcgactTcagctaTgataacatTgctcgaggatccggct
G S L R R V I D F S - - - H M L E D P A
gctaacaagccccgaaaggaagctgagTtggtgctgctgccaccgctgagcaataactagCA
A N K A R K E A E L A A A T A E Q - L A
taaccctTggggcctctaaacgggtctTgaggggTTTTTgctgaaaggaggaactata
- P L G A S K R V L R G F L L K G G T I
tccggatTcccgcaagaggccCGcagTaccggcataaccaagcctatgcctacagcat
S G Y P A R G P A V P A - P S L C L Q H

```

A4.1.1.2. γ D-crystallin S84C

Vector: pet32*

Constructed by Sean Moran

```

agaaattccctcngaataatTTTTgtttaactTTaagaaggagatataccatgggcatcat
R N S L X I F C L T L R R R Y T M G H H
catcatcatcatcatcatcatcacagcagcggccatatcgaaggTCgtatgggcaaatt
H H H H H H H H S S G H I E G R M G K I
actctgtatgaagatcgtggTTTccagggccgtcattatgagtgcagctccgatcaccCG
T L Y E D R G F Q G R H Y E C S S D H P
aacctgcagccgtacctgtctcgttgcaactccgcacgcgTTgactccggttgTTggatg
N L Q P Y L S R C N S A R V D S G C W M
ctgtatgaacagccgaactactctggTctgcagTatTTTcctgcgccgCGgactacgct
L Y E Q P N Y S G L Q Y F L R R G D Y A
gaccaccagcagTggatgggtctgtccgactccgtgcgTtTcctgccgtctgatcccgCac
D H Q Q W M G L S D S V R S C R L I P H
tgtggcagccaccgTatccgcctgtacgaacgtgaagattatcgtggTcaaTgattgag
C G S H R I R L Y E R E D Y R G Q M I E
ttcaccgaagactgctcttgCctgcaggatcgtTTTccgTTTcaacgaaTccactctctg
F T E D C S C L Q D R F R F N E I H S L
aacgTtctggaaggctcttgggtcctgtatgaactgagcaactaccgTggccgTcagTac
N V L E G S W V L Y E L S N Y R G R Q Y
ctgctgatgccgggCGactatcgtcGctaccaggactggggCGcgaccaatgctcgtgTa
L L M P G D Y R R Y Q D W G A T N A R V
ggctctctgcgccgCGTgatcgactTcagctaTgataacatTgctcGaggatccggct
G S L R R V I D F S - - - H M L E D P A
gctaacaaagccccgaaaggaagctgagTtggctgctgccaccgctgagcaataactagCa
A N K A R K E A E L A A A T A E Q - L A
taacccttggggcctctaaacgggtcttgaggggTTTTTgctgaaaggaggaactata
- P L G A S K R V L R G F L L K G G T I
tccggatTcccGcaagaggccCGcagTaccggcataaccaagcctatgcctacagcat
S G Y P A R G P A V P A - P S L C L Q H

```

A4.1.1.3. γ D-crystallin R58H

Vector: pet32*

Constructed by Sean Moran and Tianqi Zhang

```

gggaattcccctctgaatattttgtttactttaagaaggagatataccatgggcatcat
G N S P L N I L F T L R R R Y T M G H H
catcatcatcatcatcatcacagcagcggccatatcgaaggcgtatgggcaaatt
H H H H H H H H S S G H I E G R M G K I
actctgtatgaagatcgtggttccagggcgcattatgagtgcagctccgatcaccg
T L Y E D R G F Q G R H Y E C S S D H P
aacctgcagccgtacctgtctcgttgcaactccgcacgcgttgactccggttggtgatg
N L Q P Y L S R C N S A R V D S G C W M
ctgtatgaacagccgaactactctggtctgcagtatcttctgcatcgcggcgactacgct
L Y E Q P N Y S G L Q Y F L H R G D Y A
gaccaccagcagtggtgggtctgtccgactccgtgcgttccctgccgtctgatcccgcac
D H Q Q W M G L S D S V R S C R L I P H
tctggcagccaccgtatccgcctgtacgaacgtgaagattatcgtgggtcaaagtattgag
S G S H R I R L Y E R E D Y R G Q M I E
ttcaccgaagactgctcttgccctgcaggatcgtttccggttcaacgaaatccactctctg
F T E D C S C L Q D R F R F N E I H S L
aacgttctggaaggctcttgggtcctgtatgaactgagcaactaccgtggccgctcantac
N V L E G S W V L Y E L S N Y R G R X Y
ctgctgatgccgggcgactatcgtcgtaccaggactggggcgcgaccaatgctcgtgta
L L M P G D Y R R Y Q D W G A T N A R V
ngctctntgnncnncntgatcaacttcanctaataacatagctcagaggatccgggt
X S X X X X I N F X - - - H M L E D P A
gctaacaagccccgaaaggaagctgagttggctgctgccaccgctgagnaannnaannn
A N K A R K E A E L A A A T A E X X K X

```

A4.1.1.4. γ D-crystallin P23T

Vector: pet32*

Constructed by Sean Moran and Tianqi Zhang

