

# Investigating the effect of mutations on the stability and toxicity of amyloid proteins for therapeutic benefit

Mohamed Raef Smaoui

Doctor of Philosophy

School of Computer Science

McGill University

Montreal, Quebec

September 2014

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy

©Mohamed Raef Smaoui 2014

## DEDICATION

To my supportive, loving, and caring parents.

## ACKNOWLEDGEMENTS

I am forever grateful to my parents, Nejib Smaoui and Sonia Trabelsi, for their continuous support and motivation throughout my academic path. They will always be my role models in life and source of inspiration. I thank my wife, Doaa Farid, for her everlasting encouragement, patience, understanding, and teamwork in getting through this chapter of life while raising our two kids, Yousuf and Leen. I'm very thankful to my supervisor, Jérôme Waldispühl, whose mentorship, guidance, and advising in the world of bioinformatics were crucial to the accomplishment of my Ph.D.

## ABSTRACT

Amyloid aggregation is involved in the death of many cells and is believed to be the leading cause of neurodegenerative diseases such as Alzheimers, Parkinsons, Huntington, and Type II Diabetes. In order to counter their detrimental effects to the cell, it is crucial to first understand their intrinsic structural properties and their molecular dynamics. Although the topic of amyloid nucleation has been studied extensively, there are no computational studies modeling full aggregation of amyloid proteins into stable long fibril structures and analyzing their stability potentials. In addition to the computational complexity of these problems, there have been very few computational studies exploring the effect of sequence mutations on amyloid stability, amyloidegenicity and toxicity. In this thesis, we ultimately aim to construct these computational methods by modeling and simulating the dynamics of amyloid fibril build-up and assessing their sensitivity to sequence alterations. We present the tools we built to simulate the aggregation process of amyloid proteins into fibrils, explore the effect of sequence mutations on destabilizing amyloid fibrils, reveal the mutational landscape of the amylin amyloid protein involved in diabetes, and outline a novel method to construct a therapeutic agent to minimize the toxicity of amyloid oligomers in the pancreas of diabetes patients.

## ABRÉGÉ

L'agrégation d'amyloïde est impliquée dans la mort de nombreuses cellules et est considérée comme la principale cause de maladies neurodégénératives telles que la maladie d'Alzheimer, de Parkinson, de Huntington, et le diabète de type II. Afin de lutter contre leurs effets néfastes sur la cellule, il est essentiel de d'abord comprendre leurs propriétés structurelles et leurs dynamiques moléculaires. Au meilleur de notre connaissance, il n'y a pas eu d'étude de modélisation informatique qui construit l'agrégation de la protéine amyloïde dans une structure de fibrilles stables et qui analyse leurs potentiels de stabilité, ni y a-t-il d'étude informatique qui a exploré l'effet des mutations de la séquence sur la stabilité de l'amyloïde et sa toxicité. Dans cette thèse, nous visons ultimement à la construction de ces méthodes de calcul par la simulation de la dynamique des fibrilles amyloïdes au niveau moléculaire et cellulaire . Nous présentons les outils que nous avons construit pour simuler le processus d'agrégation de protéines amyloïdes dans les fibrilles, pour explorer l'effet des mutations de la séquence à déstabiliser des fibrilles amyloïdes, pour révéler le paysage de mutation de la protéine amyline amyloïde impliquée dans le diabète, et pour définir une nouvelle méthode de construire un agent thérapeutique pour réduire au minimum la toxicité des oligomères amyloïdes dans le pancréas de patients diabétiques.

## TABLE OF CONTENTS

DEDICATION	. . . . .	ii
ACKNOWLEDGEMENTS	. . . . .	iii
ABSTRACT	. . . . .	iv
ABRÉGÉ	. . . . .	v
LIST OF TABLES	. . . . .	x
LIST OF FIGURES	. . . . .	xi
1	Introduction . . . . .	1
1.1	Background . . . . .	3
1.1.1	Protein misfolding and amyloids . . . . .	4
1.1.2	Polymorphic fibrils . . . . .	6
1.2	Amyloids in various diseases . . . . .	8
1.2.1	$A\beta$ role in Alzheimer’s disease . . . . .	8
1.2.2	HET-s fungal prion . . . . .	9
1.2.3	Amylin amyloids in the pancreas . . . . .	10
1.3	Mutations affect amyloids . . . . .	12
1.4	Computational tools in studying amyloids . . . . .	13
1.4.1	Simulating protein structure & environment . . . . .	14
1.4.2	Computing solvation energy . . . . .	14
1.4.3	Molecular Dynamics . . . . .	16
1.4.4	Threading techniques . . . . .	17
1.4.5	Conformational search and structural prediction . . . . .	20
1.5	Thesis roadmap . . . . .	21
1.6	Publications and author contributions . . . . .	22
2	Computational assembly of polymorphic amyloid fibrils reveals stable aggregates . . . . .	25

2.1	Preface . . . . .	25
2.2	Abstract . . . . .	27
2.3	Introduction . . . . .	28
2.4	Materials and Methods . . . . .	30
	2.4.1 Construction & classification of polymorphic fibrils . . . . .	31
	2.4.2 Energy Minimization to filter fibril structures . . . . .	33
	2.4.3 Dipolar water model . . . . .	34
	2.4.4 Implicit water model . . . . .	38
	2.4.5 CreateFibril: Rigid Affine Transformations . . . . .	39
	2.4.6 CreateFibril: Fibril Axis Matrix . . . . .	41
2.5	Results & Discussion . . . . .	44
	2.5.1 Stability landscape . . . . .	45
	2.5.2 HET-s . . . . .	47
	2.5.3 Abeta . . . . .	49
	2.5.4 Amylin . . . . .	54
	2.5.5 Online tool & Fibril database . . . . .	61
2.6	Acknowledgments . . . . .	62
3	Computational re-engineering of Amylin sequence with reduced amyloidogenic potential . . . . .	63
3.1	Preface . . . . .	63
3.2	Abstract . . . . .	65
3.3	Introduction . . . . .	66
3.4	Materials and Methods . . . . .	69
	3.4.1 Molecular Dynamics and Energy Minimization . . . . .	75
	3.4.2 Analyzing Energy Results . . . . .	76
3.5	Results and Discussion . . . . .	76
	3.5.1 Exploring key stability regions of Amylin . . . . .	77
	3.5.2 Generating Amylin fibrils . . . . .	78
	3.5.3 Analyzing Amylin fibrils . . . . .	79
	3.5.4 Maintaining native structure & function . . . . .	89
3.6	Conclusion . . . . .	89
3.7	Acknowledgements . . . . .	90
4	Complete characterization of the mutation landscape reveals the effect on amylin stability and amyloidogenicity . . . . .	91
4.1	Preface . . . . .	91

4.2	Abstract . . . . .	92
4.3	Author Summary . . . . .	94
4.4	Introduction . . . . .	94
4.5	Methods . . . . .	98
	4.5.1 Computing the landscape . . . . .	98
	4.5.2 Mutant Structures . . . . .	101
	4.5.3 Calculating structure energy . . . . .	101
	4.5.4 Molecular Dynamics and Energy Minimization . . . . .	102
	4.5.5 Assessing structural deviations . . . . .	103
	4.5.6 Calculating energies for multiple-point mutations . . . . .	103
4.6	Results & Discussion . . . . .	106
	4.6.1 Amyloidogenicity mutational landscape analysis . . . . .	108
	4.6.2 Stable and unstable regions in native amylin . . . . .	111
	4.6.3 Amyloidogenic regions in amylin . . . . .	113
	4.6.4 Evolutionary mutations . . . . .	113
	4.6.5 Efficient estimation of stability and amyloidogenicity in multiple-point mutation landscapes . . . . .	116
	4.6.6 Mutant (A5K, A8R, G24I) is stable and less amyloidogenic	118
4.7	Conclusion . . . . .	118
4.8	Acknowledgments . . . . .	120
4.9	Supplementary material . . . . .	121
5	Probing the binding affinity of amyloids to reduce toxicity of oligomers in diabetes . . . . .	122
5.1	Preface . . . . .	122
5.2	Abstract . . . . .	124
5.3	Introduction . . . . .	125
5.4	Approach . . . . .	127
5.5	Methods . . . . .	129
	5.5.1 BASM matrix . . . . .	132
	5.5.2 Amyloidogenicity . . . . .	136
	5.5.3 Oligomer concentrations . . . . .	137
	5.5.4 Molecular Dynamics and Energy Minimization . . . . .	139
	5.5.5 Dipolar water solvent . . . . .	139
	5.5.6 Building oligomer structures . . . . .	140
	5.5.7 The effect of amyloid oligomers on the Insulin-Glucose system	140
5.6	Results . . . . .	143
5.7	Discussion . . . . .	150

5.8	Supplementary material . . . . .	155
6	Conclusion . . . . .	156
6.1	Summary . . . . .	156
6.2	Contribution to amyloid research . . . . .	158
6.3	Future direction . . . . .	158
	References . . . . .	160

