Protein secondary structure identification *in vitro* and *ex vivo* using 2D IR spectroscopy: kinetics and imaging.

By

Sidney Steven Dicke

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

DOCTOR OF PHILOSOPHY

(CHEMISTRY)

at the

UNIVERSITY OF WISCONSIN-MADISON

2023

Date of final oral examination: March 22, 2023

This dissertation is approved by the following members of the final oral committee:

Martin T. Zanni, Professor, Chemistry

Andrew R. Buller, Professor, Chemistry

Lloyd M. Smith, Professor, Chemistry

Edwin L. Sibert III, Professor, Chemistry

Abstract

Amyloidogenic proteins are implicated in over 20 human diseases. These proteins can fold into native functional structures, oligomeric intermediate structures, and extended amyloid β-sheet structures. Human islet amyloid polypeptide (hIAPP) is a 37-residue amyloidogenic protein that forms amyloid plaques during type II diabetes. These mature plaques are non-toxic; it is the oligomeric intermediate structures that disrupt cellular function and lead to pancreatic β-cell death. The scientific data presented within shows that the toxic fold ("FGAIL intermediate") of hIAPP is preserved when in contact with membrane bilayers and that this toxic fold can be stabilized using point mutations. The technical data presented within displays how protein structures are readily identified from exhumed animal tissue via 2D IR microscopy and the number of overexpressed gene segments in a type-II diabetic animal model is quantifiable using a map-then-capture sequencing approach. Future directions are presented in Chapter 7. The studies provide information on the toxic intermediates of hIAPP and one method to stabilize a toxic intermediate of hIAPP for drug screening efforts, as well as two novel experimental techniques with application in animal models.

Acknowledgements

My brothers, Sam and Tommy, have always been my best friends. They are extremely supportive and throughout graduate school I have sought their help and advice on many different occasions. I talk to both of them on the phone at least once a week and am glad that neither of them live very far away. They are so good at cheering me up and I always look forward to any vacation or weekend I get to see them.

It's been fun to watch Sam get into comedy these last few years. I always look forward to coming down for a weekend to go see a show, play basketball at Humboldt park, and hang out in Chinatown. He even started doing shows in Madison in my final year of grad school. Can't wait to wreck men's leagues together at DePaul.

Tommy always impresses me when we meet up, often for the Birkie trail run in September, basketball, or skiing over the holidays. Early in grad school, it was fun to watch your videos about dunking and, more recently, the bouldering videos. Tommy has an incredible ability to make the most of any situation. When I moved to Madison, Tommy saw that I was bummed about my tiny new apartment. After he stayed here for a week, he convinced me to believe in his mantra "the city is your oyster." After 5 years, Tommy was definitely right. Madison has lots to offer (although I still plan to move to a larger apartment). I am excited to spend more time with you both now that I am finally graduating.

My parents have also been very supportive all of my academic and extracurricular pursuits throughout my life. They usually believe in me more than I do. It is so fun to see you guys and hang out on holidays. We all seem to have caught this

cross country skiing bug a few years back and it's been a lot of fun to learn a new sport with you.

I definitely need to thank Pete Mooney from the Hoofer Sailing club. Pete has been an incredible friend ever since that first race last summer where we somehow managed to get the chute stuck under the boat on the second downwind leg. Since that day, by far the most hours I have spent with anyone in Madison have been hanging out with Pete. By the time I am writing this, the two of us have probably lost hundreds of dollars to the Qdoba on State Street. Mom, Pete always makes sure I ask for 3 scoops of tomatoes. It's good honest food at a great price. I was a bit late getting into windsurfing but have finally come around. In my defense, it is hard to sail every boat at once while also holding onto your job. Pete has a lot of wisdom that he occasionally shares, but most of his stories have no moral at all. With Pete, I have sailed hours and hours, got sunburned a billion times, and even attended some organ concerts. Only got the cops called once. Pretty good.

I have spent so many hours at the lakefront that I would be remiss not to mention some additional friends to whom I owe a lot. Keith Pierce who is an old friend of Pete's is an excellent sailor. I'd sail with Keith any day even though he'd probably rather teach tech lessons. Andy Lager, who trolls around Mendota in a vintage wooden Chris-Craft. Andy puts in a good word for us in Church (or so I'm told). Charlie Shakett, scow fleet captain and great with knots, got me in on some early scow races last season after the fleet was nearly lost. Charlie is a passionate sailor; I was surprised to learn after a race last season that he had actually dislocated his shoulder halfway through. Gretchen Ladd, when I first met Gretchen she was a horribly sunburnt die hard looking for

something to do after a stint in the engine room on a huge ocean barge. Now she runs the whole club. Jithanie Guinasera, who let me race with the women's team a few times under my secret women's alias, "Sidney." I think the women liked having me along. In the 21st century gender doesn't seem to mean much anyways. Ben Whiddon, mostly into wingfoiling now but I wouldn't mind sailing with him again in the future. Greg Stroupe, I guess, but he's mostly into dinghies. Patrick and Mike, both windsurfers, I can never remember which name goes with which person. You were right about 2 mm neoprene gloves as a one way ticket to forearm blowout in good wind; left me stranded at one of the frat docks (although at least my hands were nice and warm).

There are so many fun things to do in Madison that I can barely think of them all here. The UW Hoofers Nordic ski race team has been an additional family for me during the last 2 years. Ski racing is exhilarating and I have had my share of highs and lows. It is so much fun to always have a hot meal, a lukewarm Hammosa, and a floor to sleep on after the race. Keep sticking it to the Gophers. I hope to stay in touch with you all for future ski trips. Owen Gage, Levi Smith, Carl Branch, Anna Klein, Sophie Hicks, Katherine Koeler, thanks for always making sure I was skiing in the right direction. I have learned so much from all of you.

Of course I am also a Chemistry graduate student and need to thank everyone who has been a part of my graduate school experience. Most of my roommates are from the chemistry program and were some of my original friends in Madison. Jake and I made fast friends on our visit weekend and have remained friends throughout our time here. We put a ton of miles on some crappy old bikes during the shutdown, and both of us still bike a ton in the summer (helps that Madison has such great bike trails). Eli (who

I dragged along into the ski club), and Austin who I don't see too much anymore (but I look forward to your wedding this summer), it's been great to hang out with you the last 5 years.

Cole Sagan from the Garand group was always up for hoops, a passion I have really missed in graduate school (I have rarely played more than once or twice a week). Cole is a great basketball player with a lot of fire in pickup games. A funny thing about Cole is that when someone happens to ask about his work between games, he is not afraid to bring up deep space triplet excited states in the first sentence. I don't know how to say this to Cole, but us basketball players struggle just keeping score in games to 11. Maybe the gamers at Princeton can keep score, talk science, and chew bubble gum at the same time, but you'll have to report back to me on this one. Anyways, when Cole graduated this past January, I was super bummed out. But it helped me realize it was time to get going on my thesis.

My co workers are a huge part of my life day to day. Andy Jones early on, and later Matt Ryan have been excellent running buddies. Andy Jones also used to ski the Birkie, and it's always nice to have a friend in a race (now I know too many Birkie skiiers to count). Matt is an excellent scientist and is a blast to hang out with. It's great to start or end the workday with a run. Exploring Texas on early morning runs is still one of my favorite conference memories. Come hang out with Nan and I in Chicago once the grad school workload eases up a bit. And remember that it's just school.

Ryan and Kieran joined the group at the same time I did, and they have been really fun to work and hang out with in the office. It helps to have people moving through the same program at the same time, facing similar challenges and stresses. I can even

lump Cat into this category, even though she was a year older, I feel like we spent most of grad school hanging out. Friends at work have made graduate school way more enjoyable. Nan showed up at the end of my 3rd year and took all of us in stride. He helped me explore the world of cryptocurrencies and new cuisines (neither of which helped me make any money).

The older generation of graduate students including Cat, Jessie, and Kacie, taught me a bunch and helped me feel at home in the group. It is crazy how fast 5 years goes by, many of the students I worked hours and hours with have now been gone for years. But it warms my heart to see the excitement for our research in the attitudes of the newest students to join the group. Brynn, Harrison, Shivani, and Dasol all remind me of my early days in graduate school, you all have a super bright future.

I would like to give a big thank you to Mary Hanson, the physical chemistry path coordinator and travel coordinator for students in Marty's group. I went to a couple of conferences overseas and would inevitably end up getting lost or needing to check in to a nice hotel without floating my own credit. Mary was always ready to help in these situations, no matter the day of the week (weekends included), and no matter what time zone I was in. Additionally, Mary has been a great friend to myself and the whole group these last 5 years—she is a blast to have in the department. I wish I could stop by for coffee as much as I did during my first two years of graduate school. Best of luck in the years to come.

Last but certainly not least I need to thank my advisor, Marty, for running such a fun research group. Marty encourages students to explore all of the activities Madison has to offer. He is also super inclusive. I met Marty outside of an elevator and within 30

seconds he had invited me out to happy hour, dinner, and pool that evening. I thought “Finally, a faculty member I can actually get along with.” He runs a research group where you get to stumble through things on your own, but he is always available to help or jump in the lab if necessary. I don’t really know how, because we don’t always show up to recruiting weekends, but he seems to get all of the sharpest students. He also has a great “teacher mode” about fundamental topics and is never too busy to jump to the whiteboard and explain some funny physics thing like mode-locking. I have met undergraduates on the ski team who really liked having Marty as a general or physical chemistry professor, and I can see why.

Because we are a very social research group, the shutdown had a lasting impact—students saw each other and talked to Marty much less than was the norm when I joined the group. But after a few post-COVID conferences and group outings, I think everything is trending in the right direction.

Table of Contents

Contents

List of Figures

1 Introduction

1.1 Protein structure and function.

Proteins are long chains of amino acids responsible for many biological functions, including the transport of oxygen in our bloodstream,[1] the selective uptake and release of calcium and potassium across membranes,[2,3] and translating DNA.[4,5] They also act as signaling molecules, where the binding of one protein changes the conformation of a receptor (typically a single protein or a group of associated proteins).[6,7] Proteins fold to three dimensional shapes, referred to as a protein's structure. The native structure of a protein dictates its functionality and the loss of this structure results in an altered or dysfunctional state.[8] **Figure 1.1** depicts a protein displaying functional structure and misfolded β-sheet structure.

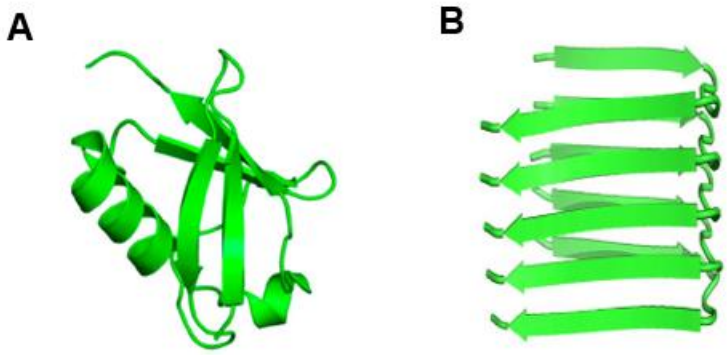


Figure 1.1: Native and aggregated protein structure.

(A) Native functional structure. (B) Loss of native structure; an aggregate rich in β-sheets.

The sequence of amino acids that compose a protein is referred to as the primary structure of that protein.[9] Amino acids located next to each other within this

sequence form shapes familiar between proteins. These shapes, known as the secondary structural elements of a protein, take the form of α-helices, β-sheets, or random coil.[9] The entire three dimensional shape of a protein composed of these secondary structural elements is referred to as tertiary structure. When multiple subunits that are not covalently attached (i.e. chemically bonded) fold together, this is referred to as globular or quaternary structure.[9] If a protein is short (<50 amino acid residues) it is often referred to as a peptide. **Figure 1.2A** displays how single amino acids are represented by the primary sequence.[10,11] **Figure 1.2B** displays the three common secondary structural elements the amino acids may form within a tertiary structure,[12] and **Figure 1.2C** shows hemoglobin, an oxygen-transporting tetrameric globular protein (the subunits, uniquely colored, are not covalently bonded).[13]

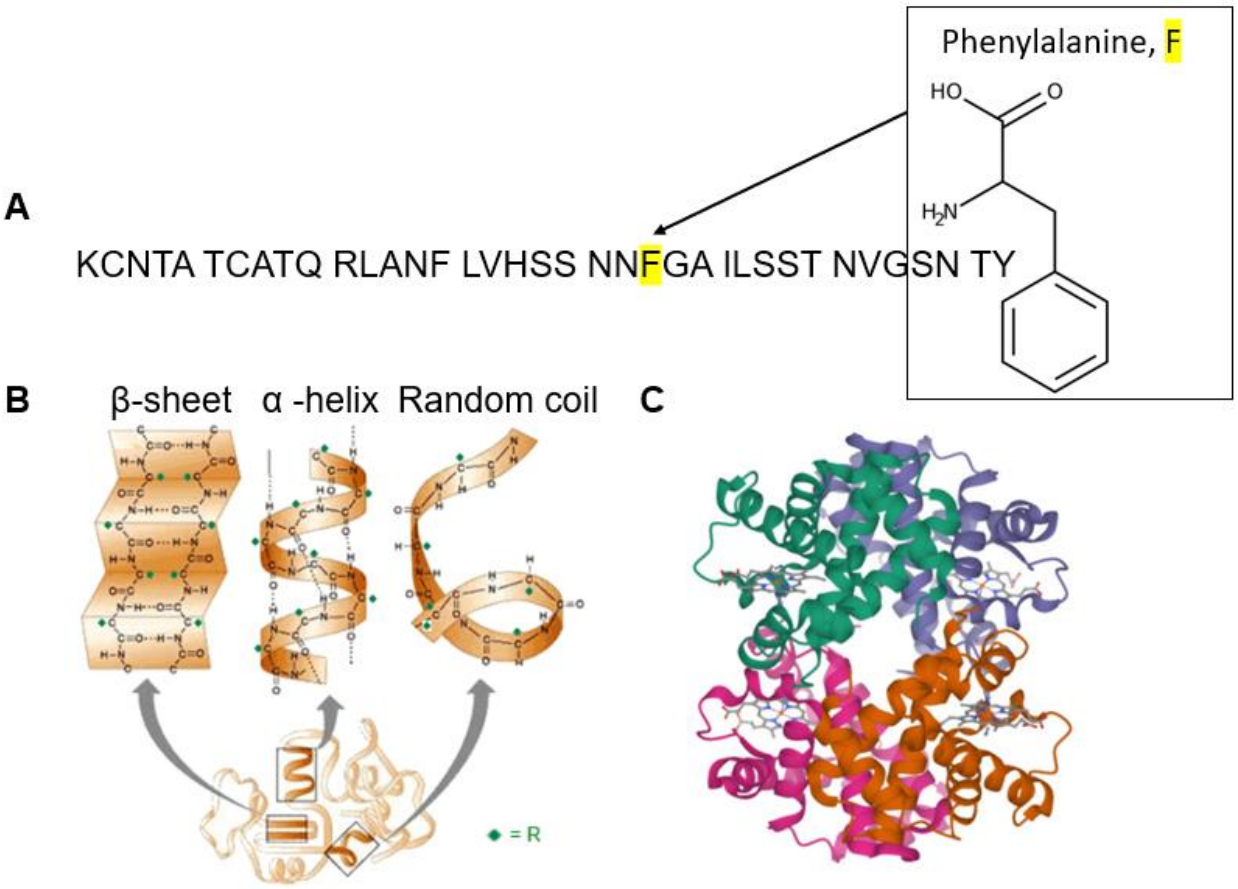


Figure 1.2: Primary, secondary, tertiary, and quaternary protein structure.

(A) The primary structure of human islet amyloid polypeptide represented by single-letter codes.[10] Shown is phenylalanine, a single amino acid, abbreviated F in the primary sequence. (B) Secondary structure elements β-sheet, α-helix, and random coil contribute to the tertiary structure of a protein.[11] (C) Quaternary or globular structure of the protein hemoglobin, responsible for oxygen transport in the blood (tertiary subunits distinguished by color).[13]

The misfolding of proteins is common in diseases.[14,15] One class of these misfolding proteins, termed amyloid proteins, are able to re-fold from their native state into extended β-sheet structures.[14] These extended β-sheet structures are also known as amyloid plaques or amyloid fibrils and offer an altered or lost function from the native fold of the protein.[16–18] There are 37 known amyloid-forming proteins associated with disease.[19] One focus of this thesis is to investigate the various folds of one such

amyloid protein, human islet amyloid protein (hIAPP or amylin), associated with type II diabetes.[10]

1.2 Pancreas function, amylin, and type II diabetes.

The human pancreas plays an important role in metabolic function. It is composed of many cell types which may be grouped into two categories: the exocrine pancreas, which secretes digestive enzymes through a network of ducts, and the endocrine pancreas, secreting hormones directly into the bloodstream.[20] β-cells located within the so-called islets of Langerhans are part of the endocrine pancreas. When food is consumed, β-cells co-secrete hormones (in this instance, the hormones are proteins) insulin and amylin. Insulin binds cell surface receptors in the liver, such as Insulin Receptor A and Insulin Receptor B, causing enhanced sugar uptake from the bloodstream (this is just one of insulin's many effects).[21] Amylin flows past the blood brain barrier and binds to receptors from the calmodulin family; a cascade process ensues resulting in satiation.[22,23] During type II diabetes--a form of diabetes acquired throughout life--repeated food consumption leads to β-cell secretion of excessive insulin (hyperinsulinemia)[24] and amylin (hyperamylinemia).[25] Amyloid plaques (introduced above) formed of misfolded amylin are found in over 95% of type II diabetes patients postmortem and are associated with the disease. The exact amino acid sequence of amylin changes depending on the species secreting it. There are conserved and non-conserved domains between species.[26] Species susceptibility to type II diabetes have been recently linked to the primary sequence of the amylin that the species secretes and the secondary structures this amylin may form (Thesis – Caitlyn Fields, 2022, unpublished); therefore we suspect the association between amylin and type II diabetes may be a causal association.

1.3 Amyloids and the toxic oligomer hypothesis.

Amyloid plaques have been observed in the islets of type II diabetes patients postmortem since 1901 and were long thought important to disease pathology.[27] The formation of amyloid plaques can be studied *in vitro* using a variety of techniques, such as Circular Dischrosim (CD) and infrared (IR) spectroscopy, which detect protein structures via their chirality and vibrational modes, respectively. These techniques monitor the folding of proteins in solution over the course of a few hours or days. It is observed that many amyloid and other protein assemblies follow sigmoidal or sigmoidal-adjacent kinetics (in the case of amyloids with metastable intermediates).[28–30] Amyloid kinetic diagrams have 3 main phases, shown in **Figure 1.3A**. In the lag phase, monomers (single proteins) encounter each other randomly in solution; they are in equilibrium with oligomers (two or more monomers in contact).[31] The growth phase occurs above a critical concentration specific to each amyloid protein, and above this concentration the equilibrium shifts towards oligomers. Above the critical concentration, oligomers encounter other oligomers and monomers faster than they dissociation back into monomers.[30] Thus they grow into much larger species (termed protofibrils) and ultimately reach a saturation phase where amyloid fibrils are observed (there may be additional folding events as this saturation is reached). Amyloid fibrils are thermodynamically stable species and do not dissociate back into oligomers and monomers. Due to their extreme stability, amyloid folding is an irreversible process in biological systems (although in the lab one can unfold amyloid fibrils using heat or a chemical denaturant).[14] A few (2-14) monomers contribute to each oligomer, and many oligomers contribute to each fibril. Oligomers and fibrils form recognizable shapes that

can be viewed using electron microscopes. These recognizable shapes are generally small but are not as microscopic as the secondary structure cartoon elements (α-helix and β-sheet) discussed above. For example, a circular oligomer may be 40-90 nanometers in diameter, whereas the individual amino acids are a few hundred picometers from one end to the other, about 0.1-0.2 nanometers per chemical bond. **Figure 1.3B** and 1.**3C** show TEM micrographs of circular oligomers during the lag phase and fibrillar amylin during the saturation phase, respectively.

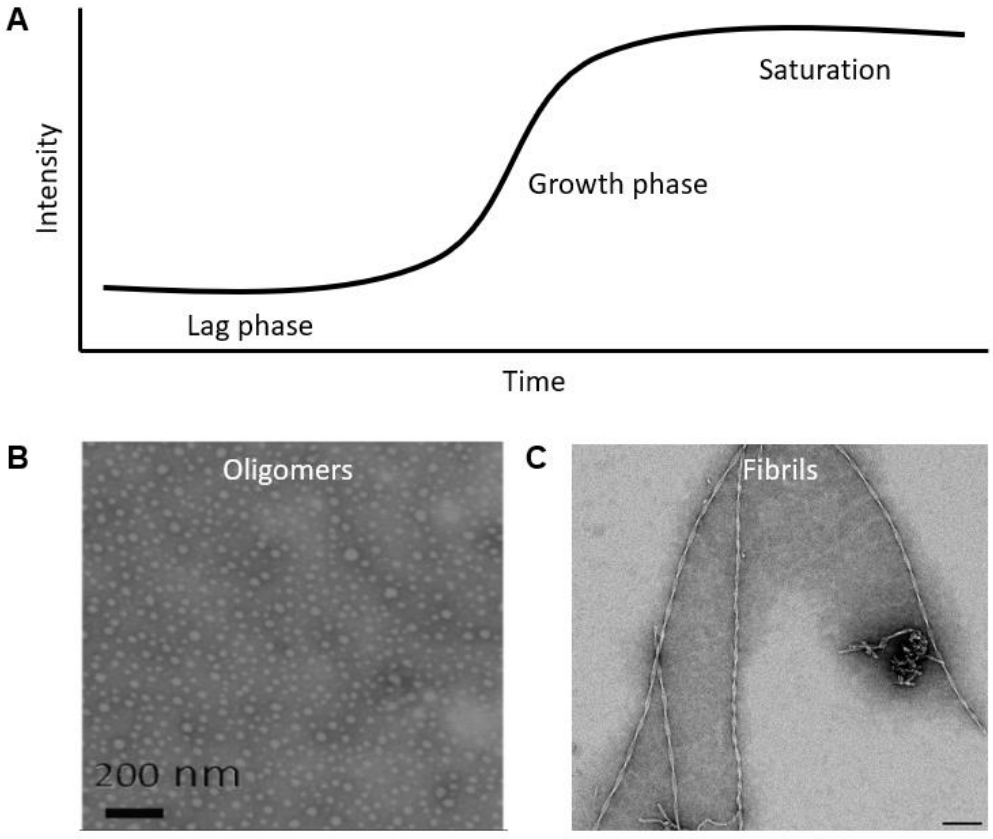


Figure 1.3: Amyloid aggregation kinetics.

(A) An example amyloid aggregation curve displaying sigmoidal kinetics, as may be observed using IR or Fluorescence intensity. (B) TEM of circular oligomers of an hIAPP mutant, 3A, collected by the Zanni group at UW-Madison SMPH facility (C) TEM of hIAPP amyloid fibrils, collected by the Zanni group at the CEMRC facility in collaboration with students from Elizabeth Wright's lab. (scale bar in figure C = 100 nm)

For many years it was assumed that amyloids, such as hIAPP, caused amyloid diseases. Recently it has been discovered that hIAPP amyloid plaques have no extracellular β-cell toxicity. Rather, “on-pathway” (during the growth phase to a mature fibril) oligomeric intermediates are capable of direct toxicity via permeabilizing the cell membrane. They have also been associated with various inflammatory mechanisms resulting in additional β-cell dysfunction and death.[32] These discoveries were foundational in developing the more recent “toxic oligomer hypothesis,” that toxic oligomers play a causal role in amyloid diseases. It is an underlying goal of this work to structurally characterize these oligomeric intermediates. A well characterized intermediate can be used as a template for pharmaceutical drug design or antibody-based therapeutics (**Figure 1.4**), but requires a deeper characterization than images from electron microscopes provide. For example, in drug targeting, the secondary structural elements and their locations must be defined. This oligomer can also be stabilized and its toxic effects can be studied in animal models, one current direction of this project (see Chapter 6).

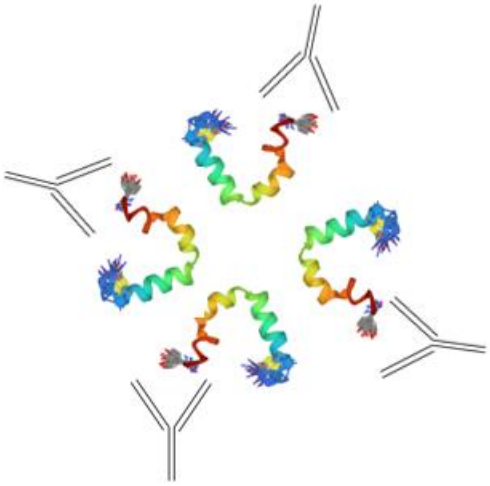


Figure 1.4: Targeting oligomers with antibodies.

Example of antibody targeting of stabilized oligomeric hIAPP tetramer. Antibodies portrayed in Y-rendition.

1.4 Previous structural characterization of hIAPP.

Using Nuclear Magnetic Resonance (NMR) spectroscopy and Cryo-Electron Microscopy (Cryo-EM), mature amyloid fibrils of hIAPP have been previously characterized. Although we are most interested in the structure of toxic intermediates, these studies can provide some insight into stable folded structures as possible final states for the aggregation mechanism. **Figure 1.5A-C** displays 3 structural models, taken from solid-state NMR, solution-state NMR, and Cryo-Electron Microscopy.

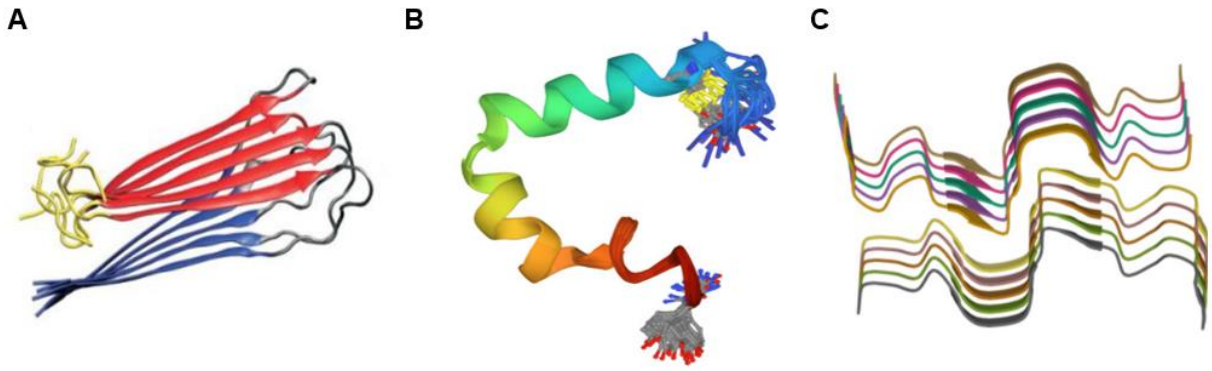


Figure 1.5: Structural characterization of hIAPP using NMR and Cryo-EM.
(A) hIAPP fibril structure solved using solid-state NMR. (B) stabilized hIAPP helix-kink-helix structure using SDS micelles solved using solution-state NMR. (C) Cryo-EM structure of hIAPP fibrils. A-C adapted from refs [33,34, and 35], respectively.

At the time of this work, these techniques (solid state NMR, solution state NMR, and Cryo-EM) have not proved capable of characterizing oligomeric intermediates. For solution-state NMR, the experimental averaging time as hIAPP evolves (changes shape) in solution is too slow, and the signal disappears after a few NMR experiments. Thus for solution state NMR experiments evolving structures must be stabilized using additional methods, such as the helix-stabilizing sodium dodecyl sulfate (SDS) micelles

approach used to solve the structure displayed in **Figure 1.5B**. This approach comes at a cost to the purpose of our protein structure study: stabilizing one particular structure conformation is not physiological; SDS micelles do not occur in animals. So antibodies raised against this conformation, when injected into a diabetic human, will probably bind nothing at all (this is also one shortfall of drug design based on non-toxic fibril structures). For solid-state NMR, experiments on soluble oligomers are only possible if they are flash-frozen, crystallized, and dried; however as with stabilized structures, this is far from physiological condition and oligomers may structurally evolve when they are dried. In Cryo-EM, the plausibility of studying a physiological sample is promising: samples are placed on a microscope grid, plunged into liquid ethane and encapsulated in vitrious ice (possibly fast enough to be frozen in native conformation). However, the smallest structure ever solved (at the time of this dissertation) is more than 50 kDa. A single hIAPP monomer is ~3.9 kDa; therefore, you would need a tridecameric (13 subunit) oligomer to have a chance at structural elucidation. Moreover, Cryo-EM relies on computer averaging of many repeating structures (or subunits, in the case of a twisting fibril); thus you would have to be able to reproducibly create the same tridecamer at a high concentration in the field of view of your microscope. This poses a further challenge: high concentrations favor fibrillization, not stable oligomers. These techniques may overcome these pitfalls in the next few decades, but this work has not yet been demonstrated.

Due to fast averaging times and bond-specific resolution, IR spectroscopy is one of the most capable techniques at studying in-solution evolving intermediates, with a trade off in global structure characterization. Typically, 1-5 isotopically labeled amino

acids are studied per IR experiment, rather than the entire protein. Previous experiments using 2D IR spectroscopy by Buchanan et al. have created a model in solution for the aggregation mechanism of hIAPP from a common metastable intermediate.[36,37]

1.5 The FGAIL intermediate of hIAPP.

In 2013, Buchanan and co-workers used two-dimensional infrared (2D IR) spectroscopy and $^{13}C^{18}O$ isotope labeling to study discrete sites within hIAPP as it evolved in solution from monomers to fibrils.[38] For the purposes of this chapter, 2D IR spectroscopy can be thought of as a black box that allows one to study the structure of one specific isotopically-labeled amino acid within a protein at a time; this technique will be discussed in greater detail in Chapter 2. The data collection and averaging of this "2D IR box" occur rapidly enough to study a species undergoing structural changes over a period of hours or days. In the previous work, an additional structure was formed during hIAPP aggregation in solution, where amino acids F-G-A-I-L form a short β-sheet prior to becoming the disordered (random coil) loop in the mature fibril. Figure 1.6 A-C displays the mechanism proposed by Buchanan and co-workers.[38] The FGAIL intermediate structure is metastable and follow-up studies discovered that hIAPP aggregation does not fit a classical sigmoidal curve. An additional stable state is needed to accurately model experimental kinetic data.[30,37] This state, referred to as the "FGAIL" intermediate, correlates well with the cytotoxicity observed in β-cell viability experiments.[37,39]

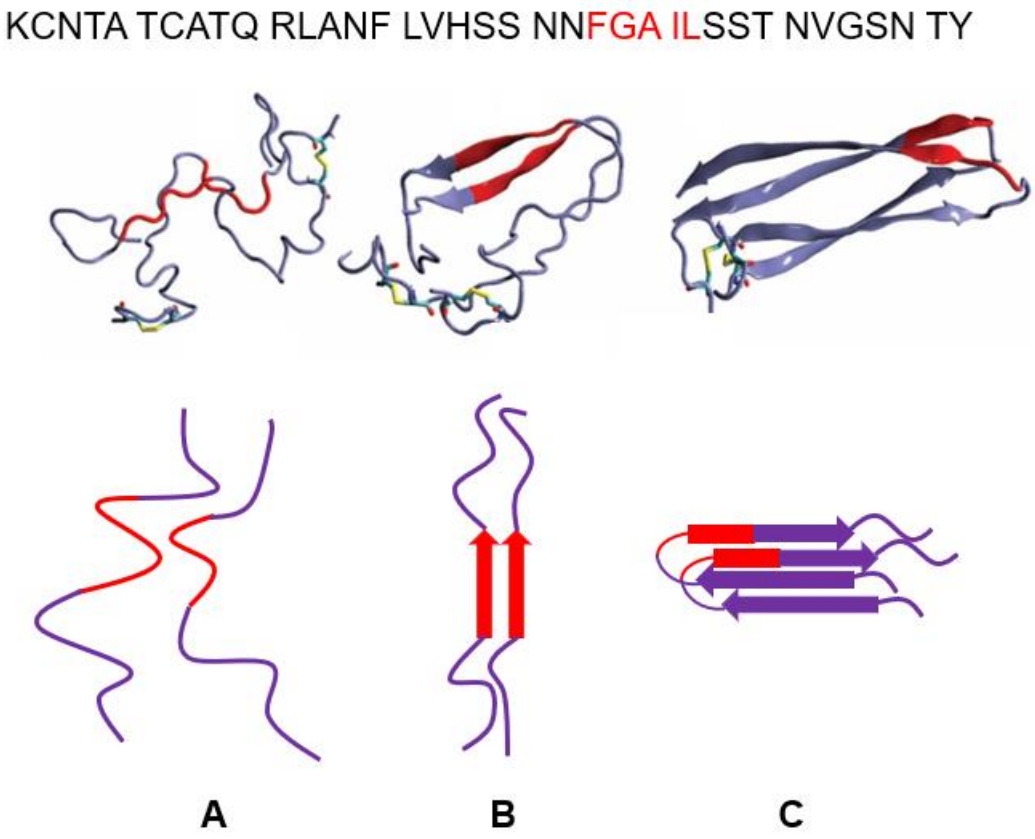


Figure 1.6: The FGAIL intermediate aggregation model.

Simplified proposed aggregation mechanism by Buchanan and co-workers. (top) Primary sequence of hIAPP with the FGAIL region highlighted in red. (A) Two monomers in close proximity in solution. (B) FGAIL β-sheet forms between two disordered regions. (C) Flanking β-sheets template out from the FGAIL region, leaving this region as a disordered loop in the fibril (as observed in fibril structures by Tycko and others). A and C are non-toxic-B is toxic to β-cells.[32,39]

As displayed in **Figure 1.7**, hIAPP is most toxic towards the end of the lag phase, when the population of oligomers in solution is the highest.

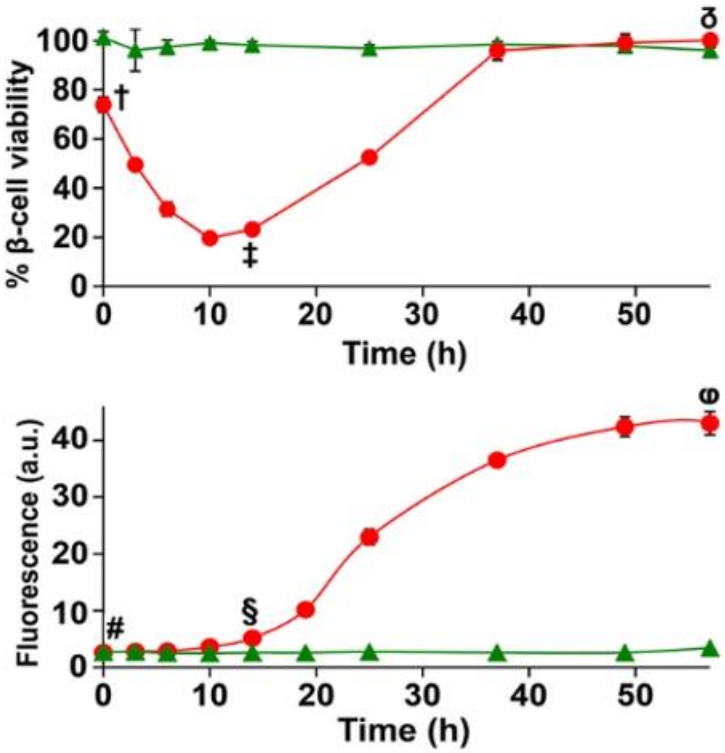


Figure 1.7: Kinetics of aggregation and cell viability.

Aggregation time course using a fluorescent probe to monitor fibril formation. The lowest β-cell viability (highest toxicity) is measured at the end of the lag phase or beginning of the growth phase, when oligomers are present in solution. Figure adapted from Abedini et al.[39]

The previous published work of Buchanan et al.,[36,37] Serrano et al.,[30,40,41] and Abedini et al.,[32,39,42,43] along with the thesis of two previous Zanni group students, Kacie Rich and Caitlyn Fields (all summarized above) provides the foundation for my graduate studies.

1.6 Thesis Overview.

The 4 projects I have worked on during graduate school are organized according to the order they were completed:

Can we characterize the protein structures occurring at different locations within an animal tissue? (Development of a moving-stage 2D IR microscope, **Chapter 3**)

How do we characterize an animal model for type II diabetes and assess if it is usable for CRISPR mutations? (Genotyping an overexpression mouse model, **Chapter 4**)

Does the FGAIL oligomer occur on membranes? (Solution phase 2D IR experiments of hIAPP with vesicles, **Chapter 5)**

Can we stabilize the toxic state of hIAPP in solution so we can get a full-length high resolution structure? (thesis and ongoing post-doctoral work, **Chapter 6**)

In **Chapter 7**, I provide a short summary and the most exciting and feasible future directions for the next generation of graduate students.

1.7 Chapter 2 References:

1. Schechter, A. N. Hemoglobin research and the origins of molecular medicine. *Blood* **112**, 3927 (2008).

2. Kratochvil, H. T. *et al.* Probing the Effects of Gating on the Ion Occupancy of the K + Channel Selectivity Filter Using Two-Dimensional Infrared Spectroscopy. (2017) doi:10.1021/jacs.7b01594.

3. Shamoo, A. E. & Ambudkar, I. S. Regulation of calcium transport in cardiac cells. *https://doi.org/10.1139/y84-002* **62**, 9–22 (2011).

4. Cozzolino, F., Iacobucci, I., Monaco, V. & Monti, M. Protein-DNA/RNA Interactions: An Overview of Investigation Methods in the -Omics Era. *J Proteome Res* **20**, 3018–3030 (2021).

5. Kozlov, A. G. *et al.* How Glutamate Promotes Liquid-liquid Phase Separation and DNA Binding Cooperativity of E. coli SSB Protein. *J Mol Biol* **434**, (2022).

6. Gerber, K. J., Squires, K. E., Hepler, J. R. & of Pharmacology KJG, D. Roles for Regulator of G Protein Signaling Proteins in Synaptic Signaling and Plasticity Programs in Molecular and Systems Pharmacology (K. *MOLECULAR PHARMACOLOGY Mol Pharmacol* **89**, 273–286 (2016).

7. Wang, M. *et al.* α2A-Adrenoceptors Strengthen Working Memory Networks by Inhibiting cAMP-HCN Channel Signaling in Prefrontal Cortex. *Cell* **129**, 397–410 (2007).

8. Kuhlman, B. & Bradley, P. Advances in protein structure prediction and design. *Nature Reviews Molecular Cell Biology 2019 20:11* **20**, 681–697 (2019).