```

tgnngaattccctctgaatattttgtttactttaagaaggagatataccatgggccatcat
X E F P L N I L F T L R R R Y T M G H H
catcatcatcatcatcatcatcacagcagcggccatatcgaaggtcgtatgggcaaatt
H H H H H H H H S S G H I E G R M G K I
actctgtatgaagatcgtggttccagggccgtcattatgagtgcagctccgatcacacc
T L Y E D R G F Q G R H Y E C S S D H T
aacctgcagccgtacctgtctcgttgcaactccgcacgcggttgactccggttggtggatg
N L Q P Y L S R C N S A R V D S G C W M
ctgtatgaacagccgaactactctggtctgcagtatcttctgcgccgcgggcgactacgct
L Y E Q P N Y S G L Q Y F L R R G D Y A
gaccaccagcagtggtgggtctgtccgactccgtgcggttccctgccgtctgatcccgcac
D H Q Q W M G L S D S V R S C R L I P H
tctggcagccaccgatatccgcctgtacgaacgtgaagattatcgtgggtcaaattgattgag
S G S H R I R L Y E R E D Y R G Q M I E
ttcaccgaagactgctcttgccctgcaggatcgtttccggttcaacgaaatccactctctg
F T E D C S C L Q D R F R F N E I H S L
aacgttctggaaggctcttgggtcctgtatgaactgagcaactaccgtggccggtcagtac
N V L E G S W V L Y E L S N Y R G R Q Y
ctgctgatgccgggcgactatcgtcgtaccaggactggggcgcgaccaatgctcgtgta
L L M P G D Y R R Y Q D W G A T N A R V
ggctctctgcgccgcntgatcnacttcagctaataacatatactcgaggatccgggt
G S L R R X I X F S - - - H M L E D P A
gctaacaaagccccgaaaggaagctgagttggctgctgccaccgctgagnaataaaaactaa
A N K A R K E A E L A A A T A E X - N -

```

A4.1.1.5. γ D-crystallin C-terminal domain

Vector: pet63*

Constructed by Sean Moran

```

gagaattccctctgaatattttgtttaactttaagaaggagatataccatgggcatcat
E N S L - I F C L T L R R R Y T M G H H
catcatcatcatcatcatcatcacagcagcggccatatcgaaggctcgttgtggcagccac
H H H H H H H H S S G H I E G R C G S H
cgtatccgcctgtacgaacgtgaagattatcgtggtcaaagattgagttcaccgaagac
R I R L Y E R E D Y R G Q M I E F T E D
tgctcttgctgcaggatcgtttccgtttcaacgaaatccactctctgaacgttctggaa
C S C L Q D R F R F N E I H S L N V L E
ggctcttgggtcctgtatgaactgagcaactaccgtggcgcgtcagtacctgctgatgccg
G S W V L Y E L S N Y R G R Q Y L L M P
ggcgactatcgtcgtaccaggactggggcgcgaccaatgctcgtgtaggctctctgcgc
G D Y R R Y Q D W G A T N A R V G S L R
cgcgtgatcgacttcagctaataacatatacgtcgcaggatccggctgctaacaagcc
R V I D F S - - - H M L E D P A A N K A
cgaaaggaagctgagttggctgctgccaccgctgagcaataactagcataacccttggg
R K E A E L A A A T A E Q - L A - P L G
gcctctaaacgggtccttgaggggttttttgcgtgaaaggaggaactatatccggatatccc
A S K R V L R G F L L K G G T I S G Y P
gcaagaggcccggcagtagcaccgataaccaagcctatgcctacagcatccagggtgacgg
A R G P A V P A - P S L C L Q H P G - R
tgccgaggatgacgatgagcgcattgtagatttcatacacgggtgctgactgcgtagc
C R G - R - A H C - I S Y T V P D C V S
aatttaactgtgataaactaccgcattaaagcttatcgatgataagctgtcaaacatgag
N L T V I N Y R I K A Y R - - A V K H E

```

A4.1.1.6. γ D-crystallin N-terminal domain

Vector: pTXB1

Constructed by Sean Moran