## LIST OF TABLES

Table	page
2-1 Predicted structural parameters for HET-s, Abeta, and Amylin fibrils produced by CreateFibril. . . . .	48
3-1 Effect of mutation choice on structural stability. . . . .	72
3-2 Amylin Mutations generated by FibrilMutant with destabilizing potential. . . . .	78
3-3 Fibril stability results. . . . .	83
3-4 TANGO results on amylin mutations. . . . .	87
4-1 Generating the mutational and amyloidogenic landscapes of amylin . .	99
4-2 Generating $\Delta\tilde{E}$ and $\Delta\tilde{G}$ for $n$ -point mutations . . . . .	105
4-3 Evolutionary amylin mutations in species . . . . .	114
4-4 Top 20 results for amylin 3-point mutations . . . . .	117
5-1 Sequences $S^*$ for various $\alpha$ and $\beta$ binding affinity values. . . . .	143
5-2 Top results solving the 4-residue parsimonious sequences for $S^{P_4}$ with $\alpha = 2$ and $\beta = 1$ . . . . .	144
5-3 The Binding Affinity Scoring Matrix (BASM) R score values of the amylin amyloid . . . . .	145
5-4 The Binding Affinity Scoring Matrix (BASM) R score values of the Pramlintide amyloid . . . . .	146
5-5 Exploring aggregation properties of top 10 results with TANGO . . . .	147

## LIST OF FIGURES

<u>Figure</u>	LIST OF FIGURES	<u>page</u>
1-1	Protein misfolding into an amyloid structure. . . . .	5
1-2	Modeling HET-s polymorphic amyloid fibril structures. . . . .	6
1-3	Short amyloid fibril structures of HET-s, A $\beta$ , and amylin. . . . .	8
1-4	Dipolar solvent model. . . . .	15
1-5	Simulating protein motion. . . . .	16
1-6	Representation of protein threading approach. . . . .	19
2-1	Classification of the polymorphic fibril structures produced by CreateFibril. . . . .	32
2-2	CreateFibril procedures and pipeline. . . . .	35
2-3	The different parameters used by CreateFibril to build structures. . . . .	46
2-4	HET-s Single fibril parameter findings. . . . .	50
2-5	Energies of ABeta and Amylin fibrils as they aggregate. . . . .	52
2-6	Cross-sectional view of the hydration shell effect on the hydrophobicity of the predominant HET-s, Abeta, and Amylin fibrils produced by AQUASOL. . . . .	53
2-7	Energies of Abeta and Amylin fibrils as they aggregate. . . . .	54
2-8	Heatmap representation of the stability landscape and the enthalpy drift of Abeta wrapped 2-Stack fibrils. . . . .	55
2-9	Abeta wrapped fibrils. . . . .	58
2-10	Convergence rate of Abeta and Amylin tweaked structures. . . . .	60

3-1	Amylin amyloids. . . . .	79
3-2	Free energies of Amylin mutated fibrils calculated with Eq. 3.1. . . . .	80
3-3	Top 3 mutation results from AmyloidMutants. Mutation A8E presents a high destabilization fibril aggregation frequency. . . . .	86
3-4	Stability of the Amylin mutants in their native fold. . . . .	88
3-5	$\delta_{rmsd}$ and RMSF plots for mutants L12E, A8E, and G33E over a 5ns simulation. . . . .	88
4-1	Stability and amyloidogenicity mutational landscapes. . . . .	107
4-2	Projections of mutational landscapes. . . . .	109
4-3	Stability analysis of mutations. . . . .	110
4-4	Residues that contribute to stability and amyloidogenicity. . . . .	112
4-5	MD results of mutation (A5K, A8R, G24I) . . . . .	119
5-1	Oligomer toxicity. . . . .	130
5-2	Binding affinity method. . . . .	131
5-3	Energies for top ten $S^{P_4}$ amylin analog oligomers. . . . .	148
5-4	Concentration of oligomers in solution. . . . .	148
5-5	RMSD and RMSF plots for the $S^{P_4}$ sequence with the T9K, L12K, S28H and T30K mutations over a 6ns simulation. . . . .	149

## CHAPTER 1

### Introduction

Some proteins that carry essential functions in cells degenerate, misfold, and aggregate into structures known as amyloid proteins. These amyloids are believed to be a major contributing factor in the progression of several neurodegenerative diseases such as Alzheimer's and diabetes. In the case of diabetes, we found that short amyloid aggregates permeate the cell membrane of cells and disrupt the electric potential across the membranes, which lead to cell death. Finding ways to limit this toxicity of amyloids is a crucial step towards the improvement of health care standards for patients. The overall aim of this thesis is to explore different methods that decrease amyloid formation and toxicity. The results of this work can help simulate accurate amyloid aggregate models that occur in a specific disease, and research novel ways to limit the harmful impact on cellular toxicity. The direct objectives of this thesis are:

1. To simulate the aggregation process of amyloid proteins into polymorphic fibrils and model, for the first time, reliable atomic structures of fibrils.
2. To use the amyloid aggregate models in exploring the effect of sequence mutations on amyloid inhibition.

3. To characterize the complete mutational landscape of the amylin amyloid in diabetes and unravel all stabilizing and amyloidogenic mutations that can affect the onset and progression of the disease.
4. To design a method towards developing a novel therapeutic agent that lowers the toxicity of amyloids in type II diabetes.

The objectives of this work are directly motivated by the following questions:

- Can we computationally simulate the process of aggregation given limited experimental data we have on amyloid structures?
- What structural parameters and physical interactions of amyloids act as potential players in the process of aggregation?
- How do we assess and confirm the correctness of aggregation models, and what tools can help us estimate their structural energy?
- How significant are amino acid mutations in affecting the emergence of amyloid proteins?
- If mutations are a significant contributing factor in affecting the emergence of amyloid proteins, can we construct a methodology to explore the effects of these mutations?
- What other methods can we utilize and explore to destabilize or inhibit amyloid fibrils from forming and affecting the rate of cell death?
- What tools can we provide to help explore the exact effect of amyloids on cell death in diabetes and can this relationship be modelled to offer a better understanding of the dynamics of the disease?

In this chapter, we provide a brief background to amyloid proteins and their involvement in various human diseases. We discuss some of the computational tools and standards used in assessing protein structure and energetics and provide a roadmap for the rest of the chapters in this thesis.

## 1.1 Background

Alzheimer's, Parkinson's, Huntington, and type II diabetes are the most common neurodegenerative diseases today affecting more than 600 million people worldwide. Alzheimer's costs the USA \$148 billion in treatment every year, Parkinson's an estimate of \$20 billion a year and type II diabetes another \$40 billion, not to mention the cost of family sorrow, grievance and hardship. The financial toll and social implications of these disease have been a driving force in motivating researchers to find cures.

These neurodegenerative diseases afflict neurons and result in the deterioration of the brain, spinal cord, and nerves. Various organs including the heart, kidney, spleen, liver, pancreas, lungs and eyes are also indirectly affected [1]. Some of the symptoms of these diseases include hallucinations, abnormalities in speech and movement, severe memory loss, depression, deterioration of cortical neurons and awkward gait [2, 3, 4, 5, 6]. Diabetes, which is one of the main focus points of this thesis, is characterized by insulin resistance [7, 8] and the failure of the pancreatic beta cells to supply required levels of insulin [9]. The failure to secrete enough insulin is caused by beta cell dysfunction and reduced beta cell mass [10, 11, 12]. Many studies have found an association between this failure to produce insulin and the emergence of

islet amyloid protein deposits in humans [13, 14, 15], in non-human primates [16, 17], and cats [18]. Interest in studying the potential pathogenic role of islet amyloid deposits in type II diabetes has been incentivized by experiments reporting that mice transgenic for human islet amyloid develop hyperglycemia [19, 20, 21], a signature condition for type II diabetes.

### 1.1.1 Protein misfolding and amyloids

For reasons not entirely known, critical proteins in the body degenerate and misfold into  $\beta$ -sheet rich structures known as amyloids. These amyloids aggregate with one another into large deposits that aggravate the conditions of neurodegenerative diseases [22]. The deposits build up over time into organs and tissues and slowly force cells into apoptosis [1]. The amyloid deposits found in various diseases share common structural properties. They are all composed of  $\beta$ -strands that run perpendicular to the axis of the aggregation. These deposits, often referred to as fibrils, possess a hidden inner hydrophobic core and are stabilized by a dense network of hydrogen bonds that create firm linear aggregates [23, 24]. The toxicity of the fibrils kills neurons in the central nervous system and infects neighbouring tissue [25, 26], which leads to the symptoms mentioned earlier. Figure 1–1 shows a snapshot of an amyloid protein in its normal native stage, a snapshot as it starts to misfold, and a snapshot in its final amyloid form.