9. Shaw, W. J. *et al.* Controls of nature: Secondary, tertiary, and quaternary structure of the enamel protein amelogenin in solution and on hydroxyapatite HHS Public Access. *J Struct Biol* **212**, 107630 (2020).

10. Akter, R. *et al.* Islet Amyloid Polypeptide: Structure, Function, and Pathophysiology. (2016) doi:10.1155/2016/2798269.

11. Amino Acids - Alanine. http://www.biology.arizona.edu/biochemistry/problem_sets/aa/phenylalanine.html.

12. The structure of proteins. http://www.chim.lu/ech1025.php.

13. Sen, U. *et al.* Crystal structures of HbA2 and HbE and modeling of hemoglobin δ4: Interpretation of the thermal stability and the antisickling effect of HbA2 and identification of the ferrocyanide binding site in Hb. *Biochemistry* **43**, 12477–12488 (2004).

14. Eisenberg, D. & Jucker, M. The amyloid state of proteins in human diseases. *Cell* **148**, 1188 (2012).

15. Chaudhuri, T. K. & Paul, S. Protein-misfolding diseases and chaperone-based therapeutic approaches. *FEBS J* **273**, 1331–1349 (2006).

16. Otzen, D. & Riek, R. Functional Amyloids. *Cold Spring Harb Perspect Biol* **11**, (2019).

17. Eisenberg, D. & Jucker, M. The amyloid state of proteins in human diseases. (2012) doi:10.1016/j.cell.2012.02.022.

18. Iadanza, M. G., Jackson, M. P., Hewitt, E. W., Ranson, N. A. & Radford, S. E. A new era for understanding amyloid structures and disease. doi:10.1038/s41580-018-0060-8.

19. Chiti, F. & Dobson, C. M. Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress Over the Last Decade. *https://doi.org/10.1146/annurev-biochem-061516-045115* **86**, 27–68 (2017).

20. Bastidas-Ponce, A., Scheibner, K., Lickert, H. & Bakhti, M. Cellular and molecular mechanisms coordinating pancreas development. *Development* **144**, 2873–2888 (2017).

21. Meyts, P. de. The Insulin Receptor and Its Signal Transduction Network. *Endotext* (2016).

22. Mohamed, L. A. *et al.* Amylin Enhances Amyloid-β Peptide Brain to Blood Efflux Across the Blood-Brain Barrier. doi:10.3233/JAD-160800.

23. Lutz, T. A. Roles of Amylin in Satiation, Adiposity and Brain Development. *Frontiers in Eating and Weight Regulation* **63**, 64–74 (2010).

24. Thomas, D. D., Corkey, B. E., Istfan, N. W. & Apovian, C. M. Hyperinsulinemia: An Early Indicator of Metabolic Dysfunction. *J Endocr Soc* **3**, 1727 (2019).

25. Ly, H. & Despa, F. Hyperamylinemia as a risk factor for accelerated cognitive decline in diabetes. *http://dx.doi.org/10.1586/14789450.2015.1104251* **12**, 575–577 (2015).

26. Noh, D., Bower, R. L., Hay, D. L., Zhyvoloup, A. & Raleigh, D. P. Analysis of Amylin Consensus Sequences Suggests that Human Amylin is Not Optimized to Minimize Amyloid Formation and Provides Clues to Factors that Modulate Amyloidogenicity HHS Public Access. *ACS Chem Biol* **15**, 1408–1416 (2020).

27. Westermark, P. Amyloid in the islets of Langerhans: Thoughts and some historical aspects. *Ups J Med Sci* **116**, 81–89 (2011).

28. Ferrone, F. A., Hofrichter, J. & Eaton, W. A. Kinetics of sickle hemoglobin polymerization: II. A double nucleation mechanism. *J Mol Biol* **183**, 611–631 (1985).

29. Xue, W. F., Homans, S. W. & Radford, S. E. Systematic analysis of nucleation-dependent polymerization reveals new insights into the mechanism of amyloid self-assembly. *Proc Natl Acad Sci U S A* **105**, 8926 (2008).

30. Serrano, A. L., Lomont, J. P., Tu, L. H., Raleigh, D. P. & Zanni, M. T. A Free Energy Barrier Caused by the Refolding of an Oligomeric Intermediate Controls the Lag Time of Amyloid Formation by hIAPP. *J Am Chem Soc* **139**, 16748–16758 (2017).

31. Morris, A. M., Watzky, M. A. & Finke, R. G. Protein aggregation kinetics, mechanism, and curve-fitting: A review of the literature. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **1794**, 375–397 (2009).

32. Abedini, A. *et al.* RAGE binds preamyloid IAPP intermediates and mediates pancreatic β cell proteotoxicity. *Journal of Clinical Investigation* **128**, 682–698 (2018).

33. Luca, S., Yau, W.-M., Leapman, R. & Tycko, R. Peptide Conformation and Supramolecular Organization in Amylin Fibrils: Constraints from Solid State NMR. *Biochemistry* **46**, 13505 (2007).

34. Nanga, R. P. R., Brender, J. R., Vivekanandan, S. & Ramamoorthy, A. Structure and membrane orientation of IAPP in its natively amidated form at physiological pH in a

membrane environment. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1808**, 2337–2342 (2011).

35. Cao, Q., Boyer, D. R., Sawaya, M. R., Ge, P. & Eisenberg, D. S. Cryo-EM structure and inhibitor design of human IAPP (amylin) fibrils. *Nature Structural & Molecular Biology 2020 27:7* **27**, 653–659 (2020).

36. Buchanan, L. E. *et al.* Structural Polymorphs Suggest Competing Pathways for the Formation of Amyloid Fibrils That Diverge from a Common Intermediate Species. *Biochemistry* **57**, 6470 (2018).

37. Buchanan, L. *et al.* Mechanism of IAPP amyloid fibril formation involves an intermediate with a transient B-sheet. *PNAS* (2013) doi:10.1073/pnas.1314481110.

38. Buchanan, L. E. *et al.* Mechanism of IAPP amyloid fibril formation involves an intermediate with a transient β-sheet. *Proceedings of the National Academy of Sciences* **110**, 19285–19290 (2013).

39. Abedini, A. *et al.* Time-resolved studies define the nature of toxic IAPP intermediates, providing insight for anti-amyloidosis therapeutics. *Elife* **5**, (2016).

40. Serrano, A. L., Ghosh, A., Ostrander, J. S. & Zanni, M. T. Wide-field FTIR microscopy using mid-IR pulse shaping. *Opt Express* **23**, 17815 (2015).

41. Ostrander, J. S., Serrano, A. L., Ghosh, A. & Zanni, M. T. Spatially Resolved Two-Dimensional Infrared Spectroscopy via Wide-Field Microscopy. *ACS Photonics* **3**, 1315–1323 (2016).

42. A, A. & DP, R. Destabilization of human IAPP amyloid fibrils by proline mutations outside of the putative amyloidogenic domain: is there a critical amyloidogenic domain in human IAPP? *J Mol Biol* **355**, 274–281 (2006).

43. Cao, P. *et al.* Islet amyloid: From fundamental biophysics to mechanisms of cytotoxicity. *FEBS Letters* vol. 587 1106–1118 Preprint at https://doi.org/10.1016/j.febslet.2013.01.046 (2013).

2 Methods:

2.1 Two-Dimensional Infrared (2D IR) Spectroscopy and Imaging.

Much of the spectroscopic results presented in this thesis are the result of interacting infrared light with samples of proteins that were either floating in solution or preserved in a wax block of paraffin.

Recall from Chapter 1 that proteins are long chains of amino acids. These amino acids contain a few different groups: an amino group, a carboxyl group, and an R group (**Figure 2.1**). Oxygen, carbon, and nitrogen contain different numbers of protons and electrons; thus they have different electronegativities. Furthermore, the electrons are mobile and slosh around. We draw these sloshing electrons on paper as a "bond" between two atoms; the bonds stretch and compress. Thus amino acids are oscillating structures along each bond axis. These interfere with each other and result in vibrational modes. Transition dipoles (just like a magnetic dipole, with a "more positive" side and a "more negative" side) exist along these modes that we can interact with using a beam of light (you could imagine this beam of light as an additional oscillating field). **Figure 2.1A** depicts a single amino acid. **Figure 2.1B** depicts an amino acid, but drawn with more geometric accuracy, and the transition dipole (black arrow) overlays the amide-I vibrational mode.[1] This mode stretches and compresses (vibrates) at different frequencies depending on the secondary structure of the protein; that is why we use this mode as a reporter to study protein folding. See how the vibrational frequency of this mode changes as the amino acids form different secondary structures in **Figure 2.2**. A pulse of light with a range of frequencies including those of the amide-I

vibration will show an absorption at the frequencies of the amide-I mode when directed through a solution of protein.

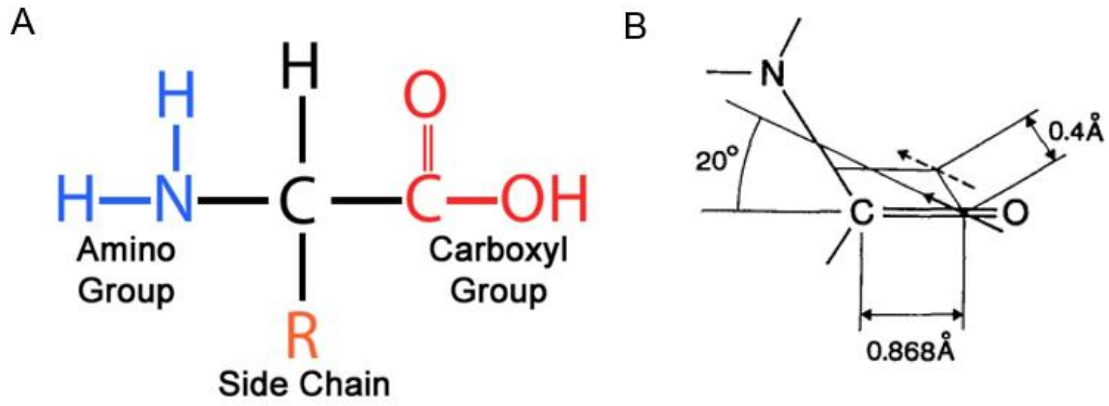
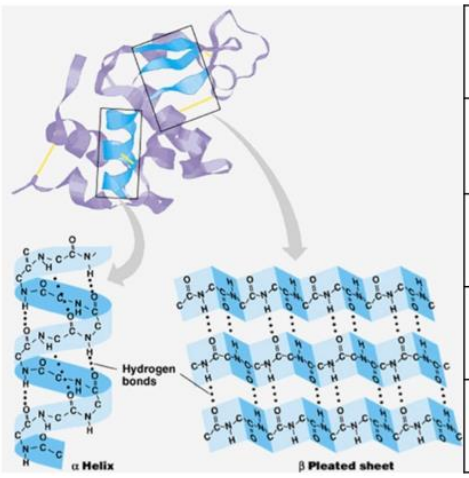


Figure 2.1: Amino acids and the amide-I transition dipole.

(A) A single amino acid containing an amino group, a carboxyl group, and a side chain ("R") group. (B) Differing electronegativities create a transition dipole, the "amide I" vibrational mode. [1]



Secondary structure element	Vibrational frequency (cm^{-1})
Random coil (single C=O oscillator)	~1660
α-helix	~1645
β-sheet	~1631
Amyloid β-sheet	~1620

Figure 2.2 Vibrational frequencies of each secondary structural element.

Carbonyl (C = O) orientations, couplings, and hydrogen bondings are unique to each protein secondary structure. [2]

[1] Image in (A) taken from: https://www.astrochem.org/sci/Amino_Acids.php, image in (B) used from reference Hamm and Zanni.[1]
[2] Image taken from: https://www.creative-proteomics.com/services/protein-secondary-structure-prediction-service.htm

2.2 The recipe for a 2D IR experiment.

As described above, to study protein samples, we need high energy pulses of light oscillating at the same frequencies as the amide-I vibrations from each secondary structure we want to observe, which occur in the mid-Infrared range. We use a commercial Solstice system to generate 1 kHz pulses of 800 nm light (the Solstice system contains Nd:YLF pump laser, a Ti:Sapph oscillator, and a regenerative amplifier). The 800 nm output is the input beam for a commercial optical parametric amplifier (OPA), which splits our beam into collinear signal and idler frequency pulses of unique user-defined frequencies, but whose sum is the frequency of the original input beam. In the home-built section of our optical table, mid-IR light is generated by transmitting the signal and idler through a Silver Gallium Sulfide ($AgGaS_2$) crystal for difference frequency generation (DFG). DFG generates a beam of light at the difference frequency between the incident beams. Now in the mid-IR, the beam is split into pump and probe pulses using a Calcium Fluoride (CaF_2) wedge (there is a back reflection off of the CaF_2 as well, which I am not using, but which a clever future student might use as a standard for shot-to-shot noise suppression). The pump pulses travel through a “pulse shaper” to create programmable time delays; this information generates the excitation, or ω_{pump} axis on a 2D IR spectrum. The pulse shaper is an Acousto-Optic Modulator (AOM) operating in the mid-IR frequency range: a crystal of Germanium (Ge) hooked up to some electrodes, the electrodes transmit an acoustic wave through the crystal, changing the properties of the crystal, which changes the properties of the light diffracted off the crystal lattice. There are some additional optics involved, such as

gratings and parabolic mirrors, but additional technical details can be found elsewhere.[1–3]

Meanwhile, the probe beam travels through a delay stage and overlaps with the pump beam at the sample location; the molecular response is generated in the same direction as our probe beam and is dispersed on a 64-pixel Mercury Cadmium Tellurium HgCdTe (MCT) array detector. This information provides the response for each of the aforementioned pulse shaper's programmed pump pulse time delays; it is the directly detected "detection", or ω_{probe}, axis on a 2D IR spectrum.

If you have not had enough yet, 2D IR spectroscopy is described *ad nauseum* in reference 1.[1] It's vital to remember that (as described above) each specific protein structure has characteristic vibrational frequency and transition dipole (μ) strength. By using a 4-pulse experiment, the signal detected on the diagonal slice of the 2D IR spectrum scale as the transition dipole of the amide I mode (which is specific to protein secondary structure) raised to the fourth (μ^4) power rather than the μ^2 scaling of linear IR spectroscopy. The diagonal features (excluding on-diagonal cross-peaks) are probably the easiest to understand in the 2D IR spectrum: they are excited and detected at the same frequency. An example 2D IR spectrum is displayed in **Figure 2.3.** Features that do not excite and detect at the same frequency are available from a 2D IR experiment as well, such as anharmonic shifts and cross-peaks. These phenomena are described in further detail in **Chapter 3**.

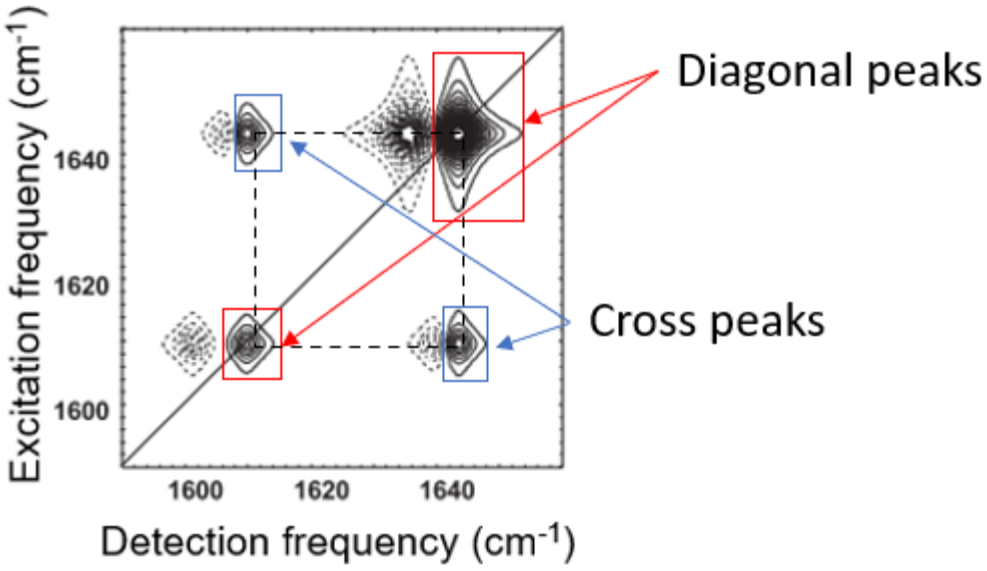


Figure 2.3: Example 2D IR spectrum.

Spectrum adapted from Hamm and Zanni.[1] Diagonal peaks scale as μ^4 and occur along the diagonal line. Cross peaks, located between each vibrational mode, scale as $\mu_a^2\mu_b^2$, where a and b are coupled vibrational modes or modes experiencing energy transfer. Excited-state absorption information is also detected (dotted lines redshifted from each vibrational mode).

Finally, by using a moving stage, one can scan a tissue sample across the laser spot between each 2D IR experiment. This is useful for creating an image of the protein structures present at each location in a sample and is used to image animal tissue sections in Chapter 3. You could also use this scanning microscope to image concentration gradients in a dissolved sample if these gradients were stable (using a flow cell similar to Guibertoni et al[4]), or to image phase-separated droplets of oligomers as was done using fluorescence microscopy by Jean et al.[5]

2.3 Isotope Labeling.

The 2D IR experiment generates a “snapshot” of all secondary structures occurring within the focus (laser spot) in the sample, for my experiments this was about 100 μm, but in principle is diffraction-limited (could be as small as 3 μm). Isotope labeling is useful for examining one or two amino acids at once, so that their 2D IR diagonal peaks are not convoluted by the rest of the secondary structures occurring within the sample.[6,7] Isotope labeling is a labeling method based on atomic mass. Recall from above that the amide I mode contains a carbon double bonded to an oxygen atom. The weight of these atoms affects the vibrational frequency of the carbon-oxygen double bond, heavier atoms vibrate more slowly. The most abundant naturally occurring carbon and oxygen nuclei have 12 and 17 neutrons, denoted ^{12}C and ^{17}O, respectively. Isotope labeling the carbon atom ($^{13}C^{17}O$) provides a 20 cm^2 redshift in the observed IR spectrum. This is often not enough to resolve from the unlabeled residues within the protein. Placing an additional isotope label on the oxygen of the same double bond ($^{13}C^{18}O$) provides an additional 40 cm^{-1} shift for a total 60 cm^{-1} redshift from a typical uncoupled residue.[3] This redshift is enough to remove the peak from the “bulk” for individual examination.[8] **Figure 2.4** compares a linear IR trace to a 2D IR diagonal trace, with labeled features for reference.

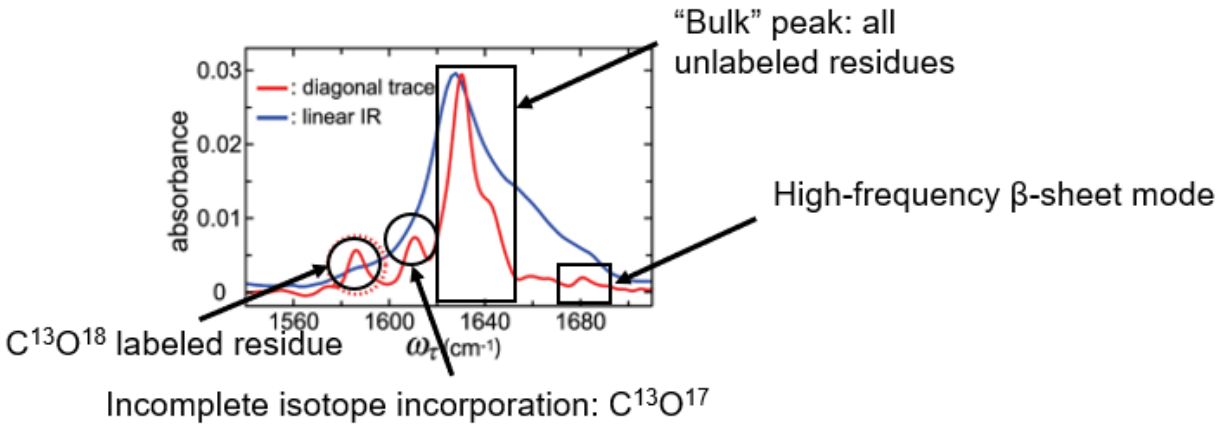


Figure 2.4: Isotope labeling; a 2D IR diagonal slice.

Isotope labeling resolves information about individual amino acids. A slice of a 2D IR spectrum is taken along the diagonal; individual features are labeled. Adapted from Kim and Hochstrasser (2009) studying Aβ, an Alzheimer's-associated amyloidogenic protein.[8]

2.4 Solid-phase peptide synthesis.

To create the proteins for 2D IR (and other) experiments, we use a commercial peptide synthesizer (CEM Liberty Blue). Solid phase peptide synthesis adds amino acids to a growing chain (attached on one end to a resin polymer bead) through cycles of deprotection, activation, and coupling (see **Figure 2.5**).

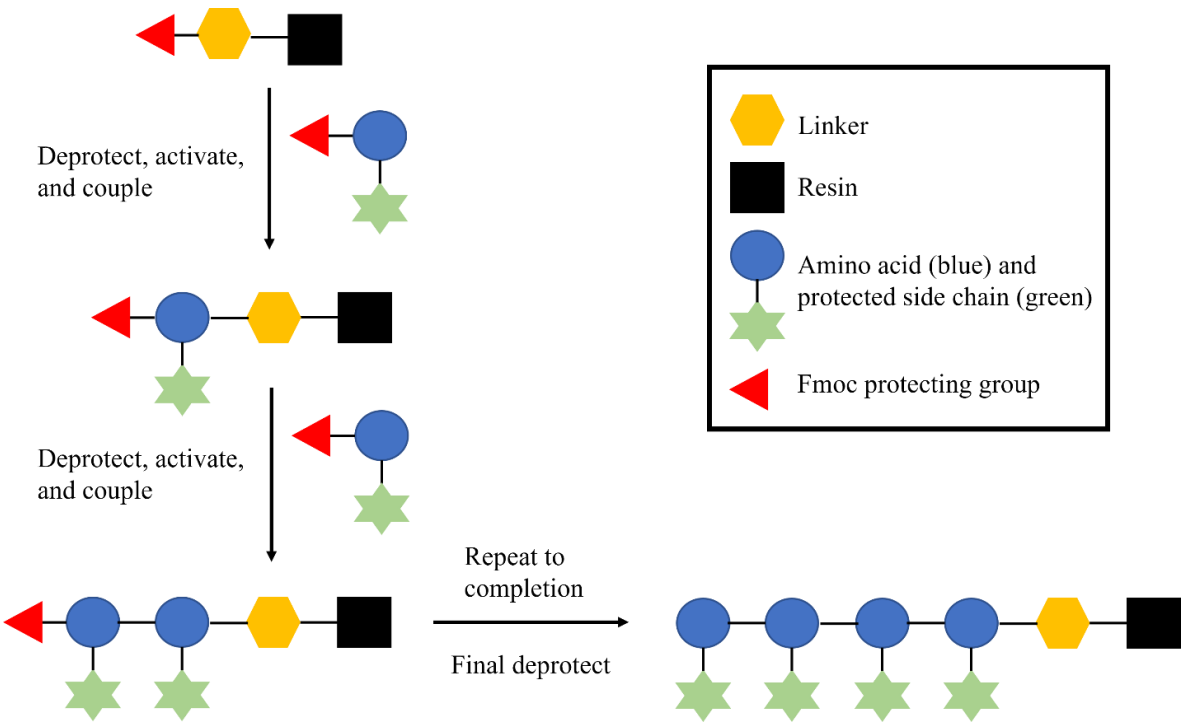


Figure 2.5 Fmoc-protected solid-phase peptide synthesis. Synthesizing a short peptide using Fmoc-based SPPS. Image taken from Caitlyn Field's Dissertation.

We use a high-swelling resin, PAL-PEG-PS, that yields an amidated C-terminus upon cleavage. We use this because mature amylin, after two cleavage events (from a Pre- and a Pre-pro-polypeptide) has an amide group ($CONH_2$) rather than a carboxillic acid (COOH). The proteins I have synthesized are hIAPP and its related variant, 3A-hIAPP, containing 3 alanine point mutations from the native hIAPP sequence. Both

hIAPP and 3A-hIAPP are both 37 residues long; yields in solid-phase peptide synthesis are greatly reduced for proteins above 50 residues.[9] PreProIAPP (prior to cleavage events in a cell), hIAPP, and 3A-hIAPP are displayed in **Figure 2.6**. The peptides are stable post-synthesis (on resin) at -20°C for up to 2 years.

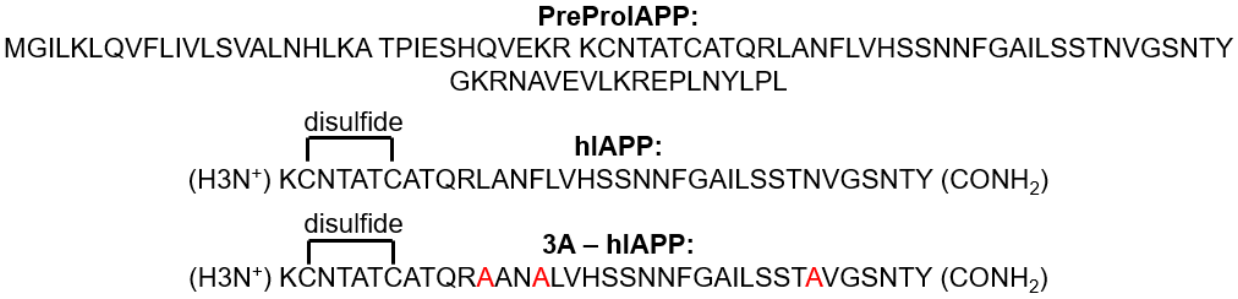


Figure 2.6: PreProIAPP, hIAPP, and 3A – hIAPP.

Sequences of PreProhIAPP, hIAPP, and 3A – hIAPP. PreProhIAPP is cleaved twice to yield hIAPP in the cell, with a disulfide bond between residues 2 and 7 and an amidated C terminus. We synthesize hIAPP and 3A – hIAPP in the lab, with the same post-translational modifications as occur in the cell. 3A has three alanine substitutions from the native sequence, shown in red.

SPPS synthesizes proteins from C-terminus to N-terminus, which is opposite direction of translation in a living cell. However, for proteins that do not require chaperones to fold, or those (such as amyloid proteins) that are intrinsically disordered, the spontaneously folded structure is biologically relevant and probably closely resembles a native fold. In SPPS, each amino acid (these are purchased from Cambridge Isotope Laboratories) has a base-labile florenylmethyloxycarbonyl (Fmoc) protecting group on the amine (H_3N^+) to prevent undesirable side reactions with this functional group. Piperidine is used as the deprotection agent. It removes the Fmoc protecting group and scavenges it from the surrounding solution; the next amino acid reacts with diisopropylcarbodiimide (DIC) to form an O-acylurea. This “activated” amino acid then reacts with the deprotected amino acid. The product is washed and

successive cycles add more amino acids to the growing chain. More information can be found in the CEM LibertyBlue user's manual or in Collins et al.[10]

2.5 Peptide cleavage.

Between the final peptide and the solid-phase resin exists an acid-labile linker. Many (potentially reactive) amino acid side chains are also protected via an acid labile protecting group as well. To remove the peptide from the resin, we introduce a cleavage cocktail containing Trifluoroacetic acid (TFA). The other ingredients are: thioanisole and ethanedithiol, which scavenge cleaved protecting groups,[11] and anisole, which prevents side reactions with native side chains (especially cysteine side chains). This mixture is microwaved for 30 minutes as described previously[12], and the acid labile chemicals are filtered from the final product. The solution is precipitated in cold ether and is ready for purification under acidic conditions.

2.6 Example cleavage protocol.

1.) Working in a fume hood, prepare cleavage cocktail. For 300 mg peptide, use 3.6 ml TFA, 0.2 ml thioanisole, 0.12 ml ethanedithiol, 0.08 ml anisole. The volume can be adjusted but the total solution must fit in the microwave vessel.

2.) Add cleavage cocktail to peptide in snugly capped microwave vessel with frit. Attach to temperature probe vessel by twisting. Remove from fume hood and insert into microwave, inserting the temperature probe until the yellow flag.

3.) Start microwave cleavage program for 30 minutes.

4.) In fume hood, disconnect microwave vessel and remove cap. Filter solution through frit at bottom of vessel into Falcon tube using Accent vacuum system.

5.) Precipitate dropwise into cold ether. Let sit for 0-48 hours. Spin down at 3800 rpm for 10 minutes, wash with cold ether, repeat spin cycle. Reconstitute in

concentrated glacial acetic acid. Working in a fume hood, create 4-6 holes in the vessel cap. Flash freeze and place under vacuum until a powder is formed.

2.7 Disulfide bond formation.

hIAPP contains two cysteines at residues 2 and 7 (1-letter code is C in Figure 2). These residues share a disulfide bond in mature hIAPP; when we synthesize hIAPP this disulfide bond does not form spontaneously. By dissolving synthetic hIAPP in 100% dimethyl sulfoxide (DMSO), we form the disulfide bond by oxidizing the cysteine residues. This reaction can take up to a week; by instead using a solution of 50% DMSO, 40% H2O, 10% glacial acetic acid, the process occurs rapidly over 12-24 hours. Since the polarity of a non-disulfide bonded peptide is different than that of a disulfide-bonded peptide, one can monitor the formation of the disulfide bond via reverse-phase high-performance liquid chromatography (rp-HPLC), as displayed in **Figure 2.7A-C**.

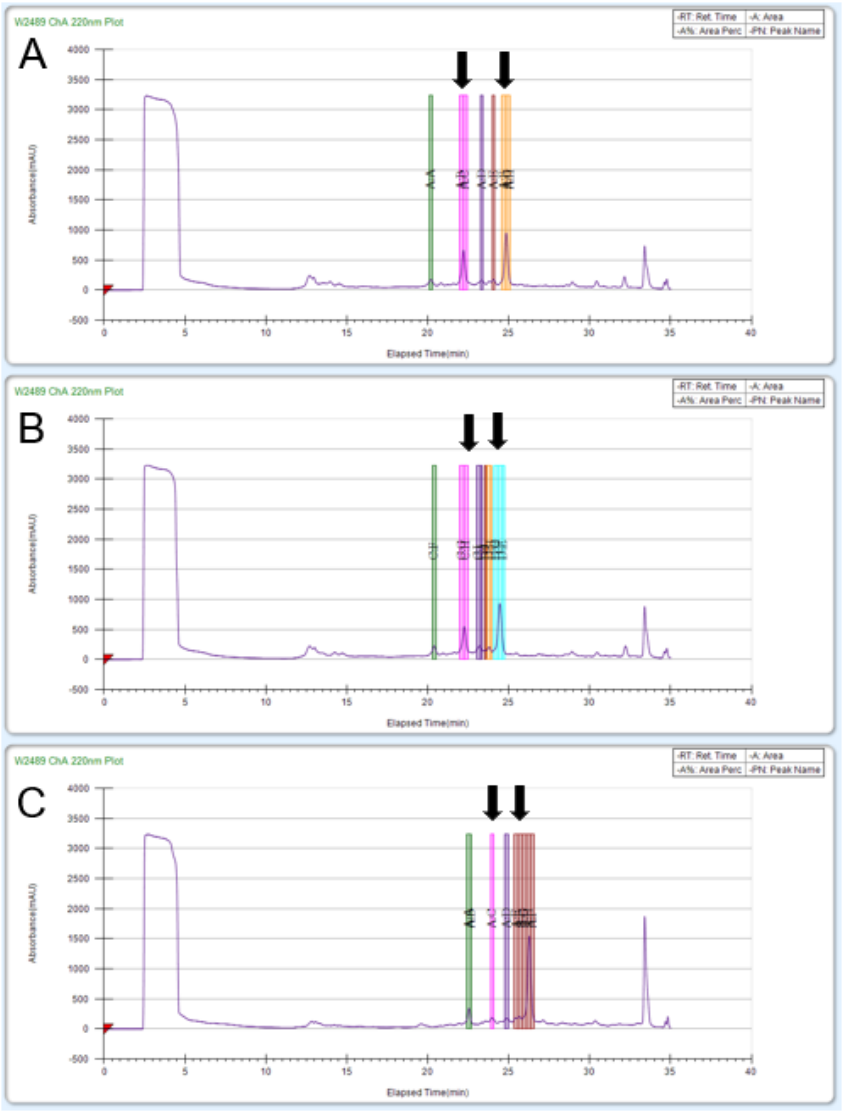


Figure 2.7: Disulfide bond formation monitored via rp-HPLC.

Monitoring disulfide bond formation via rp-HPLC. (A) After 2 hours, two peaks corresponding to hIAPP without a disulfide bonded and hIAPP with a disulfide bond visible in the chromatogram and are marked with arrows. Non-disulfide bonded hIAPP is the peak that eludes earlier. (B) After 6 hours, the peak ratio has changed and reflects a higher percentage of disulfide bonded hIAPP. (C) After 24 hours, most hIAPP is disulfide bonded. Peak identities are confirmed via MALDI-TOF mass spectrometry. Note the large DMSO peak in each chromatogram (far left) a characteristic of round 1 purification. Also note the peak drift between HPLC runs when solvents are refilled (between B and C, the main peak elution times are varied).

2.8 Reverse Phase High Performance Liquid Chromatography.

Rp-HPLC has a nonpolar stationary phase and a polar mobile phase. By adjusting the polarity of the mobile phase, one changes the retention time of the protein of interest. Our rp-HPLC is equipped with a UV-Vis spectrometer that monitors both the backbone absorbance (~220 cm^{-1}) and the tyrosine absorbance (~280 cm^{-1}) of species coming off of the column. Species of different polarity and impurities elute at different times, the purity of our sample is assessed by integrating the elution peak and dividing it by all other integrated peaks in the chromatogram. Additional rounds of HPLC can be performed to increase purity--usually only one or two purifications are required. Example rp-HPLC chromatograms for round 1 and 2 of purification are displayed in **Figure 2.8A and B**, respectively.

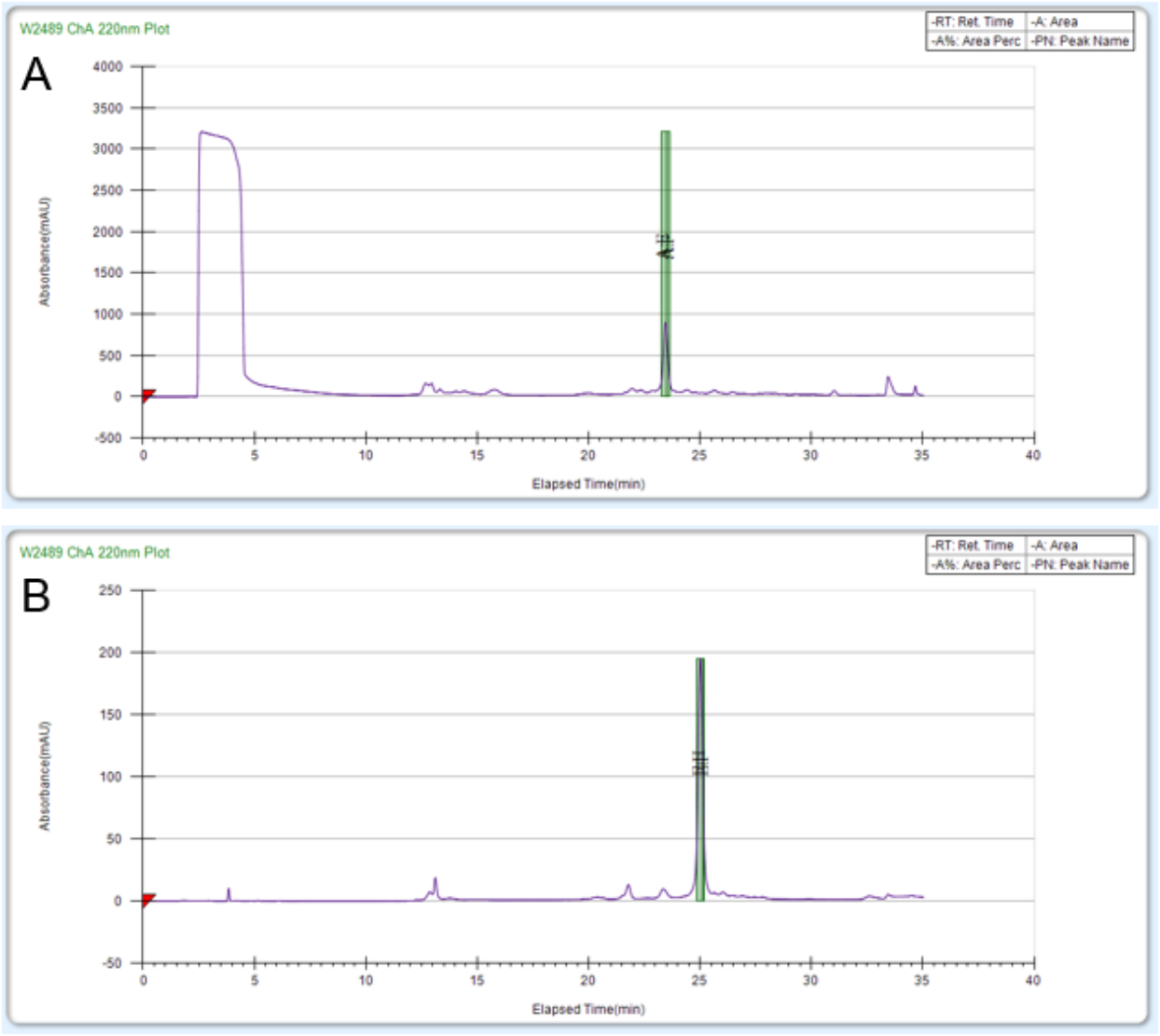


Figure 2.8 Round 1 and round 2 rp-HPLC chromatograms.

(A) Round 1 chromatogram showing early DMSO elution. (B) Round 2 chromatogram.

We use a C18 prep column with a 5 µm pore size and a flow rate of 20 mL/min. We introduce a gradient of two solvents: one is ultrapure water with 0.047% HCl (solvent A) and the other is 80% acetonitrile, 20% ultrapure water with 0.047% HCl (solvent B). For a new IAPP variant with unknown polarity, simply run a gradient incrementing the two solvents by 1%/minute and collect each peak. Analyze the peaks

(not including the initial large DMSO peak) via Matrix-Assisted Laser Desorption Ionization-Time of Flight (MALDI-TOF) spectrometry to determine which peak contains your sample of interest and create a shorter experimental protocol where this peak is collected. Previous students including myself have achieved high yields using protocols lasting 35-40 minutes (Initial column wash for about 8 minutes in 100% solvent A, increase solvent B to 10% below your desired elution threshold until 15 min, increase solvent B 1%/min until 35 minutes with desired elution occurring at 25 min, final ramp of solvent B until 40 or 42 minutes to clear the column).