```

tgaggaattccctctgaatattttgtttaactttaagaaggagatatacatatgggcaa
- G I P S E Y F V - L - E G D I H M G K
attactctgtatgaagatcgtgggttccagggccggtcattatgagtgcagctccgatcac
I T L Y E D R G F Q G R H Y E C S S D H
ccgaacctgcagccgtacctgtctcgttgcaactccgcacgcggttgactccggttgttgg
P N L Q P Y L S R C N S A R V D S G C W
atgctgtatgaacagccgaactactctgggtctgcagtatttccctgcgccgaggcgactac
M L Y E Q P N Y S G L Q Y F L R R G D Y
gctgaccaccagcagtgatgggtctgtccgactccgtgcggttccctgccgtctgatcccg
A D H Q Q W M G L S D S V R S C R L I P
cactgcatcacgggagatgcactagttgccctaccggaggcgagtcggtacgcatcgcc
H C I T G D A L V A L P E G E S V R I A
gacatcgtgccgggtgcgcggcccaacagtgacaacgccatcgacctgaaagtccttgac
D I V P G A R P N S D N A I D L K V L D
cggcatggcaatcccgtgctcgcggaccggctgttccactccggcgagcatccgggtgtac
R H G N P V L A D R L F H S G E H P V Y
acggtgcgtagcggcgaaggtctgcgtgtgacgggcaccggaaccaccggttgttgtgt
T V R T V E G L R V T G T A N H P L L C
ttggtcgacgtcgccgggggtgccgaccctgctgtggaagctgatcgacgaaatcaagccg
L V D V A G V P T L L W K L I D E I K P
ggcgattacgcggtgattcaacgcagcgcattcagcgtcgactgtgcaggttttgcccgc
G D Y A V I Q R S A F S V D C A G F A R
gggaaaccgaatttgcgcccacaacctacacagtcggcggtccctggactgggtgcgtttc
G K P E F A P T T Y T V G V P G L V R F
ttggaagcacaccaccgagacccggagcccaagctatcgccgacgagctgaccgacggg
L E A H H R D P D A Q A I A D E L T D G
cggttctactacgcgaaagtcgccagtgaccaccgacgcccggcgtgcagccgggtgtatagc
R F Y Y A K V A S V T D A G V Q P V Y S
cttcgtgtcgacacggcagaccacgcgtttatcacgaacgggttcgctcagccacgctact
L R V D T A D H A F I T N G F V S H A T
ggcctcaccgggtctgaactcaggnccctcacgacaaatcctgggtgtatccgcttggcaggt
G L T G L N S X P H D K S W C I R L A G
caacacagcttataactgctgggacaattgggtcacatatanccggcaaaacgtataaatgttt
Q H S L Y C G T I G H I X R Q N V - M F
tgcagcccnnacctcctttggcagggatngggaaacatnccaacgnttccctgccttgtg
C S P X P P L A G X G K H X N X S C L V
ggnancttfaatgactgcaggaaagggaaccngctnctaacaacaanccccaaaaggaaac
X X X N D C R K G N X X L N K X P K G N
ntaanctggctgctnccnccccctnaanaaananttancaaaaaccctnngggcctttt
X X W L X X X P X X X X X K N P X G P X

```

A4.1.2 α B-crystallin

A4.1.2.1. α B-crystallin WT

Vector: pAED4

Resource: This plasmid is gift from Professor Jonathan King's group.

```

ggtttttctntagaaataatthttgtttaantthtaagaaggagatatacatagcccatatg
G F S X E I I L F X F K K E I Y I A H M
gacatcgccatccaccaccctggatccgccccttctttcctttccactccccagc
D I A I H H P W I R R P F F P F H S P S
cgcctctttgaccagttcttcggagagcacctggtggagtctgatcttttcccagctct
R L F D Q F F G E H L L E S D L F P T S
acttcctgagtccttctaccttcggccaccctccttccctgcgggcaccagctggttt
T S L S P F Y L R P P S F L R A P S W F
gacactggactctcagagatgcgcctggagaaggacaggttctctgtcaacctggatgtg
D T G L S E M R L E K D R F S V N L D V
aagcacttctccccagaggaactcaaagttaagggtgttgggagatgtgattgagggtgat
K H F S P E E L K V K V L G D V I E V H
ggaaaacatgaagagcgccaggatgaacatggtttcatctccaggagttccacaggaaa
G K H E E R Q D E H G F I S R E F H R K
taccggatcccagctgatgtagaccctctcaccattacttcatccctgtcatctgatggg
Y R I P A D V D P L T I T S S L S S D G
gtcctcactgtgaatggaccaaggaaacaggtctctggccctgagcgcaccattcccatc
V L T V N G P R K Q V S G P E R T I P I
accctggaagagaagcctgctgtcaccgcagcccccaagaaatagatgccctttcttgaa
T R E E K P A V T A A P K K - M P F L E
ttgcattttttaaaacaagaaagtttccccaccagtgaaatgaaagtcttgtgactagtgc
L H F L K Q E S F P T S E - K S C D - C
tgaagcttgcatgcctgcaggtcgactctagaggatccccgggtaccgagctcgaattca
- S L H A C R S T L E D P R V P S S N S

```

A4.1.2.2. Histag α B-crystallin

Vector: pAED4

Constructed by Tianqi Zhang based on original α B-crystallin plasmid