It is believed that proteins of all kinds can misfold and self-assemble into amyloid fibrils [27]. Environmental stress and thermodynamic conditions facilitate the conformational change and misfolding of normal proteins into amyloid structures that

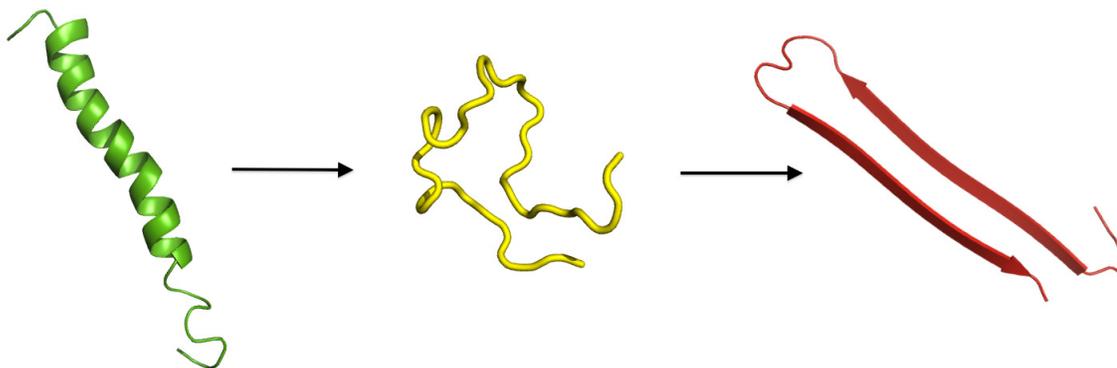


Figure 1–1: Protein misfolding into an amyloid structure. Left: the 3D structure of the amylin protein colored in green. Center: the amylin protein starting to misfold colored in yellow. Right: the misfolded, amyloid form of the amylin protein involved in diabetes.

are attributed to more than 40 pathological human conditions [28]. The amyloid aggregates usually form insoluble structures that create larger plaques and inclusion bodies that contribute to cellular inflammation [29]. Fibrils of the  $A\beta$  amyloid protein, the protein involved in Alzheimer’s disease, were observed to form plaques in human brain tissue [30]. Many believe that understanding the right conditions for misfolding can help combat this phenomenon and help advance methods to hinder the assembly of toxic amyloid aggregates and the formation of fibrils.

Researchers have used an array of methods and techniques to report the 3D structure of amyloid proteins, including electron microscopy, X-ray and neutron scattering, magnetic resonance spectroscopy, and biochemical techniques [31]. Although some of these techniques are state-of-the-art in capturing protein 3D structure, amyloid structures are extremely difficult to unravel. Their large and insoluble ultrastructural appearances prevent experimental techniques from correctly decoding their

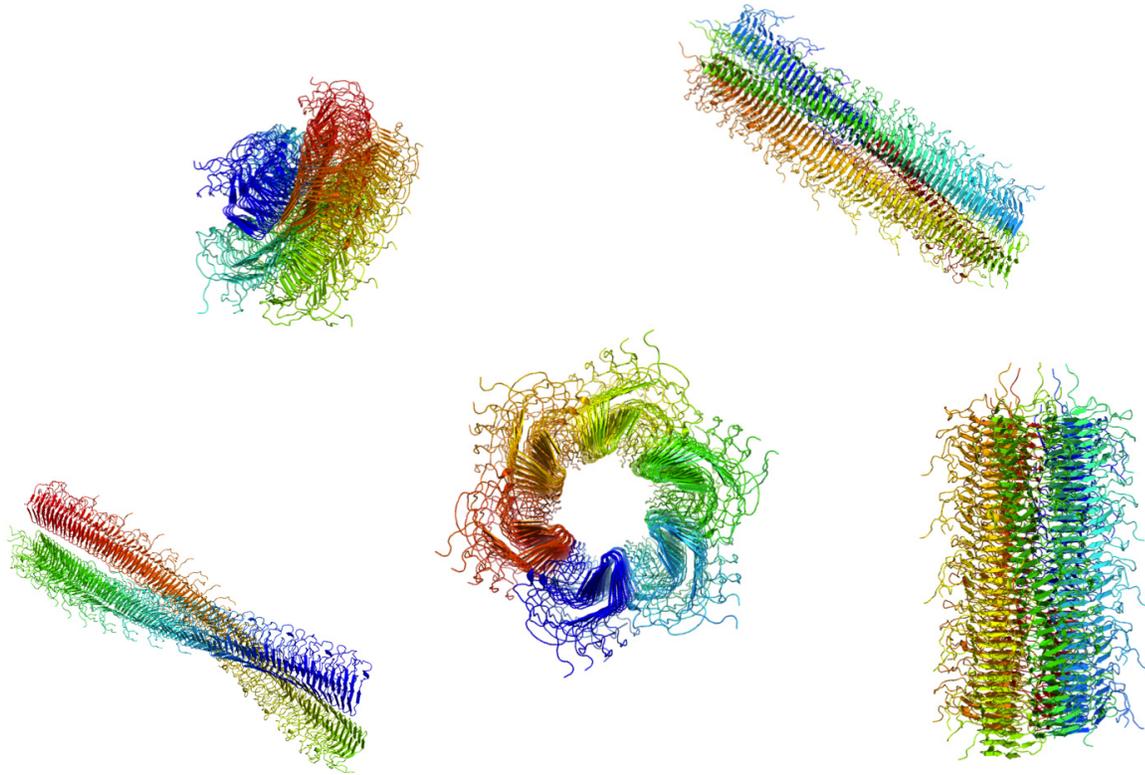


Figure 1–2: Modeling HET-s polymorphic amyloid fibril structures.

structure [32]. Nevertheless, some of the structural decoding successes include the HET-s prion [33] from the filamentous fungus *Podospora anserina*, the  $A\beta$  amyloid protein involved in Alzheimer’s disease [32], and the amylin protein in diabetes [34].

### 1.1.2 Polymorphic fibrils

Proteins that misfold into insoluble amyloid structures possess the potential to form aggregates consisting of a structural spine with cross- $\beta$  motifs. Hydrogen bonds

parallel to the spine stabilize fibril structure and determine any polymorphism properties. Depending on the alignment and strength of the hydrogen bonds, single fibril structures can assemble into higher-order forms. These higher-order polymorphisms are characterized by the packing distance between adjacent  $\beta$ -sheets on the single fibrils, the number of packed fibrils, and the degree of twisting along the main polymorphic aggregate axis. The hydrogen bonds are significant to the point that they influence the rate of polymorphic assembly and polymorphic fibril size. The various polymorphic assemblies give rise to different biological functions [35, 36] and toxicity levels in cells [37, 38]. Some studies even show that these fibrils propagate their specific polymorphisms to daughter fibrils [39, 40] to preserve their destructive potential. Some of the best known fibrils have been observed in cryo-TEM reconstruction of insulin fibrils [41], TEM analysis of amylin [42], cryo-EM analysis of HET-s [37], and TEM analysis of  $A\beta$  [43]. Figure 1–2 presents models for polymorphic fibril forms of HET-s and has been constructed by tools we developed in the next chapters.

HET-s fibrils have been observed to undergo two polymorphisms [37] and  $A\beta$  has been observed to exist in three polymorphic shapes. Many other fibrils are believed to exhibit polymorphic structures, however, the current limitations in technology and the high insolubility of these structures impede experimental validation. Efforts have turned to computational modeling and simulation to assist in revealing fibril structure and predicting aggregation dynamics.

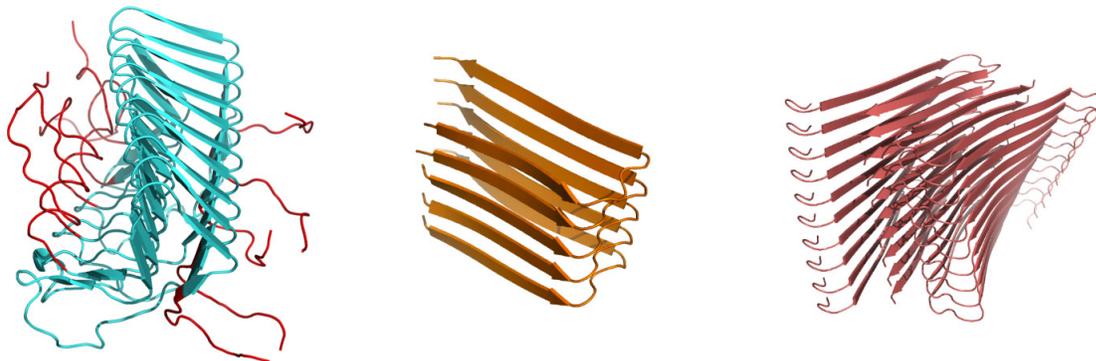


Figure 1–3: Short amyloid fibril structures of HET-s (left),  $A\beta$  (center), and amylin (right).

## 1.2 Amyloids in various diseases

Amyloids have been attributed to the progression of more than 40 diseases. In this section, we touch upon three of the best known amyloid peptides,  $A\beta$ , amylin, and HET-s, and explain their role in disease. Figure 1–3 shows 3D representations of these peptides in amyloid fibril form.

### 1.2.1 $A\beta$ role in Alzheimer’s disease

$\beta$ -amyloid peptide ( $A\beta$ ), found excessively in patients with Alzheimer’s disease, is believed to lead to neurodegeneration in humans [44]. This protein aggregates into various fibril shapes that form neuritic plaques and neurofibrillary tangles [45, 46, 47].  $A\beta$  molecules are known to form into the polymorphic 3-packed fibril [48] and 2-packed shapes [49, 35]. It is unclear whether  $A\beta$  deposits in brains of Alzheimer’s patients are the main cause of the disease, a major contributing factor, or a byproduct of reactions that degenerate neuron cells.

Polymorphic models of  $A\beta$  fibrils expose different residues on their surface to interact with water solution. 2-packed fibrils expose less hydrophobic residues on the surface and are hypothesized to be the predominant  $A\beta$  fibril structure, followed by the 3-packed fibrils and the single fibrils. The hydration shell effect on the monomers that make up these fibrils could possibly explain the stability of one or more predominant  $A\beta$  fibril structure in neuron cells and their potential toxicity.