One final improvement I have made to rp-HPLC purifications occurred due to the worldwide helium shortage. We used to continually sparge our solvents with ultrapure helium to ensure that passive carbon dioxide exchange with the laboratory air did not change the pH of our solvents. However, in response to the shortage, I determined that sparging was not necessary. When left open to laboratory air, our solvents do not appreciably change pH for about 2 weeks. Furthermore, in response to a safety inspection concern, we have rough caps on each of our solvents that should slow CO_2 exchange (the safety concern was the open storage of flammable solvents, such as acetonitrile, in the laboratory). After closing the helium tank, I have only observed minor (1-2 minute) shifts in peak elution time. We no longer sparge our solvents with helium but typically keep 1-2 helium tanks in reserve in case we use different solvents whose pH is less stable, or to use for sparging if a solvent is left exposed for a long period of time (but in this case I would suggest re-making the solvent instead of sparging).

Ideas for future students: We collaborate with Chad Reinstra's lab. The Reinstra lab students have expertise in purification of the amyloidogenic protein α-synuclein.

They use a C8 prep column and perform a single round of Fast Protein Liuqid Chromatography (FPLC) purification. Although the two proteins (hIAPP and α-synuclein) have different properties, the Reinstra group's purification is much faster than ours. A fun analytical chemistry project (possibly for a first- or second-year graduate student) would be to qualify a shorter method for protein purification that still resulted in a high yield and high purity using the (currently untouched) FPLC in our 8th-floor lab.

Following HPLC purification, the dry sample is resuspended in hexafluoroisopropanol (HFIP), a highly volatile helix promoting solvent to prevent aggregation. Small aliquots of this stock solution are dried under vacuum and resuspended in buffer to determine the concentration of the stock solution, or are used in quality control using MALDI-TOF MS.

2.9 MALDI-TOF Sample Preparation.

To assess if we have created the protein sample of interest, such as hIAPP or a related variant, we use MALDI-TOF spectrometry. Although I am not an expert in the instrumental setup (I use a Bruker microflex LRF maintained by the Chemical Instrumentation Facility), the following general protocol yields clean spectra:

2.10 Example MALDI-TOF sample preparation protocol:

1) Lyophilize 5-10 µl of the HFIP-peptide solution (MALDI-TOF requires only pico- or nano-molar amounts of peptide for detection).
2) Prepare MALDI solution using 70% acetonitrile, 30% 0.1% TFA solution.
3) Add 20-50 µl of MALDI solution to a single-use DHB sample. Allow DHB to settle to bottom of sample tube.
4) Resuspend lyophilized peptide powder in 2-10 µl of MALDI solution (no matrix).
    a. Do this for all of the peptide samples for the day
    b. Put 1 µl of each sample on to the MALDI target, recording sample locations in your laboratory notebook.
5) Add 1 µl of the DHB solution, taken from the top liquid (not the solid at the bottom of the tube), on top of each sample on the MALDI target.
6) Optional: for calibration, prepare standard samples of Insulin, Insulin β-chain, and a mix of both standard. For most hIAPP and related variants, this gives a 2-point calibration sandwiching your peptide of interest. The mix places both points on a single spectrum. Usually we purchase the insulin standards and prepare large batches in a MALDI solution before freezing them, a student then simply thaws the desired standards and adds DHB on top.

2.11 Referrals to additional sources of information.

Students in the Zanni group learn a wide variety of laboratory techniques, this is one of the factors that attracted me to the group as a rotating student. It is impossible to capture them all here in the detail required for each one. Additional information can be found in my laboratory notebook from years 4-5 (years 1-3 available in an additional notebook as well). Further information can be found in the experimental and supplementary sections of my manuscripts (Chapters 3-5 of this document), on the group server, or in the thesis of the previous students who also studied amylin and its related variants: Kacie L. Rich, Caitlyn Fields, Lauren Buchanan.

2.12 Chapter 2 References.

1. Hamm, P. & Zanni, M. *Concepts and methods of 2D infrared spectroscopy*. *Concepts and Methods of 2D Infrared Spectroscopy* vol. 9781107000056 (Cambridge University Press, 2011).

2. Shim, S.-H. & Zanni, M. T. How to turn your pump–probe instrument into a multidimensional spectrometer: 2D IR and Vis spectroscopiesvia pulse shaping. *Physical Chemistry Chemical Physics* **11**, 748–761 (2009).

3. Middleton, C. T., Woys, A. M., Mukherjee, S. S. & Zanni, M. T. 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 (2010).

4. Giubertoni, G. *et al.* Infrared Diffusion-Ordered Spectroscopy Reveals Molecular Size and Structure.

5. Pytowski, L., Lee, C. F., Foley, A. C., Vaux, D. J. & Jean, L. Liquid-liquid phase separation of type II diabetes-associated IAPP initiates hydrogelation and aggregation. *Proc Natl Acad Sci U S A* **117**, 12050–12061 (2020).

6. Buchanan, L. E., Dunkelberger, E. B. & Zanni, M. T. Examining Amyloid Structure and Kinetics with 1D and 2D Infrared Spectroscopy and Isotope Labeling. 217–237 (2012) doi:10.1007/978-3-642-22230-6_9.

7. Buchanan, L. *et al.* Mechanism of IAPP amyloid fibril formation involves an intermediate with a transient B-sheet. *PNAS* (2013) doi:10.1073/pnas.1314481110.

8. Kim, Y. S. & Hochstrasser, R. M. Applications of 2D IR Spectroscopy to Peptides, Proteins, and Hydrogen-Bond Dynamics †. doi:10.1021/jp8113978.

9. Coin, I., Beyermann, M. & Bienert, M. Solid-phase peptide synthesis: from standard procedures to the synthesis of difficult sequences. (2007) doi:10.1038/nprot.2007.454.

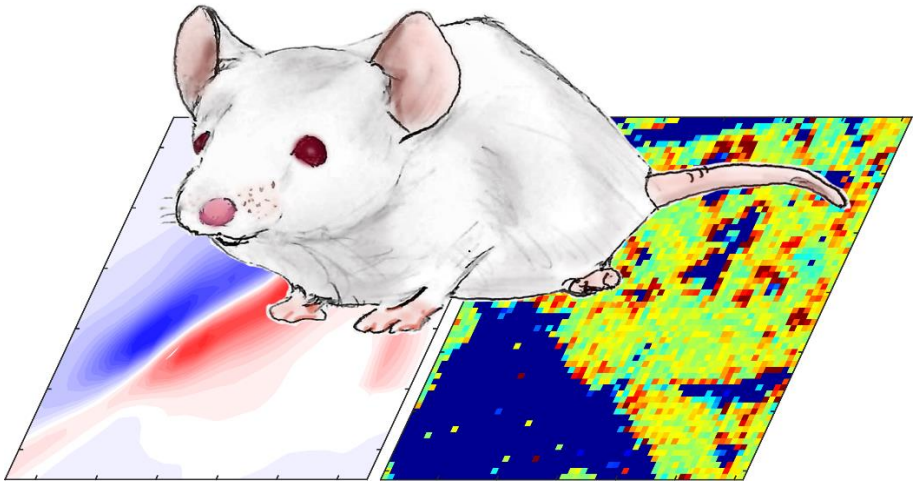
10. Collins, J. M., Porter, K. A., Singh, S. K. & Vanier, G. S. High-Efficiency Solid Phase Peptide Synthesis (HE-SPPS). (2014) doi:10.1021/ol4036825.

11. Kiso, Y. & Yajima, H. Amide Formation, Deprotection, and Disulfide Formation in Peptide Synthesis. *Peptides (N.Y.)* 39–91 (1995) doi:10.1016/B978-012310920-0/50003-6.

12. Kluczyk, A., Rudowska, M., Stefanowicz, P. & Szewczuk, Z. Microwave-assisted TFA cleavage of peptides from Merrifield resin. *Journal of Peptide Science* **16**, 31–39 (2010).

3 2D IR Bioimaging of animal tissues

Application of 2D IR Bioimaging: Hyperspectral Images of Formalin Fixed Pancreatic Tissues and Observation of Slow Protein Degradation.[3]



AUTHORS:

*Sidney S. Dicke†, Ariel M. Alperstein†^, Kathryn L. Schueler††, Donald S. Stapleton††, Shane P. Simonett††, Caitlyn R. Fields†, Farzaneh Chalyavi†, Mark P. Keller††, Alan D. Attie††, and Martin T. Zanni†**

[3] Adapted from: **Dicke SS**, Alperstein AM, Schueler KL, Stapleton DS, Simonett SP, Fields CR, Chalyavi F, Keller MP, Attie AD, Zanni MT. 2021. Application of 2D IR Bioimaging: Hyperspectral Images of Formalin-Fixed Pancreatic Tissues and Observation of Slow Protein Degradation. *J.Phys.Chem.B*. 2021, 125, 9517−9525. Headings and figure numbering have been updated for this document, but the original text remains unmolested.

AFFILIATIONS:

†Department of Chemistry, University of Wisconsin−Madison, 1101 University Avenue, Madison, WI 53706

††Department of Biochemistry, University of Wisconsin−Madison, 433 Babcock Drive, Madison, WI 53706

^Formally at Department of Chemistry, University of Wisconsin−Madison, 1101 University Avenue, Madison, WI 53706.

*Email: zanni@chem.wisc.edu

KEYWORDS:

2D IR, Bioimaging, Spectroscopy, Microscopy, Protein Structure

3.1 Abstract.

We used 2D IR Bioimaging to study the structural heterogeneity of formalin-fixed mouse pancreas. Images were generated from the hyperspectral data sets by plotting quantities associated with the amide I vibrational mode, which is created by the backbone carbonyl stretch. Images that measure the fundamental vibrational frequencies, cross peaks, and anharmonic shifts are presented. Histograms are generated for each quantity, providing averaged values and distributions around the mean that serve as metrics for protein structures. Images were generated from tissue that had been stored in formalin fixation for 3, 8, and 48 weeks. Over this period, all 3 metrics show that that the β-sheet content of the samples increase, consistent with protein aggregation. Our results indicate that formalin fixation does not entirely arrest degradation of protein structure in pancreas tissue.

3.2 Introduction.

Many techniques exist for identifying and solving the structures of proteins *in vitro*, but it is extremely difficult to monitor and image the structures of proteins in tissues. There are imaging techniques such as antibodies, fluorescent dyes, and mass spectrometry from which structure is inferred.[1–4] Electron scanning microscopy provides morphologies.[5,6] There are also spectroscopies, like NMR and MRI, which provide structural information but are difficult to apply on cellular length scales.[7–9] Most related to the topic of this manuscript are Raman and infrared (FTIR) imaging, which are techniques frequently used on tissues.[10–18] They provide secondary structure information with spatial resolutions on nanometer to micrometer scales. They can be used in conjunction with vibrational dyes[19–21] or applied label free.[22] Most often, vibrational spectroscopies utilize the amide vibrational modes to monitor secondary structure of proteins, along with the phosphodiester and methyl modes for DNA and lipids.[14,23–25] FTIR imaging can distinguish tumorous brain regions,[26,27] lesions in the aorta of cholesterol-fed rabbits,[28] and bone mineralization in wild type and density matrix protein-1 (DMP1)-knockout mice.[29] FTIR imaging is being developed for the analysis of tissue biopsies,[25,30,31] among many other applications.[13,27,32,33]

Raman or IR images can be generated at a single well-defined frequency or hyperspectral data sets generated by measuring multiple frequencies or in a Fourier transform mode.[13,26,34–36] A more recent hyperspectral vibrational imaging technique is 2D IR microscopy, which is based on ultrafast 2D IR spectroscopy.[37–40] 2D IR spectroscopy uses a series of femtosecond pulses to generate multidimensional IR spectra that correlate vibrational modes through the physics of their vibrational

couplings.[41–45] 2D IR spectroscopy is different than the analytical 2D correlation spectroscopy,[46–48] which is a means of analyzing data sets. 2D IR spectroscopy provides information about protein environments through 2D line shapes, utilizes vibrational lifetimes to discriminate against solvent exposed and disordered proteins, and enhances resolution through off-diagonal cross peaks.[44,49,50] 2D IR microscopy can be implemented in a number of ways, including point-mapping[37,39,51] and widefield approaches,[38] but either way, 2D IR microscopy is a collection of 2D IR spectra that form a hyperspectral image. 2D IR microscopy is a new technique that has mostly been applied to model systems.[37,39,40,52] The only biological systems studied so far are mouse, porcine, and human lens tissues and mouse kidney sections (stained with an IR active probe).[39,40,52]

In this work, we image formalin-fixed tissue of mouse pancreas. We present images of diagonal peaks, cross peaks, and anharmonic shifts, and use these metrics as probes of the protein structures within the sample. We also present histograms for each image displaying the range of values for each metric present in the sample. Interestingly, we have identified the appearance of features within the pancreatic tissue that are typical of protein aggregation and find that these aggregates accumulate with the time even though the tissues are formalin fixed.

3.3 Methods.

Housing and animal maintenance.

All studies performed on the mouse pancreas samples were pre-approved by the Animal Care and Use Committees at UW-Madison (Protocol #A005821) and BYU (Protocol #17-1202). The pancreas samples analyzed here were exhumed from wild type Friend Virus B/NIH Jackson (FVB/NJ) mice obtained from Jackson Laboratory (JAX stock no. 001800). FVB/NJ mice are a widely-used model organism for transgenic insertions. The mice studied here were raised in a colony where they lived with 2-5 mice per cage in a temperature and humidity-controlled room with a 12-hour light/dark cycle (6:00am-6:00pm). After weaning, the mice were on chow diet (Purina 5008) and sacrificed at 17 weeks of age.

Formalin fixed pancreas section.

Three mice were studied. A single wild type mouse pancreas is presented in the main text, and similar results were observed for two additional wild type mouse pancreas samples presented in Figure S1. The mice were euthanized via CO_2 inhalation and pancreas samples extracted as described previously[53] and immediately immersed in a 10% formalin in PBS solution. The tissues were then incubated at 4°C in the dark for 24 hours (sample is now considered formalin fixed). Formalin fixed samples were then placed in aqueous solution and the tissues were subjected to progressive dehydration in ethanol gradient prior to two final washes in 100% xylene solutions. Finally, dehydrated tissues are embedded in paraffin wax at 60°C and shaped into a block format and cooled to room temperature for slicing. The tissues are now described as “formalin fixed and

paraffin-embedded" (FFPE). Paraffin embedding was done by the veterinary medicine histology and the experimental animal pathology laboratories at UW-Madison. The pancreas removal and fixation, performed in our laboratory, takes 24 hours and the paraffin wax embedding process has a ~2-week turnaround time, thus after any routine realignments and calibrations of our optical setup we label our earliest timepoint as 3 weeks from when the sample was initially soaked in 10% formalin. FFPE tissues were stored in sealed plastic bags in the dark at room temperature between measurements. For antibody staining, paraffin needs to be removed from the tissue slices. In our study, paraffin was not removed because paraffin does not absorb in the amide-I region (as confirmed by measurements in regions of the slices that contained no tissue) and we wanted the tissue to be stable at room temperature for the duration of the experiment.

2D IR bioimaging

FTIR spectra of formalin in D_2O and 10% formalin in PBS buffer are shown in the Figure S2. 10 μm slices of FFPE mouse pancreas tissue were mounted between two CaF_2 windows. 2D IR spectra were collected as described previously.[39,52] The sample was raster scanned in steps of 50 μm to cover an area of 1-5 mm with 31-91 points per axis. A dark blue pixel indicates either there is no tissue at that location or there is excessive scatter of laser light that prevents that point from being measured. Spectra for these pixels are not included in the scatterplots nor histograms below.

3.4 Results.

2D IR bioimaging is a method of creating hyperspectral images from thousands of individual 2D IR spectra. 2D IR spectroscopy is analogous to NMR spectroscopy except that it uses pulses of infrared light to measure vibrations rather than pulsed magnetic fields for nuclear spins. Images generated from 2D IR spectra use the amide I vibrational mode, which is dominated by the carbonyl stretch motions of the peptide backbone. The frequency of the amide I band depends on the protein secondary structure. A slice along the diagonal of the 2D IR plot is analogous to an FTIR spectrum, albeit with more well-defined peaks.[54] Cross peaks appear off of the diagonal when vibrational modes are coupled to one another. More detailed information about the use of 2D IR spectroscopy to probe protein secondary structure is covered in several recent reviews.[41,42,43]

In what follows, we show 2D IR bioimages collected on the same mouse pancreas sample at 3, 8, and 48 weeks after fixation. At each time-point, we plot three different images created by analyzing the 2D IR spectra in three unique ways. The images visualize the frequencies, cross peaks, and anharmonicities, which are measures of protein structure. We note that each of the 2D IR images have a spatial resolution of about ~100 μm^2 and contain signals from all proteins within the overlapping laser beams. The 2D IR images cannot identify single proteins in the tissue, for example, but are instead providing an assessment of global populations of secondary structure. Spatial resolutions down to the diffraction limit of ~3-4 μm are possible.

Figure 3.1A shows a 2D IR spectrum of one location (~100 μm^2) within a mouse pancreas measured 3 weeks following formalin fixation. The 2D IR spectrum in Figure

3.1A is representative of the majority of locations measured from this sample. The main features are contained within black boxes. The fundamental vibrational modes (v = 0 to v = 1) appear along the diagonal, analogous to the peaks in an FTIR spectrum, at $\omega_{pump} = \omega_{probe}$ = 1635 cm^{-1} and $\omega_{pump} = \omega_{probe}$ = 1660 cm^{-1} (peaks 1 and 2). Each fundamental mode is accompanied by a 180 degree out-of-phase peak offset from the fundamental, which are the overtone peaks (v = 1 to v = 2). Peak 3 is the overtone feature for β-sheet vibrations. Peak 4 is the overtone feature for α-helix and random coil vibrations. In Figure 3.1B, a diagonal cut through the overtones is plotted that can be interpreted similarly to an FTIR spectrum, knowing that the intensities scale with the square instead of linearly with the absorption coefficient.[39,54,55] We cut through the overtone because the tissue scatters laser light, which causes a background that interferes with the fundamental transition but not the overtone. The diagonal cut easily illustrates the relative intensities of the peaks at ω_{pump} = 1635 cm^{-1} and 1660 cm^{-1}, which correspond to the peaks in boxes 1 and 2 respectively. In lens (unpublished result), liver, and heart tissues, formalin fixation has previously been reported to shift the amide I vibrational frequencies by about 4 cm^{-1} to higher values,[23] because it is more non-polar than water (several control spectra of preservation chemicals are given in Figure S2: formalin in D_2O, formalin in PBS-D_2O buffer, and formalin in PBS-H_2O buffer). We find that subtracting 4 cm^{-1} from the two features in boxes 1 and 2 puts their frequencies at about 1631 cm^{-1} and 1656 cm^{-1}, which are typical frequencies measured *in vitro* for β-sheet and random coil secondary structures, respectively, consistent with the previously reported 4 cm^{-1} shift upon fixation.[23]

Figure 3.1C, 3.1D, and 3.1E show images generated by plotting the ratio of the intensity at the maximum of peaks 3 and 4. Figure 3.1C, 3.1D, and 3.1E were collected 3, 8, and 48 weeks following formalin fixation, respectively (images are from the same slice of pancreas, but not the same position within the tissue. See Methods.). Figure 3.1F, 3.1G, and 3.1H bins the values of each image into a histogram, again representing timepoints of 3, 8, and 48 weeks after formalin fixation (histograms displayed below image containing the same data values). The average value in Figure 3.1C is 1.5, indicating that most positions have a β-sheet peak that is 50% larger than the random coil/α-helix peak, similar to the chosen representative spectra (Fig. 3.1A and 3.B). Some regions have much lower β-sheet content with a few locations having a very high percentage of β-sheet proteins. The tissue contains a variety of cell types and biological structures (hematoxylin and eosin stained section available in Figure S3) including acinar cells, α, β, and γ cells of the islets, blood vessels, and secretory ducts. In the images presented here, we are not scanning at a spatial resolution that can distinguish between cell types, but the heterogeneity of the images suggests that cell types might be distinguished with higher resolution. The images collected at 8 and 48 weeks after fixation and slicing, along with their respective histograms, reveal that the ratio of the intensities becomes larger with age, indicating that the amount of β-sheet secondary structure relative to the random coil/α-helix feature is more prominent. The average ratio is about 4 at 48 weeks with a larger distribution of values, indicating that the tissue is becoming more structurally heterogeneous. Each pixel in these images represents a measurement that encompasses ~100 μm^2 and each pixel is spaced by 50 μm. Some of the images are more “pixelated” because a smaller area was measured and thus contain fewer individual measurements.

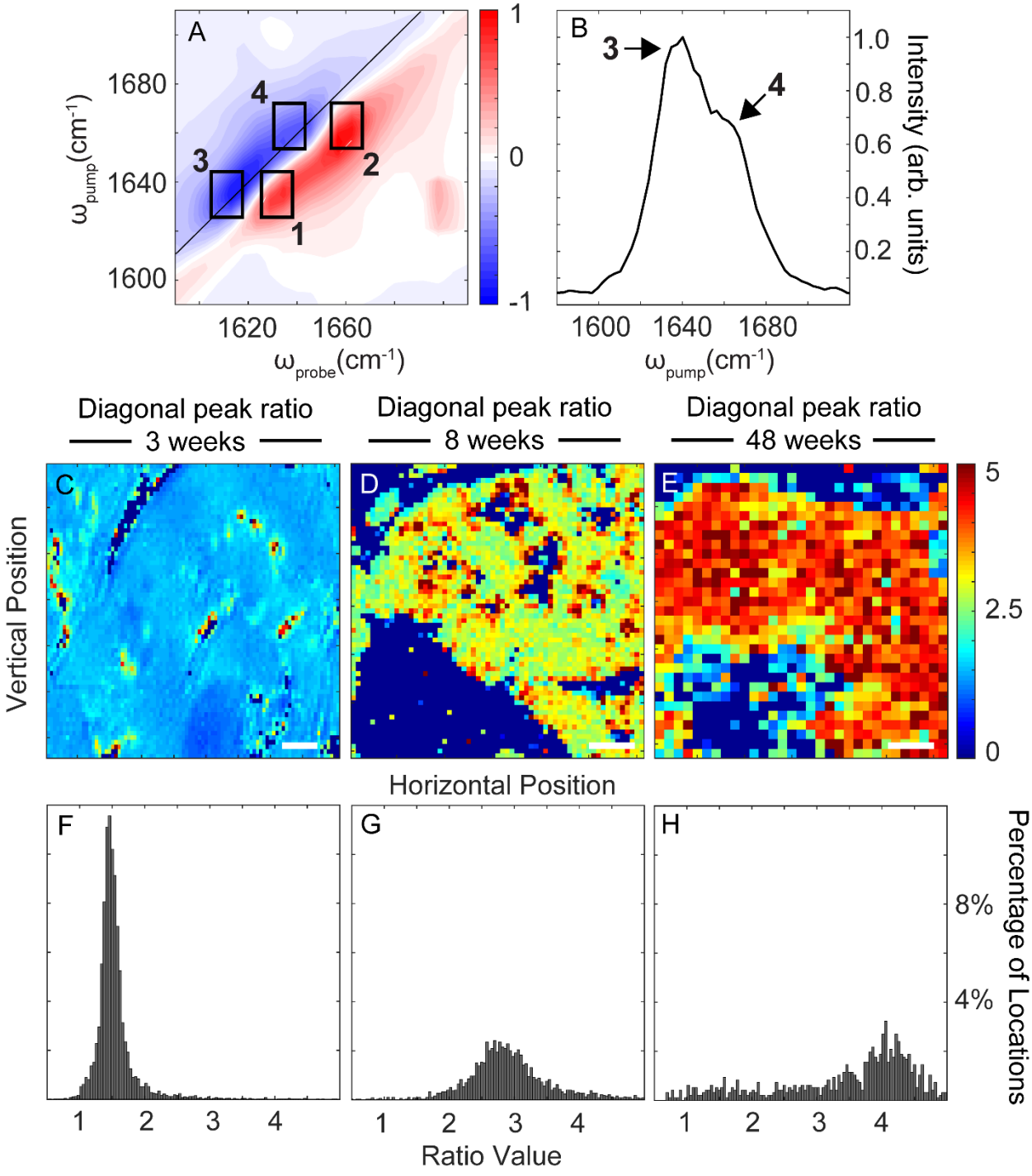


Figure 3.1: Hyperspectral images using overtone slices.

Hyperspectral images created using the intensities of the overtone absorptions. (A) Representative 2D IR spectrum. Peaks labeled by boxes 1 and 2 result from the $v = 0$ to $v = 1$ excitation and bleach for β-sheet ($\omega_{pump} = \omega_{probe} = 1635$ cm^{-1}) and random coil structures, respectively. Peaks labeled by boxes 3 and 4 result from the $v = 1$ to $v = 2$ excited state absorptions. (B) A slice through the overtone features taken from the 2D spectrum (A, thin line) with peaks 3 and 4 labeled. Hyperspectral images created by taking the ratio of the maximum intensity in box 3 and 4 for an image measured at (C) 3 weeks, (D) 8 weeks, and (E) 48 weeks after formalin fixation. Pixels are spaced 50 μm apart. Scale bars represent 1.06 mm, 0.5 mm, and 0.25 mm, for total areas of 4.25 mm^2, 3.1 mm^2, and 1.6 mm^2, respectively. F, G, and H are histograms of the respective images.

Figure 3.2A shows the same 2D IR spectrum as Figure 3.1A, but with the cross-peak region marked by a box, where the intensity has been magnified by a factor of 3. Figure 3.2B was processed in the same manner as Figure 3.2A, but it is a spectrum from the tissue measured at 8 weeks after formalin fixation. Figure 3.2C is a close-up of the cross peak region that has the two spectra overlaid, thereby comparing the cross peaks at 3 and 8 weeks after fixation (black and red contours, respectively). The cross peaks appear at different frequencies in the two spectra. In the 3-week sample, the cross peak is at $\omega_{pump} = 1635$ cm^{-1}, $\omega_{probe} = 1697$ cm^{-1} (peak 5). The 8-week sample has its cross peak at $\omega_{pump} = 1629$ cm^{-1} and $\omega_{probe} = 1700$ cm^{-1} (peak 6).[39] The cross peaks correlate coupled vibrational modes, indicating that the β-sheet in the 8-week sample has a 6 cm^{-1} lower frequency than at 3 weeks.

The two cross peak spectra in Fig. 3.2C are representative spectra of the cross peaks in the 3- and 8-week samples. Of the thousands of 2D IR spectra collected in these images, some have peak 5, some have peak 6, and many contain both. To visualize the heterogeneity, we show images next to their corresponding histograms in Figures 3.2D (image) and 3.2E (corresponding histogram), and Figures 3.2F (image) and 3.2G

(corresponding histogram). Because the cross peaks overlap with one another, we do not plot their absolute intensities, but use the ratio between the two peaks instead. For the 3-week tissue, the average ratio is about 0.5. A value of 0.5 means that the cross peak is solely created by peak 5. A value of 1 means that there are equal contributions from the peaks 5 and 6. For the 8-week tissue, there are many locations where the ratio is 1 or larger, indicating that the tissue at that position is dominated by peak 6.

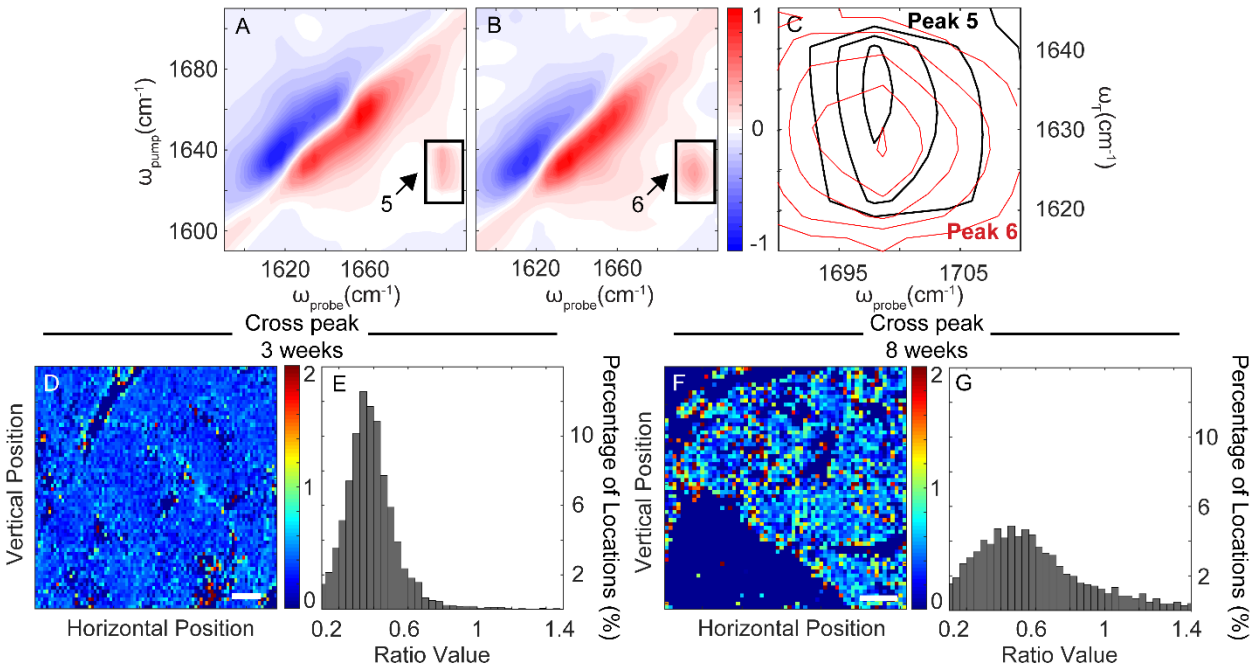


Figure 3.2: Hyperspectral images created from cross peak intensities.

2D IR spectrum collected from the mouse pancreas (A) 3 weeks (peak 5) and (B) 8 weeks (peak 6) after formalin fixation with an arrow indicating the cross peak, scaled by a factor of 3. (C) An overlay of the two cross peaks. Peak 5 is displayed in black and peak 6 is displayed in red. (D) A tissue image created by taking the ratio of the intensity at the peak 5 position compared to the peak 6 position for the 3-week-old pancreas tissue. (E) The same ratio values in (D) represented by a histogram. (F) A tissue image created by taking the ratio of the intensity at the peak 5 position compared to the peak 6 position for the 8-week-old pancreas tissue. (G) The same ratio values in (F) represented by a histogram. Scale bars represent 1.06 mm and 0.5 mm, for total areas of 4.25 mm^2 and 3.1 mm^2 in (D) and (F), respectively.

We also analyzed the anharmonicities of the β-sheet and random coil/α-helix peaks. Figure 3.3A shows the same 2D IR spectrum from the 3-week-old pancreas as in the previous figures, but is marked with a thin horizontal cut that passes through peaks 1 and 3. The intensity along the horizontal cut is displayed in Fig. 3.3B. The fundamental transition has a positive intensity and the overtone is negative. To extract the anharmonic shift, we fit each cut to two Gaussians (dashed lines) whose frequency and widths are varied to best reproduce the spectrum (dashed black line). Details of the fitting method are given in Equation S1. The frequency difference is the anharmonic shift. Fits were performed individually for each 2D IR spectrum measured, creating the images and corresponding histograms shown in Figures 3.3C (image) and 3.3D (corresponding histogram) for the 3-week tissue, and Figures 3.3E (image) and 3.3F (corresponding histogram) for the 8-week tissue.[39] A similar procedure was performed along the thin line passing through peaks 2 and 4 (Fig. 3.3A) producing the images and histograms for the anharmonic shift of the random coil/α-helix peaks. For the 3-week tissue, figure 3.3G is the image and 3.3H is the corresponding histogram, and for the 8-week tissue, 3.3I is the image and 3.3J is the corresponding histogram.

The distribution of values is narrow in the 3-week tissue, with an average value of 19 cm^{-1} for the β-sheet mode and 22 cm^{-1} for the random coil/α-helix mode. The anharmonic shift depends on the extent that the vibrational modes are delocalized; the larger the delocalization, the smaller the anharmonic shift. Typically, β-sheets have vibrational modes delocalized over 5-15 amide bonds,[56–58] whereas the vibrational modes of α-helix are about 5 amide bonds and random coils are delocalized over a maximum of two amino acids.[54] The images and histograms presented here are consistent with those physics,

giving us confidence in the consistency of the fits. We do note that these values are larger than the anharmonic shift of a single amide group, which is between 12-15 cm^{-1},[59,60] which we attribute to the fact that these tissues contain many proteins and so the extracted values are highly averaged quantities.

For the 8-week tissue, the average anharmonic shift of the β-sheet mode is 18 cm^{-1} and the random coil/α-helix mode is 24 cm^{-1}. Thus, the average β-sheet anharmonic shift is 2 cm^{-1} *smaller* at 8 than 3 weeks, and the average random coil/α-helix mode is 2 cm^{-1} *larger*. A smaller anharmonic shift indicates a larger delocalization, consistent with formation of larger β-sheets. The larger anharmonic shift is consistent with a higher population of random coil structures. The two observations taken together, larger β-sheets and more random coil, is consistent with protein aggregation. The distributions of the anharmonic shift at 8 weeks is also larger than at 3 weeks, indicating a greater degree of variations in secondary structure, as would be expected for aggregation.

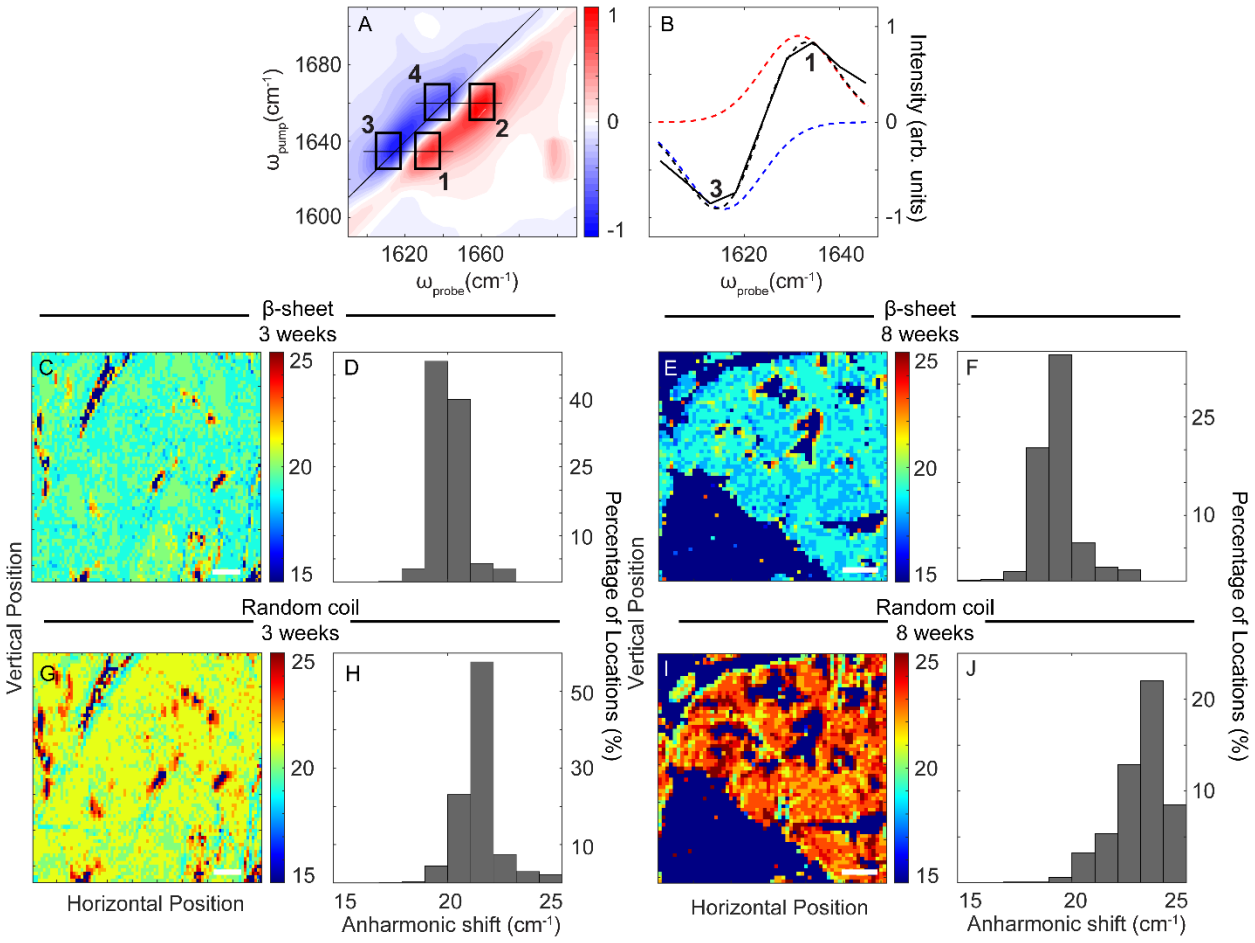


Figure 3.3: Hyperspectral images created from anharmonic shifts.

(A) A 2D IR spectrum of pancreas tissue 3 weeks after fixation with two cuts displayed (black lines) at 1635 cm^{-1} and 1660 cm^{-1}, respectively. The horizontal cuts pass through an overtone and a fundamental vibration at their respective pump frequencies. (B) Slice at 1635 cm^{-1}. Positive and negative Gaussian fits (dashed red and blue lines, respectively) and the sum of the two fits (dashed black line) fit to the β-sheet pump slice from the spectrum displayed in (A). (C) is an image created from the anharmonic shift value of the β-sheet cut taken in 50 μm steps across a mouse pancreas 3 weeks after fixation, and (D) is the corresponding histogram representing these values. (E) is an image created from the anharmonic shift value of the β-sheet cut taken in 50 μm steps across a mouse pancreas 8 weeks after fixation, and (F) is the corresponding histogram representing these values. (G) is an image created from the anharmonic shift value of the random coil/α-helix cut taken in 50 μm steps across a mouse pancreas 3 weeks after fixation, and (H) is the corresponding histogram representing these values. (I) is an image created from the anharmonic shift value of the random coil/α-helix cut taken in 50 μm steps across a mouse pancreas 8 weeks after fixation, and (J) is the corresponding histogram representing these values. (Histogram axis held between 15 and 25 for consistency with the images).