```

gnaatccntnaatattttgtttacttttagaaggagatatacatagtcccatatgcatcat
X I X X Y F V Y F R R R Y T - S H M H H
caccatcaccacctcgagatcgagggtcgtatggacatcgccatccaccacccctggatc
H H H H L E I E G R M D I A I H H P W I
cgccgccccttctttcctttccactccccagccgcctctttgaccagttcttcggagag
R R P F F P F H S P S R L F D Q F F G E
cacctgttggagtctgatcttttcccagcgtctacttccctgagtccttctaccttcgg
H L L E S D L F P T S T S L S P F Y L R
ccaccctccttctcgcgggcacccagctggtttgacactggactctcagagatgcgctg
P P S F L R A P S W F D T G L S E M R L
gagaaggacagggttctctgtcaacctggatgtgaagcacttctccccagaggaactcaa
E K D R F S V N L D V K H F S P E E L K
gttaagggtgttgggagatgtgattgaggtgcatggaaaacatgaagagcgccaggatgaa
V K V L G D V I E V H G K H E E R Q D E
catggtttcatctccaggaggttccacaggaataaccggatcccagctgatgtagaccct
H G F I S R E F H R K Y R I P A D V D P
ctcaccattacttcatccctgtcatctgatggggctcctcactgtgaatggaccaaggaaa
L T I T S S L S S D G V L T V N G P R K
caggctcttggccctgagcgcaccattcccatcaccctggaagagaagcctgctgtcacc
Q V S G P E R T I P I T R E E K P A V T
gcagcccccaagaaatagatgcccttcttgaattgcatttttttaaacaagaaagtctc
A A P K K - M P F L E L H F L K Q E S F
cccaccagtgaatgaaagtcttgtgactagtgtgtaagcttgcattgctgcaggtcgact
P T S E - K S C D - C - S L H A C R S T
ctagaggatccccgggtaccgagctcgaattcatcgatgatatcagatccggctgctaac
L E D P R V P S S N S S M I S D P A A N
aaagcccgaaggaagctgagttggctgctgccaccgctgagcaataactagcataacc
K A R K E A E L A A A T A E Q - L A - P
cttggggcctctaaacgggtcttgaggggttttttgctgaaaggaggaactatatccgga
L G A S K R V L R G F L L K G G T I S G
tatccacaggacgggtgtggctgcctatgatcgcgtagtcgatagtggtccaagtancga
Y P Q D G C G R H D R V V D S G S K X R
aacgagcaggactgggcggcggccaaaagcggctcggacantgctccnaanaacggggggc
N E Q D W A A A K S G R T X L X X T G G
ccatanaaattgcatcaacgcatatanncgctanantcgactgcatnaatgantcngcca
P X K L H Q R I X X L X S T A X M X X P

```

A4.1.3 hIAPP

A4.1.3.1. WT hIAPP

Vector: pTXB1

Constructed by: Maxim Grechko, Sean Moran and Tianqi Zhang

```

tgaggaattncctctagaatattttgtttaactttaagaaggagatatacatatgaaatgc
E E X P L E Y F V - L - E G D I H M K C
aacactgccacttgtgcaacgcagcgcctggcaaatttttagttcattccagcaacaac
N T A T C A T Q R L A N F L V H S S N N
tttgggtgccattctctcatctaccaacgtgggatccaatacatactgcatcacgggagat
F G A I L S S T N V G S N T Y C I T G D
gcactagttgcacctaccgagggcgagtcggtacgcacgcgcgacatcgtgccgggtgcg
A L V A L P E G E S V R I A D I V P G A
cggcccaacagtgacaacgccatcgacctgaaagtcttgaccggcatggcaatcccgtg
R P N S D N A I D L K V L D R H G N P V
ctcgccgaccggctgttccactccggcgagcatccgggtgtacacgggtgcgtacggtcgaa
L A D R L F H S G E H P V Y T V R T V E
ggtctgctgtgtgacgggcaccgcgaaccaccgcttgttgtgtttggtcgacgtcncggg