Recently observed structures, termed as wrapped  $A\beta$  fibrils, are believed to further stabilize 2 packed polymorphs. Stroud *et al.* [49] used X-ray powder diffraction to observe that some  $A\beta$  fibrils compose of laterally associated protofilaments that twist around more than one internal helical axis. Together, the internal axes also twist (in phase) around a main fibril axis, creating multiple wrappings in structures. Stroud *et al.* showed that larger twisting angles result in greater fibril curvature and larger holes in fibrils, suggesting that  $A\beta$  fibril toxicity could be related to their polymorphic shapes and potential for forming pores.

### 1.2.2 HET-s fungal prion

HET-s is a fungal prion involved in the programmed cell death of filamentous fungi *Podospora anserina*. When it is in its misfolded amyloid form, it is involved in a self recognition process known as heterokaryon incompatibility; when two cells respectively contain amyloid HET-s and native HET-s fuse, programmed cell death occurs [37]. Unraveling the structure of HET-s (in both amyloid and native form) and the mechanism of fibril propagation and cell toxification might prove beneficial

in understanding the reasons behind amyloid misfold and disease propagation and progression.

HET-s fibrils have been studied extensively as a result of their available high resolution atomic structures [33]. Its fibrils exhibit a left handed swirling orientation of HET-s amyloid monomers around their axis. The experimentally observed predominant form of HET-s fibrils are the single and 3-packed fibrils [37]. It is believed that at pH values lower than 3, single fibrils come together and pack to form 3-packed structures. The most energetically favorable HET-s fibrils possess a helical pitch of about 410-Å, beta sheet aggregation of 4.8-Å, and left-handed twist of the fibril [37].

### 1.2.3 Amylin amyloids in the pancreas

Islet amyloid polypeptide (amylin) deposits in the pancreas are believed to be a contributing factor to the onset and progression of type II diabetes [50, 51]. Polymerized amylin fibrils *in vitro* showed a diverse ensemble of polymorphic shapes [52, 42], but unfortunately to date, no one has been able to crystallize full-length human amylin. Several models have been proposed, one of which by Wiltzius *et al.* [34] built using biochemical and structural data along with the fibril NNFGAIL and SSTNNG crystallized regions of amylin to formulate a structure with atomistic details for the protein in the form of a 2-packed fibril. It is our understanding at the time of this thesis that this is the best model in the literature regarding amylin in its monomeric amyloid form. This model claims that the 2-packed fibril hides most hydrophobic residues from water and hence is very stable.

Normally after a meal, the body digests food into glucose for energy consumption. When glucose levels increase in the blood (after a meal), the  $\beta$ -cells in the pancreas generate insulin and amylin proteins at a ratio of 1:100 as a response. Together, insulin and amylin lower glucose levels in the blood and play a crucial role in maintaining glucose homeostasis [53]. Although amylin is one of the most crucial proteins in the body, it has a tendency to misfold into destructive amyloid forms and aggregates to create insoluble fibrils that damage the  $\beta$ -cells that produce it. This tendency increases when a person stresses their body by eating too much, forcing high spikes of glucose levels in the blood and an increased production of insulin and amylin. The amylin peptide forms small and large deposits in the intra and extra-cellular chambers of  $\beta$ -cells over the time span of years [54, 55]. Although it remains obscure why amylin would misfold and convert into insoluble, aggregating amyloid fibrils, these deposits are among the most common pathological features of type II diabetes found diabetic subjects [56, 57].

The deposits can create pores in the membranes of  $\beta$ -cells that disrupt ion flow and kill cells. Once a  $\beta$ -cell dies, it can no longer produce insulin. If many  $\beta$ -cells die, the pancreas cannot produce enough insulin to lower glucose levels in the body and the person becomes diabetic (needs insulin injections: type II diabetes). Finding a way to limit the aggregation and toxicity of the amylin fibrils could help lengthen the lifespan of the  $\beta$ -cells and maintain glucose-insulin homeostasis. This would be of critical importance for attempting to find therapeutics to combat the growth of amylin fibrils in pancreas [51].

### 1.3 Mutations affect amyloids

Proteins have been observed to undergo natural mutations that change their amyloidogenicity potential and rate formation. Several cases of Parkinson's disease are associated with amino acid mutations of the alpha-synuclein ( $\alpha$ S) protein [58, 59, 60]. The A30P mutation in  $\alpha$ S decreases the overall rate of fibril formation [61, 62], while the H50Q, H50A, and G51D mutants aggregate more quickly than the wild type but more slowly than A53T and E46K mutants [63]. Single-point mutations have been observed to be sufficient to affect the landscape of the  $A\beta_{42}$  protein in Alzheimer's and change the internal dynamics between microstates [64]. R5A mutation studies showed a decrease both in the tendency towards  $A\beta$  aggregate formation and a reduced toxicity in Alzheimer's [65]. A mutation in amino acid position 25 of  $A\beta$ , the loop area connecting two beta strands has been shown to destabilize  $A\beta$  fibrils [66]. Furthermore, a single mutation of serine-to-glycine at position 20 in amylin in Chinese and Japanese populations [67] is associated with early onset of type II diabetes [68, 69] and amplified amyloid formation [70, 71, 72]. The Q10R mutation of the amylin gene in the New Zealand Maori population is believed to explain the high susceptibility and prevalence of diabetes in Maori patients [73]. Rodents and mice exhibit six point mutations in their amylin peptide that inhibit fibril formation [74, 75, 76]. Moreover, more than 120 single point mutations have been associated with the systemic disorder FAP [77].

As a result of the frequently observed natural mutations, research into deliberately mutating amino acids of amyloids has been proposed as an explorative method to destabilize fibrils and reduce toxicity. Computational methods coupled and steered

with Molecular Dynamics simulations have proved to be a viable strategy to study the impact of mutations, however, they are unscalable and expensive in resources and time [78]. Advancements in understanding the effects of amyloid sequence mutations on fibril toxicity and formation rate has paved the way for the development of therapeutic agents to replace highly amyloidogenic species. Re-engineering the genetic code of proteins and administering them as substitute agents for patients is a promising strategy for drug development. Pramlintide, a mutated protein version of amylin, is used as a drug replacement in type I and type II diabetes and has been shown to produce less fibrils and cause less  $\beta$ -cell death in the pancreas [79]. It is still unclear how point mutations alter the pathway of oligomerization and the kinetics of fibril conformational transitions [64], however, it is clear that proteins aggregate through nucleation-dependent polymerization [22, 63]. Hence, exploring mutations that affect the nucleation of amyloid monomers has the potential to facilitate the development of more therapeutic agents to inhibit oligomer formation and reduce the effect of disease.

### 1.4 Computational tools in studying amyloids

Determining the high resolution structure of amyloid proteins is a crucial step towards understanding the mechanisms of amyloid aggregation, causes of fibril toxicity, and perpetuation of fibril structures in membranes. Studying high resolution structures of amyloid proteins is also believed to enhance the possibilities of developing new reagents to inhibit fibril formation. However, due to the current limitations of experimental methods, researchers have not been very successful in discovering

amyloid structures. Many have turned to computational efforts in hope of correctly deciphering structure [80]. Computational researchers in the field have attempted to model amyloid proteins and simulate their aggregation, but are facing simultaneously scalability and complexity challenges [80, 81, 82, 83].

#### **1.4.1 Simulating protein structure & environment**

Because simulating protein structures is a computational intensive task, most researchers have used heuristic approaches to create coarse grain models and capture only the most important details affecting their systems [84, 85, 86]. In addition, solvents are usually never explicitly added in a simulation, rather, implicitly modelled with potential functions and limited performance due to the complexity that solvents add. However, to accurately model and quantify the dynamics and stability of amyloid proteins, such approximations are limiting and do not provide accurate results. To perform acceptable simulations, one needs a detailed molecular structure model and a tool to accurately compute solvation energy, coulomb interactions, and electrostatic potentials.

#### **1.4.2 Computing solvation energy**

Solvation energy is a term that takes into account hydrogen bonds and interactions between a molecule and its surrounding water solvent. It is the free energy that is required to transfer a solute molecule from a solvent environment to vacuum and is highly dependent on pH and temperature [87, 88]. The accurate and efficient calculation of solvation energy is difficult [89]. In standard practice, solvation energy

of a molecule is calculated by computing the Poisson-Boltzmann (PB) second order, elliptic, nonlinear partial differential equation. This method quickly increases in complexity as the number of molecules increase in a system, making exact computation unfeasible. The most recent method in approximating solvation energy has been introduced by the AQUASOL framework. Currently, this tool is the state-of-the-art at solving the dipolar Poisson-Boltzmann-Langevin equation (DPBL) instead of PB to estimate solvation energy [89] (see Figure 1–4). Solving the DPBL improved the precision and performance of solvation energy calculations. According to its authors, AQUASOL is fast, accurate and robust. In this work, we rely on AQUASOL to provide the solvation energy calculations for the molecules we simulate.