Images of the cross peaks and anharmonic shifts are not shown for the 48-week tissue because excessive laser scatter prevented accurate analysis of the cross peaks and anharmonic shifts. Analysis of the overtone peaks was still possible because those features are large and lie away from the scatter. We found, in general, that older tissues scattered the laser light more strongly than younger tissues. That qualitative observation is also consistent with protein aggregation; in several studies on cataract tissues, laser scatter correlated with aggregation.[39,52]

3.5 Discussion.

The above images spatially map 3 quantities that reflect protein secondary structure: frequencies, cross peaks, and anharmonic shifts. Data collected on tissues 3, 8, and 48 weeks after formalin fixation show that all 3 quantities change, indicating that the protein structures are slowly altered within the tissue over the course of the year. The 8 and 48 week images are more pixelated, but still contain over 950 individual spectra. Moreover, multiple locations across the pancreas were measured, and the observations confirmed in 2 other mice (see SI). Thus, we conclude that degradation occurs uniformly throughout the tissue.

The peak associated with β-sheets becomes larger, the lower-frequency cross peak gains intensity, the anharmonic shift in the β-sheet region becomes smaller, and the anharmonic shift of the random coil/α-helix region becomes larger. These observations are consistent with a decrease in native β-sheet and α-helical structure and the increase in random coil structures and non-native β-sheets. We know that it is non-native β-sheet structures, because the cross peak that forms is at lower frequency. We also see increased laser scatter in aged samples, indicating that the tissues are more spatially heterogeneous.

From these observations, we conclude that the proteins within the tissue are slowly aggregating. Amyloid fibrils are a common type of aggregate, but their frequencies are 10-20 cm^{-1} lower than that of native β-sheets.[39,52,55,57] The frequency difference we observe here is only a few wavenumbers. The anharmonic shift of amyloid β-sheets can

be as small as 5 cm^{-1},[61] which also does not agree with the data here. Thus, we assign the features here to small or amorphous protein aggregation.

Formalin is used to fix tissues because it cross links proteins and other biological molecules creating a gel-like state.[62] Formalin reacts most readily with the amino acids cysteine and lysine, creating covalent bonds within and across proteins.[62–64] Many *in vitro* studies have determined the mechanism of methylene bridge formation between proteins when exposed to formalin.[62,65,66] Cross linking is an established method for retaining cellular structures and organization of organelles.[67,68] Although the structures are now cross-linked, formalin is thought to preserve native protein secondary structure components.[69,70] However, previous studies using circular dichroism and gel electrophoresis studies that have also observed structural changes occurring due to fixation and/or wax embedding processes.[70,71]

The results presented here indicated that formalin fixation does not entirely prevent protein degradation within pancreas tissues. In our prior work using 2D IR spectroscopy to study cataracts in lens tissues extracted from mice, we did not observe fixation related protein structure changes (unpublished results). Lens tissues have the highest protein density of any major organ in the body (ca. ~35%) of which 90% are crystallin proteins, which are extremely stable.[39] Pancreas tissues contain proportionally less protein, at ca. ~50 mg protein/gram compared to ca. ~240 mg/ml in the lens cortex (also reported at higher densities depending on location).[72,73] Thus, fixation may be more effective on lens than pancreas tissues.[39,74] Another factor is that the pancreas contains many enzymes that digest proteins and so formalin fixation is done quickly following animal sacrifice to prevent enzymatic degradation.[75] Thus, the change in secondary structure here might be

due to slow aggregation of partially digested proteins or the incomplete deactivation of enzymes.

It is interesting to note that antibody stains continue to function after fixation, (if the tissue is dehydrated, as is the case for paraffin-embedded fixed tissues, heat- or chemical-induced antigen retrieval is usually necessary prior to staining).[71,76,77] One might expect that aggregation would prevent antibody binding. It is difficult for us to quantify the amount of protein that has aggregated, but we believe it is <10%, suggesting that antibody binding would still be >90% effective. It would be interesting to perform a series of time elapsed 2D IR and antibody imaging studies to determine if aggregation is anticorrelated to antibody binding. A study of that nature might help determine which specific proteins are aggregating. The samples in this study were stored in plastic bags in the dark (see Methods), similar to the common practice of storing FFPE tissues on benchtops or equivalent spaces. Future studies might test if storage under nitrogen slows and prevents degradation.

3.6 Conclusion.

This study reports 2D IR bioimages of formalin fixed pancreas tissues collected over nearly a 1-year timespan. The data was analyzed with three different quantities that are sensitive to protein secondary structure: frequencies, anharmonicities, and cross peaks. All three quantities are consistent with protein aggregation into amorphous small aggregates with small amounts of non-native β-sheet. Thus, formalin fixation does not completely arrest the secondary structure of the proteins within the pancreatic tissue. It may be important to take our results into account when performing chemical, antibody, or spectroscopic analyses of formalin fixed pancreas tissues. It would be interesting to perform a year-long series of antibody stains alongside 2D IR bioimaging to look for a correlation in antibody binding and secondary structure changes.

ASSOCIATED CONTENT:

Supplementary figures and information on Gaussian fitting available in SI (Microsoft Word document).

AUTHOR INFORMATION

Corresponding Author:

*Martin Zanni, Department of Chemistry, University of Wisconsin−Madison, Madison, WI 53706. Email: zanni@chem.wisc.edu

Present Addresses:

†Sidney S. Dicke, Caitlyn R. Fields, Farzaneh Chalyavi, and Martin T. Zanni - Department of Chemistry, University of Wisconsin−Madison, Madison, WI 53706

††Kathryn L. Schueler, Donald S. Stapleton, Shane P. Simonett, Mark P. Keller, Alan D. Attie - Department of Biochemistry, University of Wisconsin−Madison, Madison, WI 53706

†^Ariel M. Alperstein completed her PhD at the University of Wisconsin-Madison during which she contributed to this work. She is now a Postdoctoral Researcher at the Department of Chemistry, University of Minnesota, Minneapolis, MN 55455

Author Contributions:

All authors contributed to this work. SSD developed and performed 2D IR bioimaging experiments. SSD and MTZ wrote the paper.

Funding Sources:

This work was supported by the National Institutes of Health (1R01DK101573-01, 1R01DK101573-06, 1R01DK102948-01A1 (A.D.A.), and R01DK079895 (M.T.Z.)).

Notes:

The authors declare the following competing financial interest(s): Martin Zanni is co-owner of PhaseTech Spectroscopy, Inc., which sells mid-IR and visible pulse shapers and 2D spectrometers like those used in this publication.

ACKNOWLEDGMENTS

The authors thank the Histology Laboratory at the University of Wisconsin-Madison school of Veterinary Medicine and the Experimental Animal Pathology Laboratory (EAPL) at the University of Wisconsin-Madison for tissue embedding and slicing. The authors would also like to thank Kieran M. Farrell for composing the mouse featured in the TOC graphic.

3.7 Chapter 3 References.

(1) Biancalana, M.; Koide, S. Molecular Mechanism of Thioflavin-T Binding to Amyloid Fibrils. *Biochim Biophys Acta.* **2010**, *1804*(7):1405-12.

(2) Howie, A. J.; Brewer, D. B.; Howell, D.; Jones, A. P. Physical Basis of Colors Seen in Congo Red-Stained Amyloid in Polarized Light. *Lab. Investig.* **2008**, *88* (3), 232–242.

(3) Yakupova, E. I.; Bobyleva, L. G.; Vikhlyantsev, I. M.; Bobylev, A. G. Congo Red and Amyloids: History and Relationship. *Biosci Rep*. 2019, *39*(1)

(4) Kayed, R.; Glabe, C. G. Conformation-Dependent Anti-Amyloid Oligomer Antibodies. *Methods Enzymol*. **2006**, *413*: 326–44.

(5) Franken, L. E.; Grünewald, K.; Boekema, E. J.; Stuart, M. C. A. A Technical Introduction to Transmission Electron Microscopy for Soft-Matter: Imaging, Possibilities, Choices, and Technical Developments. *Small* **2020**, *16* (14).

(6) Qiang, W.; Yau, W. M.; Lu, J. X.; Collinge, J.; Tycko, R. Structural Variation in Amyloid-β Fibrils from Alzheimer's Disease Clinical Subtypes. *Nature* **2017**, *541* (7636), 217–221.

(7) Tycko, R. Solid-State NMR Studies of Amyloid Fibril Structure. **2011**. *Annu. Rev. Phys. Chem.* 2011;62:279-99.

(8) Reif, B.; Ashbrook, S. E.; Emsley, L.; Hong, M. Solid-State NMR Spectroscopy. *Nat. Rev. Methods Prim.* **2021**, *1* (1), 1–23.

(9) Molugu, T. R.; Lee, S.; Brown, M. F. Concepts and Methods of Solid-State NMR Spectroscopy Applied to Biomembranes. *Chem. Rev*. **2017**, *117*, 19, 12087–12132.

(10) Noreen, R.; Moenner, M.; Hwu, Y.; Petibois, C. FTIR Spectro-Imaging of Collagens for Characterization and Grading of Gliomas. *Biotechnol Adv.* **2012**;*30*(6):1432-46.

(11) Ali, M. H. M.; Rakib, F.; Al-Saad, K.; Al-Saady, R.; Goormaghtigh, E. An Innovative Platform Merging Elemental Analysis and Ftir Imaging for Breast Tissue Analysis. *Sci. Rep.* **2019**, *9* (1), 18–20.

(12) Ali, M. H. M.; Rakib, F.; Abdelalim, E. M.; Limbeck, A.; Mall, R.; Ullah, E.; Mesaeli, N.; McNaughton, D.; Ahmed, T.; Al-Saad, K. Fourier-Transform Infrared Imaging Spectroscopy and Laser Ablation -ICPMS New Vistas for Biochemical Analyses of Ischemic Stroke in Rat Brain. *Front. Neurosci.* **2018**, *12*(647).

(13) Bhargava, R. Infrared Spectroscopic Imaging: The next Generation. *Appl Spectrosc.* **2012**, Oct; *66*(10): 1091–1120.

(14) Baker, M. J.; Trevisan, J.; Bassan, P.; Bhargava, R.; Butler, H. J.; Dorling, K. M.; Fielden, P. R.; Fogarty, S. W.; Fullwood, N. J.; Heys, K. A.; et al. Using Fourier Transform IR Spectroscopy to Analyze Biological Materials. *Nat. Protoc.* **2014**, *9* (8), 1771–1791.

(15) Stewart, S.; Priore, R. J.; Nelson, M. P.; Treado, P. J. Raman Imaging. *Annu. Rev. Anal. Chem.* **2012,** 5:337-360.

(16) Opilik, L.; Schmid, T.; Zenobi, R. Modern Raman Imaging: Vibrational Spectroscopy on the Micrometer and Nanometer Scales. *Annu. Rev. Anal. Chem.* **2013**, *6*, 379–398.

(17) Zhang, Y.; Hong, H.; Cai, W. Imaging with Raman Spectroscopy. *Curr. Pharm. Biotechnol.* **2010**, *11* (6), 654–661.

(18) Jones, R. R.; Hooper, D. C.; Zhang, L.; Wolverson, D.; Valev, V. K. *Nanoscale Research Letters* **2019**. 14,(231)1–34.

(19) Wei, L.; Hu, F.; Shen, Y.; Chen, Z.; Yu, Y.; Lin, C. C.; Wang, M. C.; Min, W. Live-Cell Imaging of Alkyne-Tagged Small Biomolecules by Stimulated Raman Scattering. *Nat. Methods.* **2014**,*11*(4),410–412.

(20) Hu, F.; Zeng, C.; Long, R.; Miao, Y.; Wei, L.; Xu, Q.; Min, W. Supermultiplexed Optical Imaging and Barcoding with Engineered Polyynes. *Nat. Methods* **2018**, *15* (3), 194–200.

(21) Chen, Z.; Wei, L.; Zhu, X.; Min, W. Extending the Fundamental Imaging-Depth Limit of Multi-Photon Microscopy by Imaging with Photo-Activatable Fluorophores. *Opt. Express* **2012**, *20* (17), 18525.

(22) Fernandez, D. C.; Bhargava, R.; Hewitt, S. M.; Levin, I. W. Infrared Spectroscopic Imaging for Histopathologic Recognition. *Nat. Biotechnol*. **2005**, *23*(469–474).

(23) Zohdi, V.; Whelan, D. R.; Wood, B. R.; Pearson, J. T.; Bambery, K. R.; Black, M. J. Importance of Tissue Preparation Methods in FTIR Micro-Spectroscopical Analysis of Biological Tissues: “Traps for New Users.” *PLoS One* **2015**, *10* (2).

(24) Choo, L. P. in.; Wetzel, D. L.; Halliday, W. C.; Jackson, M.; LeVine, S. M.; Mantsch, H. H. In Situ Characterization of β-Amyloid in Alzheimer’s Diseased Tissue by Synchrotron Fourier Transform Infrared Microspectroscopy. *Biophys. J.* **1996**, *71* (4), 1672–1679.

(25) Gazi, E.; Dwyer, J.; Gardner, P.; Ghanbari-Siahkali, A.; Wade, A. P.; Miyan, J.; Lockyer, N. P.; Vickerman, J. C.; Clarke, N. W.; Shanks, J. H.; et al. Applications of Fourier Transform Infrared Microspectroscopy in Studies of Benign Prostate

and Prostate Cancer. A Pilot Study. *J. Pathol.* **2003**, *201* (1), 99–108.

(26) Kole, M. R.; Reddy, R. K.; Schulmerich, M. V.; Gelber, M. K.; Bhargava, R. Discrete Frequency Infrared Microspectroscopy and Imaging with a Tunable Quantum Cascade Laser. *Anal. Chem.* **2012**, *84* (23), 10366–10372.

(27) Petibois, C.; Desbat, B. Clinical Application of FTIR Imaging: New Reasons for Hope. *Trends Biotechnol.* **2010**, *28*, 495–500.

(28) Palombo, F.; Cremers, S. G.; Weinberg, P. D.; Kazarian, S. G. Application of Fourier Transform Infrared Spectroscopic Imaging to the Study of Effects of Age and Dietary L-Arginine on Aortic Lesion Composition in Cholesterol-Fed Rabbits. *J. R. Soc. Interface*. **2009**, *6*(37):669-80.

(29) Ling, Y.; Rios, H. F.; Myers, E. R.; Lu, Y.; Feng, J. Q.; Boskey, A. L. DMP1 Depletion Decreases Bone Mineralization in Vivo: An FTIR Imaging Analysis. *J. Bone Miner. Res.* **2005**, *20* (12), 2169–2177.

(30) Tiwari, S.; Kajdacsy-Balla, A.; Whiteley, J.; Cheng, G.; Hewitt, S. M.; Bhargava, R. INFORM: INFrared-Based ORganizational Measurements of Tumor and Its Microenvironment to Predict Patient Survival. *Sci. Adv.* **2021**, *7* (6), eabb8292.

(31) Reddy, R. K.; Bhargava, R. Accurate Histopathology from Low Signal-to-Noise Ratio Spectroscopic Imaging Data. *Analyst* **2010**, *135* (11), 2818–2825.

(32) Su, W. H.; Sun, D. W. Fourier Transform Infrared and Raman and Hyperspectral Imaging Techniques for Quality Determinations of Powdery Foods: A Review. *Compr. Rev. Food Sci. Food Saf.* **2018**, *17* (1), 104–122.

(33) Glassford, S. E.; Byrne, B.; Kazarian, S. G. Recent Applications of ATR FTIR Spectroscopy and Imaging to Proteins. *Biochim Biophys Acta* **2013**, *1834*(12),

2849-2858.

(34) Wrobel, T. P.; Mukherjee, P.; Bhargava, R. Rapid Visualization of Macromolecular Orientation by Discrete Frequency Mid-Infrared Spectroscopic Imaging. *Analyst* **2017**, *142*, 75.

(35) Lewis, E. N.; Levin, I. W.; Treado, P. J.; Reeder, R. C.; Story, G. M.; Dowrey, A. E.; Marcott, C. Fourier Transform Spectroscopic Imaging Using an Infrared Focal-Plane Array Detector. *Anal. Chem.* **1995**, *67* (19), 3377–3381.

(36) Wrobel, Thomasz P., Kole, Matthew R., Bhargava, R. Emerging Trends and Opportunities in Discrete-Frequency Infrared and Raman Spectroscopic Imaging. *Spectroscopy* **2016**, *31* (6), 28–45.

(37) Baiz, C. R.; Schach, D.; Tokmakoff, A. Ultrafast 2D IR Microscopy. *Opt. Express* **2014**, *22* (15), 18724.

(38) Ostrander, J. S.; Serrano, A. L.; Ghosh, A.; Zanni, M. T. Spatially Resolved Two-Dimensional Infrared Spectroscopy via Wide-Field Microscopy. *ACS Photonics* **2016**, *3* (7), 1315–1323.

(39) Alperstein, A. M.; Ostrander, J. S.; Zhang, T. O.; Zanni, M. T. Amyloid Found in Human Cataracts with Two-Dimensional Infrared Spectroscopy. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116* (14), 6602–6607.

(40) Fournier, F.; Guo, R.; Gardner, E. M.; Donaldson, P. M.; Loeffeld, C.; Gould, I. R.; Willison, K. R.; Klug, D. R. Biological and Biomedical Applications of Two-Dimensional Vibrational Spectroscopy: Proteomics, Imaging, and Structural Analysis. *Acc. Chem. Res.* **2009**. *42*, 9, 1322–1331,.

(41) Ghosh, A.; Ostrander, J. S.; Zanni, M. T. Watching Proteins Wiggle: Mapping

Structures with Two-Dimensional Infrared Spectroscopy. *Chem. Rev.* **2017**, *117* (16), 10726–10759.

(42) Petti, M. K.; Lomont, J. P.; Maj, M.; Zanni, M. T. Two-Dimensional Spectroscopy Is Being Used to Address Core Scientific Questions in Biology and Materials Science. *J. Phys. Chem. B* **2018**, *122*, 6, 1771–1780.

(43) Middleton, C. T.; Woys, A. M.; Mukherjee, S. S.; Zanni, M. T. Residue-Specific Structural Kinetics of Proteins through the Union of Isotope Labeling, Mid-IR Pulse Shaping, and Coherent 2D IR Spectroscopy. *Methods*. **2010**;*52*(1):12-22.

(44) Hamm, P.; Zanni, M. *Concepts and Methods of 2D Infrared Spectroscopy*; Cambridge University Press, 2011.

(45) Baiz, C. R.; Błasiak, B.; Bredenbeck, J.; Cho, M.; Choi, J. H.; Corcelli, S. A.; Dijkstra, A. G.; Feng, C. J.; Garrett-Roe, S.; Ge, N. H.; et al. Vibrational Spectroscopic Map, Vibrational Spectroscopy, and Intermolecular Interaction. *Chem. Rev.* **2020**, *120*, 15, 7152–7218.

(46) Lasch, P.; Noda, I. Two-Dimensional Correlation Spectroscopy (2D-COS) for Analysis of Spatially Resolved Vibrational Spectra. *Appl. Spectrosc.* **2019**, *73*(4) 359–379.

(47) Murayama, K.; Czarnik-Matusewicz, B.; Wu, Y.; Tsenkova, R.; Ozaki, Y. Comparison between Conventional Spectral Analysis Methods, Chemometrics, and Two-Dimensional Correlation Spectroscopy in the Analysis of near-Infrared Spectra of Protein. *Appl. Spectrosc.* **2000**, *54* (7), 978–985.

(48) Morita, S. I.; Miura, Y. F.; Sugi, M.; Ozaki, Y. New Correlation Indices Invariant to Band Shifts in Generalized Two-Dimensional Correlation Infrared Spectroscopy.

Chem. Phys. Lett. **2005**, *402* (1–3), 251–257.

(49) Fayer, M.D. *Ultrafast Infrared Vibrational Spectroscopy*. 1st Ed, CRC Press, 2013.

(50) Cho, M. *Two-Dimensional Optical Spectroscopy*. 1st Ed, CRC Press, 2009.

(51) Tracy, K. M.; Guchhait, B.; Tibbetts, C.; Luther, B. M.; Krummel, A. Visualizing Chemical Dynamics in an Ionic Liquid Microdroplet Using Ultrafast 2D IR Microscopy. *ChemRxiv* (2019).

(52) Zhang, T. O.; Alperstein, A. M.; Zanni, M. T. Amyloid β-Sheet Secondary Structure Identified in UV-Induced Cataracts of Porcine Lenses Using 2D IR Spectroscopy. *J. Mol. Biol.* **2017**, *429* (11), 1705–1721.

(53) Li, D. S.; Yuan, Y. H.; Tu, H. J.; Liang, Q. Le; Dail, L. J. A Protocol for Islet Isolation from Mouse Pancreas. *Nat. Protoc.* **2009**, *4* (11), 1649–1652.

(54) Grechko, M.; Zanni, M. T. Quantification of Transition Dipole Strengths Using 1D and 2D Spectroscopy for the Identification of Molecular Structures via Exciton Delocalization: Application to α-Helices. *J. Chem. Phys.* **2012**, *137* (18), 184202.

(55) Fields, C. R.; Dicke, S. S.; Petti, M. K.; Zanni, M. T.; Lomont, J. P. A Different Hiapp Polymorph Is Observed in Human Serum than in Aqueous Buffer: Demonstration of a New Method for Studying Amyloid Fibril Structure Using Infrared Spectroscopy. *J. Phys. Chem. Lett.* **2020**, *11* (15), 6382–6388.

(56) Hahn, S.; Kim, S.-S.; Lee, C.; Cho, M. Characteristic Two-Dimensional IR Spectroscopic Features of Antiparallel and Parallel-Sheet Polypeptides: Simulation Studies. *J. Chem. Phys.* **2005**, *123*, 84905.

(57) Lomont, J. P.; Ostrander, J. S.; Ho, J. J.; Petti, M. K.; Zanni, M. T. Not All β-Sheets Are the Same: Amyloid Infrared Spectra, Transition Dipole Strengths, and

Couplings Investigated by 2D IR Spectroscopy. *J. Phys. Chem. B* **2017**, *121* (38), 8935–8945.

(58) Strasfeld, D. B.; Ling, Y. L.; Gupta, R.; Raleigh, D. P.; Zanni, M. T. Strategies for Extracting Structural Information from 2D IR Spectroscopy of Amyloid: Application to Islet Amyloid Polypeptide. *J. Phys. Chem. B*. **2009**, *113*, 47, 15679–15691.

(59) Zanni, M. T.; Gnanakaran, S.; Stenger, J.; Hochstrasser, R. M. Heterodyned Two-Dimensional Infrared Spectroscopy of Solvent-Dependent Conformations of Acetylproline-NH 2. *J. Phys. Chem. B* **2001**, *105*, 28, 6520–6535.

(60) Hamm, P.; Lim, M.; Hochstrasser, R. M. Structure of the Amide I Band of Peptides Measured by Femtosecond Nonlinear-Infrared Spectroscopy. *J. Phys. Chem. B* **1998**, *102*, 31, 6123–6138.

(61) Dunkelberger, E.B.; Grechko, M.; Zanni, M.T. Transition Dipoles from 1D and 2D Infrared Spectroscopy Help Reveal the Secondary Structure of Proteins: Application to Amyloids. *J. Phys. Chem. B.* **2015,** *119,* 44, 14065-14075.

(62) Thavarajah, R.; Mudimbaimannar, V. K.; Elizabeth, J.; Rao, U. K.; Ranganathan, K. Chemical and Physical Basics of Routine Formaldehyde Fixation. *J. Oral. Maxillofac. Pathol.* **2012**, *16*(3): 400–405.

(63) Kamps, J. J. A. G.; Hopkinson, R. J.; Schofield, C. J.; Claridge, T. D. W. How Formaldehyde Reacts with Amino Acids. *Commun. Chem.* **2019**, *2* (1), 1–14.

(64) Michiels, T. J. M.; Schöneich, C.; Hamzink, M. R. J.; Meiring, H. D.; Kersten, G. F. A.; Jiskoot, W.; Metz, B. Novel Formaldehyde-Induced Modifications of Lysine Residue Pairs in Peptides and Proteins: Identification and Relevance to Vaccine Development. *Mol. Pharm.* **2020**, *17* (11), 4375–4385.

(65) Lu, K.; Ye, W.; Zhou, L.; Collins, L. B.; Chen, X.; Gold, A.; Ball, L. M.; Swenberg, J. A. Structural Characterization of Formaldehyde-Induced Cross-Links between Amino Acids and Deoxynucleosides and Their Oligomers. *J. Am. Chem. Soc.* **2010**, *132* (10), 3388–3399.

(66) Metz, B.; Kersten, G. F. A.; Hoogerhout, P.; Brugghe, H. F.; Timmermans, H. A. M.; De Jong, A.; Meiring, H.; Ten Hove, J.; Hennink, W. E.; Crommelin, D. J. A.; et al. Identification of Formaldehyde-Induced Modifications in Proteins: Reactions with Model Peptides. *J. Biol. Chem.* **2004**, *279* (8), 6235–6243.

(67) Graham, L.; Orenstein, J. M. Processing Tissue and Cells for Transmission Electron Microscopy in Diagnostic Pathology and Research. *Nat. Protoc.* **2007**, *2* (10), 2439–2450.

(68) Wang, N. S.; Minassian, H. The Formaldehyde-Fixed and Paraffin-Embedded Tissues for Diagnostic Transmission Electron Microscopy: A Retrospective and Prospective Study. *Hum. Pathol.* **1987**, *18* (7), 715–727.

(69) Mason, J. T.; O'leary2, T. J. Effects of Formaldehyde Fixation on Protein Secondary Structure: A Calorimetric and Infrared Spectroscopic Investigation; *J. Histochem. Cytochem*. **1991;** *39*(2):225-9.

(70) Fowler, C. B.; O'Leary, T. J.; Mason, J. T. Modeling Formalin Fixation and Histological Processing with Ribonuclease A: Effects of Ethanol Dehydration on Reversal of Formaldehyde Cross-Links. *Lab. Investig.* **2008**, *88* (7), 785–791.

(71) Fowler, C. B.; Evers, D. L.; O'Leary, T. J.; Mason, J. T. Antigen Retrieval Causes Protein Unfolding: Evidence for a Linear Epitope Model of Recovered Immunoreactivity. *J. Histochem. Cytochem.* **2011**, *59* (4), 366–381.

(72) Gu, C.; Xu, H. Effect of Oxidative Damage Due to Excessive Protein Ingestion on Pancreas Function in Mice. *Int. J. Mol. Sci.* **2010**, *11* (11), 4591–4600. https://doi.org/10.3390/ijms11114591.

(73) Zhao, H.; Magone, M. T.; Schuck, P. The Role of Macromolecular Crowding in the Evolution of Lens Crystallins with High Molecular Refractive Index. *Phys. Biol.* **2011**, *8* (4), 046004.

(74) Truscott, R. J. W. Eye Lens Proteins and Cataracts. *Protein Misfolding, Aggregation, Conform. Dis.* **2007**, 435–447.

(75) Hudkins, K. L.; Pichaiwong, W.; Wietecha, T.; Kowalewska, J.; Banas, M. C.; Spencer, M. W.; Mühlfeld, A.; Koelling, M.; Pippin, J. W.; Shankland, S. J.; et al. BTBR Ob/Ob Mutant Mice Model Progressive Diabetic Nephropathy. *J. Am. Soc. Nephrol.* **2010**, *21* (9), 1533–1542.

(76) Shi, S. R.; Shi, Y.; Taylor, C. R. Antigen Retrieval Immunohistochemistry: Review and Future Prospects in Research and Diagnosis over Two Decades. *J Histochem Cytochem*. **2011**, *59*(1):13-32.

(77) Bogen, S.; Vani, K.; Sompuram, S. Molecular Mechanisms of Antigen Retrieval: Antigen Retrieval Reverses Steric Interference Caused by Formalin-Induced Cross-Links. *Biotech. Histochem.* **2009**, *84* (5), 207–215.

3.8 Supporting Information.

Supporting Information for:

Application of 2D IR Bioimaging: Hyperspectral Images of Formalin Fixed Pancreatic Tissues and Observation of Slow Protein Degradation

Sidney S. Dicke†, *Ariel M. Alperstein*†^, *Kathryn L. Schueler*††, *Donald S. Stapleton*††, *Shane P. Simonett*††, *Caitlyn R. Fields*†, *Farzaneh Chalyavi*†, *Mark P. Keller*††, *Alan D. Attie*††, *and Martin T. Zanni*†*

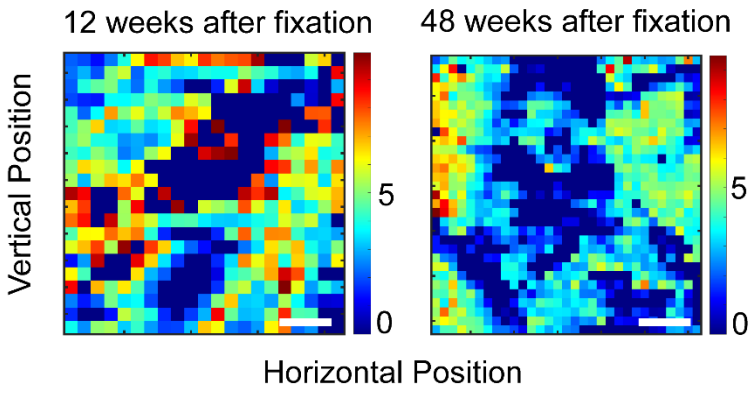
AFFILIATIONS:

†Department of Chemistry, University of Wisconsin−Madison, 1101 University Avenue, Madison, WI 53706

††Department of Biochemistry, University of Wisconsin−Madison, 433 Babcock Drive, Madison, WI 53706

^Formally at Department of Chemistry, University of Wisconsin−Madison, 1101 University Avenue, Madison, WI 53706.

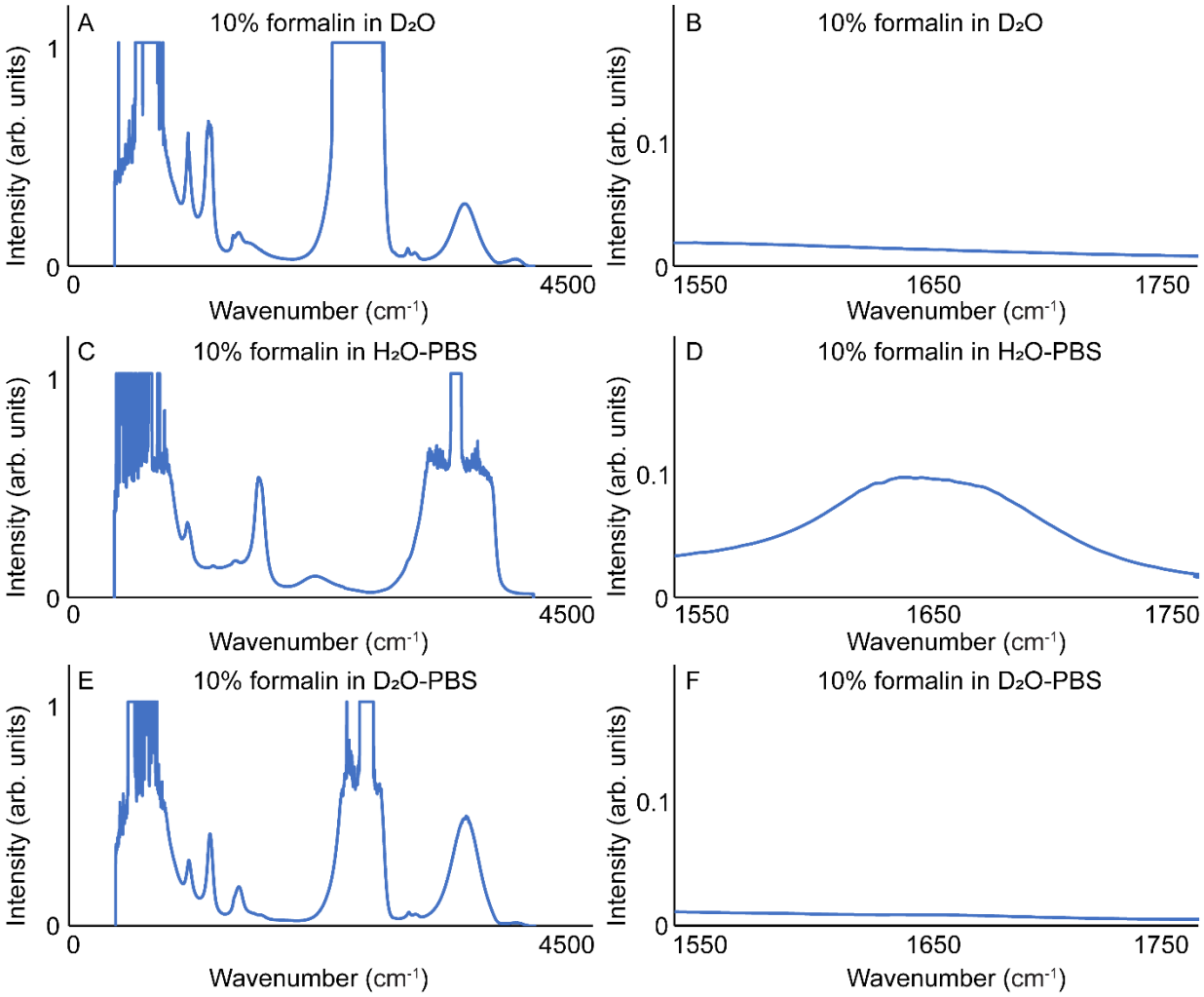
*Email: zanni@chem.wisc.edu



Supplementary Figure 3.4 (S1): 2D IR images of pancreas tissue from 2 additional mice.

(Left) Image from additional wild type mouse 12 weeks after formalin fixation. (Right) Image from an additional wild type mouse 1 year after fixation. Ratio uses peak 3 and peak 4 as discussed in text. Scale bars = 300 μm.

As shown in SI-figure 1, degradation was observed for the additional 2 pancreas tissues studied. However, the degradation rate is not expected to be consistent between tissues as the ratios for the 48-week-old sample were observed to be higher than the ratios observed for the pancreas in the main text.

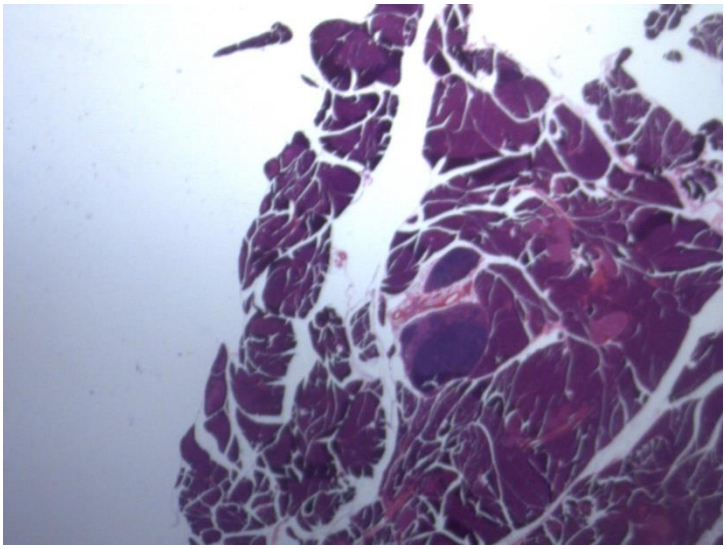


Supplementary Figure 3.5 (S2) – FTIR spectra of formalin solutions.

FTIR spectra of formalin solutions. (A) FTIR spectrum of 10% formalin in D_2O. (B) Amide-I region of FTIR spectrum of 10% formalin in D_2O.(C) FTIR spectrum of 10% formalin in H_2O-PBS buffer. (D) Amide-I region of FTIR spectrum of 10% formalin in H_2O-PBS buffer. (E) FTIR spectrum of 10% formalin in D_2O-PBS buffer. (F) Amide-I region of FTIR spectrum of 10% formalin in D_2O-PBS buffer.

As shown in SI-Figure 2, the fixation solution contains background FTIR signal in the Amide-I region from the O-H bend, but there is no amide-I signal from the formalin. O-H

is dehydrated from the sample during prior to embedding in paraffin wax, whereas formalin remains active within the tissue.



Supplementary Figure 3.6 (S3) – Hematoxylin and Eosin stain of a slice of mouse pancreas tissue.

Hematoxylin and Eosin stain of a slice of mouse pancreas tissues. Various shades of pink to blue are observed for a variety of cellular and subcellular structures. This slice was taken from the same wild type mouse pancreas studied in the main text and was sequential to (~10 μm away from) the unstained slice used for 2D IR bioimaging.

Details of the Gaussian fitting for calculation of anharmonic shift values:

Two Gaussian curves were fit to the data at fixed pump frequencies (for β-sheet anharmonic shift, ω_{pump} = 1635 cm^{-1} , for random coil anharmonic shift, ω_{pump} = 1660 cm^{-1}), using the method described by Alperstein et al.[1] It was found that summation of the two Gaussians consistently overestimated the anharmonic shift, thus an additional horizontal scaling factor was applied to each curve until the total Gaussian accurately represented the data (see Figure 3.3B in the main text, positive Gaussian is horizontally

redshifted and negative Gaussian is horizontally blueshifted until the sum of the two Gaussians accurately represents the data). Thus the final equation used was:

Equation S1:

y = (a1(i)*exp(-(((x+/-SF1)-b1(i))/c1(i)).^2)))+(a2(i)*exp(-(((x+/-SF2)-b2(i))/c2(i)).^2))

where a1, b1, c1, a2, b2, and c2 are found iteratively using the Matlab function “gauss2” where a1 and a2 are the peak heights, b1 and b2 are the peak centers, and c1 and c2 are the peak widths for the positive and negative Gaussians, respectively and SF1 and SF2 are the horizontal scaling factors applied to the positive and negative Gaussian curves after fitting. The sign of SF1(-) and SF2(+) is set so that the positive peak redshifts and the negative peak blueshifts on the probe axis.

Parameters used for the β-sheet anharmonic shift fit were:

Lower bounds: [-3 0 0 0.9 1635 0]

Upper bounds: [1.05 1620 16 1 1635 11]

Initial conditions: [-1 1617 6 1 1635 6]

Conditions used for the random coil anharmonic shift fit were:

Lower bounds: [-3 0 0 0.4 1660 0]

Upper bounds: [1.05 1650 16 1 1660 11]

Initial conditions: [-1 1640 6 1 1660 6]

Given in [a1 b1 c1 a2 b2 c2] format as described above. During the fitting, the fundamental maximum is held constant (as the lower and upper bounds are the same). The widths,

peak amplitudes, and frequency corresponding to the overtone maxima are allowed to vary.

3.9 Reference for Chapter 3 Supporting Information:

[1] A.M. Alperstein, J.S. Ostrander, T.O. Zhang, M.T. Zanni, Amyloid found in human cataracts with two-dimensional infrared spectroscopy, Proc. Natl. Acad. Sci. 116 (2019) 6602–6607. https://doi.org/10.1073/PNAS.1821534116.