G L R V T G T A N H P L L C L V D V X G
gtgccgaccctgctgtggaagctgatcgacgaaatcaagccgggcgattacgcggtgatt
V P T L L W K L I D E I K P G D Y A V I
caacgcagcgcattcagcgtcgactgtgcaggttttgcccgcgggaaaccgaaatttgcg
Q R S A F S V D C A G F A R G K P E F A
cccacaacctacacagtcggcgtccctggactgggtgcgtttcttgaagcacaccaccga
P T T Y T V G V P G L V R F L E A H H R
gaccgggacgcccagctatcgccgacgagctgaccgacgggcggttctactacgcgaaa
D P D A Q A I A D E L T D G R F Y Y A K
gtcgccagtgtcaccgacgcccggcgtgcagccgggtgtatagccttcgtgtcgacacggca
V A S V T D A G V Q P V Y S L R V D T A
gaccacgcgtttatcacgaacgggttcgctcagccacgctactggcctcaccggttctgaa
D H A F I T N G F V S H A T G L T G S E
ctcangcctcacgacaaatcctgggtgatccgcttggcaggtcaacacagcttatactgc
L X P H D K S W C I R L A G Q H S L Y C
gggaacaattgggttcacatataacggncnagacgtataaatgttttgcagccccacacc
G N N W F T Y N X X D V - M F C S P T P
tccttggcaggatgggaaacccttccaacgnttcctgccttgtggggcanncttnaatgac
S L A G W E T L P T X P A L W A X X N D

```

A4.1.3.2. hIAPP L12P

Vector: pTXB1

Constructed by: Ariel Alperstein based on the original hIAPP WT plasmid

natnagnaaattccctctagaaataatTTTgTTtaactTTtaagaaggagatatacatatg
 X X X I P S R N N F V - L - E G D I H **M**
 aaatgcaacactgccacatgtgcaacgcagcgcccggcaaattTTTtagttcattccagc
 K C N T A T C A T Q R P A N F L V H S S
 aacaactTTTggtgccattctctcatctaccaacgtgggatccaatacatactgcatcagc
 N N F G A I L S S T N V G S N T Y C I T
 ggagatgcactagttgccctacccgagggcgagtcggtacgcacatcgccgacatcgtgccg
 G D A L V A L P E G E S V R I A D I V P
 ggtgcgcgggcccaacagtgacaacgccatcgacctgaaagtccttgaccggcatggcaat
 G A R P N S D N A I D L K V L D R H G N
 cccgtgctcgccgaccggctgTTccactccggcgagcatccgggtgtacacgggtgctgacg
 P V L A D R L F H S G E H P V Y T V R T
 gtcgaaggtctgctgtgacgggcaccgcgaaccaccgTTgTTgTgTTTggtcgacgTc
 V E G L R V T G T A N H P L L C L V D V
 gccggggtgccgaccctgctgtggaagctgatcgacgaaatcaagccggggcgattacgCG
 A G V P T L L W K L I D E I K P G D Y A
 gtgattcaacgcagcgcatcagcgtcgactgtgcaggtTTTgcccgcgggaaacccgaa
 V I Q R S A F S V D C A G F A R G K P E
 tTTgCGcccacaacctacacagtcggcgTccctggactggtgCGtTTcttTggaagcacac
 F A P T T Y T V G V P G L V R F L E A H
 caccgagaccCGgacgcccagctatcgccgacgagctgaccgacggggcggttctactac
 H R D P D A Q A I A D E L T D G R F Y Y
 gCGaaagtcGCCagTgtcaccgacgCCggcgTgcagcCGgtgtatagccttCGgtgTcGac
 A K V A S V T D A G V Q P V Y S L R V D
 acggcagaccacgCGtTTatcacgaacgggttcgTcagccacgctactggcctcaccggt
 T A D H A F I T N G F V S H A T G L T G
 ctgaactcaggcctcagcagaaaatcctggtgtatccgcttggcaggtcaacacagcttat
 L N S G L T T N P G V S A W Q V N T A Y

A4.1.3.3. hIAPP V32P

Vector: pTXB1

Constructed by: Kaarin Evens based on the original hIAPP WT plasmid