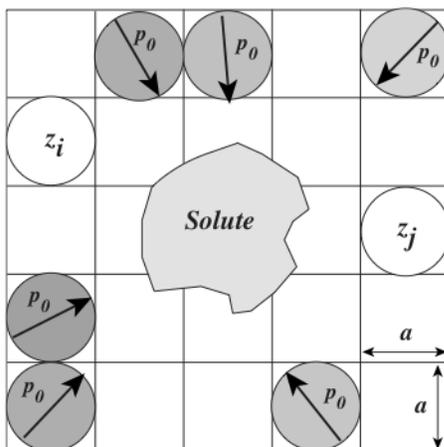


Figure 1–4: Dipolar solvent model. Illustration of the lattice gas model for the DPBL equation. Each lattice cell may be empty, occupied by one ion or occupied by a water dipole of constant magnitude  $p_0$  but variable orientation. This example shows multiple sites occupied by water dipoles and two sites occupied by ions with valence  $z_i$  and  $z_j$ . The lattice size  $a$  sets the size of the ions and dipoles. (figure and caption taken from an article by Patrice Koehl et al. [89].)

### 1.4.3 Molecular Dynamics

We resort to Molecular Dynamics (MD) to simulate the motion of molecules we model throughout this work. In particular, we use the GROMACS package [90, 91] to perform MD simulations and track the motion of molecules over time (see Fig. 1–5). GROMACS simulates protein motion using Newtonian physics and uses force fields and thermodynamics parameters to assess potential energies of structures as they change over time. The package essentially attempts to mimic how proteins would behave in a natural environment. In our simulations, we use the GROMACS package to study how the behavior and energetics of our proteins change under varying conditions. We do not determine the solvation energy term with GROMACS as it is expensive to compute compared to AQUASOL.



Figure 1–5: Simulating protein motion. Three snapshots of the structure of the amylin protein as it moves in solution (hidden) over a time frame of 30 picoseconds.

#### 1.4.4 Threading techniques

Traditional experimental methods, such as x-ray crystallography, electron microscopy, and nuclear magnetic resonance have not seen great success in elucidating the structures of most amyloids, due to insolubility and noncrystalline nature of amyloid proteins. To circumvent these challenges, computational techniques have been introduced and widely used to build structural motifs of amyloid fibrils, test their stability, and reveal the mechanisms behind fibril formation. One prominent approach is to generate amyloid models with threading tools and test their structural stability using MD simulations. Although this approach potentially enables exploring structural conformations that stabilize proteins [92], threading amyloid structures remains difficult and inaccurate. Figure 1–6 shows a drawing of this process.

From the sparse scientific data gathered about amyloids, it is established that these proteins possess rich  $\beta$ -strand morphologies that are held by hydrogen bonds running perpendicular to their main fibril axis. This can be clearly observed for the A $\beta$ , amylin, and HET-s peptides. Using this information and data provided by other statistical knowledge relating the potential of sequences to fold into known structures, threading tools and software have been intensively used to generate and deduce realistic structural models for unknown proteins [93, 94, 95].

Amyloids are mostly insoluble proteins that cannot be observed by experimental methods. However, since their amino acid sequences can easily be determined, threading tools are a first good step in modelling and predicting their unknown structures by using statistical information obtained from published structures in the

Protein Data Bank (PDB). Structural threading is powerful when it correctly identifies structural homologs by aligning target amino acid sequences onto structural templates [96, 97]. TANGO [98] and AmyloidMutants [99] are among the current threading tools that explore protein (including amyloid) stability and analyze the effect of secondary structure modifications on properties such as amyloidogenicity and protein aggregation. Such tools use coarse-grained models that enable them to perform high-throughput screenings, but cannot achieve the accuracy of higher resolution models that take into account long-range interactions in structure [78]. Various energy functions are used to empirically evaluate the validity of models. One such function, the Ramachandran Plot, shows the acceptable range of phi-psi angles for different amino acids in a 3D structure. This plot assesses protein tertiary structures and confirms or rejects threaded models.

After the threading and initial structural assessment of a few amyloid protein models, MD simulations are used to measure with great detail the stability of these structures. MD simulations undergo intensive numerical calculations of physical movements of atoms within molecules. Forces between particles and potential energies are computed and updated in a continuous time frame using minuscule time intervals. Each threaded protein is initially prepared in a box of water molecules at time  $t = 0$ . As time progresses, the simulation calculates the effect of the forces produced by each atom on the total stability of the amyloid structure. At the end of the run, simulation results either reveal proteins that break up and denature (due to high unstable internal forces) or proteins that have reshaped slightly to optimize their stability. Only those proteins that have maintained their stability after the run

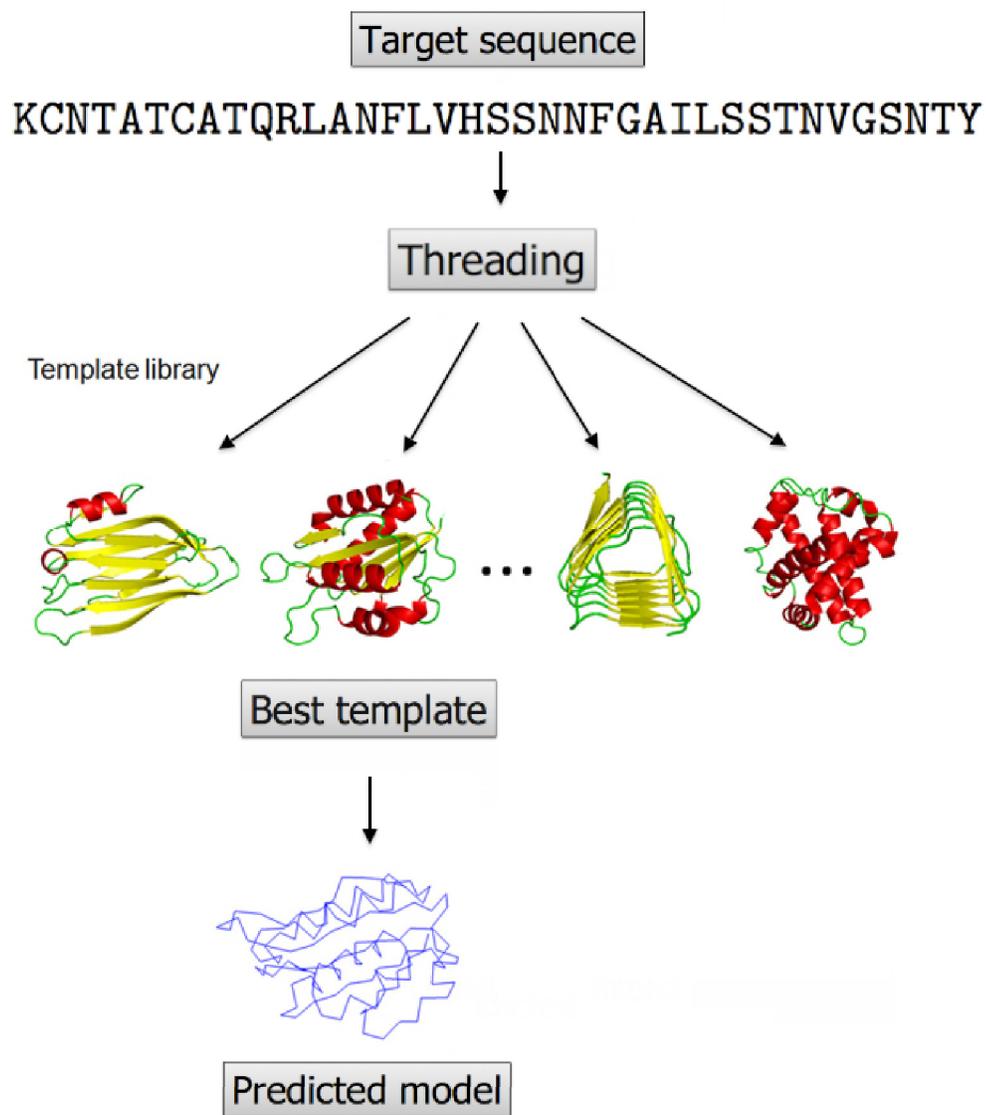


Figure 1–6: Representation of protein threading approach. Target sequence is the sequence of a protein for modeling. The sequence is “threaded” against several protein structure templates. The template that returns the best fit is used as a predictive model for the sequence.

are chosen for further assessment and believed to be closer candidates to the real unknown amyloid structures.

#### **1.4.5 Conformational search and structural prediction**

Protein structure prediction has been of utmost interest to researchers attempting to correlate structure with protein function. The theory in the field claims that structure is tightly linked to function. Amyloid proteins exist in native and amyloid forms, each exhibiting a different biological function. For several years, scientists have tried to explore the unknown structure of some amyloids without much success, due to the current limitations in biological instruments and to the insolubility of amyloid aggregates. Nevertheless, the primary sequence of those proteins can be easily determined. With the known primary sequence, the challenge is in developing computational methods to infer protein structure from primary sequence.

Starting with a protein's primary sequence of amino acids, scientists aim to determine the most energetic 3D fold for the protein. This task usually involves creating a large number of possible folds to explore the structural landscape of the protein and assessing each fold's potential using some energy function [30]. Ideally, this search process would find the structure with the lowest energy in the landscape and return it as a potential result. However, since no knowledge is known about the folding landscape and assessing the stability of each possible fold is too expensive, intelligent algorithms scan parts of the landscape for minima points and usually use variations of steepest descent algorithms [100] in their search. The search problem in the continuous landscape is difficult, usually exhibits no clear energetic pattern,

is composed of many minima, and unsurprisingly is NP complete. In the case where researchers re-engineering mutations into a molecule to alter stability, they assume that a few single-point mutations will unlikely alter a structural fold. Prior knowledge regarding a folding route of a sequence, or parts of a sequence, can be used to assist in conformational searchers and structure prediction.