4 Genotyping an overexpression animal model for type II diabetes.

TITLE:

A protocol for locating and counting transgenic sequences from laboratory animals using a map-then-capture (MapCap) sequencing workflow: procedure and application of results.[4]

AUTHORS:

*Sidney S. Dicke†, C. Dustin Rubinstein††, James M. Speers††, Mark E. Berres††, Derek M. Pavelec††, Kathryn L. Schueler†††, Donnie S. Stapleton†††, Shane P. Simonett†††, Mark P. Keller†††, Alan D. Attie†††, and Martin T. Zanni†**

AFFILIATIONS:

† Department of Chemistry, University of Wisconsin−Madison, 1101 University Avenue, Madison, WI 53706, U.S.A.

†† UW-Madison Biotechnology Center, University of Wisconsin−Madison, 425 Henry Mall, Madison, WI 53706, U.S.A.

††† Department of Biochemistry, University of Wisconsin−Madison, 433 Babcock Drive, Madison, WI 53706, U.S.A.

*Corresponding Author: Martin T. Zanni, Email: zanni@chem.wisc.edu

[4] Adapted from: **Dicke SS**, Rubinstein CD, Speers JM, Berres ME, Pavelec DM, Schueler KL, Stapleton DS, Simonett SP, Keller MP, Attie AD, and Zanni MT. 2022. A protocol for locating and counting transgenic sequences from laboratory animals using a map-then-capture (MapCap) sequencing workflow: procedure and application of results. *bioRxiv.* Posted 19 January 2022.

Keywords : Transgenic model, DNA sequencing, genotyping, phenotyping.

4.1 Abstract.

Transgenic rodent models for human diseases have been widely used over the past 50 years and are a mainstay of many biomedical research programs. Oftentimes the sequence of the transgenic segment of DNA is carefully designed but incorporation of this DNA into the host genome is less well understood. Structural variation and insertional mutagenesis may occur at transgenic insertion sites. Here, we present a robust workflow including identification of the transgene locus via selective Illumina sequencing followed by Cas9-mediated target DNA enrichment of the locus, which successfully identified beginning and end sites of a large transgenic insertion into a murine model for human amylin-induced type II diabetes. Enriched sequences were mapped via Oxford Nanopore sequencing. Although the insertion was too long for a single mapped genetic sequence to encompass, the method provided multiple insights relevant to the animal model: a minimum number of forward- and reverse-facing transcript copies as well as characterization of an inversion point within the insertion site. The insertion start point containing both murine and human DNA was used to identify and separate animals hemizygous for the transgenic insertion from homozygous animals. This identification could be performed early in the rodent life cycle prior to maturation (i.e. breeding age), thus allowing for management of colony phenotypes and eliminating the need to "genotype by phenotype" later on (onset of amylin-induced type II diabetes does not occur until ~8-10 weeks of age for this model). We further confirmed our homozygous diabetic mice function the same as colonies established in other labs and present full antibody and fluorescent-staining protocols (available in SI). Lastly, we note that, due to our genotyping, a novel animal was identified within our

colony: non-diabetic homozygous mice. Indeed, only 37% of homozygous mice bred in our colony became diabetic.

4.2 Author summary (for broad audience).

Transgenic rodent models are important to studying human diseases. When creating a new rodent model, one may insert new DNA into a well-characterized background genome. However, it is oftentimes not known where the new DNA was incorporated, how many times it was incorporated, or if any coding sequences or regulatory elements within native DNA were disrupted. Here, we have developed a method to characterize transgenic animals, and have applied it to a popular model for studying human amylin-induced type II diabetes.

4.3 Introduction.

Understanding mutated genomes of transgenic animals.

Almost 50 years has passed since the first successful introduction of transgenic material into a mouse(1–4), and to this day transgenic mouse alleles remain an indispensable biomedical tool from basic research to development of preclinical therapeutics. Transgenic mouse alleles are traditionally created by microinjecting recombinant DNA into the pronuclei of fertilized eggs and identifying integration events of the transgenic fragment into a random locus or loci of the genome. Neither the number nor location of insertion segments are reliable; thus native genes may be disrupted and expression levels of transgenic DNA vary widely among these models.(5,6) Continued breeding of mouse lines often results in mutagenesis over time, occasionally resulting in phenotype-enhancing or phenotype-suppressing effects.(7) Although more precise genome editing strategies have since become available, the transgenic method continues to reliably generate many animal models for human diseases, some of which have been commercialized. While phenotypes of these alleles are often carefully described, precise molecular characterization of transgenic alleles is seldomly reported. In fact, reports of molecular characterization of transgenic alleles indicate that transgenic alleles are often more complex than expected.(8)

One example of an animal model produced in this fashion that is now widely available is the *RIPHAT hIAPP +/-* mouse (FVB/N-Tg(Ins2-IAPP)RHFSoel/J, The Jackson Laboratory stock no. 008232). Phenotypically, these mice are highly valuable: when bred to homozygosity, some mice experience human amylin-induced type II diabetes (T2D) after ~10 weeks of age (T2D penetrance has previously been

undefined). Molecularly, the location of the transgenic DNA is not known, although using primers designed based on the non-native transgene promoter sequence easily distinguishes animals carrying the transgenic allele from those not carrying it. However, this method fails to distinguish hemizygous transgenic from homozygous transgenic animals (SI-1). The distinction between hemi- and homozygous mice is typically only possible after the mice become diabetic (using phenotype to estimate the genotype). This approach is suboptimal financially (as all transgenic mice must be raised to ~10 weeks of age before their phenotype is known) and further assumes that mature phenotype perfectly reflects genotype (i.e., 100% penetrance).

While detailed molecular characterizations of transgenic alleles are infrequent, great improvements have been made in our ability to study whole genomes of organisms. This includes substantial decreases in costs of short-read sequencing and development of powerful long-read sequencing technologies that allow transgenes to be mapped and characterized, respectively (9,10). In this study, we use these technologies to genetically characterize a colony of *RIPHAT hIAPP (+/+)* mice and study their resulting phenotypes. Here we report a map-then-capture sequencing workflow (hereafter referred to as ‘MapCap’) to easily permit the characterization of transgenic alleles. Briefly, the first step uses selective amplification of transgene-containing Illumina inserts to map the locus of integration. The second step uses Cas9 to selectively sequence that locus using Oxford Nanopore Technology. This method of molecularly characterizing transgenic mouse alleles can be applied to any transgenic mouse model with a known transgenic sequence, even if the insertion locus is unknown. This method

yields the insertion locus, the structure of the insertion, and a rapid PCR-based genotyping assay to distinguish alleles with and without the transgene.

4.4 Description of the method.

DNA sequencing.

DNA was isolated from mouse tail and blood of wild type, hemizygous, and homozygous mice and provided to the UW-Madison Biotechnology Center DNA sequencing facility (double-blinded to sample ID). DNA target sequence construct was designed according to the patent submitted by Soeller et al (US patent 6187991 B1) describing *RIPHAT* transgenic construct (US 6187991 B1 sequence ID no. 7 encompassing rat insulin II promoter and 5' untranslated leader, IAPP coding region, albumin intron I, and GAPDH polyadenylation region. See SI for full primer design). Illumina and Oxford Nanopore sequencing was performed and analyzed at the UW-Madison Biotechnology Center (UWBC) using protocols provided by manufacturer and a Cas9 enrichment protocol developed internally at UWBC.

Mice.

Breeding pairs were purchased form Jackson Laboratories (stock no. 008232). Initial sequencing was performed on 18-week-old homozygous mouse tail snip DNA. Follow-up rounds of sequencing were performed on blood DNA from 10-week-old wild type FVB, hemizygous, and homozygous mice. Mice were handled and sacrificed according to approved UW-Madison Research Animal Resources and Compliance (RARC) protocols.

Illumina sequencing of transgene insertion site.

Identification of the DNA junction between the integrated *RIPHAT hIAPP* transgene and the *mouse strain* genome was performed following a modified High-throughput Insertion Tracking by Deep Sequencing (HITS) method (11). Libraries were

prepared using TruSeq Nano DNA Library Prep (Illumina). DNA was fragmented to an average size of 400bp using Covaris M220 Focused-ultrasonicator (Covaris, Inc). No size selection was performed prior to adapter ligation. Adapter ligation was performed using 15μM duplex oligos (5'-ACACTCTTTCCCTACACGACGCTCTTCCGATC*T and 5'-[Phos]GATCGGAAGAGC*C*A). Enrichment of adapter-ligated library containing the *RIPHAT hIAPP* transgene was performed, targeting the IAPP region. Custom oligos (5'-ACACTCTTTCCCTACACGACGC-3' and 5'-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTCACAGTTGCCATGTAGACC-3') were substituted for PCR Primer Cocktail at 0.2 μM final and 16 cycles of PCR were performed. A second round of PCR was performed to incorporate indexes and TruSeq universal adapter sequences. PCR was performed using KAPA HiFi HotStart Ready Mix (Roche Diagnostics) and custom oligos (5'-TGATACGGCGACCACCGAGATCTACAC[55555555]ACACTCTTTCCCTACACGACGCTCTTCCGATCT-3' and 5'-AAGCAGAAGACGGCATACGAGAT[77777777]GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT -3') at 0.2 μM final, where bracketed sequences are 8nt indexes. The following conditions were used for amplification: 95°C for 3 minutes, followed by 8 cycles of 95°C for 30 seconds, 55°C for 30 seconds, and 72°C for 30 seconds, with a final extension at 72°C for 5 minutes. Libraries were sequenced on a MiSeq System at 10pM final using MiSeq Reagent Nano Kit v2, 300-cycles (Illumina) with paired-end reads of 250 cycles and 50 cycles. Asymmetric cycles were used to increase the chances of sequencing through the junction site given the library insert size and

orientation. Sequence reads in FASTQ format were obtained with bcl2fastq (version v2.20.0.422), part of the standard Illumina Pipeline.

High molecular weight DNA extraction.

Genomic DNA was extracted from 200 µL of whole blood obtained via retro-orbital bleed using Circulomics Nanobind CBB Big DNA Kit (Circulomics, Baltimore, MD) following the manufacturer's recommendations. DNA was size selected using Circulomics Short Read Eliminator Kit.

Cas9-mediated target enrichment.

Once the transgenic insertion locus is mapped to the mouse genome, pairs of Cas9 target sites were chosen 700-750 bp flanking the mapped insertion site, with PAMs oriented inwards. Oxford Nanopore adapters were selectively ligated to the target region according to previously published methods.(12) High Molecular Weight (HMW) DNA was dephosphorylated, inactivating DNA ends. Cas9-mediated double-strand breaks were introduced, A-tailed, and adapters were ligated using Ligation Sequencing Kit (SQK-LSK109, Oxford Nanopore Technologies Ltd, Oxford, UK). Libraries were sequenced with a R9.4.1 flowcell (FLO-MIN106D) on a GridION platform (Oxford Nanopore Technologies Ltd). Basecalling was performed with Guppy 3.2.8+bd67289 using the high-accuracy basecalling model.

Fluorescence microscopy.

FFPE tissues were sliced at 10 µm and stained using Congo Red, Thioflavin T, anti-Insulin or anti-hIAPP (T4157 polyclonal antibody raised against mature hIAPP amino acids 25-37, Peninsula Laboratories International) prior to viewing on a Nikon Intensilite Microscope (ThT, anti-Insulin, anti-hIAPP), a Nikon STORM/PALM

microscope with TIRF illuminator (Congo Red) or Nikon Fluorescence Microscope equipped with light polarization sliding control (Congo Red). Amyloid staining protocols have been described previously(13–15) or are freely available online from StainsFile (©2019). Images were processed using FIJI (ImageJ software). Full step-by-step protocols used to stain tissues presented here are available in SI.

4.5 Application of results.

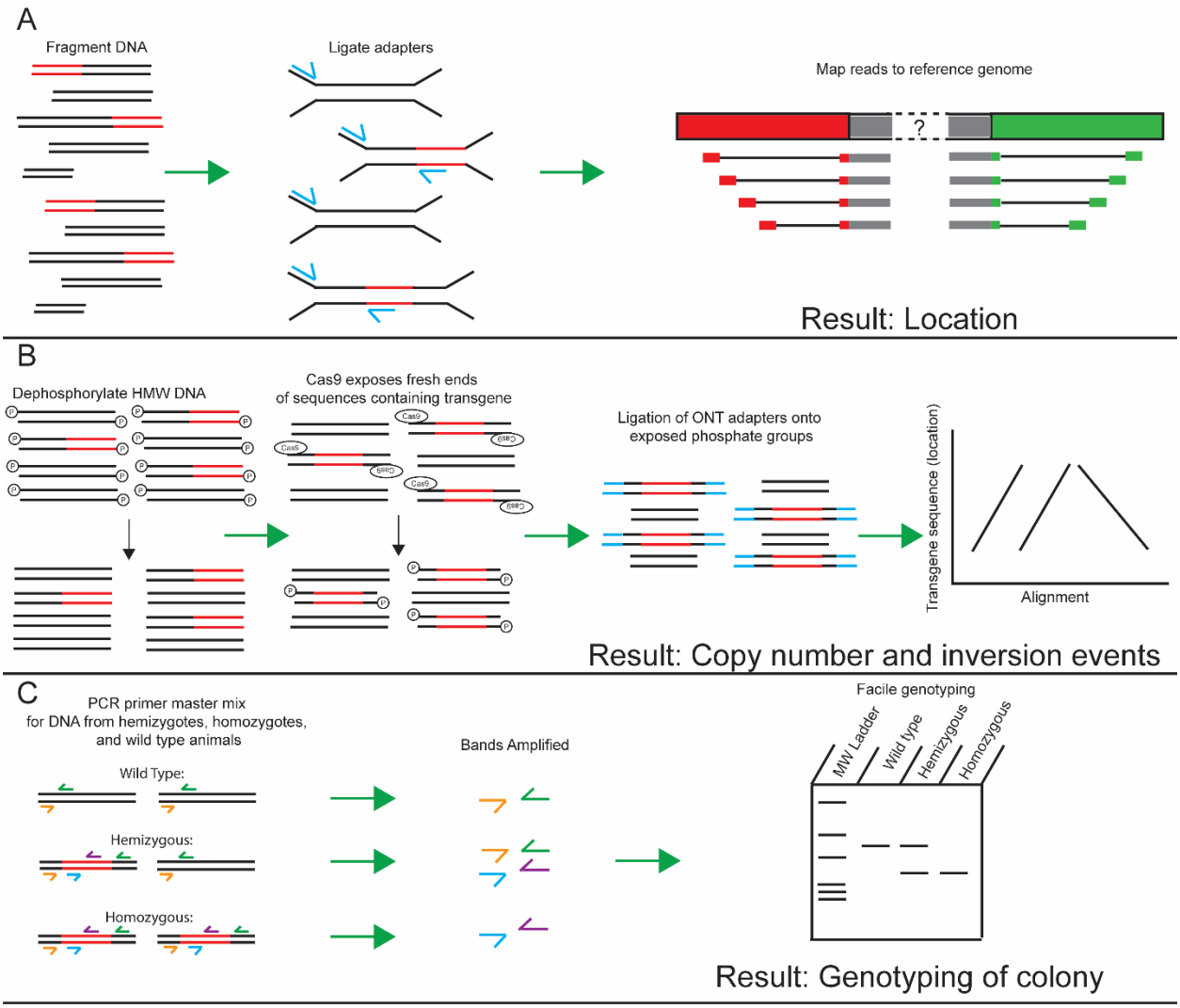


Figure 4.1: MapCap sequencing of transgenic animals.

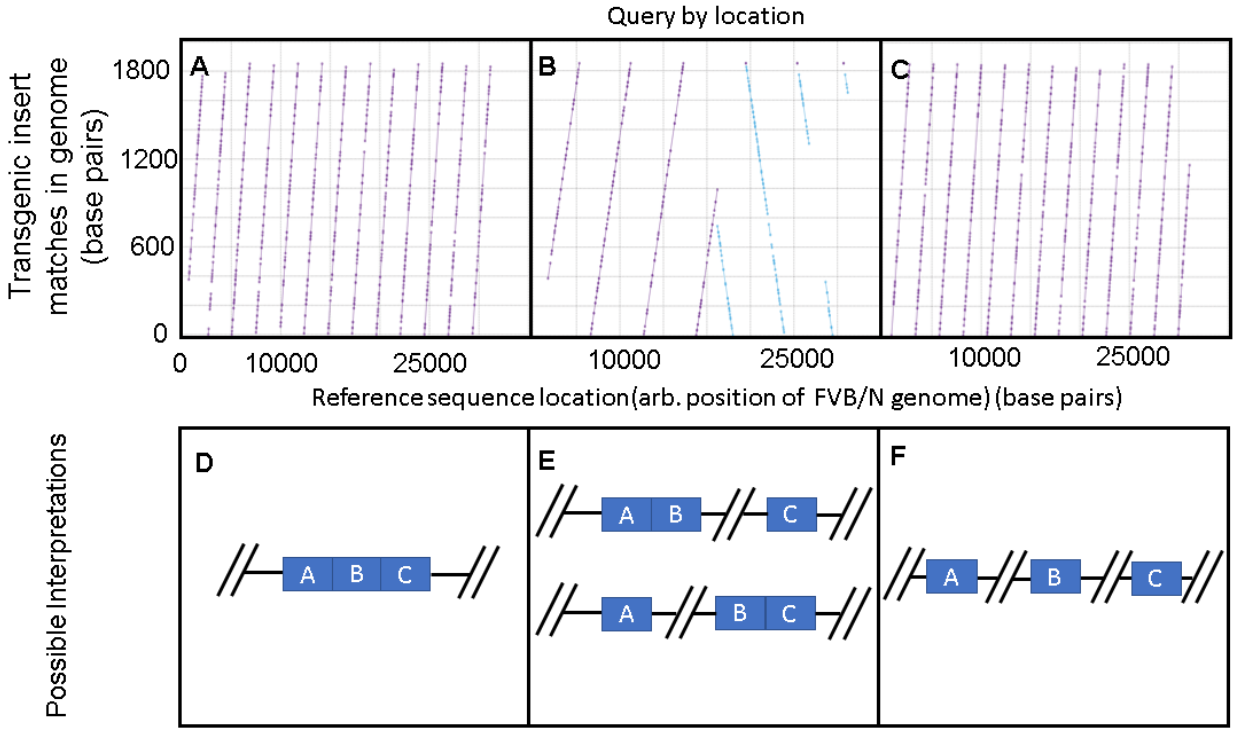
3-step workflow for MapCap sequencing of transgenic animals. A) Fragmenting DNA, Ligation of Adapters, and read mapping to a reference genome provides the location of a transgenic insertion sequence. B) Dephosphorylation of high molecular weight DNA, Cas9-mediated exposure of transgenic ends, ligation of ONT adapters and alignment provide the copy number and inversion events. C) Using the results from A) and B), a PCR primer mix is created for distinguishing wild-type, hemizygous, and homozygous animals.

First, we wanted to define both the chromosomal location and insertion site of the RIPHAT hIAPP transgene. Genomic DNA was isolated from tail punches of

heterozygous mice followed by TruSeq Nano DNA Library Prep. Transgene-specific primers were designed using the plasmid sequence used to generate the RIPHAT transgene. An additional primer was used against the Illumina adapters, as shown in Figure 4.1A, allowing selective amplification of inserts containing the transgene. Mapping this library back to mouse genome yielded the precise insertion location of the transgene.

After mapping the transgene to the precise location within the mouse genome, a Cas9-enrichment long-read DNA assessment was performed (Figure 4.1B), as shown in Figure 4.1C, results from A) and B) are used to genotype the colony. A pair of guide RNAs flanking the insertion site, PAM facing inwards, were designed. Reads passing quality thresholds (guppy defaults) were aligned to the RIPHAT hIAPP transgene sequence with fasta36. Those reads containing any portion of the transgene were aligned to the mouse genome (FVB_NJ_v1, build GCA_001624535.1) with minimap2. Reads spanning the insertion sites on both the 5’ and 3’ ends were identified. Additional reads containing full and partial portions of the transgene, but lacking overlap with the native genomic sequence, were also identified. Inspection of all genomic alignments indicated that no read (longest = 31,520 bp) spanned the entire insertion. Dotplot analysis allowed us to identify at least 26 tandem copies of the expected transgenic insertion, separated by at least one inversion event. A single read revealed at least 13 tandem copies on the centrometic/telomeric side (Figure 4.2A). Another read revealed an inversion between 4 tandem repeats (Figure 4.2B). A third read revealed at least 13 tandem copies on the telomeric/centromeric side (Figure 4.2C).

Our evidence indicates one of 3 scenarios (Figure 4.2D-F) (although there are two possibilities for scenario E). In one scenario, all of the highest-quality reads overlap, and the transgenic region contains 26 copies of the original transgenic construct: 13 forward-facing copies and 13 reverse-facing copies. The two scenarios in Figure 4.2E display possibilities where 2 high-quality reads overlap, but are not overlapped with the third read. Lastly, Figure 4.2F displays a final scenario where none of the reads overlap. In all scenarios, there is a minimum of 26 copies of the transgene at the insertion site, indicating that homozygous mice carry 52 copies of the transgenic construct. Our results place the transgenic construct in a noncoding region of chromosome 15; proximity to the two nearest-neighbor coding sequences is displayed in Figure 4.2G.



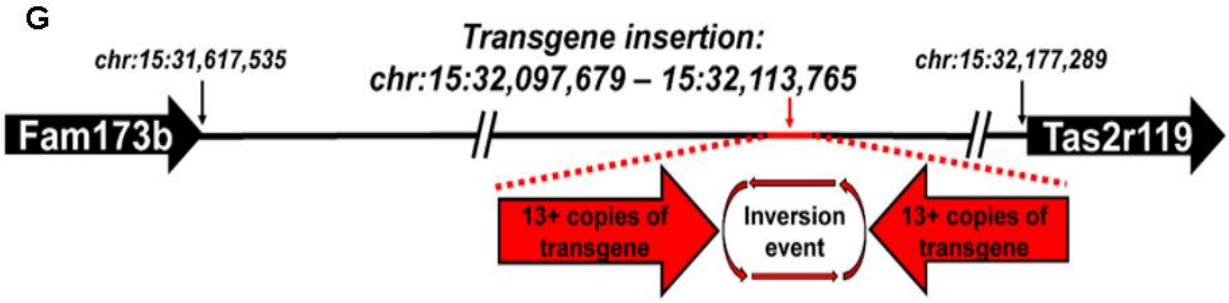


Figure 4.2 Transgenic material in RIPHAT hIAPP (+/-) mice.

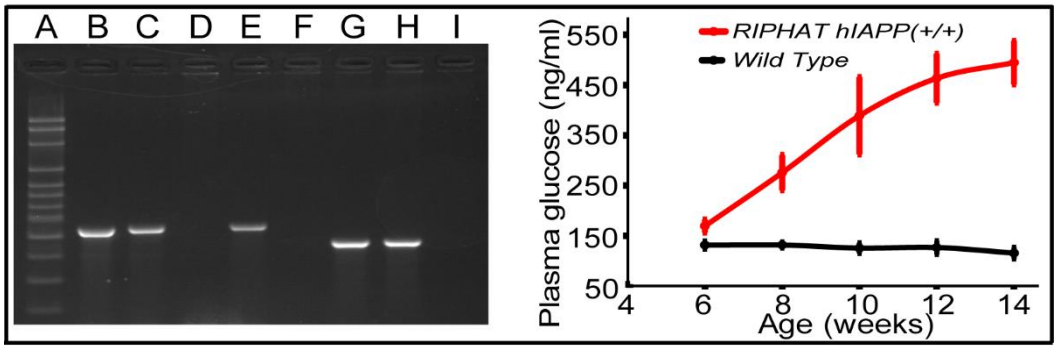
A minimum of 26 copies of transgenic material exist in *RIPHAT hIAPP (+/-)* transgenic mice. Our longest read in the forward direction contained 13 repeat segments (A), followed by an inversion point (B), followed by 13 repeat segments in the opposite orientation (C). The minimum number of copies at the transgenic site is 26, a scenario displayed in (D), where all reads overlap. Scenarios also exists where only 2 of the 3 reads overlap (E), or where none of the reads overlap (F). The transgene insertion site in relation to other coding sequences on mouse chromosome 15 (G).

Using PCR sequences complementary to the rat insulin 2 promoter, it is not possible to distinguish the mice in our colony (SI-1). This is presumably due to high sequence similarity between the rat insulin promoter and the mouse insulin promoter sequences; thus wild-type mice are also positive for the ca. ~500 b.p. band (SI-1). Using a complementary sequence to human GAPDH, located near the end of the transgenic sequence, it is possible to distinguish wild type mice from transgenic mice, but not hemizygous from homozygous mice (SI-1). Shown in Figure 4.3 (top left panel) are results of PCR analysis of wild type mice, hemizygous *RIPHAT hIAPP(+/-)* mice and homozygous *RIPHAT hIAPP(+/+)* mice using primers complementary to part-mouse part-transgene genetic material, developed using our method. Each mouse is examined in 2 lanes of the resulting gel, one lane testing for the presence of native sequence and one for the presence of the transgene. PCR-based genotyping enabled us to identify a 460-bp product for mIAPP versus a 361-bp product for the hIAPP transgene (Figure 4.3)

Control mice and homozygous mice yield the 460-bp and 361-bp products, whereas hemizygous mice yielded both products. Figure 4.3 (top right panel) illustrates averaged fasting blood glucose levels for male *RIPHAT hIAPP(+/+)* mice that became diabetic and wild-type mice (black curve) over 6 to 14 weeks of age.

Male *RIPHAT hIAPP(+/+)* mice spontaneously develop diabetes (T2D).(16) Significant differences in blood glucose between *RIPHAT hIAPP(+/+)* and wild type mice are present at all ages. A few female *RIPHAT hIAPP(+/+)*mice were monitored. They developed diabetes, although not until ~25 weeks of age (results not shown).

Regardless of sex, not all *RIPHAT hIAPP(+/+)* mice develop diabetes by 12 weeks of age. Among 41 *RIPHAT hIAPP(+/+)* male mice, only 15 (~37%) developed diabetes by 12 weeks of age (Figure 4.3, bottom chart), which to our knowledge has not been reported previously. Only *RIPHAT hIAPP(+/+)* homozygous mice that achieve blood glucose levels >300 mg/dl at 12 weeks of age were used in the studies described below.



	-/- (total)	+/- (total)	+/+ (total)	+/+ (T2D)
F1 progeny	10	18	14	3
Total homozygotes	N/A	N/A	41	15

Figure 4.3 Genotyping and phenotyping of wild-type and transgenic RIPHAT mice.

Genotyping and phenotyping of wild-type and transgenic *RIPHAT* mice. Top, left panel: PCR distinguishes wild-type, hemizygous *RIPHAT hIAPP(+/-)*, and homozygous *RIPHAT hIAPP(+/+)* mice. A 100 bp ladder (A). Wild-type FVB/N and C57BL/6J mice display 460-bp band (B and E, respectively) for the native sequence, but do not display

the 361-bp band for the insertion sequence (F and I, respectively). Hemizygous *RIPHAT hIAPP(+/-)* mice display 460-bp band for native sequence (C) and 361-bp band (G) for the insertion sequence, whereas homozygous *RIPHAT hIAPP(+/+)* mice do not display the 460-bp native sequence (D) and are positive for the insertion sequence (H). Top, right panel: Averaged blood glucose levels for wild-type and diabetic *RIPHAT hIAPP(+/+)* mice. Bottom, chart: F1 Progeny of *RIPHAT hIAPP(+/-)* mice. Continued breeding yielded 41 total homozygotes, 15 of which had T2D. Data within chart measured at 15 weeks of age.

Islets from wildtype and *RIPHAT hIAPP(+/+)* mice are both amylin-positive according to Anti-hIAPP antibodies (T4157 polyclonal antibody raised against hIAPP amino acids 25-37, Peninsula Laboratories International) that according to our experiments recognize both mIAPP and hIAPP (SI-2, A2 and A4). However, in contrast to wild type mice, islets in *RIPHAT hIAPP(+/+)* mice show a dramatic loss of β-cells, as judged by reduced insulin immunoreactivity (SI-2, A1 and A3). Islet amyloid using Thioflavin T (SI-2, B1, B2, and B3) and Congo red under polarized light (SI-2, C3) was identified in *RIPHAT hIAPP(+/+)* mice but not wild-type mice (SI-2, B4 and C4, for Thioflavin T and Congo Red under polarized light, respectively). These results suggest that expression of human IAPP results in amyloid formation and, because of the loss of insulin-producing β-cells, inhibits normal islet function, in agreement with previous work on these mice.(16–18)

4.6 Discussion and conclusions.

In summary, we have developed a robust method to genotype colonies of transgenic animals and applied this method to a colony of wild-type and *RIPHAT hIAPP(+/- and +/+)* mice. We also characterized the animals using previous methods, such as blood glucose tracking and fluorescence microscopy. In so doing, we observed that the prevalence of type II diabetes amongst homozygous animals is only 37%, and discovered a previously unidentified subset of animals—those homozygous for the transgenic insertion, but who did not get type II diabetes. We expect other transgenic animal models may have uncharacterized genetic penetrance as well, if the animals are only characterized by phenotype. This method improves the current understanding of the genotype-versus-phenotype relationship in animal models of human diseases.

ACKNOWLEDGEMENTS:

The authors thank the University of Wisconsin-Madison Biotechnology Center for help designing the workflow shown in Figure 1, comparing multiple available sequencing methods, and designing the Cas9-mediated target enrichment.

Funding Sources:

This work was supported by the National Institutes of Health (1R01DK101573-01, 1R01DK101573-06, 1R01DK102948-01A1 (A.D.A.), and R01DK079895 (M.T.Z.)).

4.7 Chapter 4 references.

1. Gordon JW, Scangos GA, Plotkin DJ, Barbosa JA, Ruddle FH. (1980) Genetic transformation of mouse embryos by microinjection of purified DNA. Proc Natl Acad Sci U S A, 77(12):7380–4.
2. Jaenisch R, Mintz B. (1974) Simian virus 40 DNA sequences in DNA of healthy adult mice derived from preimplantation blastocysts injected with viral DNA. Proc Natl Acad Sci U S A, 71(4):1250–4.
3. Jaenisch R. (1976) Germ line integration and Mendelian transmission of the exogenous Moloney leukemia virus. Proc Natl Acad Sci U S A, 73(4):1260–4.
4. Palmiter RD, Brinstert RL. (1985) Transgenic Mice Minireview. Cell, 41:343–5.
5. Naoto H, Cho A, B Kulkarni A. (2010) Engineering transgenic constructs and mice. Curr Protoc Cell Bio, 1–12.
6. Goodwin LO, Splinter E, Davis TL, Urban R, He H, Braun RE, et al. (2019) Large-scale discovery of mouse transgenic integration sites reveals frequent structural variation and insertional mutagenesis. Genome Res, 29(3): 494–505.
7. Pereira R, Halford K, Sokolov BP, Khillan JS, Prockop DJ. (1994) Phenotypic variability and incomplete penetrance of spontaneous fractures in an inbred strain of transgenic mice expressing a mutated collagen gene (COL1A1). J Clin Invest, 93(4):1765–9.
8. Laboulaye MA, Duan X, Qiao M, Whitney IE, Sanes JR. (2018) Mapping transgene insertion sites reveals complex interactions between mouse transgenes and neighboring endogenous genes. Front Mol Neurosci, 11: 385.
9. Heather JM, Chain B. (2016) The sequence of sequencers: The history of

sequencing DNA. Genomics, 107(1):1–8.

10. Hafford-Tear NJ, Tsai YC, Sadan AN, Sanchez-Pintado B, Zarouchlioti C, Maher GJ, et al. (2019) CRISPR/Cas9-targeted enrichment and long-read sequencing of the Fuchs endothelial corneal dystrophy–associated TCF4 triplet repeat. Genet Med, 21:2092-102.
11. Wong SMS, Gawronski JD, Lapointe D, Akerley BJ. (2011) High-throughput insertion tracking by deep sequencing for the analysis of bacterial pathogens. Methods Mol Biol, 733:209-22.
12. Gilpatrick T, Lee I, Graham JE, Raimondeau E, Bowen R, Heron A, et al. (2020) Targeted nanopore sequencing with Cas9-guided adaptor ligation HHS Public Access Author manuscript. Nat Biotechnol, 38(4):433–8.
13. Mukherjee A, Morales-Scheihing D, Salvadores N, Moreno-Gonzalez I, Gonzalez C, Taylor-Presse K, et al. (2017) Induction of IAPP amyloid deposition and associated diabetic abnormalities by a prion-like mechanism. J Exp Med, 214(9):2591–610.
14. Moreno-Gonzalez I, Estrada LD, Sanchez-Mejias E, Soto C. (2013) Smoking exacerbates amyloid pathology in a mouse model of Alzheimer's disease. Nat Commun, 4:1495.
15. Castilla J, Saá P, Hetz C, Soto C. (2005) In vitro generation of infectious scrapie prions. Cell, 121(2):195–206.
16. Janson J, Soeller WC, Roche PC, Nelson RT, Torchia AJ, Kreutter DK, et al. (1996) Spontaneous diabetes mellitus in transgenic mice expressing human islet amyloid polypeptide. Proc Natl Acad Sci, 93(14):7283-8.

17. Matveyenko A V, Butler PC. (2006) Islet Amyloid Polypeptide (IAPP) Transgenic Rodents as Models for Type 2 Diabetes. ILAR J, 47(3):225-33.

18. Couce M, Kane LA, O'Brien TD, Charlesworth J, Soeller W, McNeish J, et al. (1996) Treatment with growth hormone and dexamethasone in mice transgenic for human islet amyloid polypeptide causes islet amyloidosis and β-cell dysfunction. Diabetes, 45(8):1094–101.

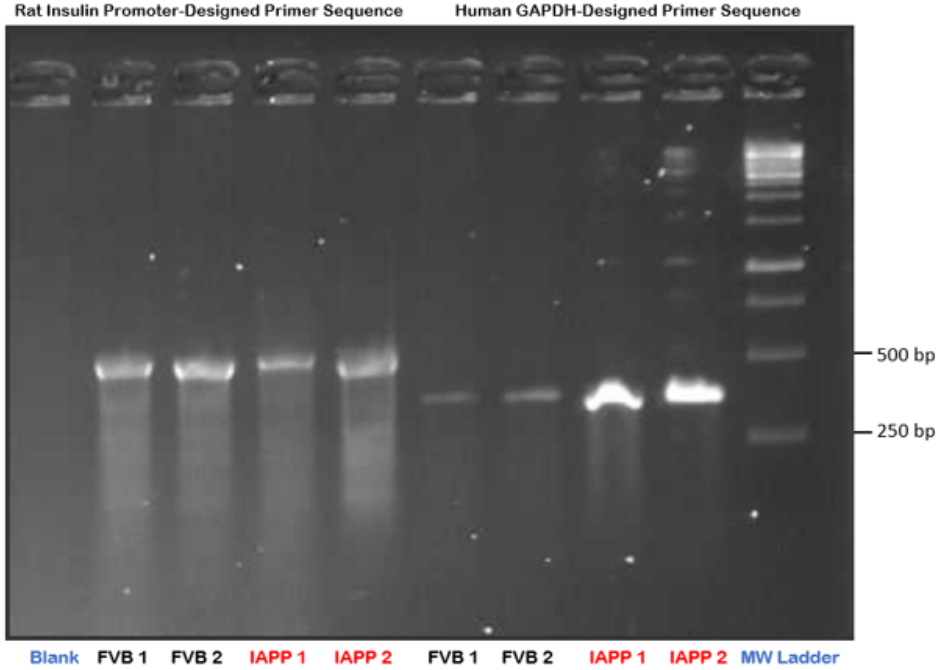
4.8 Supporting information.

Supporting information for “A protocol for locating and counting transgenic sequences from laboratory animals using a map-then-capture (MapCap) sequencing workflow: procedure and application of results.”

Authors: Sidney S. Dicke†, C. Dustin Rubinstein††, James M. Speers††, Mark E. Berres††, Derek M. Pavelec††, Kathryn L. Schueler†††, Donnie S. Stapleton†††, Shane P. Simonett†††, Mark P. Keller†††, Alan D. Attie†††, and Martin T. Zanni†*

SI-1: Information from original transgene construct separates transgenic from wild-type animals, but not hemizygous from homozygous animals

Rat insulin promoter-designed primer distinguishes transgenic from wild-type mice, but not homozygous mice from hemizygous mice:



PCR test results
F1/R1 520 bp strong band in fvb and IAPP
F2/R2 361 bp strong band in IAPP only
F/R2 1451 bp band in IAPP only
F1/R 1026 bp strong band in IAPP, light band in fvb

Supplementary Figure 4.4 (SI – 1): Genotyping mice in the RIPHAT hIAPP colony.

SI-2: Microscopy of mouse islets indicates diabetic homozygous mice from our colony exhibit amyloid plaques, whereas wild-type mice do not.

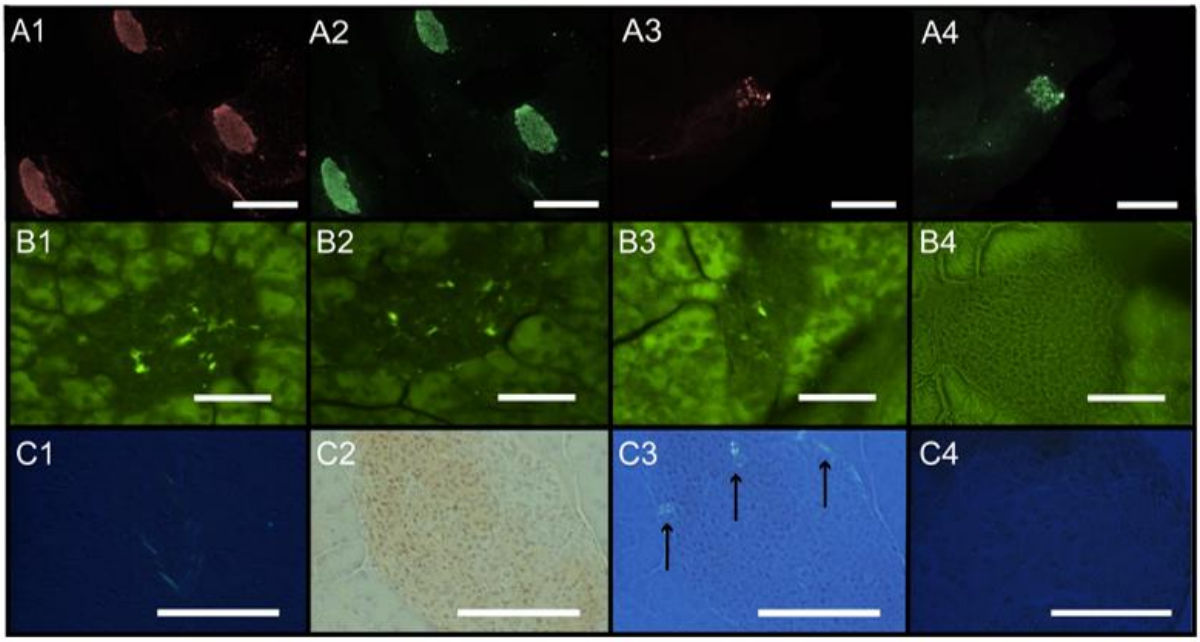
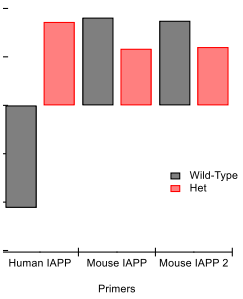


Figure 2: Antibody and plaque staining of murine tissues. Wild-type mice stain positive for both insulin (A1) and amylin (A2). Diabetic mice stain positive for insulin and amylin (A3 and A4, respectively) as well, but show a reduction in islet-positive insulin area (A3) as reported previously (scale bars=150 µm). Diabetic *RIPHAT hIAPP(+/+)* mice stain positive for islet amyloid using Thioflavin T (B1, B2, and B3), whereas wild-type mice (B4) do not (scale bars=50 µm). (C1) Positive control for amyloid fibrils using polarized light. Congo red-stained islet from diabetic *RIPHAT hIAPP(+/+)* mouse under light microscope (C2) and under polarized light microscope (C3), where apple-green birefringence indicates presence of islet amyloid. No apple-green birefringence is observed in wild-type mouse islets (C4) (scale bars=50 µm).