```
atTTTgTTTactTTtagaaggagatatacatatgaaatgcaacactgccacatgtgcaacg
I L F T L E G D I H M K C N T A T C A T
cagcgcctggcaaatttttagttcattccagcaacaactttggtgccattctctcatct
Q R L A N F L V H S S N N F G A I L S S
accaaccgggatccaatacatactgcatcacgggagatgcactagttgccctaccgag
T N P G S N T Y C I T G D A L V A L P E
ggcgagtcggtagcatcgccgacatcgtgcccgggtgcgcggcccaacagtgacaacgcc
G E S V R I A D I V P G A R P N S D N A
atcgacctgaaagtccttgaccggcatggcaatcccgtgctcgccgaccggctgtccac
I D L K V L D R H G N P V L A D R L F H
tccggcgagcatccgggtgtacacgggtgcgtacgggtcgaaggtctgcgtgtgacgggcacc
S G E H P V Y T V R T V E G L R V T G T
ggaaccaccggttgtgtgtttggtcgacgtcgccgggtgccgaccctgctgtggaag
A N H P L L C L V D V A G V P T L L W K
ctgatcgacgaaatcaagccgggcgattacgcgggtgattcaacgcagcgcattcagcgtc
L I D E I K P G D Y A V I Q R S A F S V
gactgtgcaggttttggccgcgggaaaccgaatttgcgcccacaacctacacagtcggc
D C A G F A R G K P E F A P T T Y T V G
gtccctggactggtgcgttttcttgggaagcacaccaccgagaccgggacgcccagctatc
V P G L V R F L E A H H R D P D A Q A I
gccgacgagctgaccgacgggcggttctactacgcgaaagtgcgccagtgtcaccgacgcc
A D E L T D G R F Y Y A K V A S V T D A
ggcgtgcagccgggtgtatagccttcgtgctgacacggcagaccacgcggttatcacgaac
G V Q P V Y S L R V D T A D H A F I T N
gggttcgtcagccacgctactggcctcaccgggtctgaactcaggcctcacgacaaatcct
G F V S H A T G L T G L N S G L T T N P
```

A4.1.3.4. hIAPP L12P V32P

Vector: pTXB1

Constructed by: Tianqi Zhang based on the original hIAPP WT plasmid