Recent developments in computer power and modelling approaches have started to pave the way for deciphering amyloid structure and characterizing fibril morphology. Algorithms to search structural energy landscapes, predict folding patterns, and screen millions of fibril inhibitors are providing advances in the understanding of fibril formation and in the prevention of disease. In this thesis, we build on these developments and explore novel techniques to accurately model fibril polymorphisms and morphology. We use the models to study the impact of mutations on fibril stability, amyloidogenicity, and therapeutic capacity.

### **1.5 Thesis roadmap**

This chapter introduced the biological background and clinical importance of amyloid proteins, and discussed the current computational methods available to study them. The following four chapters describe novel computational methods in modeling amyloid fibrils and in finding ways to destabilize their structures. Each chapter presents a specific research project, with the addition of a Preface section that motivates the research question addressed in it. Chapter 2 introduces a tool that accurately builds amyloid fibril models by assessing structural and landscape energetics. Chapter 3 describes a method to analyze an amyloid structure and infer

optimal single-point mutations that weaken amyloid fibrils. The tool from Chapter 2 is used to build the mutated amyloid fibrils and assess their stability. In Chapter 4, we use the techniques developed in the previous chapter to characterize the complete mutational landscape of the amylin amyloid to find mutations that improve the efficacy of the current Pramlintide drug in diabetes. Chapter 5 introduces a mathematical model relating amyloid production to the emergence and progression of diabetes and establishes a novel computational method to reduce the toxicity of amyloids in the disease. Finally, Chapter 6 summarizes these research contributions and presents discussions on future works.

## 1.6 Publications and author contributions

This thesis comprises the full text and figures of four scientific articles, one of which has been published and three have been submitted for publication. These articles are listed below in the order they appear in this thesis. I am the first author of each one of them.

- **Chapter 2:** M. R. Smaoui, F. Poitevin, M. Delarue, P. Koehl, H. Orland, and J. Waldispühl. “Computational Assembly of Polymorphic Amyloid Fibrils Reveals Stable Aggregates”. *Biophysical Journal*, vol. 104, pp. 683–693, 2013

*In this work, I designed and implemented the CreateFibril tool, the stability landscape method, performed all experiments, and wrote the paper under the supervision of Dr. Jérôme Waldispühl and Dr. Henri Orland. Dr. Jérôme Waldispühl and Dr. Henri Orland and I designed the study. Marc Delarue,*

*Frederic Poitevin and Patrice Koehl updated their AQUASOL framework to work with proteins. Finally, Frederic Poitevin helped in generating the results for Figure 5 and wrote the “Dipolar water model” paragraphs in the methods sections.*

- **Chapter 3:** M. R. Smaoui and J. Waldispühl. “Computational re-engineering of Amylin sequence with reduced amyloidogenic potential”. *Submitted for review in June 2014.*

*Dr. Jérôme Waldispühl and I designed the study for this manuscript. I performed all experiments, designed and implemented the computational tool, and wrote the paper.*

- **Chapter 4:** M. R. Smaoui and J. Waldispühl. “Complete characterization of the mutation landscape reveals the effect on amylin stability and amyloidogenicity”. *Submitted for review in Sept 2014.*

*Dr. Jérôme Waldispühl and I designed the study for this manuscript. I performed all experiments, designed and implemented the computational tool, and wrote the paper.*

- **Chapter 5:** M. R. Smaoui, H. Orland and J. Waldispühl. “Probing the binding affinity of amyloids to reduce toxicity of oligomers in diabetes”. *Submitted*

*for review in Sept 2014.*

*Dr. Jérôme Waldispühl and I designed the study for this manuscript. I performed all experiments, designed and implemented the computational tool, and wrote the paper. Dr. Henri Orland wrote the “Oligomer concentration” part in the methods section.*

## CHAPTER 2

### Computational assembly of polymorphic amyloid fibrils reveals stable aggregates

#### 2.1 Preface

We discussed in Chapter 1 the difficulty faced by experimental scientists to decode the shapes of amyloid aggregate structures. The shapes of many amyloid aggregates are believed to be polymorphic and highly dependent on environment conditions such as temperature, pressure, and pH. Amyloids form in various sizes, shapes, and structural forms. The  $A\beta$  fibrils in Alzheimer's disease form in fibril pairs [101, 35] and triplets [48, 102]. The HET-s amyloids form single and triple fibrils depending the pH of their surrounding system [37]. The different amyloid structures and shapes influence the infectivity of fibrils in various diseases [103].

Although fibrils share many similar structural characteristics including beta-strands that run perpendicular to a main fibril axis [23, 24], the insolubility and large sizes of these aggregates have hindered our understanding of their interaction dynamics and their detailed compositions. The few experimental successes in capturing the structures of amyloid fibrils include identifying short segments of the HET-s fibril [33], determining an incomplete structure for the  $A\beta$  amyloid [32], and a computationally assisted reconstruction of amylin from biological data [34]. These results

are a good start and could be pieced together to reveal structural information about their higher-order polymorphic fibrils.

Predicting how an amyloid protein will aggregate and what polymorphic shape it would assume is a challenging task given the sparse experimental data on amyloid proteins. The idea behind this chapter is to introduce a computational method to tackle this challenge and predict full atomistic models of amyloid aggregates in various polymorphic forms. Many steps were needed to accomplish this. First, we surveyed the literature on amyloid fibrils and gathered all experimental figures and illustrations showing low-resolution visualizations of fibrils. With this data, we identified fibril aggregation patterns and constructed a classification of possible fibril polymorphic assemblies. Second, we built a tool that takes a single amyloid protein (PDB file) and a set of structural parameters to construct the polymorphic assemblies outlined in our classification. The assemblies were constructed using transformation matrices and the careful design of a fibril axis. The transformation matrices moved 3-dimensional amyloid protein structures in space and aggregated them into fibril structures. To address the complexity of determining which parameters produce a reasonable structure and to reduce the search space for various degrees of freedom involved in constructing polymorphic fibrils, we developed the stability landscape method as a quick brute force approach to analyze energies returned by a selection of parameters. Finally, after successfully building a set of polymorphic fibrils for a single amyloid protein, determining the likelihood of each polymorph to occur in nature was not straightforward. This is only possible with the use of an energy function that accurately and quickly computes the free energies of the large fibril

structures. We expanded the function of the AQUASOL [89] tool to compute the free energies of proteins and large assemblies. AQUASOL computationally validated experimental results for amyloids and helped us determine which polymorphs would potentially occur for a certain protein.

Decoding amyloid and fibril structures are important for therapeutic engineering of molecules to inhibit their formation or alter environment conditions to weaken and hinder their toxicity. In this work, our aim is to build a computational framework that creates polymorphic fibril ensemble structures for a given amyloid protein and to test the likelihood of each polymorph to occur. We validated our approach with experimental data for the HET-s and A $\beta$  proteins and predicted fibril results for amylin. Experimental data supported our predictions for amylin fibrils [104].

The remaining content of this chapter is reprinted with permission from:

- M. R. Smaoui, F. Poitevin, M. Delarue, P. Koehl, H. Orland, and J. Waldispühl. “Computational Assembly of Polymorphic Amyloid Fibrils Reveals Stable Aggregates”. *Biophysical Journal*, vol. 104, pp. 683-693, 2013

Copyright (2013) Biophysical Society.

## 2.2 Abstract

Amyloid proteins aggregate into polymorphic fibrils that damage tissues of the brain, nerves, and heart. Experimental and computational studies investigated the structural basis and the nucleation of short fibrils, but the prediction and the precise quantification of the stability of larger aggregates remain elusive. We established a

complete classification of fibril shapes and developed a tool, called CreateFibril, to automatically build these complex polymorphic modular structures. Stability landscapes, a technique we introduced to reveal reliable fibril structural parameters, was applied to assess fibril stability. CreateFibril constructed HET-s, Abeta, and Amylin fibrils up to 17nm in length, and utilized a novel dipolar solvent model that captured the effect of dipole-dipole interactions between water and very large molecular systems to assess their aqueous stability. Our results validated experimental data for HET-s and Abeta, and suggested novel findings for Amylin. In particular, we predicted the correct structural parameters for the one and three predominant HET-s protofilaments. We revealed and structurally characterized all known Abeta polymorphic fibrils including structures recently classified as wrapped fibrils. Finally, we elucidated the predominant Amylin fibrils and asserted that native Amylin is more stable than its amyloid form. CreateFibril and a database of all stable polymorphic fibril models we tested along with their structural energy landscapes are available on <http://amyloid.cs.mcgill.ca>.

### **2.3 Introduction**

Amyloid proteins are believed to be associated either in partial causality or complete aggravation with the severity of neurodegenerative diseases such as Alzheimer's, Parkinson's, Huntington, and Type II diabetes [30, 1]. These mis-folded proteins form stable aggregates, known as fibrils, that damage tissues of the brain, nerves, and heart, leading to symptoms of severe memory loss, deterioration of cortical neurons, fatigue, muscular rigidity, and depression [22, 2, 3, 4, 5, 6]. Apart from their

common central cross- $\beta$  spinal core [105], fibrils assemble in polymorphic structures and pack in several orientations giving rise to different biological functions [35, 36] and toxicity levels in cells [37, 38]. Furthermore, studies showed that fibrils are capable of propagating their specific polymorphisms to daughter fibrils [39, 40] to preserve their destructive potential. Moreover, it is hypothesized that proteins of all kinds can self-assemble into amyloid fibrils under optimum conditions [106]. Some of the best known fibrils have been observed in cryo-TEM reconstruction of insulin fibrils [41], TEM analysis of Amylin [42], cryo-EM analysis of HET-s [37], and TEM analysis of Abeta [43].