Supplementary Figure 4.5 (SI - 2): Antibody and plaque staining of murine tissues.

SI-3: mRNA measurement for WT versus het mice reveals high expression of transgene relative to B-actin, a housekeeping gene.



Supplementary Figure 4.6 (SI - 3): Relative expression of the RIPHAT transgene.

SI-4: Primer sequences for distinguishing WT, hemizygous, and homozygous via PCR:

CHR15-F2 ctgagtcacagatgcagggat

CHR15-R2 tcgttcacttggcttgcatc

IAPP-F gcatcctgaagctgcaagta

IAPP-R gcattcctcttgccatatgtattg

Supplementary Figure 4.7 (SI-4): Genotyping the RIPHAT mice: distinguishing hemizygous from homozygous mice.

SI-4: Rat insulin-2 designs primer sequence

Rat Ins 2

CCTGAGGAGAACCTCTCCACACTGCCCTGGTCTTCCCACCCTGGTGTCCCAGATA

CCCGGAGTGTGAGTGGCTGCAGCACTTTCTGGGGGACAAGAAGTAGGGAGCAAG

GGGCTCACAGTTCAAGTCTGGTGGCTATAAAGCCCTGCATAGGGTAGAGTTCTCGC

TCATGCAACGACACCAAGGGTTTTTGCTGTCTGCTCGGGGAACAGGGCAGTACCA

AATCAGGAACAGAAAGAGTCAAGGATCCCCCAACCACTCCAAGTGGAGGCTGAGA

AAGGTTTTGTAGCTGGGTAGAGTATGTACTAAGAGATGGAGACAGCTGGCTCTGAG

CTCTGAAGCAAGCACCTCTTATGGAGAGTTGCTGACCTTCAGGTGCAAATCTAAGAT

ACTACAGGAGAATACACCATGGGCTTCAGCCC

AGTTGACTCCCGAGTGGGCTATGGGTTTGTGGAAGGAGAGATAGAAGAGAAGGGA

CCTTTCTTCTTGAATTCTGCTTTCCTTCTACCTCTGAGGGTGAGCTGGGGTCTCAGC

TGAGGTGAGGACACAGCTATCAGTGGGAACTGTGAAACAACAGTTCAAGGGACAA

AGTTACTAGGTCCCCCAACAACTGCAGCCTCCTGGGGAATGATGTGGAAAAATGCT

CAGCCAAGGACAAAGAAGGCCTCACCCTCTCTGAGACAATGTCCCCTGCTGTGAA

CTGGTTCATCAGGCCACCCAGGAGCCCCTCTTAAGACTCTAATTACCCTAAGGCTAA

GTAGAGGTGTTGTTGTCCAATGAGCACTTTCTGCAGACCTAGCACCAGGCAAGTGT

TTGGAAACTGCAGCTTCAGCCCCTCTGGCCATCTGCTGATCCACCCTTAATGGGAC

AAACAGCAAAGTCCAGGGGTCAGGGGGGGGGTGCTTTGGACTATAAAGCTAGTGG

GGATTCAGTAACCCCCAGCCCTAAGTGACCAGCTACAGTCGGAAACCATCAGCAAG

CAGGTATGTACTCTCCAGGGTGGGCCTGGCTTCCCCAGTCAAGACTCCAGGGATTT

GAGGGACGCTGTGGGCTCTTCTCTTACATGTACCTTTTGCTAGCCTCAACCCTGAC

TATCTTCCAGGTCATTGTTCCAACATGGCCCTGTGGATCCGCTTCCTGCCCCTGCT

GGCCCTGCTCATCCTCTGGGAGCCCCGCCCTGCCCAGGCTTTTGTCAAACAGCAC

CTTTGTGGTTCTCACTTGGTGGAAGCTCTCTACCTGGTGTGTGGGGAGCGTGGATT

CTTCTACACACCCATGTCCCGCCGCGAAGTGGAGGACCCACAAGGTAAGCTCTGC

TCCTGAATTCTATCCCAAGTGCTAACTACCCTGTTTGTCTTTCACCCTTGAGACCTT

GTAAATTGTGCCCTAGGTGTGGAGGGTCTCAGGCTAACCAGTGGGGGGCACATTT

CTGTGGGCAGCTAGACATATGTAAACATGGTAGCTGCCAAGAAGGAGTGAGAATCC

TTCCTTAAGTCTCCTAGGTGGTGACGGGTGGCCAGGCCCCAGGATAGGTACCCATT

TGGGGACCCCATAGAGCACCGCACCGACCGAGGGATGGTAACAGGATGTGTAGGT

TTTGGAGGCCCATATGTCCATTCATGACCAGTGACTTGTCTCACAGCCATGCAACCC

TTGCCTCCTGTGCTGACTTAGCAGGGGATAAAGTGAGAGAAAGCCTGGGCTAATCG

GGGGGTCGCTCGGCTCCTCCTAACTGGATTGTCCTATGTGTCTTTG

CTTCTGTGCTGCTGATGCTCTGCCCTGTGCTGACATGACCTCCCTGGCAGTGGCA

CAACTGGAGCTGGGTGGAGGCCCGGGGGCCGGTGACCTTCAGACCTTGGCACTG

GAGGTGGCCCGGCAGAAGCGCGGCATCGTGGATCAGTGCTGCACCAGCATCTGC

TCTCTCTACCAACTGGAGAACTACTGCAACTAGGCCCACCACTACCCTGTCCACCC

CTCTGCAATGAATAAAACCTTTGAAAGAGCACTACAAGTTGTGTGTACATGCGTGCA

TGTGCATATGTGGTGCGGGGGGAACATGAGTGGGGTCGGCTGGAGTGGTCGCGG

CTTAATCTATCTGTCCAGCAGCAATATCCCTAAAGGGACAGCCCGAGACTTCTTTGG

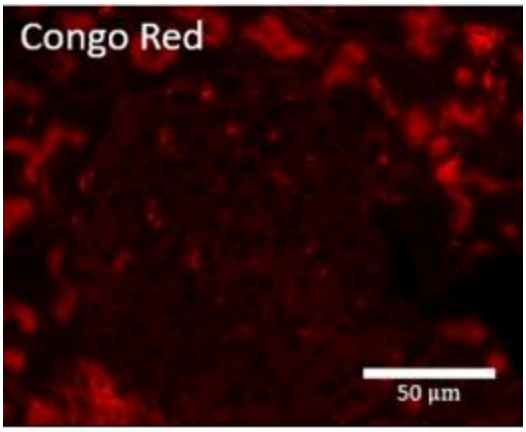
CTTTCTCTGAATCCCCGGTTTTCTCCTGACTCC

CTGGCACCTCTCACTGGTGCCTGTTAGACCGAACCTGAGCAAAGGCAGTTCCTTC
CCCCGAGATAGCTAACGCCTTCGGTTGTCTGGGAGGACCGTGACCCCTACCCCTG
ATGCTTCTAGCCGGCTAGAGAAGAGTTAGAGGTCTTTGGAATGCATGGGTGATGGC
ATCTAGGAACTAATAATTCGTTGCTCTATAGCCCTGGAGGATCCTCAAGGGCCCTTAT
TACTTTTATAAGGAAGAAGACCAAAATATCCCAACCACAGCTTTCACCTAGCCCTCCT
TAGTTCTTTTTCCAGAGCTATTTTAGAGGAGTATCTGACTGGGAAGAAATTGGGCTT
GGTACCTTGAGCTGGAAGGCCATGGAGTCATTCTTAAAGAGCTTATCCNAGCTCTG
GGAGGACAGAGAACAGCCCCAACTGCTCTCAGACTATCCAATGACCTTTAGTGCTC
TGAACAGGACAGGCACCCCACACAGGTGAATAACATACTGAAGGACTGGGCAGGA
GCAGAACTCCACTTCTCAGGAATGCCAGTTGCAAGTTCTAAGCAAGGTAGCCATGA
GAGAGGTCCAGGGCTGGGGTCTCCTATTACTTCTCAAGTCCCTCCTCCCAAGACAG
GTAGGGCCTCTCATCAAGATTTTCTTGAAACTTAAACTGAAAAGCCACATAACATCTA
AGATCT

Supplementary Figure 4.8 (SI - 5): Primer design via the Rat Inuslin II Promoter.

SI-6: Congo red stained *RIPHAT hIAPP* +/- mouse islet (STORM microscope with

TIRF illuminator):



Supplementary Figure 4.9 (Figure SI - 6): Congo red stained RIPHAT hIAPP (+/-) mouse islet.

SI-7:

Primer Design for Oxford Nanopore DNA sequencing (for full RIPHAT transgenic insertion sequence see US patent #6187991 B1):

GATCCCCCAACCACTCCAAGTGGAGGCTGAGAAAGGTTTTGTAGCTGGGTAGAGTA

TGTACTAAGAGATGGAGACAGCTGGCTCTGAGCTCTGAAGCAAGCACCTCTTATGG

AGAGTTGCTGACCTTCAGGTGCAAATCTAAGATACTACAGGAGAATACACCATGGG

GCTTCAGCCCAGTTGACTCCCGAGTGGGCTATGGGTTTGTGGAAGGAGAGATAGA

AGAGAAGGGACCTTTCTTCTTGAATTCTGCTTTCCTTCTACCTCTGAGGGTGAGCT

GGGGTCTCAGCTGAGGTGAGGACACAGCTATCAGTGGGAACTGTGAAACAACAGT

TCAAGGGACAAAGTTACTAGGTCCCCCAACAACTGCAGCCTCCTGGGGAATGATGT

GGAAAAATGCTCAGCCAAGGACAAAGAAGGCCTCACCCTCTCTGAGACAATGTCC

CCTGCTGTGAACTGGTTCATCAGGCCACCCAGGAGCCCCTATTAAGACTCTAATTAC

CCTAAGGCTAAGTAGAGGTGTTGTTGTCCAATGAGCACTTTCTGCAGACCTAGCAC

CAGGCAAGTGTTTGGAAACTGCAGCTTCAGCCCCTCTGGCCATCTGCTGATCCACC
CTTAATGGGACAAACAGCAAAGTCCAGGGGTCAGGGGGGGGTGCTTTGGACTATA
AAGCTAGTGGGGATTCAGTAACCCCCAGCCCTAAGTGACCAGCTACAGTCGGAAAC
CATCAGCAAGCAGGTATGTACTCTCCAGGGTGGGCCTGGCTTCCCCAGTCAAGACT
CCAGGGATTTGAGGGACGCTGTGGGCTCTTCTCTTACATGTACCTTTTGCTAGCCT
CAACCCTGACTATCTTCCAGGTCATTGTTCCACCATGGGCATCCTGAAGCTGCAAG
TATTTCTCATTGTGCTCTCTGTTGCATTGAACCATCTGAAAGCTACACCCATTGAAAG
TCATCAGGTGGAAAAGCGGAAATGCAACACTGCCACATGTGCAACGCAGCGCCTG
GCAAATTTTTTAGTTCATTCCAGCAACAACTTTGGTGCCATTCTCTCATCTACCAACG
TGGGATCCAATACATATGGCAAGAGGAATGCAGTAGAGGTTTTAAAGAGAGAGCCA
CTGAATTACTTGCCCCTTTAGGTGCACGTAAGAAATCCATTTTTCTATTGTTCAACTT
TTATTCTATTTTCCCAGTAAAATAAAGTTTTAGTAAACTCTGCATCTTTAAAGAATTATT
TTGGCATTTATTTCTAAAATGGCATAGCATTTTGTATTTGTGAAGTCTTACAAGGTTAT
CTTATTAATAAAATTCAAACATCCTAGGTAAAAAAAAAAGGTCAGAATTGTTTAGTGAC
TGTAATTTTCTTTTGCGCACTAAGGAAAGTGCAAAGTAACTTAGAGTGACTGAAACT
TCACAGAATAGGGTTGAAGATTGAATTCATAACTATCCCAAAGACCTATCCATTGCAC
TATGCTTTATTTAAAAACCACAAAACCTGTGCTGTTGATCTCATAAATAGAACTTGTAT
TTATATTTATTTACATTTTAGTCTGTCTTCTTGGTTGCTGTTGATAGACACTAAAAGAG
TATTAGATATTATCTAAGTTTGAATATAAGGCTATAAATATTTAATAATTTTTAAAATAGTA
TTCTTGGTAATTGAATTATTCTTCTGTTTAAAGGCAGAAGAAATAATTGAACATCATCC
TGAGTTTTTCTGTAGGAATCAGAGCCCAATATTTTGAAACAAATGCATAATCTAAGTC
AAATGGAAAGAAATATAAAAAGTAACATTATTACTTCTTGTTTTCTTCAGTATTTAACAA
TCCTTTTTTTTTCTTCCCTTGCCCAGACAAGCTTCTAGTGACCCCTGGACCACCAGC

CCCAGCAAGAGCACAAGAGGAAGAGAGAGACCCTCACTGCTGGGGAGTCCCTGC

CACACTCAGTCCCCCACCACACTGAATCTCCCCTCCTCACAGTTGCCATGTAGACC

CCCTGAAGAGGGGAGGGGCCTAGGGAGCCGCACCTTGTCATGTACCATCAATAAA

GTACCCTGTGCTCAACCAGTTACTTGTCCTGTCTTATTCTAGGGTCTGGGGCAGAG

GGGAGGGAAGCTGGGCTTGTGTCAAGGTGAGACATTCTTGCTGGGGAGGGACCT

GGTATGTTCTCCTCAGACTGAGGGTAGGGCCTCCAAACAGCCTTGCTTGCTTCGAG

AACCATTTGCTTCCCGCTCAGACGTCTTGAGTGCTACAGGAAGCTGGCACCACTAC

TTCAGAGAACAAGGCCTTTTCCTCTCCTCGCTCCAGTCCTAGGCTATCTGCTGTTG

GCCAAACATGGAAGAAGCTATTCTGTGGGCAGCTCCAGGGAGGCTGACAGGTGGA

GGAAGTCAGGGCG

Human IAPP

RIP promoter

Sequenced between red blocks

IAPP-R1 TTGGAAACTGCAGCTTCAGC

IAPP-R2 TATCTGCTGTTGGCCAAACA

IAPP-F1 GGAGACAGCTGGCTCTGAGC

IAPP-F2 TCACAGTTGCCATGTAGACC

IAPP-F AGCTGCAAGTATTTCTCATTGTG

IAPP-R CAATACATATGGCAAGAGGAATGC

Supplementary Figure 4.10 (SI - 7): Primer design for Oxford Nanopore sequencing.

Protocols used for tissue staining:

Immunohistochemistry:

1) Dewax and Rehydrate slides:

10 minutes submerged in Xylenes

Move to new solution of Xylenes, submerge for 10 more minutes

2) Rehydrate in solutions of decreasing ethanol concentration (3 min incubation/solution):

100% EtOH 3 min

95% EtOH 3 min

85% EtOH 3 min

75% EtOH 3 min

50% EtOH 3 min

DI H20 3 min

PBS 3 min

PBS 3 min

3) Block non-specific binding sites for 1 hour (Blocking solution is 10% Donkey Serum in PBS +/+):

Remove slide from PBS (the final 3 minute incubation)

Blot the BACK of the slide on a napkin. Use WATman paper to clean off non tissue parts of the front of the slide (do NOT wipe tissue itself)

Lightly pipet 500 ul of blocking solution over the tissue section

Incubate in blocking solution for 1 hr.

4) Incubate in primary antibody overnight (Antibody solution is 1% Donkey Serum in PBS +/+, the antibody dilution is usually given on the spec sheet but you may have to contact the manufacturer)

Blot the BACK of the slide on a napkin. Use WATman paper to clean off non tissue parts of the front of the slide (do NOT wipe tissue itself)

Lightly pipet 300 ul of primary antibody solution over the tissue section

Let sit overnight in cold (4 degree), wet, dark place. If you have an empty pipet container, fill the bottom up with water, lay the slides on the shelf that usually holds the tips, and put the cover on. Do not let the slides dry out

5) Wash and stain with secondary Antibody (Again, secondary antibody is prepared at desired dilution in 1% Donkey Serum in PBS +/+. I recommend ALSO adding a DAPI stain or something to help visualize the different cell types—for example, in the pancreas, we also use insulin stains to label the B-cells. But, IF you are adding another stain, first do the stains separately and then together to make sure they do not cross-react)

Wash once with PBS (+/+). Blot the BACK of the slide as before, and clean off the non-tissue sections of the slide with WATman paper.

Add secondary antibody solution (300 ul/slide)

Let sit for 30 minutes in the cold, wet, dark place as before

6) Wash and mount for visualization

Wash once with PBS (+/+). Blot the BACK of the slide as before, and clean off the non-tissue sections of the slide with WATman paper.

Blot the BACK of the slide on a napkin. Use WATman paper to clean off non tissue parts of the front of the slide (do NOT wipe tissue itself)

Add 20 ul of mounting media in 3 drops across the tissue (20 ul per drop)

Drop a cover slide over the top

'Use WATman paper to press out from the inside of the cover slide gently to the outside—so that no air bubbles remain in the tissue section

Seal the slide with nail polish around the outside

7) storage

Store fluorescent stains in the dark

Congo Red:

(from AFIP Laboratory methods in Histology)

Need: 1% Congo red in DI water, 1% NaOH in DI water, and 1% NaOH+50%EtOH in DI water (alkaline alcohol solution), Meyer's hematoxylin solution

1.) Hydrate tissue sections to water (same as for IHC)

2.) Stain in filtered Congo Red solution

3.) Rinse in DI water briefly; proceed immediately to step 4

4.) Place in alkaline alcohol solution

5.) Wash in running tap water for 5 minutes

6.) Place in Meyer's hematoxylin solution for 5 minutes

7.) Wash in running tap water for 15 minutes

8.) Dehydrate with 95% ethyl alcohol for 2 minutes (x2), absolute ethyl alcohol 2 minutes (x2), and xylenes 2 minutes (x2)

9.) Mount (as for IHC).

Thioflavin T:

1.5 g ThT_300 ml 0.1 N HCL; i.e. 0.5% ThT in 0.1 N HCL (pH ~1.4, see Burns, Pennock, Stoward 1967)

Acetic acid, 1%

1.) Hydrate tissue sections (same as for IHC) into 100% H20.

2.) Place in alum hematoxylin for 2 minutes (suppress background fluorescence)

3.) Rinse with water

4.) Place in ThT solution for 3 minutes

5.) Rinse with water.

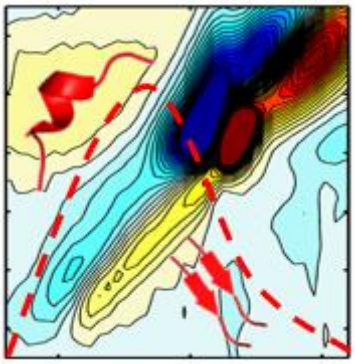
6.) Place in 1% acetic acid for 20 minutes

7.) Rinse with water.

8.) Mount on fluorescence medium as in IHC.

5 FGAIL intermediate detected on membranes via 2D IR spectroscopy.

Metastable intermediate during hIAPP aggregation catalyzed by membranes as detected with 2D IR spectroscopy.[5]



Experiments using 2D IR and 13C18O isotope labeling establish that amyloid fibril formation of hIAPP when catalyzed by membranes includes a metastable intermediate and that this intermediate has a similar structure at G24A25 in the FGAIL region as the corresponding intermediate in solution, thought to be the toxic species.

AUTHORS:

*Sidney S. Dicke†⁺, Michał Maj†^⁺, Caitlyn R. Fields†, and Martin T. Zanni†**

[5] Adapted from: **Dicke SS,** Maj M, Fields CR, Zanni MT. 2022. Metastable intermediate during hIAPP aggregation catalyzed by membranes as detected with 2D IR spectroscopy. *RSC Chem. Biol.* 2022, 3, 931–940.

AFFILIATIONS:

†Department of Chemistry, University of Wisconsin−Madison, 1101 University Avenue, Madison, WI 53706

^Formally at Department of Chemistry, University of Wisconsin−Madison, 1101 University Avenue, Madison, WI 53706

[+]These authors contributed equally

*Email: zanni@chem.wisc.edu

KEYWORDS:

2D IR, Amyloids, Aggregation, Protein Structure

5.1 Abstract.

The aggregation of human islet amyloid polypeptide (hIAPP) into amyloid fibrils involves formation of oligomeric intermediates that are thought to be the cytotoxic species responsible for β-cell dysfunction in type 2 diabetes. hIAPP oligomers permeating or disrupting the cellular membrane may be one mechanism of toxicity and so measuring the structural kinetics of aggregation in the presence of membranes is of much interest. In this study, we use 2D IR spectroscopy and $^{13}C^{18}O$ isotope labeling to study the secondary structure of the oligomeric intermediates formed in solution and in the presence of phospholipid vesicles at sites L12A13, L16V17, G24A25 and V32G33. Pairs of labels monitor the couplings between associated polypeptides and the dihedral angles between adjacent residues. In solution, the L12A13 residues form an oligomeric β-sheet in addition to an α-helix whereas with the phospholipid vesicles they are α-helical throughout the aggregation process. In both solution and with DOPC vesicles, L16V17 and V32G33 have disordered structures until fibrils are formed. Similarly, under both conditions, G24A25 exhibits 3-state kinetics, created by an oligomeric intermediate with a well-defined β-sheet structure. Amyloid fibril formation is often thought to involve intermediates with exceedingly low populations that are difficult to detect experimentally. These experiments establish that amyloid fibril formation of hIAPP when catalyzed by membranes includes a metastable intermediate and that this intermediate has a similar structure at G24A25 in the FGAIL region as the corresponding intermediate in solution, thought to be the toxic species.

5.2 Introduction.

Human islet amyloid polypeptide (hIAPP) is a hormone synthesized with insulin in the β-cells of pancreatic islets that plays a critical role in controlling glucose homeostasis and related metabolic processes.[1,2] Despite its vital function, it is also known to form amyloid fibril aggregates that are deposited at high concentrations in pancreatic islets of type 2 diabetes (T2D) patients.[3–7] The mechanism of cytotoxicity remains largely unknown, and several mechanisms may be at work. One of the proposed mechanisms for cytotoxicity involves hIAPP oligomers interacting with cellular membranes.[8–13] Evidence has been found that hIAPP forms discrete pores in the membrane, similar to those formed by transmembrane ion channels.[14–18] hIAPP may also damage membranes through a detergent-like effect or, in the process of forming fibrils, cause leakage by ripping lipids from the bilayers.[19–21] Regardless of the mechanism, time-lapse toxicity studies suggest that it is oligomeric intermediates that are the toxic species and not the mature fibril.[22–24]

There exists little direct information about the structure of hIAPP oligomers, and even less information about oligomer structure in the presence of membranes. Monomeric hIAPP is largely disordered in solution,[25–27] but may adopt a helix conformation on phospholipid membranes or SDS micelles;[28] residues 5-20 have been observed to form a helix upon introduction of HFIP[29] via NMR and CD spectroscopy; residues 8-18 have been predicted to form a helical core upon interacting with membranes preceding aggregation events;[30] and residues 10-17 have the highest helical propensity[31] in hIAPP according to MD simulations and EPR spectroscopy. Additional work by Miranker and co-workers suggest helices are a stable structure for rat IAPP

(rIAPP), which they concluded using small molecule foldamers that stabilize helix dimers.[32] There is also evidence of β-sheets in the oligomer, primarily observed via 2D IR spectroscopy, as discussed below. Recent work by Roderiguez Camargo et al. using nanodiscs to stabilize hIAPP intermediates has found further evidence of oligomeric β-sheets.[33] In contrast to oligomer structure, the fibrillar structures are relatively well characterized with solid-state NMR and cryoEM. In fibrillar states, residues 20-29 have been observed as a disordered loop with flanking β-sheets.[34] The structural properties also change with pH, and recent cryo-EM and NMR structures have identified β-sheets within the 20-29 region, at pH 6, whereas this region is a loop at pH 7.4.[34–37]

One reason that oligomers are difficult to study is because they are transiently populated, forming intermediates in the aggregation pathway. In classical nucleation theory, often used to describe amyloid fibril formation, oligomers are an unstable species that undergo a nucleation event when they are present above some critical concentration, upon which the fibrils are then grown. According to this model, as well as recent work on the secondary nucleation of hIAPP,[38,39] intermediates should be difficult to identify because they grow into fibrils almost as soon as they form and so exist at low concentrations.[40,41] Seemingly at odds with classical nucleation theory, we have experimentally identified an intermediate of hIAPP that is stable during the lag phase. It has a critical concentration of between 150 and 250 μM[41] and a β-sheet-like structure between the FGAIL residues that span 23 to 27,[42] and a helical structure at residues L12A13.[43] The β-sheet-like region may extend as far as residues S20 to S29.[44–48] The intermediate is an oligomer predicted to be formed by at least 5 associated polypeptides[41] and correlates with cellular toxicity.[22] The oligomers are kinetically stabilized because the

FGAIL region must disorder to form the structure of the fibril, causing a free energy barrier between 3-20 kcal/mol, depending on the size of the oligomer, which is the reason why this intermediate is observed throughout the lag phase.[41] If proline mutations are introduced into the FGAIL region to disrupt the oligomer structure, fibrils do not form, consistent with an on-pathway intermediate.[49]

Here, we study hIAPP secondary structure in solution using Tris buffer at pH 7.37 to mimic the cytosol or extracellular space with and without the addition of phospholipid vesicles. The primary goal of our study is to identify differences in the aggregation pathway when aggregation is membrane catalyzed. The study is designed to test regions of hIAPP whose secondary structure we know is well defined in buffer. We are particularly interested in determining if the “FGAIL intermediate” described above also forms when aggregation takes place in solution with vesicles, since we think that the structure in the FGAIL region stabilizes the oligomers and is the cause of toxicity.

5.3 Materials and Methods.

2D IR Spectroscopy

The 2D IR spectrometer setup used in this work consists of a 1 kHz amplifier (Solstice, Spectra-Physics) coupled to a commercial optical parametric amplifier (TOPAS, Light Conversion Ltd.) that produces signal and idler beams in the near-IR. The signal and idler beams are then focused onto a 0.5 mm thick $AgGaS_2$ crystal to generate mid-IR light at 6 µm through a difference-frequency generation process. The mid-IR light is split into pump and probe pulses with a beam-splitter and the pump light is directed into a Ge-based acousto-optic modulator (AOM) that enables phase stabilization, rapid data acquisition, and phase cycling. Rapid acquisition makes it possible to collect 2D IR spectra continuously during peptide aggregation with high signal-to-noise ratio while phase cycling effectively removes pump scatter.[50,51] The pump beam leaving the AOM is focused with the probe beam onto a sample at a cross geometry and the self-heterodyned signal is measured with a LN_2-cooled MCT array detector. For a more detailed description of a pulse-shaping assisted 2D IR spectrometer, we refer readers to previous publications.[51–56]

Peptide synthesis and purification

Isotope-labeled hIAPP samples were synthesized by Fmoc-based chemistry using an automated microwave peptide synthesizer (LibertyBlue, CEM) on PAL-PEG-PS resin according to published protocols.[51,57] Peptides were cleaved off resin with a microwave-assisted cleavage system (Accent, CEM) using TFA:TIS:H_2O (18:1:1 v/v/v) solution. To form a disulfide bond between the N-terminal cysteines, crude peptide was dissolved in DMSO and kept at room temperature for 24 hours. Following the oxidation, the peptide

was purified with reversed-phase HPLC using C18 preparative column (Waters XSelect) using a two-buffer purification gradient. Buffer 1 was composed of 0.045% HCl in H_2O, and buffer 2 of 80% CH_3CN, 20% H_2O and 0.045% HCl (v/v). Purification was run with a gradient, increasing the percentage of buffer 2 by 1% per minute. Purity of the peptides were assessed by integrating the 220 nm peptide backbone peak versus all other impurities visible on the HPLC chromatogram; peptide identity was confirmed via MALDI. All peptides used for experiments were >95% purity. All measurements were carried out at 1 mM peptide concentration.

Preparation of phospholipid vesicles

DOPC (1,2-dioleoyl-sn-glycero-3-phosphocholine) was purchased from Avanti Polar Lipids and stored at -20°C until use. Cholesterol was acquired from Sigma-Aldrich and used without further purifications. Liposome samples were prepared through the standard thin-film rehydration procedure. Initially, a mixture of DOPC and cholesterol was prepared in the ratio 75:25, dissolved in chloroform, and evaporated over a stream of nitrogen. Dry film was then placed under vacuum to remove any residual solvent. The film was rehydrated with a Tris buffer solution at pH 7.37 and vortexed at room temperature for 2 hours. The solution then underwent several freeze-thaw cycles and extruded 12 times through a 200 nm membrane. Dynamic light scattering measurements post-extrusion confirmed the average size of the vesicles to be between 140 and 160 nm. The final concentration of the liposomes was set to 75 mM.

Non-negative matrix factorization

2D IR spectra in this work were analyzed using a non-negative matrix factorization algorithm[58] as previously applied by Maj et al[43]. Briefly, a non-negative input matrix *I*

consisting of time-resolved diagonal traces extracted from the 2D spectra is fit to a set of non-negative *W* and *H* matrices, containing vibrational eigenspectral and time-dependent intensity changes, respectively. Matrix multiplication of *W* by *H* yields a matrix similar to *I*, differing by a residual matrix *U*, giving the quality of the fit (Equation 1)

$$I = WH + U.$$

The quality of approximation is determined by the multiplicative update algorithm. 16 replicates of factorization were obtained per data set.

5.4 Results.

Amyloid aggregation is a difficult problem to study with standard structural biology techniques because it involves kinetics, large assemblies, and (in this case) DOPC vesicles. With two-dimensional infrared (2D IR) spectroscopy it is possible to monitor kinetically evolving structures via on-the-fly acquisition of 2D IR spectra.[51,54,59,60] The information on the peptide structure is extracted from time-resolved vibrational spectra measured in the amide I band region (~1610-1710 cm^{-1}) and the $^{13}C^{18}O$ isotope-labeled region (~1570-1610 cm^{-1}). The amide I band arises primarily due to carbonyl stretching vibration of the peptide backbone, making it a sensitive probe of secondary structure, while introduction of $^{13}C^{18}O$ isotope labels is an efficient way of extracting structural information at a specific site in the peptide sequence.[50,51,53] We utilize an isotope labeling method, called “dihedral indexing,”[43] that increases the sensitivity to secondary structure by taking advantage of the vibrational coupling between adjacent amino acids, which was how we detected α-helices in the monomeric state of hIAPP at physiological pH.[43]

In what follows, we apply 2D IR spectroscopy and the dihedral indexing approach to investigate the structure and kinetics of transient intermediates formed during the aggregation of hIAPP in neat Tris buffer and in the presence of phospholipid vesicles. We choose DOPC/Cholesterol mixtures as our model membranes as mimics of the β-cell plasma environment although physiological plasma membranes are much more varied in composition and contain negatively-charged headgroups such as phosphoserine.[61,62] We introduce isotope labels into four distinct sites of the hIAPP sequence, including the most amyloidogenic FGAIL region. As we show below, we resolve vibrational spectra of each species formed on the aggregation pathway, including α-helical vesicle-associated

species, oligomers, and fully-formed fibrils. The data shows that the FGAIL intermediate formed with the presence of vesicles has the same secondary structure in the FGAIL region as to the one that forms in solution without vesicles at these locations. The results strengthen the hypothesis of the FGAIL being the most critical region for the amyloidogenic properties and toxicity of hIAPP.[45]

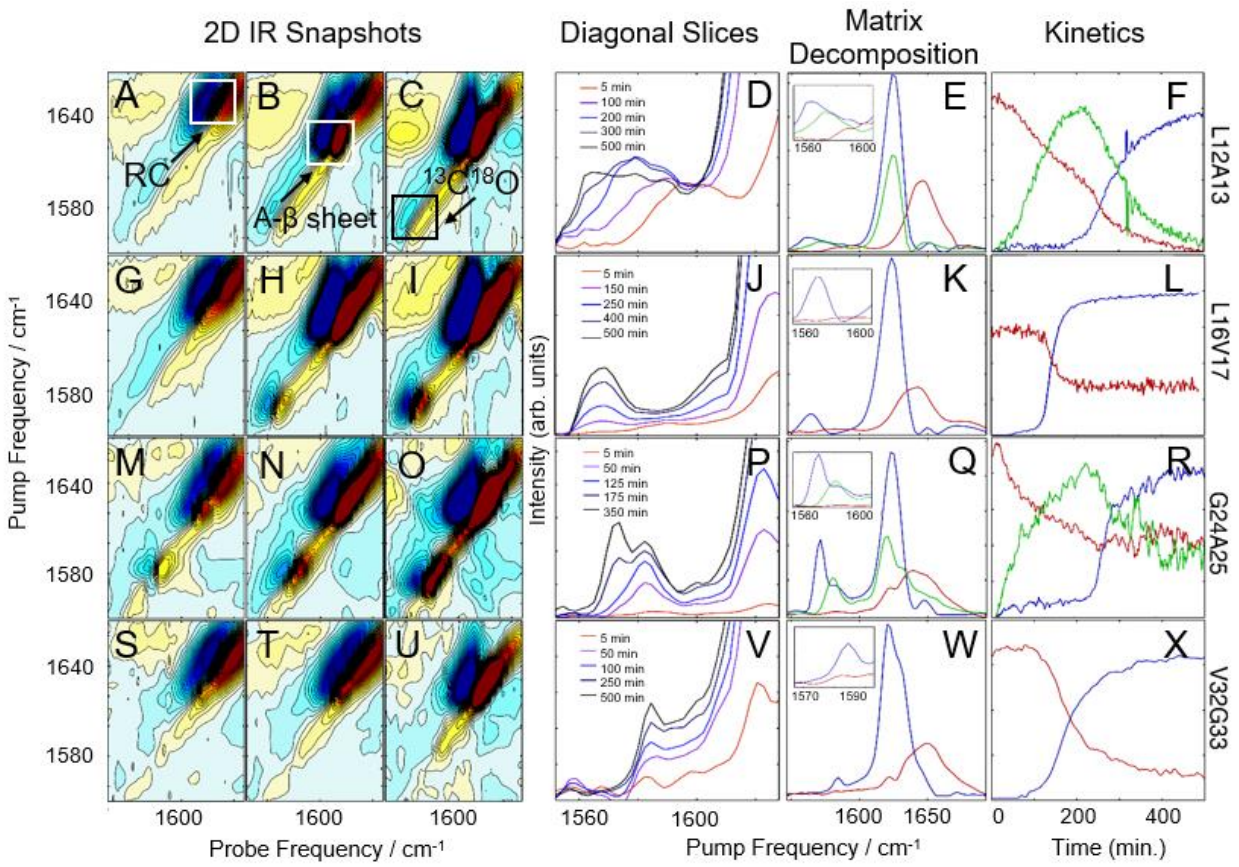


Figure 5.1: 2D-IR spectra and kinetics of labeled hIAPP in Tris buffer.

2D IR snapshots of L12A13-labeled hIAPP aggregation at 5 minutes (**A**), 150 minutes (**B**), and 500 minutes (**C).** (**D**)The diagonal of 2D IR spectra in the isotope labeled region of L12A13-labeled hIAPP at 5 time points during aggregation. Matrix decomposition of the time-dependent diagonal trace reveals 3 state, shown in red, green, and blue in **E. F** displays kinetics plots of the intensities of the matrix-decomposition traces shown in **E**.
2D IR snapshots of L16V17-labeled hIAPP aggregation at 5 minutes (**G**), 250 minutes (**H**), and 500 minutes (**I).** (**J**)The diagonal of 2D IR spectra in the isotope labeled region of L16V17-labeled hIAPP at 5 time points during aggregation. Matrix decomposition of the time-dependent diagonal trace reveals 2 states, shown as a red and green trace in **K. L** displays kinetics plots of the intensities of the matrix-decomposition traces shown in **K**.

2D IR snapshots of G24A25-labeled hIAPP aggregation at 5 minutes (**M**), 125 minutes (**N**), and 350 minutes (**O).** (**P**)The diagonal of 2D IR spectra in the isotope labeled region of G24A25-labeled hIAPP at 5 time points during aggregation. Matrix decomposition of the time-dependent diagonal trace reveals 3 states, shown in red, green, and blue in **Q. R** displays kinetics plots of the intensities of the matrix-decomposition traces shown in **Q**. 2D IR snapshots of V32G33-labeled hIAPP aggregation at 5 minutes (**S**), 100 minutes (**T**), and 500 minutes (**U).** (**V**)The diagonal of 2D IR spectra in the isotope labeled region of V32G33-labeled hIAPP at 5 time points during aggregation. Matrix decomposition of the time dependent diagonal trace reveals 2 states, shown as red and blue traces in **W. X** displays kinetics plots of the intensities of the matrix-decomposition traces shown in **W**. Data presented in A-L was previously published by Maj et al[43] and is reproduced here for ease of comparison.

hIAPP in buffer: Two- and Three-State Kinetics at different locations

Shown in Figure 5.1 are the results of 2D IR kinetics experiments of hIAPP labeled at L12A13, L16V17, G24A25, and V32G33, at a concentration of 1 mM in Tris buffer at pH 7.37, with no vesicles present. Data in Figure 5.1 A-L were previously published[43]; we include them here for ease of reference. Representative 2D IR spectra are shown at the beginning of aggregation (5 minutes after initiating aggregation), at the end of aggregation whereafter no discernable changes to the spectra occurred, and at a mid-point between the two. A typical 2D IR spectrum consists of a diagonal signal arising due to transitions between the ground state and the first vibrationally excited states, and a negative excited state absorption signal that is shifted from the diagonal line due to anharmonicity.[50,53]

The unlabeled residues will have absorptions between 1610 and 1710 cm^{-1}, whereas the isotope labeled residues will fall between 1565 and 1605 cm^{-1}.[42,43,51] To help interpret the data, the unlabeled portion of the spectra in Figure 5.1A, 5.1B, and 5.1C are labeled “RC” for ‘random coil’ and “A-β sheet” for ‘amyloid B-sheet’, while the isotope labeled portion is labeled “$^{13}C^{18}O$”. The unlabeled residues are useful for monitoring the transition from monomer to fibrils. At 5 minutes, the spectrum in Figure 5.1A is dominated

by a broad peak at 1650 cm^{-1}, indicating that hIAPP is largely random coil, as would be expected at the beginning of the lag phase.[42,63] By 150 minutes, the amyloid β-sheet peak is the largest feature and continues to grow throughout the experiment (Figure 5.1C). All of the samples measured, regardless of where hIAPP is isotope labeled, have similar unlabeled features because there are 35 unlabeled residues and only 2 labeled residues.