```

atTTTgTTTactTTtagaaggagatatacatatgaaatgcaacactgccacatgtgcaacg
I L F T L E G D I H M K C N T A T C A T
cagcgcCCcgGcaaatttttagttcattccagcaacaactttggtgccattctctcatct
Q R P A N F L V H S S N N F G A I L S S
accaaccCGggatccaatacatactgcatcacgggagatgcactagttgccctaccCGag
T N P G S N T Y C I T G D A L V A L P E
ggcgagtcGGtacgcatcgccgacatcgtgCCgggtgcgcggcccaacagtgacaacGCC
G E S V R I A D I V P G A R P N S D N A
atcgacctgaaagtccttgaccGGcatggcaatccCGtgctcgccgaccGGctgttccac
I D L K V L D R H G N P V L A D R L F H
tccggcgagcatccGGgtgtacacGGtgCGtacGGtcgaaggTctgCGtgTgacGGgcacc
S G E H P V Y T V R T V E G L R V T G T
gCGaaccaccCGttgTtGtGtttGGtcgacgTcgCCgggtGCCgaccctgctgTggaag
A N H P L L C L V D V A G V P T L L W K
ctgatcgacgaaatcaagccGGgcgattacgCGgtgattcaacgcagcgcattcagcGtc
L I D E I K P G D Y A V I Q R S A F S V
gactgtgcaggttttGCCgcGGgaaaccCGaatttgcGCCcacaacctacacagtcGGc
D C A G F A R G K P E F A P T T Y T V G
gtccctggactGGtgCGtttcttGgaagcacaccCGagaccCGgacGCCcaagctatc
V P G L V R F L E A H H R D P D A Q A I
gccgacgagctgaccgacGGgCGgttctactacgCGaaagtCGccagTgtcaccgacGCC
A D E L T D G R F Y Y A K V A S V T D A
ggcgtgcagccGGgtgtatagccttCGtgTcgacacGGcagaccacgCGtttatcCGaac
G V Q P V Y S L R V D T A D H A F I T N
gggttCGtcagccacgctactGGcctcaccGGtctgaaactcaggcctcagcacaatcct
G F V S H A T G L T G L N S G L T T N P

```

A4.1.3.5. Histag WT hIAPP

Vector: pTXB1

Constructed by: Tianqi Zhang

tgaggaattccctctgaaatattttgtttactttaagaaggagatatacatatgcaccac
 - G I P S E I F C L L - E G D I H **M** H H
 catcatcaccacatgaaatgcaacactgccacatgtgcaacgcagcgcctggcaaatttt
 H H H H M K C N T A T C A T Q R L A N F
 ttagttcattccagcaacaactttggtgccattctctcatctaccaacgtgggatccaat
 L V H S S N N F G A I L S S T N V G S N
 acatactgcatcacgggagatgcactagttgccctacccgagggcgagtcggtacgcac
 T Y C I T G D A L V A L P E G E S V R I
 gccgacatcgtgccgggtgcgcggcccaacagtgacaacgccatcgacctgaaagtcctt
 A D I V P G A R P N S D N A I D L K V L
 gaccggcatggcaatcccgtgctcgcggatcggctggtccactccggcgagcatccgggtg
 D R H G N P V L A D R L F H S G E H P V
 tacacgggtgcgtacggtcgaaggtctgcgtgtgacgggcaccgcgaaccaccggtggtg
 Y T V R T V E G L R V T G T A N H P L L
 tgtttggtcgcagtcgccggggtgccgaccctgctgtggaagctgatcgacgaaatcaag
 C L V D V A G V P T L L W K L I D E I K
 ccgggcatcgcggtgattcaacgcagcgcattcagcgtcgcactgtgcaggttttgcc
 P G D Y A V I Q R S A F S V D C A G F A
 cgcgggaaaccgcaatttgcgcccacaacctacacagtcggcgtccctggactggtgcgt
 R G K P E F A P T T Y T V G V P G L V R
 ttcttgggaagcacaccaccgagaccgggacgcccaagctatcgccgacgagctgaccgac
 F L E A H H R D P D A Q A I A D E L T D
 gggcggttctactacgcgaaagtcgccagtgtcaccgacgcggcggtgcagccgggtgtat
 G R F Y Y A K V A S V T D A G V Q P V Y
 agccttcgtgtcgcacacggcagaccacgcgtttatcacgaacgggttcgtcagccacgct
 S L R V D T A D H A F I T N G F V S H A
 actggcctcaccgggtctgaactcaggcctcacgacaaatcctggtgtatccgcttggcag
 T G L T G L N S G L T T N P G V S A W Q

A4.1.3.6. Histag L12P V32P hIAPP

Vector: pTXB1

Constructed by Tianqi Zhang

gagnaattccctctgaataatTTTgtTTaactTTaagaaggagatatacatatgcaccac
 E X F P L N N F V - L - E G D I H **M** H H
 catcatcaccacatgaaatgcaacactgccacatgtgcaacgcagcgccccggcaaatttt
 H H H H M K C N T A T C A T Q R P A N F
 ttagttcattccagcaacaactTTTggtgccattctctcatctaccaacccgggatccaat
 L V H S S N N F G A I L S S T N P G S N
 acatactgcatcacgggagatgcactagTTgcccTaccCGaggGcgagtcggTaccgcatc
 T Y C I T G D A L V A L P E G E S V R I
 gccgacatcgtgccgggtgcgcggcccaacagtgacaacgccatcgacctgaaagtcctt
 A D I V P G A R P N S D N A I D L K V L
 gaccggcatggcaatcccgtgctcgcggaccggctgTTccactccggcgagcatccggTg
 D R H G N P V L A D R L F H S G E H P V
 tacacggTgcgtacggTcgaaggTctgcgtgtgacgggcaccgcgaaccacccgtTgTg
 Y T V R T V E G L R V T G T A N H P L L
 tgtTTggtcgcagTcgcggggTgcccgaccctgctgtggaagctgatcgacgaaatcaag
 C L V D V A G V P T L L W K L I D E I K
 ccgggcgattacgcggTgattcaacgcagcgcattcagcgtcgcactgtgcaggtTTTgcc
 P G D Y A V I Q R S A F S V D C A G F A
 cgcgggaaacccgaattTgcgcccacaacctacacagTcggcgtccctggactggtgcgt
 R G K P E F A P T T Y T V G V P G L V R
 ttctTggaagcacaccacccgagacccggacgccaagctatcgccgacgagctgaccgac
 F L E A H H R D P D A Q A I A D E L T D
 gggcggttctactacgcgaaagTcgccagTgtcaccgacgcggcgTgcagccggTgtat
 G R F Y Y A K V A S V T D A G V Q P V Y
 agccttctgTgtcgacacggcagancacgcgtttatcacgaacgggttcgTcagccacgct
 S L R V D T A X H A F I T N G F V S H A
 actggcctcaccggTctgaaactcaggcctcacgacaaatcctggTgtatccgctTggcag
 T G L T G L N S G L T T N P G V S A W Q
 gtcaacacagcttatactgcgggacaattggTcacatataacgccaagacgtataaatgt
 V N T A Y T A G Q L V T Y N A K T Y K C

A4.1.3.7. Full protein sequence of intein in pTBX1 plasmid

The intein is expressed fused to the target protein at the cysteine underlined in the sequence below. The molecular weight of the protein sequence from cysteine is 27858.3 Da. The protein sequence shown here is the sequence in original pTBX1 plasmid without gene insertion.