Recent experimental studies enabled to describe with high atomic resolution the molecular structures of HET-s and Abeta fibrils [33, 48], but to date, computational modelling and simulation studies are limited to the analysis of the nucleation phase that involve only a few polypeptide chains [107, 108, 109]. It follows that the prediction and the understanding of the aggregation process remains obscure. More importantly, although the importance of water in forming and stabilizing fibrils has been widely acknowledged, current simulation frameworks are restricted to implicit solvent models with limited performance. Computational simulation of amyloid fibrils is facing simultaneously scalability and complexity challenges. The modelling and precise quantification of the stability of the molecular structure of polymorphic amyloid fibrils are key to understanding the toxicity potential and the self-propagation mechanisms of these proteins.

In this paper, we introduce CreateFibril, a computational framework to build polymorphic fibrils of amyloid proteins and explore their stability by means of stability landscapes. We developed these landscapes to reveal reliable fibril structural parameters and assist CreateFibril in building realistic structures. Upon creating fibrils, their structural stability in water is assessed by a novel dipolar solvent model that captures the effect of dipole-dipole interactions and computes the hydration shell that forms around proteins – an insight that cannot be reached with implicit solvent models. More specifically, we expanded the AQUASOL framework [89, 110] to compute the solvation, Coulomb, and Van der Waals energies of molecular systems with up to 36,180 atoms. CreateFibril explored the architectural landscape of HET-s, Abeta, and Amylin proteins and captured with great accuracy the properties of helical pitch, packing distance for multi-meric polymorphs, and hydrogen bond distance on  $\beta$ -sheets between amyloid monomers as they aggregated. We predicted correct structural parameters for HET-s fibrils, revealed and characterized all Abeta fibril polymorphs, and promoted a fibril model for the most common Amylin polymorph. We discovered that native Amylin is more stable than its amyloid form.

## 2.4 Materials and Methods

We aimed to analyze the stability and dynamics of large polymorphic amyloid fibrils using force field calculations. 3D models of these fibril structures were required for analysis but unfortunately, all protein databases only provide very short fibril segments of a few monomers in length. Thus, we designed an automated tool to bridge this gap and create longer fibril models from amyloid fragments and simulate

the structure of various polymorphic fibrils. The tool resorted to the computational technique of Rigid Affine Transformations [111] to construct the fibril models. Fig. 2–2 presents a flowchart of CreateFibril’s pipeline that we describe next.

### 2.4.1 Construction & classification of polymorphic fibrils

Our automated tool, called CreateFibril, produced an ensemble of stable polymorphic fibril structures from a monomer amyloid in the PDB [112]. Single fibrils (Fig. 2–1) were constructed by assembling copies of a monomer amyloid side by side to mimic the oligomerization result and elongation of fibrils. This assemblage was stabilized by the alignment of the monomers’ hydrogen bonds and  $\beta$ -sheets that twist around a helical fibril axis, emulating the natural assembly of amyloids [113]. Higher-order structures (Rings, Stacks, and Polygons) included harmonic combinations of Single fibrils packed in different orientations and distances. The key to create good structures was to pick “parameter” values that build architecturally stable fibrils. The following are the “parameters” our tool provides to create fibrils:

1. Protein PDB file of an amyloid fragment
2. Fibril class: Ring, Polygon, or Stack
3. Number of filaments and their packing distance perpendicular to fibril axis
4. Fibril axis location and direction
5. Rotation angle of amyloid monomers along the fibril axis
6. Hydrogen bond distance of  $\beta$ -sheets along the fibril axis
7. Length of fibril

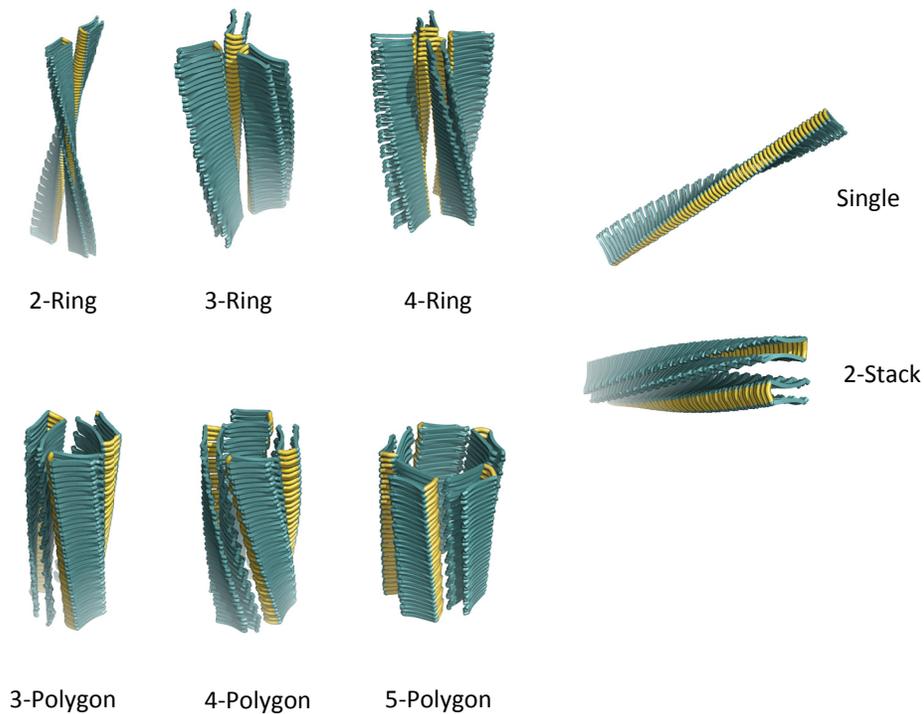


Figure 2–1: Classification of the polymorphic fibril structures produced by CreateFibril. Ring fibrils pack and join at turns, Polygon fibrils join to create a polygon shaped core, and Stack fibrils pack laterally.

The algorithm that builds polymorphic amyloid fibrils performs numerous Translation and Rotation Affine Transformations and is explained in detail in the SI Materials and Methods. Although the algorithm may seem complex, CreateFibril’s interface is built to be user-friendly and intuitive. Moreover, CreateFibril is the only automated tool for building fibrils to date.

### 2.4.2 Energy Minimization to filter fibril structures

It is expected that only a specific set of “parameter” values would create realistic fibril models. A certain choice of “parameter” values could assemble monomers too close to one another or too far from one another, while another choice might assemble monomers that rotate abruptly with respect to the main fibril axis. We were interested in finding “parameter” values that CreateFibril could use to create stable conformations. We exhaustively built structures with reasonable distances and angles to search proper “parameter” values for HET-s, Abeta, and Amylin. An energy function,  $E_{D,\theta}$  based on light Energy Minimization runs and stability landscape results for each “parameter” value guided the search. “Parameter” values that returned the most stable structures were returned for each fibril type ( $n$ -Polygon,  $n$ -Ring, and  $n$ -Stack) and were kept for further analysis by the dipolar water model.

Light Energy Minimization was performed upon creating a structure for two reasons. First, we performed minimization to relieve the system from any clashes that could have occurred during the fibril elongation process due to our rigid placement of monomers and second, we used the minimization data to gather stability information of the structure. Through minimization we calculated the LJ Potential and Coulomb Forces of the structure before, during and after minimization and used them to construct stability landscapes and assess the quality of the “parameter” values we started with. Structures that drifted significantly from their initial configuration and contained high LJ and Coulomb energies were not believed to live in local minima, and hence the initial choice of “parameter” values do not produce stable conformations.

To perform Energy Minimization and potential calculations, we used the GRO-MACS [91] Molecular Dynamics and Energy Minimization package with the following parameters: CHARMM force field and the SPC model [114], a TIP3P box [115, 116] with a minimum distance of 15 Å from any edge of the box to any amyloid atom. The box was kept empty of water and simulations were done in vacuum for two reasons. First, water molecules tend to restrict significant conformational change of unstable fibrils; they added noise to our calculations by stabilizing all systems. Second, minimization to pinpoint the unstable structures occurred orders of magnitude faster without explicit water molecules in the simulation box. We were not interested in a full MD run, hence water was not of utter importance here. The systems were energy minimized for 1000 steepest gradient descent steps with an energy step of 0.01. Long-range electrostatic interactions were calculated using the particle mesh Ewald with a cutoff of 1.0 Å for all simulations [117, 118].