The more insightful portion of the spectrum is the $^{13}C^{18}O$ region encompassed by the black box in Figure 5.1C for which there are clear differences between labeled amino acids. In Figure 5.1D, a diagonal trace through the isotope labeled spectral region of the combination band is tracked over the course of the experiment.[43,64] Diagonal slicing can be interpreted similarly to FTIR spectra, although 2D IR spectra are more sensitive to secondary structures.[50,64–66] When a secondary structure forms, like an α-helix or a β-sheet, the backbone carbonyl (amide I) vibrations from multiple amino acids couple to one another, creating delocalized vibrational modes.[43,50] That delocalization also increases the transition dipole strength, μ, of the delocalized mode. Higher sensitivity arises because 2D IR spectra scale as μ^4, whereas FTIR scale as μ^2.[50,64,65] Thus, unlike an FTIR spectrum, one does not expect the integrated area of a 2D IR spectrum to remain constant.[66,67] After 5 minutes (red trace), a peak is visible in the labeled region above 1600 cm^{-1}, indicative of α-helical secondary structure. After 100-300 minutes (purple trace, blue trace, and grey trace), multiple features appear in the isotope region, with the lower-frequency peak increasing in intensity over time, consistent with a transition from α-helix to β-sheet structure. This data, determining that 27-38% of residues L12A13 form a helical structure during the lag phase of aggregation, prior to forming oligomeric β-sheets, was previously published.[43] After 500 minutes (black trace), an isotope-labeled

feature has grown in at 1570 cm^{-1}, indicative of ordered β-sheet structure. Thus, L12A13 exhibits 3-states.

It is clear from the raw data that 3 states are present at L12A13, but their spectra are overlapping. To determine the spectra for each of the 3-states and better resolve their kinetics, we apply a non-negative matrix factorization (referred to as “matrix factorization” in the text) method previously applied by Maj et al[43] (See Methods). Matrix factorization uses the thousands of spectra measured during the kinetics to find either 2 or 3 eigenvectors corresponding to distinct spectra and their respective populations as a function of time. The raw experimental data are reproduced by carrying out matrix multiplication of the resulting vectors. Figure 5.1E displays three state decomposition of the diagonal slices of L12A13 yielding the vibrational eigenspectra of monomers (red), intermediates (green), and fibrils (blue). Plotting the intensities of the eigenspectra over time yields the kinetics traces shown in Figure 5.1F. The monomer trace (red) starts high during the lag phase, and decays throughout the experiment as amyloid fibrils form (blue). The monomer eigenspectrum corresponding to the monomer trace displays an isotope-region frequency above 1600 cm^{-1} (Figure 5.1E, zoom, red), consistent with a population of monomeric hIAPP polypeptides with helices in this region. The oligomer trace (green) displays a lower frequency in the isotope region consistent with small β-sheets (Figure 5.1E, zoom, green). The fibril formation eigenspectrum (Figure 5.1E, blue) follows sigmoidal kinetics (Figure 5.1F), as is typical for amyloid fibril formation. The intermediate state (green) rises as the monomers decay and falls as fibrils form, consistent with a transiently populated structure during the lag phase.

Figure 5.1 G, H, and I present analogous spectra for isotope labeling at L16V17. For L16V17, only monomers or fibrils are observed; thus, as the fibrils form, the monomer decreases and the fibril signal increases, indicative of two-state kinetics. The matrix factorization identifies 2 states, consistent with the visual inspection of Figure 5.1J. The intensities of the peaks corresponding to each state are plotted in Figure 5.2L, showing the inverse correlation in intensities between monomers (red trace) and fibrils (blue trace) over time. In contrast to Figure 5.1E, the monomeric trace does not display a prominent isotope-labeled peak indicating that unaggregated hIAPP is largely disordered in solution.

The next set of labels, G24A25, is within the so-called “FGAIL” region of hIAPP. The FGAIL region spans residues 23 to 27 and has been a matter of extensive research because sequence comparisons between hIAPP variants distinguish it as the primary reason for the amyloidogenic properties of the peptide.[45,68–71] We have studied the FGAIL region previously using 2D IR spectroscopy,[41,42,56,72] but this is the first time that we use double isotope labels. The spectra of G24A25 peptide are presented in Figure 5.1M-O. At 5 minutes, a small isotope signal is visible at ~1580 cm^{-1}. By 175 minutes, the 1580 cm^{-1} peak has become very intense. By the end of aggregation, the 1580 cm^{-1} isotope-labeled peak has lessened in intensity and a new intense spectral feature is seen at 1568 cm^{-1}. Thus, G24A25 exhibits 3-state kinetics, which is confirmed by matrix decomposition in Figure 5.1Q and R. In Figure 5.1R, the intensity of the 3 states is plotted over time, showing the rise and fall of an intermediate, which is present during the lag time. The intensity of the intermediate peak indicates that it is caused by coupling between multiple polypeptides, consistent with an oligomer adopting a regular secondary structure. The

oligomeric and fibril traces exhibit isotope peaks in the 1570-1590 cm^{-1} region, consistent with oligomeric polymorphs[73] or separate β-sheet populations.[43]

The last set of double labels are V32G33, whose spectra are shown in Figure 5.1S-X. 2-state kinetics are observed in solution, indicating a transition from monomer to fibril with little to no regular secondary structure present during the lag phase when oligomers are present.

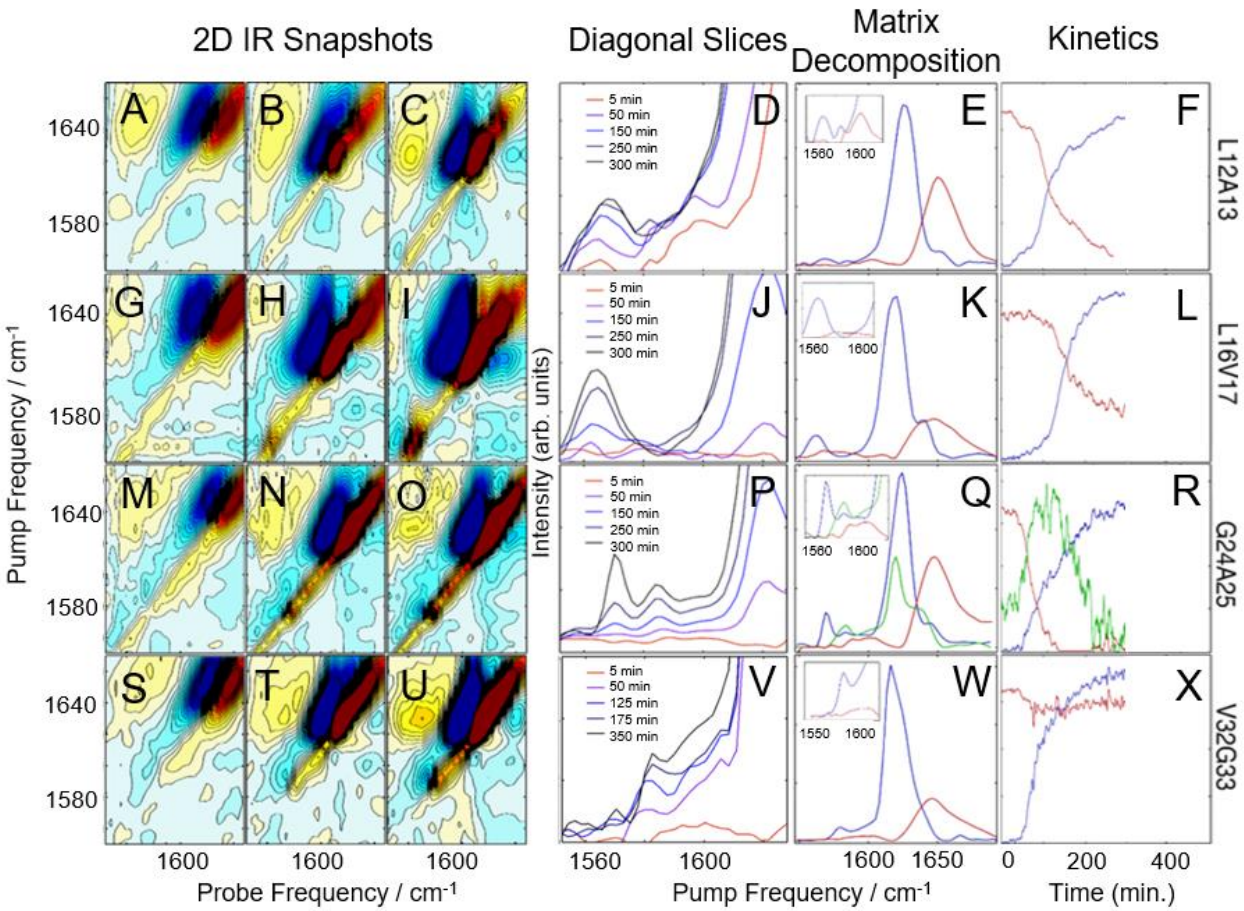


Figure 5.2: 2D-IR spectra and kinetics hIAPP in the presence of DOPC vesicles prepared with cholesterol.

2D IR snapshots of L12A13-labeled hIAPP aggregation at 5 minutes (**A**), 150 minutes (**B**), and 350 minutes (**C).** (**D**)The diagonal of 2D IR spectra in the isotope labeled region of L12A13-labeled hIAPP at 5 time points during aggregation. Matrix decomposition of the time-dependent diagonal trace reveals 2 states, shown in red and blue in **E. F** displays kinetics plots of the intensities of the matrix-decomposition traces shown in **E**.

2D IR snapshots of L16V17-labeled hIAPP aggregation at 5 minutes (**G**), 150 minutes (**H**), and 350 minutes (**I).** (**J**)The diagonal of 2D IR spectra in the isotope labeled region of L16V17-labeled hIAPP at 5 time points during aggregation. Matrix decomposition of the time-dependent diagonal trace reveals 2 states, shown as a red and green trace in **K. L** displays kinetics plots of the intensities of the matrix-decomposition traces shown in **K**. 2D IR snapshots of G24A25-labeled hIAPP aggregation at 5 minutes (**M**), 150 minutes (**N**), and 350 minutes (**O).** (**P**)The diagonal of 2D IR spectra in the isotope labeled region of G24A25-labeled hIAPP at 5 time points during aggregation. Matrix decomposition of the time-dependent diagonal trace reveals 3 states, shown in red, green, and blue in **Q. R** displays kinetics plots of the intensities of the matrix-decomposition traces shown in **Q**. 2D IR snapshots of V32G33-labeled hIAPP aggregation at 5 minutes (**S**), 125 minutes (**T**), and 350 minutes (**U).** (**V**)The diagonal of 2D IR spectra in the isotope labeled region of V32G33-labeled hIAPP at 5 time points during aggregation. Matrix decomposition of the time dependent diagonal trace reveals 2 states, shown as red and blue traces in **W. X** displays kinetics plots of the intensities of the matrix-decomposition traces shown in **W**.

Data in Figure 5.2 A-X is analogous to that presented in Figure 5.1 A-X except experiments in Figure 5.2 were performed in the presence of DOPC vesicles (see Methods).

Figure 5.2 A-F display results for L12A13-labeled hIAPP. The diagonal slices and matrix decomposition analysis in Figure 5.2 D and E, respectively, reveals 2-state behavior. Interestingly, this is not the behavior without vesicles, for which 3-state kinetics were observed (Figure 5.1, D and E). Based on the frequencies these states appear at, the data presented here suggests that while aggregating in the presence of vesicles, L12A13 sample either α-helix or fibrillar states, but we see no evidence of the short β-sheets that were observed in solution.

Figure 5.2 G-L display results for L16V17 labeled hIAPP in the presence of vesicles. For this set of labels, 2-state kinetics are observed, as was observed in buffer. Thus, the vesicles do not induce a structural change in the L16V17 region of the oligomer.

Figure 5.2 M-R display results for isotope-labeled hIAPP at G24A25, residing in the FGAIL region that forms a parallel β-sheet in oligomers in buffer. These results show

that G24A25 proceeds in a 3-state kinetic model with 2D IR spectra that exhibit a clearly resolved isotope labeled peak at 1580 cm^{-1}; the spectra and kinetics of G24A25 with vesicles is very similar to what was seen in buffer. Therefore, the FGAIL oligomer, as monitoring by G24A25, is still an intermediate in the aggregation pathway of hIAPP whether or not amyloid formation occurs in the presence of DOPC vesicles.

Data for the final hIAPP label, V32G33, is shown in Figure 5.2S-X. It undergoes 2-state kinetics, similar to the data taken without vesicles, although some subtle differences exist in the spectra (such as the intensity of the random coil peak).

We also note that aggregation in the presence of DOPC vesicles is more rapid. The lag time, as measured by the half-rise time of the sigmoidal kinetics in Figures 5.1D and 5.2D, is about twice as short in the presence of vesicles than without, taking about 200 minutes without vesicles and 100 minutes with vesicles. Because aggregation is faster, data was only collected for 300 minutes in Figure 5.2 instead of 500 minutes, as in Figure 5.1. The kinetics in Figures 5.1 and 5.2 are plotted on the same x-axis to illustrate the increase in aggregation rate. Reproducibility of kinetics is difficult in amyloid aggregation experiments, but the trend holds for all data collected at each set of sites, and is consistent with membrane enhanced aggregation kinetics reported previously.[74,75] Due to convection currents when hIAPP is mixed with the vesicle solution, all hIAPP will come into contact with a vesicle during the time that it takes to place the sample in the spectrometer. Thus, even the earliest measured 2D IR spectra will already be impacted by interactions with vesicles.[76] That being said, we cannot determine if hIAPP remains bound to the membranes during the course of the experiment, but the change in structure

at L16V17 and the increase in aggregation rate (decreased lag time) proves that the membranes does alter the hIAPP structure and the kinetics.

5.5 Discussion.

In the **Results** above, we presented data from 4 sites of hIAPP over the course of aggregation both in the absence- and presence- of vesicles. The primary question that we are trying to address is whether the FGAIL oligomer that we have extensively studied and modeled in buffer is also involved in hIAPP aggregation in the presence of vesicles. In buffer, the proteins evolve from monomers to oligomers and then to amyloid fibrils.[22,42] The ability to spectroscopically resolve each of the 3 states depends on the couplings between the isotope labeled residues.[43] We have previously established that all residues studied here have sizeable inter-molecular couplings once the fibrils have formed because of the close proximity between β-strands.[42,43,65] Even residues G24A25 that are at the edge of a disordered loop still couple to the adjacent hIAPP polypeptides above and below it in the fibril.[34,43] Thus, the spectra for all 4 sets of amino acids here are expected to exhibit a distinct spectrum when fibrils are formed. That being said, not all amino acids will necessarily report on the formation of the oligomers. If a residue has a disordered structure in both the monomer and the oligomer, then its spectrum will be unchanged during the lag phase and new spectral features will not be observed until fibrils are formed. We expected G24A25 to exhibit a distinct oligomer spectrum when aggregation proceeded in buffer, because we knew from our extensive prior work that the FGAIL region adopts a parallel β-sheet in the lag phase prior to fibrils formation.[42] Indeed, G24A25 in buffer did indeed exhibit 3-state kinetics. Thus, some residues will exhibit 2-state kinetics and others 3-state kinetics, depending on whether those particular residues have a unique secondary/quaternary structure in the oligomer.

Shown in Figure 5.3 is a graphical summary of the structural conclusions drawn from the 2D IR spectra and kinetics presented above. We break the aggregation kinetics into monomer, oligomer and fibril, reflecting their respective timepoints during the kinetic traces. The amino acids measured in this study are labeled by the secondary structure drawn from the interpretation of the 2D IR spectra. Secondary structure assignments are also given to other residues as measured in prior studies. Shown in Figure 5.4 is a graphical cartoon of the 3-state aggregation pathway, highlighting the similarities and differences in the aggregation mechanism in buffer and in the presence of DOPC vesicles.

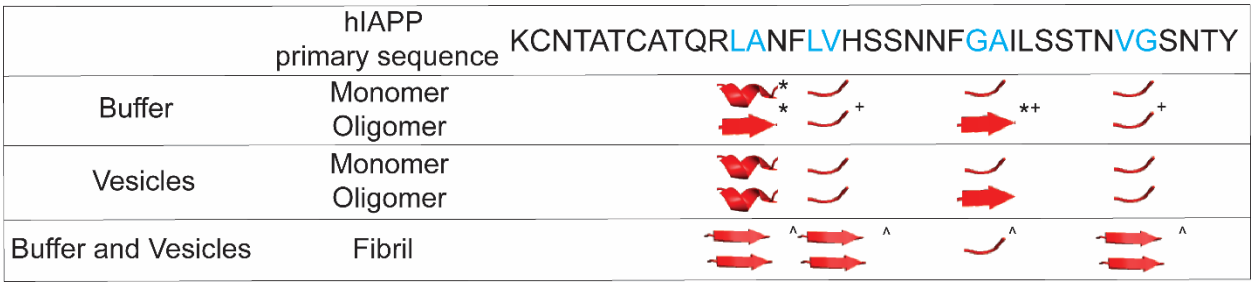


Figure 5.3: Structural information during aggregation for hIAPP monomers, oligomers, and fibrils, in vesicles and in neat buffer.

Double – label structural information for hIAPP monomers, oligomers, and fibrils. The hIAPP primary sequence (top) displays the sites of each double-label, colored cyan. *,+,and ^ denote structural assignments given in [43],[42], and [34], respectively.

The most important result of this study is that, when hIAPP aggregates in solution with DOPC vesicles, oligomers with a parallel β-sheet at G24A25 are observed like they are in buffer. As indicated in Figure 3, in buffer the parallel β-sheet in the oligomer extends from at least F23 to L27, spanning the FGAIL sequence.[42] Therefore, at least a portion of the oligomeric β-sheet is present when aggregation is impacted by DOPC vesicles.

This finding is of interest because, it is thought, that the FGAIL β-sheet is the reason that the oligomers have a large and stable population during the lag phase.

Amyloid aggregation is typically thought to be a nucleation event with the formation of the oligomers acting as the rate limiting step. That traditional nucleation model is at odds with the observations here and our previous work in buffer,[41] because if the oligomers are the rate limiting step to fibril formation, then their population should be very low since they are unlikely to form and once they do form, they nucleate fibrillization. But, instead, oligomers are readily observed during the lag phase. In buffer, we have concluded that the rate limiting step in the free energy landscape of aggregation is the unstructuring of the FGAIL β-sheet in the oligomer in order to adopt the fold of the fibril.[41] The need to unstructure the FGAIL β-sheet creates a barrier that kinetically traps the oligomers, which accounts for why they can be easily observed experimentally. The results presented in this article indicate that the FGAIL β-sheet of the oligomers still forms even when aggregation occurs with DOPC vesicles. Thus, the oligomers are a meta-stable species regardless of whether they form in buffer or in the presence of DOPC vesicles. Figure 5.4 provides a cartoon rendition of aggregation in buffer (top panel) and in buffer with DOPC vesicles present (bottom panel) consistent with our results presented here and those presented in previous work.[34,42,43]

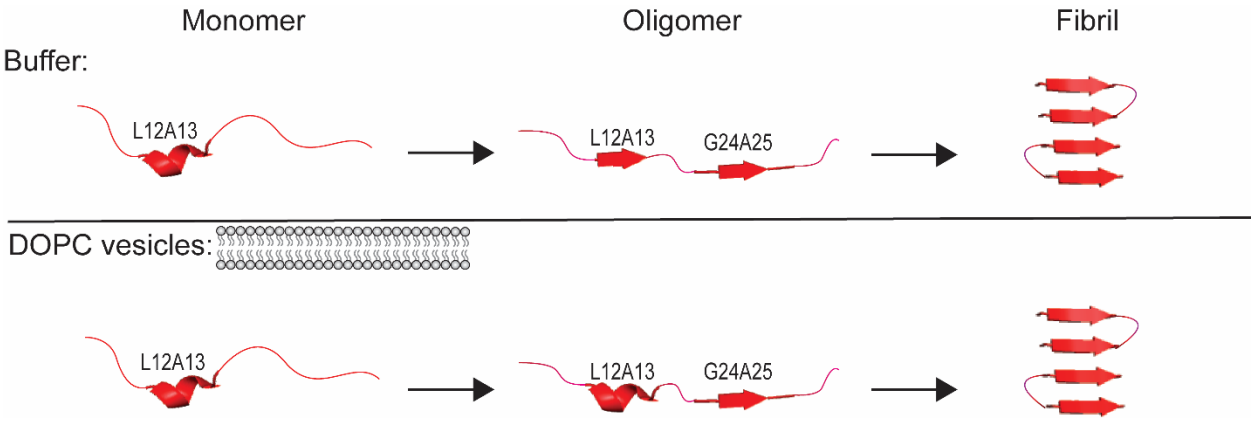


Figure 5.4: Proposed mechanism of fibril assembly with and without vesicles.

Proposed mechanism of assembly of amyloid fibrils in solution and in solution with vesicles consistent with the results of the double labels used in this study and previous work.[34,42,43]

We cannot determine from these experiments whether the oligomer assembles on the vesicles and remains bound or whether the vesicles serve to nucleate the oligomer with subsequent aggregation occurring in solution. hIAPP is positively charged and binds strongly to negative vesicles, including DOPC[8,10,77]. The N-terminus of hIAPP is amphipathic and forms α-helices on anionic lipids as established by CD spectroscopy[30] and site-specific EPR labeling.[31] Interactions between hIAPP and zwitterionic vesicles like DOPC have previously been observed although there is limited high resolution structural data available.[61] One postulate is that the N-terminal helices coalesce into helical bundles on the membranes, enabling the β -sheets of the amyloid fibrils.[30,31,74,78,79] Helical bundles may serve as Ca^{2+} ion channels that initiate an apoptotic cascade.[16,74,78–80] The atomic structure is only known in SDS micelles by NMR[28] and on POPS vesicles by EPR[31] as mentioned above. An additional structure solved by Ramamoorthy and co-workers found residues A5-V17 and S20-L23 of rat IAPP in a helical conformations on DPC micelles.[81] Thus, the helical structure that we observe in the oligomer might be caused by hIAPP binding to the vesicles, albeit creating a shorter helix than seen previously by NMR or EPR (because we observe helicity at L12A13 but not L16V17).[30,31]

Alternatively, rather than the polypeptides binding to the membrane and then organizing into oligomers, the membrane might instead be nucleating oligomer formation similar to the way that many small molecules alter amyloid aggregation.[42,82–85] It is well-known that helical inducing agents catalyze hIAPP aggregation.[86,87] For example, a few percent of hexafluoroisopropanol or trifluoroethanol added to buffer causes faster amyloid

aggregation.[75,86] Helical polypeptides added to hIAPP also induces amyloid fiber formation.[87] Thus, the vesicles might increase the proportion of alpha-helices, which then serve to nucleate oligomers. Indeed, we have previously pointed out the similarity of leucine rich repeat proteins to the known secondary structure of hIAPP, and hypothesized that the tertiary structure of hIAPP has a similar fold.[43] Leucine rich repeat proteins have helices that stabilize short, 3-4 residue β-sheets.[43,88] If the oligomers have a structure similar to the Leucine rich repeat proteins, then inducing helicity might stabilize the FGAIL β-sheet, whether or not the oligomer remains bound to the vesicles.

5.6 Conclusions.

In this manuscript we reported the structure throughout aggregation of hIAPP at four regions. Two sets of labels were placed near the N-terminus, one set of labels was placed in the FGAIL region, and one set of labels was placed in the C-terminus. Experiments were performed *in vitro*, either in buffer or when mixed with vesicles made of DOPC, which is a zwitterionic lipid. In buffer, residues L12A13 and G24A25 exhibited 3-state kinetics, reflecting an α-helix to short β-sheet to fibrillar β-sheet for L12A13 and reflecting a random coil to short β-sheet to fibrillar β-sheet for G24A25. In buffer with DOPC vesicles, the parallel FGAIL oligomer sheet is still observed, but the L12A13 β-sheet is not. Our experiments cannot determine if the oligomers are bound to the vesicles or if the vesicles are nucleating/stabilizing oligomer formation by inducing helices, for example, but either way, they are impacting the kinetics, the structure at L12A13, and retaining the G24A25 β-sheet. Since this oligomer has a well-defined parallel β-sheet, it is hopeful that it might be targeted by compounds that recognize its structure. Indeed, in buffer, artificial macrocycles pre-programmed with the FGAIL sequence altered the aggregation kinetics by stabilizing the oligomer's β-sheet.[42] In the future, lipids of more complex structural composition including negatively-charged lipids,[89–92] could be studied using 2D IR spectroscopy to further examine the effects of sterics and electrostatics on hIAPP aggregation kinetics.

AUTHOR INFORMATION

Corresponding Author

*Martin Zanni, Department of Chemistry, University of Wisconsin−Madison, Madison, WI 53706. Email: zanni@chem.wisc.edu

Present Addresses

†Sidney S. Dicke, Caitlyn R. Fields, and Martin T. Zanni - Department of Chemistry, University of Wisconsin−Madison, Madison, WI 53706

†^Michał Maj completed his postdoctoral research at University of Wisconsin-Madison. He is currently a Professor at the University of Gothenburg, Box 462, 405 30 Gothenburg, Sweden.

Author Contributions

S.S.D, M.M., and C.R.F. conducted data curation, formal analysis, investigation, and validation. M.M. coded software and performed non-negative matrix factorization for spectral decomposition. M.T.Z. provided project administration, supervision, and visualization. S.S.D. and M.T.Z. wrote the paper.

Funding Sources

This work was supported by the National Institutes of Health (R01DK079895).

Conflict of interest statement:

To ameliorate conflicts of interest, the University of Wisconsin-Madison requires us to disclose that M.T.Z. is part owner of PhaseTech Spectroscopy, Inc, a company that manufactures 2D spectrometers such as those used in this report.

5.7 Chapter 5 references.

1 D. L. Hay, S. Chen, T. A. Lutz, D. G. Parkes and J. D. Roth, *Pharmacol. Rev.*, 2015, **67**, 564–600.

2 W. A. Scherbaum, *Exp. Clin. Endocrinol. Diabetes*, 1998, **106**, 97–102.

3 P. Westermark, A. Andersson and G. T. Westermark, *Physiol. Rev.*, 2011, **91**, 795–826.

4 A. Clark, G. J. Cooper, C. E. Lewis, J. F. Morris, A. C. Willis, K. B. Reid and R. C. Turner, *The Lancet*, 1987, **330**, 231–234.

5 G. J. Cooper, A. C. Willis, A. Clark, R. C. Turner, R. B. Sim and K. B. M. Reid, *Proc. Natl. Acad. Sci. U. S. A.*, 1987, **84**, 8628–8632.

6 R. L. Hull, G. T. Westermark, P. Westermark and S. E. Kahn, *J. Clin. Endocrinol. Metab.*, 2004, **89**, 3629–3643.

7 A. Lorenzo, B. Razzaboni, G. C. Weir and B. A. Yankner, *Nature*, 1994, **368**, 756–760.

8 K. Sasahara, *Biophys. Rev.*, 2018, **10**, 453-462.

9 K. Sasahara, K. Morigaki and K. Shinya, *FEBS J.*, 2014, **281**, 2597–2612.

10 K. Sasahara, K. Morigaki, T. Okazaki and D. Hamada, *Biochemistry*, 2012, **51**, 6908–6919.

11 L. Caillon, A. R. F. Hoffmann, A. Botz and L. Khemtémourian, *J. Diabetes Res.*, 2016, 2016: 5639875.

12 L. Khemtémourian, J. A. Killian, J. W. M. Höppener and M. F. Engel, *Exp. Diabetes Res.*, 2008, 2008:421287.

13 M. F. M. Engel, L. Khemtémourian, C. C. Kleijer, H. J. D. Meeldijk, J. Jacobs, A. J.

Verkleij, B. de Kruijff, J. A. Killian and J. W. M. Höppener, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 6033–6038.

14 Y. Hirakura, W. W. Yiu, A. Yamamoto and B. L. Kagan, *Amyloid*, 2000, **7**, 194–199.

15 A. Quist, I. Doudevski, H. Lin, R. Azimova, D. Ng, B. Frangione, B. Kagan, J. Ghiso and R. Lal, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 10427–10432.

16 T. A. Mirzabekov, M. Lin and B. Kagan, *J. Biol. Chem.*, 1996, **271**, 1988–1992.

17 M. Anguiano, R. J. Nowak and P. T. Lansbury Jr, *Biochemistry*, 2002, **41**, 11338–11343.

18 J. R. Brender, S. Salamekh and A. Ramamoorthy, *Acc. Chem. Res.*, 2012, **45**, 454–462.

19 J. D. Green, L. Kreplak, C. Goldsbury, X. L. Blatter, M. Stolz, G. S. Cooper, A. Seelig, J. Kistler and U. Aebi, *J. Mol. Biol.*, 2004, **342**, 877–887.

20 S. A. Jayasinghe and R. Langen, *Biochim. Biophys. Acta*, 2007, **1768**, 2002–2009.

21 M. F. M. Sciacca, S. A. Kotler, J. R. Brender, J. Chen, D. Lee and A. Ramamoorthy, *Biophys. J.*, 2012, **103**, 702–710.

22 A. Abedini, A. Plesner, P. Cao, Z. Ridgway, J. Zhang, L. H. Tu, C. T. Middleton, B. Chao, D. J. Sartori, F. Meng, H. Wang, A. G. Wong, M. T. Zanni, C. B. Verchere, D. P. Raleigh and A. M. Schmidt, *Elife*, 2016, 5:e12977.

23 J. Janson, R. H. Ashley, D. Harrison, S. McIntyre and P. C. Butler, *Diabetes*, 1999, **48**, 491–498.

24 L. Milanesi, T. Sheynis, W. Xue, E. V. Orlova, A. L. Hellewell, R. Jelinek, E. W.

Hewitt, S. E. Radford and H. R. Saibil, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 20455–20460.

25 N. F. Dupuis, C. Wu, J. E. Shea and M. T. Bowers, *J. Am. Chem. Soc.*, 2011, **133**, 7240.

26 C. Wu and J. E. Shea, *PLOS Comput. Biol.*, 2013, **9**, e1003211.

27 M. B. Erlach, H. R. Kalbitzer, R. Winter and W. Kremer, *Biophys. Chem.*, 2019, **254**, 106239.

28 R. P. R. Nanga, J. R. Brender, S. Vivekanandan and A. Ramamoorthy, *Biochim. Biophys. Acta*, 2011, **1808**, 2337-2342.

29 J. Cort, Z. Liu, G. Lee, S. M. Harris, K. S. Prickett, L. S. Gaeta and N. H. Andersen. *Biochem. Biophys. Res. Commun.*, 1994, **204**, 1088–1095.

30 J. D. Knight, J. A. Hebda and A. D. Miranker, *Biochemistry*, 2006, **45**, 9496–9508.

31 M. Apostolidou, S. A. Jayasinghe and R. Langen, *J. Biol. Chem.*, 2008, **283**, 17205-17210.

32 S. Kumar and A. D. Miranker, *Chem. Commun.*, 2013, **49**, 4749-4751.

33 D. C. Rodriguez Camargo, K. J. Korshavn, A. Jussupow, K. Raltchev, D. Goricanec, M. Fleisch, R. Sarkar, K. Xue, M. Aichler, G. Mettenleiter, A. K. Walch, C. Camilloni, F. Hagn, B. Reif and A. Ramamoorthy, *Elife*, 2017, **6**, e31226.

34 S. Luca, W. Yau, R. Leapman and R. Tycko, *Biochemistry*, 2007, **46**, 13505-13522.

35 C. Röder, T. Kupreichyk, L. Gremer, L. U. Schäfer, K. R. Pothula, R. B. G. Ravelli, D. Willbold, W. Hoyer and G. F. Schröder, *bioRxiv*, 2020, doi:10.1101/2020.02.11.944546.

36 Q. Cao, D. R. Boyer, M. R. Sawaya, P. Ge and D. S. Eisenberg, *Nat. Struct. Mol. Biol.,* 2020*,* **27**, 653–659.

37 F. Weirich, L. Gremer, E. A. Mirecka, S. Schiefer, W. Hoyer and H. Heise, *PLoS One*, 2016, **11**, e0161243.

38 P. Arosio, T. P. J. Knowles and S. Linse, *Phys. Chem. Chem. Phys.*, 2015, **17**, 7606–7618..

39 D. C. Rodriguez Camargo, S. Chia, J. Menzies, B. Mannini, G. Meisl, M. Lundqvist, C. Pohl, K. Bernfur, V. Lattanzi, J. Habchi, S. I. Cohen, T. P. J. Knowles, M. Vendruscolo and S. Linse, *Front. Mol. Biosci.*, 2021, **8**, 757425.

40 L. Zhang and J. D. Schmit, *Isr. J. Chem.*, 2017, **57**, 738-749.

41 A. L. Serrano, J. P. Lomont, L. H. Tu, D. P. Raleigh and M. T. Zanni, *J. Am. Chem. Soc.*, 2017, **139**, 16748–16758.

42 L. Buchanan, E. Dunkelberger, H. Tran, P. Cheng, C. Chiu, P. Cao, D. Raleigh, J. J. De Pablo, J. Nowick and M. T. Zanni, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 19285-19290.

43 M. Maj, J. P. Lomont, K. L. Rich, A. M. Alperstein and M. T. Zanni, *Chem. Sci.*, 2018, **9**, 463–474.

44 C. Betsholtz, L. Christmansson, U. Engström, F. Rorsman, V. Svensson, K. H. Johnson and P. Westermark, *FEBS Lett.*, 1989, **251**, 261–264.

45 P. Westermark, U. Engstrom, K. H Johnson, G. T. Westermark and C. Betsholtz, *Proc. Natl. Acad. Sci. U. S. A.*, 1990, **87**, 5036–5040.

46 K. Tenidis, M. Waldner, J. Bernhagen, W. Fischle, M. Bergmann, M. Weber, M. L. Merkle, W. Voelter, H. Brunner and A. Kapurniotu, *J. Mol. Biol.*, 2000, **295**, 1055–

1071.

47 T. T. Ashburn, M. Auger and P. T. Lansbury Jr., *J. Am. Chem. Soc.*, 1991, **114**, 790–791.

48 G. G. Glenner, E. D. Eanes and C. A. Wiley, *Biochem. Biophys. Res. Commun.*, 1988, **155**, 608–614.

49 Y. Kiriyama and H. Nochi, *Cells*, 2018, **7**, 95.

50 P. Hamm and M. Zanni, *Concepts and methods of 2D infrared spectroscopy*, Cambridge University Press, 2011.

51 C. T. Middleton, A. M. Woys, S. S. Mukherjee and M. T. Zanni, *Methods*, 2010, 52, 12–22.

52 L. E. Buchanan, E. B. Dunkelberger and M. T. Zanni, in *Protein Folding and Misfolding*, ed. H. Fabien and Dieter Naumann, Springer, Berlin, Heidelberg, 1st edn, 2012, vol. 1, ch. 9, 217–237.

53 M. K. Petti, J. P. Lomont, M. Maj and M. T. Zanni, *J. Phys. Chem. B*, 2018, **122**, 1771-1780.

54 S.-H. Shim and M. T. Zanni, *Phys. Chem. Chem. Phys.*, 2009, **11**, 748–761.

55 D. B. Strasfeld, Y. L. Ling, R. Gupta, D. P. Raleigh and M. T. Zanni, *J. Phys. Chem. B,* 2009, **113**, 15679-15691.

56 S.-H. Shim, R. Gupta, Y. L. Ling, D. B. Strasfeld, D. P. Raleigh and M. T. Zanni, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 6614–6619.

57 T. O. Zhang, M. Grechko, S. D. Moran and M. T. Zanni, *Methods Mol. Biol.*, 2016, **1345**, 21–41.

58 M. W. Berry, M. Browne, A. N. Langville, V. P. Pauca and R. J. Plemmons,

Comput. Stat. Data Anal., 2007, **52**, 155–173.

59 W. Rock, Y.-L. Li, P. Pagano and C. M. Cheatum, *J. Phys. Chem. A*, 2013, **117**, 6073–6083.

60 S. K. Karthick Kumar, A. Tamimi and M. D. Fayer, *J. Chem. Phys.*, 2012, **137**, 184201.

61 X. Zhang, J. R. St. Clair, E. London and D. P. Raleigh, *Biochemistry*, 2017, **56**, 376–390.

62 T. Skotland and K. Sandvig, *Nat. Comm.*, 2019, **10**, 2752.

63 R. Akter, P. Cao, H. Noor, Z. Ridgway, L. H. Tu, H. Wang, A. G. Wong, X. Zhang, A. Abedini, A. M. Schmidt and D. P. Raleigh, *J. Diabetes Res.*, 2016;2016:2798269.

64 S. S. Dicke, A. M. Alperstein, K. L. Schueler, D. S. Stapleton, S. P. Simonett, C. R. Fields, F. Chalyavi, M. P. Keller, A. D. Attie and M. T. Zanni, *J. Phys. Chem. B*, 2021, **125**, 9517–9525.

65 C. R. Fields, S. S. Dicke, M. K. Petti, M. T. Zanni and J. P. Lomont, *J. Phys. Chem. Lett.*, 2020, **11**, 6382–6388.

66 M. Grechko and M. T. Zanni, *J. Chem. Phys.*, 2012, **137**, 184202.

67 E. B. Dunkelberger, M. Grechko and M. T. Zanni, *J. Phys. Chem. B*, 2015, **119**, 14065–14075.

68 C. T. Middleton, P. Marek, P. Cao, C. C. Chiu, S. Singh, A. M. Woys, J. J. De Pablo, D. P. Raleigh and M. T. Zanni, *Nat. Chem.*, 2012, **4**, 355–360.