```

taactttaagaaggagatatacatatggctagctcgcgagtcgacggcgccgccaattc
- L - E G D I H M A S S R V D G G R E F
ctcgagggctcttcctgcatcacgggagatgcactagttgccctacccgagggcgagtcg
L E G S S C I T G D A L V A L P E G E S
gtacgcatcgccgacatcgtgcccgggtgcgcgcccaacagtgacaacgccatcgacctg
V R I A D I V P G A R P N S D N A I D L
aaagtccttgaccggcatggcaatcccgtgctcgccgaccggctgttccactccggcgag
K V L D R H G N P V L A D R L F H S G E
catccgggtgtacacgggtgcgtacgggtcgaaggtctgcgtgtgacgggcaccgccaaccac
H P V Y T V R T V E G L R V T G T A N H
cggttgttgtgtttggtcgacgctcgccgggtgcccgaccctgctgtggaagctgatcgac
P L L C L V D V A G V P T L L W K L I D
gaaatcaagccgggcgattacgcggtgattcaacgcagcgcattcagcgtcgactgtgca
E I K P G D Y A V I Q R S A F S V D C A
ggttttgcccgcgggaaacccgaatttgcgcccacaacctacacagtcggcgctccctgga
G F A R G K P E F A P T T Y T V G V P G
ctgggtgcgtttcttgaagcacaccaccgagaccgggacgccaagctatcgccgacgag
L V R F L E A H H R D P D A Q A I A D E
ctgaccgacgggcggttctactacgcgaaagtcgccagtggtcaccgacgcccggcggtgcag
L T D G R F Y Y A K V A S V T D A G V Q
ccgggtgatagccttcgtgctcgacacggcagaccacgcgtttatcacgaacgggttcgctc
P V Y S L R V D T A D H A F I T N G F V
agccacgctactggcctcaccgggtctgaactcaggcctcacgacaaatcctgggtgatcc
S H A T G L T G L N S G L T T N P G V S
gcttggcaggtcaacacagcttatactgcccggacaattgggtcacatataacggcaagacg
A W Q V N T A Y T A G Q L V T Y N G K T
tataaatgtttgcagccccacacctccttggcaggatgggaaccatccaacggttcctgcc
Y K C L Q P H T S L A G W E P S N V P A
ttgtggcagcttcaatgactgcaggaaggggatccggctgctaacaagcccgaagggaa
L W Q L Q - L Q E G D P A A N K A R K E

```

A4.2 Peptides

Below is a table of the peptides, with their form (purified or crude, solid or in solution), and the person the samples are given to, or where they are stored. Most of these peptides do not have isotope labels unless stated.

Name of the peptide	Isotope label	# of AA	Sequence	Form	Given to/Stored at
Unlabeled, amidated FGAIL	N/A	5	FGAIL (-NH ₂)	Purified, powder	-20°C 8 th F
Amidated FGAIL, ¹³ C Labeled F	1- ¹³ C label on F	5	¹³ C FGAIL	Purified, powder	-20°C 8 th F
Unlabeled, Free-FGAIL	N/A	5	FGAIL (-COO ⁻)	Purified, powder	-20°C 8 th F
Amidated CFGAILSS	N/A	8	CFGAILSS (-NH ₂)	Crude, powder	-20°C 8 th F
Amidated NNFGAILSS	N/A	9	NNFGAILSS (-NH ₂)	Purified, powder	-20°C 8 th F
Amidated NNFGAILSS	N/A	9	NNFGAILSS (-NH ₂)	On resin	-20°C 7 th F
Acetylated A6K	N/A	7	(Acetyl-) AAAAAAK (-COO ⁻)	Purified, in HFIP-d	Josh Ostrander
Acetylated A6K	N/A	7	(Acetyl-) AAAAAAK (-COO ⁻)	Crude, powder	-20°C 8 th F
¹³ C hIAPP WT	¹³ C All	37	Expressed		
¹³ C hIAPP L12P V32P	¹³ C All	37	Expressed	Purified, powder	-20°C 8 th F
¹² C WT full length αB-crystallin		175	Expressed	Purified, powder	-20°C 8 th F

A5. Appendix 5

List of publications

1. **Tianqi O. Zhang**, Ariel M. Alperstein and Martin T. Zanni. “In vitro and in vivo fiber formation using recombinant crystallin proteins and porcine lens tissues.” *Submitted*. **2016**.
2. Yoshiaki Yano, Kotaro Kondo, Yuta Watanabe, **Tianqi O. Zhang**, Jia-Jung Ho, Shinya Oishi, Nobutaka Fujii, Martin T. Zanni, and Katsumi Matsuzaki. “GXXXG-mediated parallel and antiparallel dimerization of transmembrane helices and its inhibition by cholesterol: single-molecule FRET and 2D IR studies”. *Angewandte Chemie*. *Submitted*. **2016**
3. **Tianqi O. Zhang**, Maksim Grechko, Sean D. Moran and Martin T. Zanni. “Isotope-Labeled Amyloids via Synthesis, Expression, and Chemical Ligation for Use in FTIR, 2D IR, and NMR Studies.” *Protein Amyloid Aggregation, Methods and Protocols*. **2016**. Vol 1345, pp 21-41. DOI 10.1007/978-1-4939-2978-8-2
4. Gary K. L. Chan, Andrzej Witkowski, Donald L. Gantz, **Tianqi O. Zhang**, Martin T. Zanni, Shobini Jayaraman, and Giorgio Cavigliolo. “Myeloperoxidase-mediated Methionine Oxidation Promotes an Amyloidogenic Outcome for Apolipoprotein A-I.” *J. Biol. Chem.* **2015**. DOI 10.1074/jbc.M4.63.442
5. Ghosh, Jia-Jung Ho, Arnaldo L. Serrano, David R. Skoff, **Tianqi O. Zhang**, Martin T. Zanni. “Two-dimensional sum-frequency generation (2D SFG) spectroscopy: Summary of principles and its application to amyloid fiber monolayers.” *Faraday Discuss.* **2015**. DOI 10.1039/C4FD00173G
6. **Tianqi O. Zhang**, Lauren E. Buchanan, Martin T. Zanni. “Insights into Amylin Aggregation by 2D IR Spectroscopy.” *Biomedical Spectroscopy and Imaging*. **2014**, 3, pp189-196. DOI 10.3233/BSI-140078
7. Ivan Peran, Tracey Oudenhoven, Ann Marie Woys, Matthew D. Watson, **Tianqi O. Zhang**, Isaac Carrico, Martin T. Zanni, and Daniel P. Raleigh. “A General Strategy for the Bio-orthogonal Incorporation of Strongly Absorbing, Solvation Sensitive Infrared Probes into Proteins.” *J. Phys. Chem. B*. **2014**, DOI: 10.1021/jp5008279
8. Sean D. Moran *, **Tianqi O. Zhang** *, Martin T. Zanni. “An Alternative Structural Isoform in Amyloid-like Aggregates Formed from Thermally Denatured Human γ D-crystallin.” *Protein Science*. **2014**, 23, pp321-331. DOI: 10.1002/pro.2422 (*equal contribution)
9. Sean D. Moran, **Tianqi O. Zhang**, Sean M. Decatur, Martin T. Zanni. “Amyloid Fiber Formation in Human γ D-Crystallin Induced by UV-B Photodamage.” *Biochemistry*. **2013**, 52, pp 6169-6181. DOI:10.1021/bi4008353.
10. Alfonso R. Lam, Sean D. Moran, Nicolas K. Preketes, **Tianqi O. Zhang**, Martin T. Zanni, Shaul Mukamel. “Study of the γ D-Crystallin Protein Using Two-Dimensional Infrared (2DIR) Spectroscopy: Experiment and Simulation.” *J. Phys. Chem. B*, **2013**, 117 (49), pp 15436–15443. DOI: 10.1021/jp405159v