69 A. Abedini and D. P. Raleigh, *J. Mol. Biol.*, 2006, **355**, 274–281.

70 Z. Ridgway, K. H. Lee, A. Zhyvoloup, A. Wong, C. Eldrid, E. Hannaberry, K.

Thalassinos, A. Abedini and D. P. Raleigh, *Biophys. J.*, 2020, **118**, 1142–1151.

71 J. S. Fortin, M. Benoit-Biancamano, *Amyloid*, 2015, **22**, 194–202.

72 J. P. Lomont, K. L. Rich, M. Maj, J. J. Ho, J. S. Ostrander and M. T. Zanni, *J. Phys. Chem. B*, 2018, **122**, 144–153.

73 L. E. Buchanan, M. Maj, E. B. Dunkelberger, P. N. Cheng, J. S. Nowick and M. T. Zanni, *Biochemistry*, 2018, **57**, 6470–6478.

74 D. Raleigh, X. Zhang, B. Hastoy and A. Clark, *J. Mol. Endocrinol.*, 2017, **59**, R121–R140.

75 J. A. Williamson, J. P. Loria and A. D. Miranker, *J. Mol. Biol.*, 2009, **393**, 383–396.

76 Y. L. Ling, D. B. Strasfeld, S. H. Shim, D. P. Raleigh and M. T. Zanni, *J. Phys. Chem. B*, 2009, **113**, 2498–2505.

77 J. Lesma, F. Bizet, C. Berardet, N. Tonali, S. Pellegrino, M. Taverna, L. Khemtémourian, J.-L. Soulier, C. van Heijenoort, F. Halgand, T. Ha-Duong, J. Kaffy and S. Ongeri, *Front. Cell Dev. Biol.*, 2021, **9**, 2531.

78 M. Pannuzzo, A. Raudino, D. Milardi, C. La Rosa and M. Karttunen, *Sci. Reports,* 2013*,* **3**, 2781.

79 X. Dong, Q. Qiao, Z. Qian and G. Wei, *Biochim. Biophys. Acta. Biomembr.*, 2018, **9**, 1826-1839.

80 S. Casas, A. Novials, F. Reimann, R. Gomis and F. M. Gribble, *Diabetologia*, 2008, **51**, 2252–2262.

81 R. P. R. Nanga, J. R. Brender, J. Xu, K. Hartman, V. Subramanian and A. Ramamoorthy, *J. Am. Chem. Soc.,* 2009, **131**, 8252-8261.

82 Y. Xu, R. Maya-Martinez, N. Guthertz, G. R. Heath, I. W. Manfield, A. L. Breeze, F.

Sobott, R. Foster and S. E. Radford, *Nat. Commun.*, 2022*,* **13**, 1040.

83 L. M. Young, J. C. Saunders, R. A. Mahood, C. H. Revill, R. J. Foster, L. H. Tu, D. P. Raleigh, S. E. Radford and A. E. Ashcroft, *Nat. Chem.*, 2015, **7**, 73-81.

84 A. Nath, D. E. Schlamadinger, E. Rhoades and A. D. Miranker, *Biochemistry*, 2015, **54**, 3555-3564.

85 S. Kumar, D. E. Schlamadinger, M. A. Brown, J. M. Dunn, B. Mercado, J. A. Hebda, I. Saraogi, E. Rhoades, A. D. Hamilton and A. D. Miranker, *Chem. Biol.*, 2015, **22**, 369–378.

86 K. Yanagi, M. Ashizaki, H. Yagi, K. Sakurai, Y. H. Lee and Y. Goto, *J. Biol. Chem.*, 2011, **286**, 23959–23966.

87 I. Saraogi, J. A. Hebda, J. Becerril, L. A. Estroff, A. D. Miranker and A. D. Hamilton, *Angew. Chem. Int. Ed. Engl.*, 2010; **49**, 736–739.

88 J. Bella, K. L. Hindle, P. A. Mcewan and S. C. Lovell, *Cell. Mol. Life Sci.*, 2008, **65**, 2307-2333.

89 S. A. Jayasinghe and R. Langen, *Biochemistry*, 2005, **44**, 12113–12119.

90 J. D. Knight and A. D. Miranker, *J. Mol. Biol.*, 2004, **341**, 1175–1187.

91 P. E. S. Smith, J. R. Brender and A. Ramamoorthy, *J. Am. Chem. Soc.*, 2009, **131**, 4470-4478.

92 D. H. J. Lopes, A. Meister, A. Gohlke, A. Hauser, A. Blume and R. Winter, *Biophys. J.*, 2007, **93**, 3132-41.

6 A brief foray into 3A.

6.1 Research focus.

Structure of stabilized soluble toxic oligomeric amylin studied at biological pH using Electron Microscopy, solution state NMR spectroscopy and Thioflavin T fluorescence assays.

Creating a stable toxic oligomer of hIAPP has been an attractive project focus for many students in the group for nearly a decade. Much information has been collected on a variety of mutants. Here, I add to the work of previous students using electron microscopy, Thioflavin T fluorescence assays, and solution NMR spectroscopy. This is an unpublished, ongoing project that I will continue to complete as a post doctoral researcher in the Zanni group.

Contributing students/scientists:

Sidney S. Dicke, Harrison J. Esterly, Brynn LeMasters (Graduate students, Martin Zanni group)

Caitlyn R. Fields, Kacie L. Rich (Previous graduate students, Martin Zanni group)

Joseph Kim, Juan C. Sanchez (Graduate students, Elizabeth Wright group)

Thirupathi Ravula (Scientist, Chad Reinstra/Katherine Henzler-Wildman group)

6.2 Abstract.

The aggregation of human amylin (hIAPP, or amylin) into amyloid fibrils involves a metastable intermediate thought responsible for both direct and receptor-mediated β-cell cytotoxicity. This toxic oligomer exists transiently at low concentrations and has eluded high-resolution structural determination. Point mutations within residues 22-27 of mature amylin (FGAIL) prevent the formation of this oligomer and have recently been linked to species propensity for type II diabetes. Here, we have stabilized the toxic oligomeric state using 3 alanine substitutions outside of the FGAIL region (L12A, F15A, and T30A) chosen to disrupt favorable steric packing and noncovalent interactions thought to stabilize fibril formation. Our stabilized oligomer has a lag phase of over 2 days and a near-identical oligomer in the FGAIL region as shown using $^{13}C^{18}O$ labeled 3A-hIAPP via two-dimensional infrared spectroscopy. We find 3A-hIAPP has comparable cytotoxicity to human amylin. The structure of this oligomer was additionally examined using solution NMR and Electron Microscopy. The 3A mutant is suitable for future structural studies as a model for the toxic oligomer of hIAPP and could be used for screening in oligomer-specific antibody development or drug-discovery efforts.

6.3 Background.

Many studies have elucidated high-resolution structural information about fibrillar species. Using EM and solid state NMR, Luca and co-workers presented a protocol to develop synthetic fibrils with a single predominant morphology of 4 layered β-sheets.[1] These layers were formed by two symmetric layers of amylin fibrils, with two sheets contributed per subunit. Interestingly, the sequence predicted to determine species susceptibility to type II diabetes (the "FGAIL" sequence) was discovered to form the disordered loop between β-strands from the same subunit.[1] Another study by Weirich and colleagues observed the FGAILS segment to form a part of the β-sheet core despite using the free acid form of hIAPP.[2] Röder et al used Cryo-EM to solve 3 structures of hIAPP fibrils formed at pH 6; their main polymorph (~90% of fibrils observed) resolved residues 13-37 to 4.2Å resolution.[3] The segment 21-NNFGAIL-27 were the majority of residues contributing to the protofilament interface.[3] This polymorph and an additional polymorph (~10% of fibrils observed) solved at similar resolution (although not unambiguously modeled) found residues 26-32 to form a β-sheet core between individual subunits.[3] The third polymorph, only solved at 8.1Å resolution, did not allow for development of an atomic model.[3] More recently, Cao and colleagues have reported four Cryo-EM structures of hIAPP fibrils seeded by patient-extracted hIAPP aggregates, each of which were solved at 3.8-4.1Å resolution, 3 of which were previously unreported.[4] They elucidated 3 distinct protofilament cores amongst the fibril polymorphs, longer than had been previously determined. Core folds 1 and 2 reported in this study both contained residues 20-29.[4]

More recent studies using solution state NMR have provided information potentially relevant to oligomeric assemblies, using additional buffer components. Nanga et al showed that segment 1-19 of hIAPP forms helices on detergent micelles, potentially relevant to membrane-mediated toxicity.[5] Rodriguez Camargo and co workers used solution state NMR to study nanodisc-stabilized oligomers; these oligomers exhibited β-sheet structure at I26-S29, although the N22FGAIL27 region exhibited flexibility and increased solvent accessibility.[6] The authors note that this increased availability for interaction may allow this segment to template further aggregation.[6] It should be noted that at physiological pH, hIAPP aggregates too fast for solution state NMR structure determination, and additional stabilizing components such as detergents[5,7] or nanodiscs[6] may alter the native structure formed by hIAPP. Thus, we set out to structurally resolve an oligomeric species with minimal alterations or additional components in the media.

Little detailed information is known about these oligomeric assemblies largely because the low concentrations and transient nature of oligomers make them difficult to characterize structurally. Most amyloid fibrils form through a nucleated growth mechanism.[8] According to that mechanism, monomers and oligomers exist in equilibrium until small assemblies of proteins form nucleating seeds that grow into fibrils by templating additional proteins. Once a significant number of seeds form, the population of monomers and oligomers is lost, as nearly all protein converts to amyloid fibrils. Thus there is a low concentration of oligomers that disappears rapidly as proteins aggregate into amyloid fibrils.

In this study, we use three point mutations to slow aggregation and thereby trap hIAPP into an oligomeric state structurally similar to the oligomeric species formed by wild type hIAPP. We place these mutations at sites that prevent fibril formation but do not appreciably alter the structure of wild-type oligomeric assemblies. Because multiple fibril structures have been submitted, we chose these 3 mutations to prevent any expected polymorphs formed at pH 7.4. Since formation of the fibrils are inhibited, the equilibrium between monomers and oligomers is retained for longer periods of time. Using this system, we study the soluble oligomeric structure of a three-alanine mutant of hIAPP (3A-hIAPP) at biological pH using 2D IR spectroscopy, solution NMR spectroscopy, Electron Microscopy, and ThT fluorescence.

6.4 Current research directions.

We have attempted to make fibril formation less favorable by replacing 3 residues involved in favorable steric or hydrophobic interactions within the mature fibril with alanine, a small amino acid. Thus we expect to have shifted the equilibrium towards soluble oligomers. Previous research showed that isotope labels placed within the FGAIL region of the 3A-hIAPP mutant shared the same vibrational frequencies as isotope labels placed within the FGAIL region of human amylin during the lag phase, thus we believe we have preserved the toxic oligomer structure in this region (results presented in thesis of Kacie Rich and Caitlyn Fields). We also know from these previous students that 3A-hIAPP displays similar extracellular toxicity to Ins-1 cells (an insulin-producing β-cell line) as hIAPP. **Figure 6.1A** shows the primary sequence of hIAPP and the amino acid substitutions made for the 3A mutant, **Figure 6.1B** highlights these residues in one model of hIAPP fibrils, created using constraints from solid state NMR.

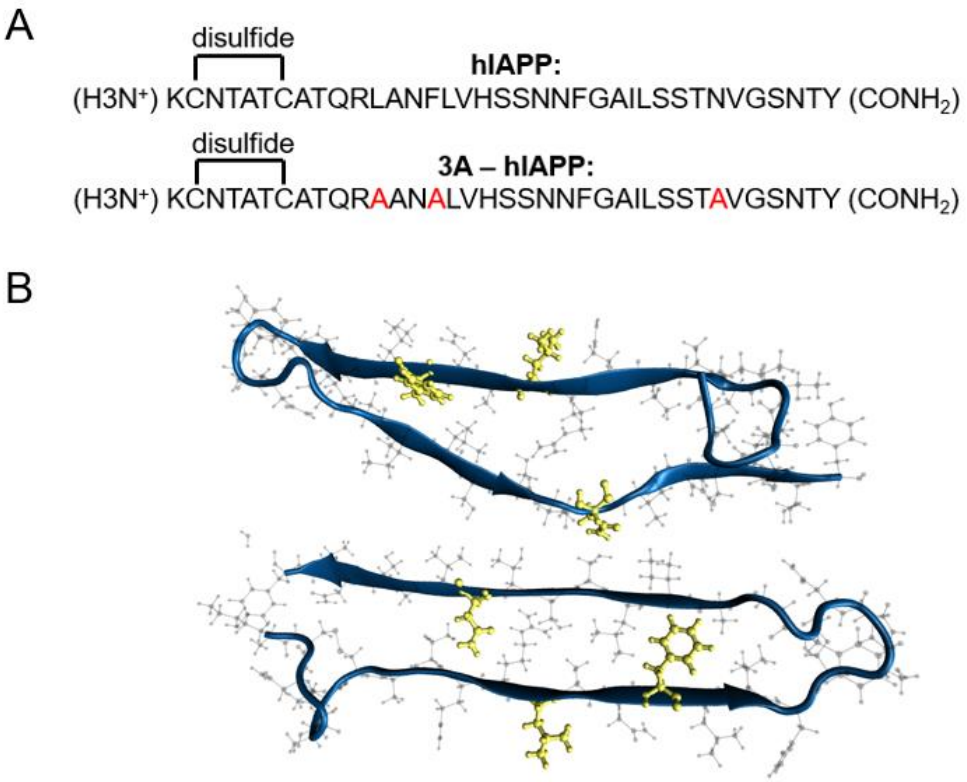


Figure 6.1 The primary sequences of hIAPP and 3A-hIAPP, and the Tycko model with alanine substitution locations highlighted.

(A) The primary sequences of hIAPP and 3A-hIAPP. (B) The locations of these substitutions within an amyloid structure solved via solid state NMR.

In the 3A mutant, we make the following replacements from the hIAPP primary sequence: L12A, F15A, and N31A. In all cases, an amino acid with a large side chain is replaced by an amino acid, alanine, that has a smaller side chain. **Figure 6.2** displays Phenylalanine and Alanine side-by-side as an example of one such substitution.

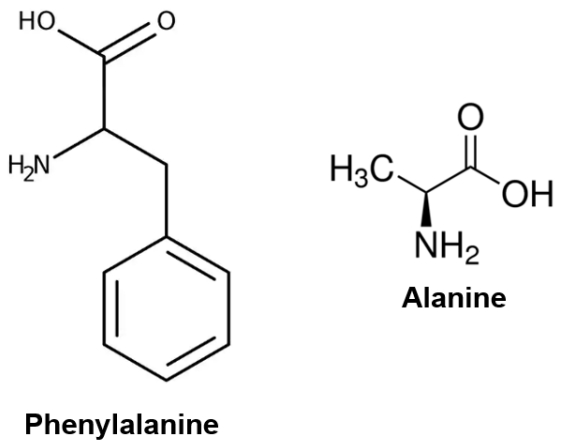


Figure 6.2 Side-chain size matters: phenylalanine and alanine.

Example of a large amino acid, phenylalanine, replaced with an alanine residue in the 3A mutant. This replacement may eliminate effects such as steric zippering, pi-pi stacking, or other noncovalent interactions that make amyloid fibrils of hIAPP thermodynamically favorable.

In collaboration with Elizabeth Wright's group, samples of 3A -hIAPP and hIAPP were pre-aggregated for 18 hours, stained at pH 7.2 using phosphotungstic acid (PTA), and imaged. **Figure 6.3** shows hIAPP and 3A-hIAPP at 1 mM and 250 μM, respectively. At 250 μM, 3A-hIAPP showed small oligomers, whereas hIAPP displayed amyloid fibrils. **Figure 6.4A and B** shows the same small oligomers of 3A under enhanced magnification. The preparation was repeated an additional time to yield similar results, although follow-up studies showed that is was also possible for the 3A-hIAPP sample to become pre-aggregated, an often observed phenomenon in the amyloid field when preparing sample of hIAPP. Cryo-EM preparations of 3A also displayed some circular structures, but these were not possible to distinguish from commonly observed ethane artifacts (liquid ethane is not involved in the preparation of the negative stain electron microscopy samples).

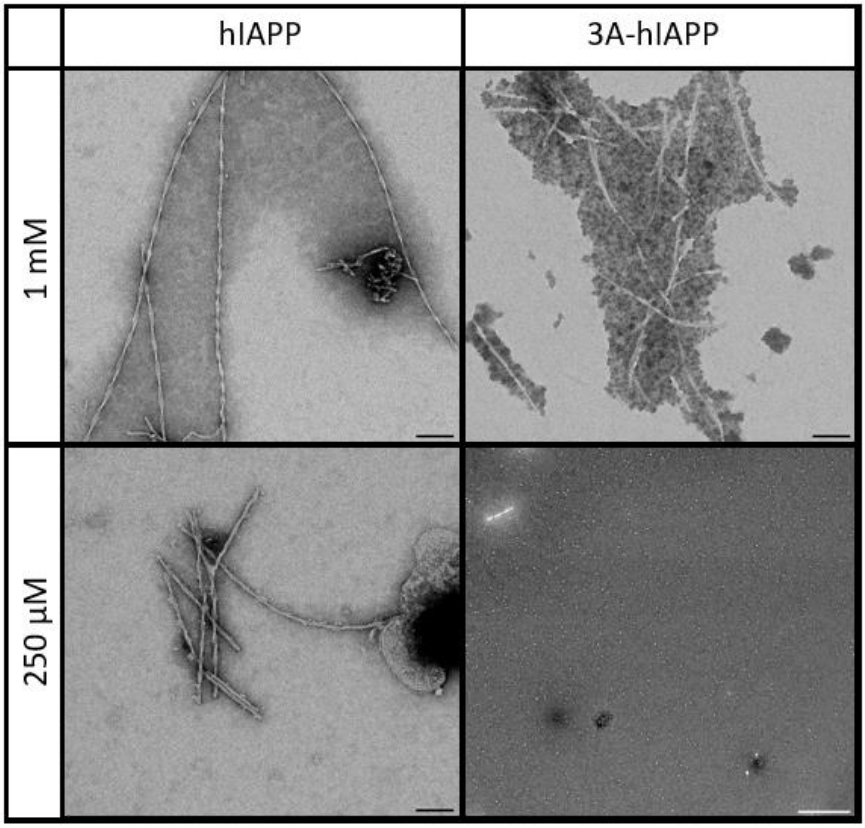


Figure 6.3: Negative-stain EM images of hIAPP and 3A-hIAPP.

Negative-stain EM images of hIAPP and 3A-hIAPP prepared at matching concentrations at pH 7.2 after an 18-hour pre-incubation. In this sample, fibrils are not observed for 3A at the 250 µM concentration. Stain used here is PTA. Scale bars = 100 µm.

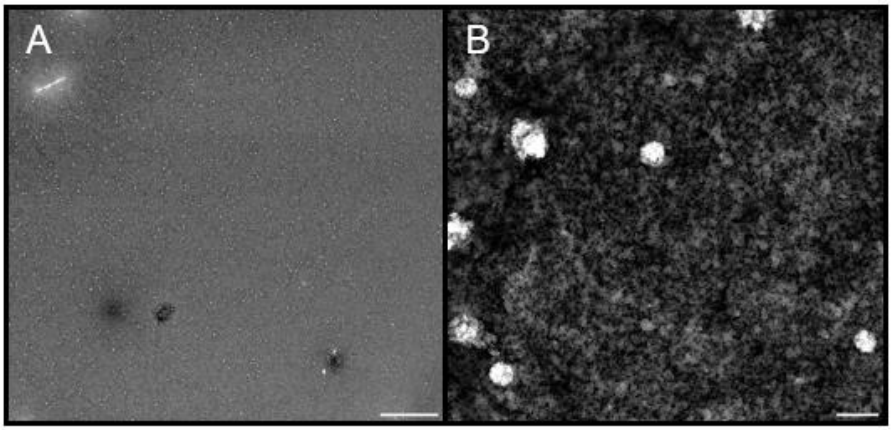


Figure 6.4: 3A oligomers.

(A) The same 3A-hIAPP image at 250 μM displayed in Figure 3. (B) Increased magnification of the same sample enhances viewing ability of the small white structures, about 40-70 nanometers in diameter. Scale bars = 100 μm.

3A has also been studied by many students using 2D IR spectroscopy. **Figure 6.5A-C** display aggregation data collected using a 100 kHz 2D IR spectrometer. **Figure 6.5D** tracks the intensity of each peak, corresponding to random coil, native β-sheet, and amyloid β-sheet. When compared to a Thioflavin T assay (**Figure 6.5E**), it is apparent that the sample used for 2D IR was seeded or pre-aggregated, as the lag phase observed was much shorter. The curves displayed in **Figure 6.5D** are not smooth, one explanation is a population of fibrils started early in the lag phase and merged with other fibril populations later in the experiment. However, the cause of abnormal individual kinetic traces are difficult to elucidate. One of the current project directions is to achieve sample preparation methods that yield a consistent, reproducible lag phase to compare to the short (30 min-1 hr) lag phases often reported for hIAPP.

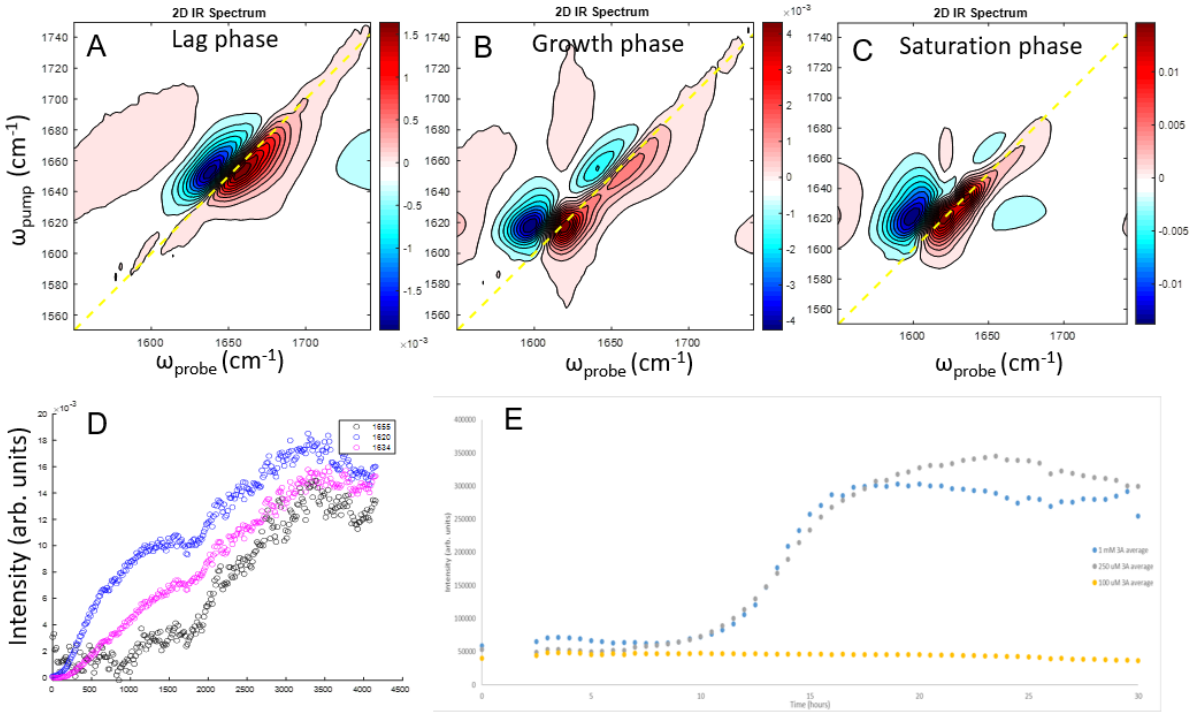


Figure 6.5: Kinetics of 3A aggregation studied with 100 kHz 2D IR spectroscopy and Thioflavin T fluorescence.

100 kHz 2D IR spectra of 3A-hIAPP aggregation taken during the (A) lag phase, (B) growth phase, and (C) saturation phase. (D) Amyloid peak (1620 cm^{-1}), Native β-sheet peak (1634 cm^{-1}), and random coil peak (1660 cm^{-1}) intensity plotted vs. time. (E) Kinetic trace of Thioflavin-T binding to amyloid fibrils of 3A-hIAPP at 100 μM (orange, no binding), 250 μM (gray), and 1 mM (blue).

During the final year of my PhD, I also managed a collaboration with Chad Reinstra and Katie Henzler-Wildman's lab in Biochemistry. With higher field magnets than before and a newly minted scientist, Thirupathi Ravula (who studied hIAPP during his PhD), we believed that solution NMR could finally yield a high-resolution structure of the 3A oligomer. The original plan was to express hIAPP and 3A-hIAPP in bacteria to provide Thiru with fully labeled $^{13}C^{15}N$ samples, but I was never able to fully purify the mixture. After 6 months of work, the best result I obtained was hydrogel-forming His-tagged 3A-hIAPP. The formation of hydrogels was interesting because mature hIAPP has been observed to phase-separate, but these hydrogels were unusable for spectroscopic

experiments (or even for mass determination via MALDI). We decided to return to solid-phase synthesis, choosing individual isotope labels carefully. The labeling scheme we came up with, ultimately resulting in >95% peak assignment in a 2D NMR spectrum, was $^{13}C^{15}N$-labeled residues FGAILSNTV, displayed in **Figure 6.6** below.

1	K	Lys
2	C	Cys
3	N	Asn
4	T	Thr
5	A	Ala
6	T	Thr
7	C	Cys
8	A	Ala
9	T	Thr
10	Q	Gln
11	R	Arg
12	A	Ala
13	A	Ala
14	N	Asn
15	A	Ala
16	L	Leu
17	V	Val
18	H	His
19	S	Ser
20	S	Ser

21	N	Asn
22	N	Asn
23	F	Phe
24	G	Gly
25	A	Ala
26	I	Ile
27	L	Leu
28	S	Ser
29	S	Ser
30	T	Thr
31	A	Ala
32	V	Val
33	G	Gly
34	S	Ser
35	N	Asn
36	T	Thr
37	Y	Tyr

Figure 6.6: $^{13}C^{15}N$ isotope-labeling scheme for 95% 2D NMR peak assignment of 3A-hIAPP.

Using this isotope-labeling scheme, HN and HA have been assigned for each amino acid except for the cysteine at position 2. Excitingly, it was determined early on that the FGAIL peaks were shifted out of the random coil region of the NMR spectra, the peak shifts and linewidths were consistent with a short ordered structure similar to a β-sheet.

Below is a list of experimental data collected to date on 3A-hIAPP. Should Thiru leave the facility for any reason, these experiments are on the UW-Madison NMRFAM server and a student could learn to analyze the data.

1: 2D-15H-HSQC

2: 3D HNCO

3: 3D HN(CA)CO

4: 3D HNCACB

5: 3D CA/CB (CO)NH

6: 3D-15N-HSQC-TOCSY

7: 3D-15N-HSQC-NOESY

8: 2D HSQC-TOCSY

9: 3D HNHA

10: 2D HNCOhb

11: 13C-HSQC-CT

12: x-filtered 2D-TOCSY

13: x-filtered 2D NOESY

14: 13C-HSQC aliphatic

15: 1D PROTON

To demonstrate the resolution of the high field NMR experiments, a spectrum is from an HSQC experiment with assigned peaks in **Figure 6.7**.

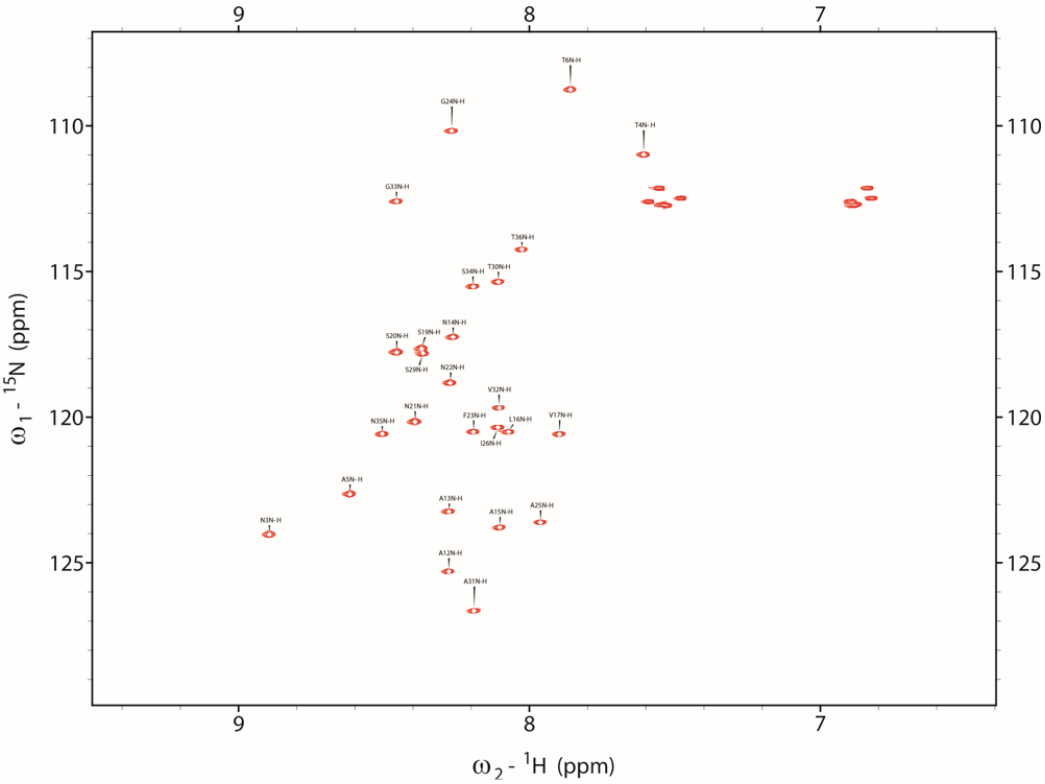


Figure 6.7: 2D-1H-^{15}N-HSQC of ^{15}N^{13}C-labeled 3A-hIAPP.

6.5 Perspectives and winnowing.

The 3A-hIAPP project is at an exciting stage. We have enough data to believe that we have created a model for the toxic FGAIL oligomer. During my tenure as a postdoctoral researcher in the Zanni group, I will help Harrison Esterly and Brynn LeMasters (younger graduate students) collect a series of kinetic data—2D IR of 3A-hIAPP with double labels similar to those used in Chapter 5 (to resolve helical segments from β-sheet), along with additional TEM and ThT experiments. Additionally, previous student Caitlyn Fields has left some exciting gel results to compare the oligomer size of a 3A-hIAPP oligomer to a wild type oligomer. By the time I leave UW-Madison, I would like this project to finally be in a publishable form.

6.6 Chapter 6 references.

1. Luca, S., Yau, W. M., Leapman, R. & Tycko, R. Peptide Conformation and Supramolecular Organization in Amylin Fibrils: Constraints from Solid State NMR. *Biochemistry* **46**, 13505 (2007).

2. Weirich, F. *et al.* Structural Characterization of Fibrils from Recombinant Human Islet Amyloid Polypeptide by Solid-State NMR: The Central FGAILS Segment Is Part of the β-Sheet Core. *PLoS One* **11**, e0161243 (2016).

3. Röder, C. *et al.* Amyloid fibril structure of islet amyloid polypeptide by cryo-electron microscopy reveals similarities with amyloid beta. doi:10.1101/2020.02.11.944546.

4. Cao, Q. *et al.* Cryo-EM structures of hIAPP fibrils seeded by patient-extracted fibrils reveal new polymorphs and conserved fibril cores. *Nature Structural & Molecular Biology 2021 28:9* **28**, 724–730 (2021).

5. Nanga, R. P. R., Brender, J. R., Vivekanandan, S. & Ramamoorthy, A. Structure and membrane orientation of IAPP in its natively amidated form at physiological pH in a membrane environment. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1808**, 2337–2342 (2011).

6. Rodriguez Camargo, D. C. *et al.* Stabilization and structural analysis of a membrane-associated hIAPP aggregation intermediate. *Elife* **6**, (2017).

7. Prakash, R. *et al.* Three-Dimensional Structure and Orientation of Rat Islet Amyloid Polypeptide Protein in a Membrane Environment by Solution NMR Spectroscopy. doi:10.1021/ja9010095.

8. Serrano, A. L., Lomont, J. P., Tu, L. H., Raleigh, D. P. & Zanni, M. T. A Free Energy Barrier Caused by the Refolding of an Oligomeric Intermediate Controls the Lag Time of Amyloid Formation by hIAPP. *J Am Chem Soc* **139**, 16748–16758 (2017).

7 Conclusions and future directions.

Spectroscopic dissertations for today's tomorrow.

I hope the research in this thesis provides the foundation for future studies in the Zanni group. Here, I seek to inspire and excite.

7.1 Future directions: 2D IR imaging.

In Chapter 3, a new imaging technique was developed using 2D IR to study formalin-preserved tissue slices. Because we discovered that the protein structures within the tissues changes over time, it is important that future raster-scan methods image over the course of a few hours rather than a few days. Although averaging-on-a-chip and 100 kHz lasers are now available with reduced spectral acquisition time, focusing to and past the diffraction limit also challenge students to collect data wisely rather than in bulk (because so many 2D spectra could be recorded in a single small sample). One way future students could improve this technology is by performing fast transient absorption images of much larger samples using a single time delay; when a structure of interest is detected, one could then record a full 2D IR spectrum. Using Kieran's new polarization scheme (0,0,60,-60) for measuring cross-peaks in TA spectra, this fast-image method can still examine useful features in addition to those that occur on the diagonal (while still in the pump-probe geometry), such as vibrational couplings. One final experimental challenge was kind of silly, but with our current setup it is extremely difficult to image the same area twice. Electron microscopes typically overcome this by creating a low magnification map of a sample, with TA imaging, we could do the same thing if a molecular reporter were attached to the CaF_2 window in some regular pattern.

7.2 Future directions: Genotyping animal models and studying diseases.

Chapter 4 provides a method for evaluating overexpression animal models of human diseases. This was a collaborative effort involving many individuals, and I am not sure how to improve the sequencing technology. But, working with Alan Attie's lab and the UW Biotechnology Center, any future student working with hIAPP animal models or Cataract animal models could use the shared facilities in the same manner. Additionally, any lab using the *RIPHAT hIAPP* mouse model could use our primers to genotype their mice and study the new phenotype our study elucidated (non-diabetic homozygote). Indeed, Mandar Muzumdar's laboratory at Yale has contacted us with their intention to phenotype their colony using our methods. They are working to elucidate the role of obesity in pancreatic cancer.

Perhaps the most exciting development to come from my work on the Chapter 4 project was that we decided to start from scratch and create our own mouse model for type II diabetes, using a 1 for 1 exchange of mouse amylin DNA for 3A amylin DNA. All of the early 3A mice died, perhaps unsurprisingly because it is toxic, but we have since produced an inverted 3A mouse model, for whom the DNA flips orientation when bred with another mouse model. Future students could isolate pancreatic islets and use fluorescence and 2D IR imaging to study the development of oligomers and plaques within these tissues.

7.3 Future directions: The FGAIL oligomer in biological environments.

Chapter 5 outlines an in-depth aggregation mechanism for hIAPP interacting with membranes at 4 different domains; the FGAIL intermediate we previously studied in solution was conserved, whereas the structures at other regions of hIAPP were not. However, these membranes were simple vesicles composed of DOPC lipids and cholesterol. An easy next step is to employ the same double-labeling schemes that distinguish α-helix from β-sheet and study interactions with more biologically relevant membranes such as those containing membrane associated receptors or various lipid headgroups. One might also change the pH of the solution to study aggregation mechanisms in secretory vesicles. This scheme is also readily applied to other aggregation-prone proteins, such as amyloid-β.

7.4 Future directions: Stabilizing the toxic oligomer for high-resolution structural determination.

Chapter 6 contains the progress I made on an extremely challenging project that has baffled Zanni group students for about a decade. The TEM and ThT results for 3A-hIAPP show similarities to aggregation mechanisms for human IAPP. Excitingly, the solution NMR data presents an ordered FGAIL region (similar to a short β-sheet structure predicted by IR spectroscopy). Chapter 6 is an ongoing project that I will help current students complete as a post-doctoral researcher in the Zanni group. Long term future directions involve drug and antibody screens against the toxic 3A oligomer, as well as new animal models (mentioned previously) to study the effects of longer lived toxic oligomers.

8 Publications.

7. **S. S. Dicke**, M. Maj, C. R. Fields, M. T. Zanni. Metastable intermediate during hIAPP aggregation catalyzed by membranes as detected with 2D IR spectroscopy. *RSC Chem. Biol.* 2022, 3, 931–940.

6. **S. S. Dicke**, C. D. Rubinstein, J. M. Speers, M. E. Berres, D. M. Pavelec, K. L. Schueler, D. S. Stapleton, S. P. Simonett, M. P. Keller, A. D. Attie, and M. T. Zanni. A protocol for locating and counting transgenic sequences from laboratory animals using a map-then-capture (MapCap) sequencing workflow: procedure and application of results. *bioRxiv.* Posted 19 January 2022.

5. A. M. Alperstein, K. S. Molnar, **S. S. Dicke**, K. M. Farrell, L. N. Makley, and M. T. Zanni. Analysis of amyloid-like secondary structure in the Cryab-R120G knock-in mouse model of hereditary cataracts by two-dimensional infrared spectroscopy. *PLoS ONE.* 2021; 16(9).

4. **S. S. Dicke**, A. M. Alperstein, K. L. Schueler, D. S. Stapleton, S. P. Simonett, C. R. Fields, F. Chalyavi, M. P. Keller, A. D. Attie, and M. T. Zanni. Application of 2D IR Bioimaging: Hyperspectral Images of Formalin-Fixed Pancreatic Tissues and Observation of Slow Protein Degradation. *J. Phys. Chem. B.* 2021, 125, 9517−9525.

3. K. M. Farrell, J. S. Ostrander, A. C. Jones, B. R. Yakami, **S. S. Dicke**, C. T. Middleton, P. Hamm, and M. T. Zanni. 2020. Shot-to-shot 2D IR spectroscopy at 100 kHz using a Yb laser and custom-designed electronics. *Opt. Express.* 2020, 28(22), 33584 - 33602.

2. C. R. Fields, **S. S. Dicke**, M. K. Petti, M. T. Zanni, and J. P. Lomont. A Different hIAPP Polymorph Is Observed in Human Serum Than in Aqueous Buffer: Demonstration of a New Method for Studying Amyloid Fibril Structure Using Infrared Spectroscopy. *J.Phys. Chem. Lett.* 2020, 11, 15, 6382–6388.

1. **S. S. Dicke**, L. Tatge, P. E. Engen, and L. R. Masterson. Isothermal titration calorimetry and vesicle leakage assays highlight the differential behaviors of tau repeat segments upon interaction with anionic lipid membranes. *Biochem. Biophys. Res. Comm*. 2017, 493 (4), 1504-1509.