### 2.4.3 Dipolar water model

At this stage we had “parameter” values that structurally encoded an ensemble of the most stable fibril structures of each type of Polygon, Ring, and Stack class. To infer the most stable polymorphs likely to form in nature, we needed to assess the stability of these structures and compare them in water. The main challenge in estimating the energy of large polymers was calculating of the Solvation energy, the energy associated with the formation of a structure in water. For this, we extended the development of a fast and detailed dipolar water model introduced by AQUASOL [89, 110], the first tool to succeed in computing the solvation free energy of molecules

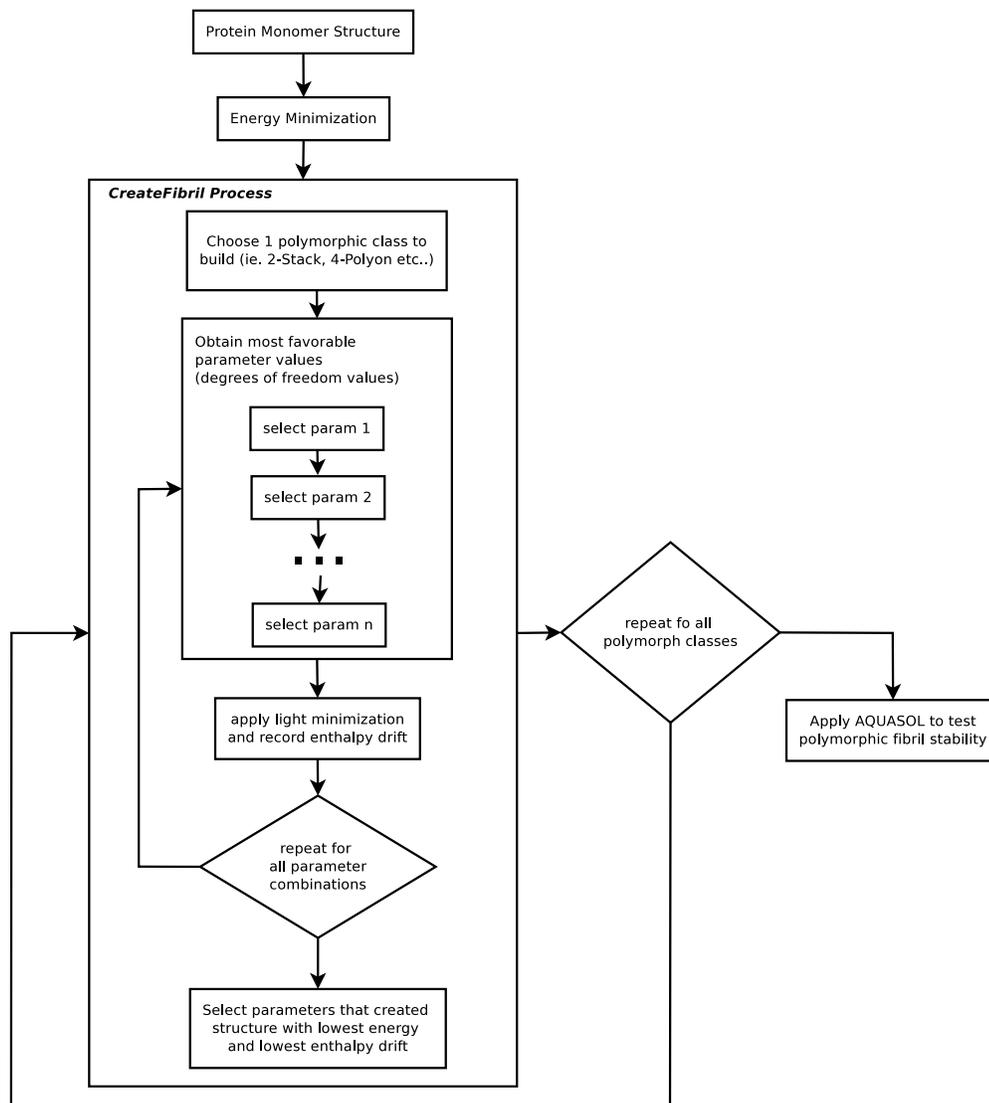


Figure 2–2: CreateFibril procedures and pipeline.

by solving the dipolar non-linear Poisson-Boltzmann-Langevin equation. Our model took into account dipole-dipole interactions and treated water explicitly to accurately calculate the Solvation energy, Coulomb and Van der Waals forces of our fibrils, as seen in Eq. 2.1. CreateFibril used the result of adding these terms as a measure to assess the total stability of the fibrils it created. The lower the sum of these three values, the more stable a fibril is. Our choice of method was orders of magnitude faster than full Molecular Dynamics simulations.

Once a solution was found by AQUASOL, the free energy of the system was computed a posteriori. Nevertheless, a problem inherent to all grid approaches was the emergence of an artifactual term in the energy, the so-called grid energy. In order to correct for this term and get the true free energy  $F_E$ , one needs to compute the free energy of the system with the solvent concentration defined at a given value  $F_{(p_0, C_{dip})}$ , and subtract the free energy of the system with the solvent concentration set to zero  $F_{(0,0)}$ . This leads to minimization of the artifactual term, but also takes out the Coulomb energy of the solute, which needs to be brought back afterwards. Finally, one wants to subtract the solvent energy term  $N_w \mu_w$ , which is linear with the number of water molecules within the grid, and can also be computed analytically. If one is interested in the solvation free energy  $F_{solv}$ , then the Coulomb energy must not be brought back.

The results we obtained compare the stability of the best structures across all polymorphic types we designed. In particular, we compared the solvation energies and the total energies of all polymorphs at different fibril lengths to analyze the stability behaviour of the polymorphs.

$$F_E = F_{solv} + F_{coulomb14} + F_{vdw} + F_{bnd} \quad (2.1)$$

where,

$$\begin{aligned} F_{solv} &= F_{(p_0, C_{dip})} - F_{(0,0)} - N_w \mu_w \\ F_{bnd} &= F_{bond} + F_{angle} + F_{dihed} + F_{imp} \\ \mu_w &= k_B T \frac{\ln(1 - N_A C_{dip} a^3)}{N_A C_{dip} a^3} \\ N_w &= \int_{solvent} d\mathbf{r} \rho_{dip}(\mathbf{r}) \end{aligned}$$

The atomic charges and radii were assigned with PDB2PQR using CHARMM force field at neutral pH. Optimization of Hydrogen-bonding was performed. The following parameters were used in AQUASOL: Grid of 257 points per edge spaced by 1 Å, temperature 300K, surface definition: Solvent Accessible Surface (Rprobe = 1.4 Å), trilinear interpolation protocol for projection of fixed charges on the grid, lattice grid size for the solvent: a = 2.8 Å, solvent made of dipoles of moment  $p_0 = 3.00D$  at a concentration of  $C_{dip} = 55M$ , no salt was added to the solution, electrostatic potential set at zero at the boundaries, and stopping criteria for residual:  $1.10^{-6}$  (when possible).

It is key to note that the solvent and the small ions are correctly treated in all our calculations, both as far as enthalpy and entropy are concerned. Since we use frozen models for proteins, two scenarios arise: First, if the protein is very well

structured, we neglect the small contribution of vibrational entropy, and second, if the protein has unstructured parts such as loops, we neglect the conformational entropy of these unstructured parts. However, as far as we know, there is no rigorous way to take conformational entropy into account other than by performing long MD simulations and thermodynamic integrations, a hopeless approach for the problem we are interested in. It should therefore be checked that the monomers under study are well structured, with minimal unstructured parts.

#### **2.4.4 Implicit water model**

The second method we used to assess the stability of the polymorphs was based on an implicit model of water molecules in solution. We used the tool FAMBE-pH [119] to calculate the Solvation energy and enthalpy of our structures. FAMBE-pH solves the Poisson equation with an optimized fast adaptive multigrid boundary element (FAMBE) method that implicitly models water.

The energy values returned by the implicit model and our explicit model naturally differed in magnitude. The explicit model returned Solvation energy values much higher than the implicit model because it considered dipole-dipole and atomic interactions with explicit water molecules in its calculations. Since the water volume has been kept relatively constant in all the explicit simulations we were able to compare the energies among different fibrils modelled explicitly, and compare their trends with the implicit simulations.

### 2.4.5 CreateFibril: Rigid Affine Transformations

Affine Transformations are represented by mathematical matrices which we use to act on the atom positions of proteins stored in PDB files. The translation matrix is capable of moving atoms, and when applied to a protein's atom positions it produces the effect of moving the protein in 3D space. We represent an atom in the following homogenous representation,

$$a = \begin{bmatrix} a_x \\ a_y \\ a_z \\ 1 \end{bmatrix} \quad (2.2)$$

where  $a_x$ ,  $a_y$ , and  $a_z$ , are how far  $a$  lies in the direction of the x-axis, y-axis, and z-axis, respectively. To move the atom  $a$  by a distance  $d_x$  in the x-axis direction,  $d_y$  in the y-axis direction, and  $d_z$  in the z-axis direction, we apply the following translation matrix  $T(d)$ ,

$$T(d) = \begin{bmatrix} 1 & 0 & 0 & d_x \\ 0 & 1 & 0 & d_y \\ 0 & 0 & 1 & d_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (2.3)$$

onto the homogenous representation of  $a$  in the following way

$$a' = T(d) * a = \begin{bmatrix} a_x + d_x \\ a_y + d_y \\ a_z + d_z \\ 1 \end{bmatrix}. \quad (2.4)$$

Ignoring the 1 at the very bottom of  $a'$ , we observe that the point  $a'$  is in fact a translation of a distance of  $d_x$  in the x-axis direction,  $d_y$  in the y-axis direction, and  $d_z$  in the z-axis direction, as wanted. This way, we can move a whole protein in space a distance  $d$  by applying  $T(d)$  onto all its atom positions. Furthermore, we apply  $T(d)$  to copies of an amyloid protein and force them to aggregate side-by-side and simulate the assembly of fibrils.

The second type of Rigid Affine Transformation matrices we find useful is rotation matrices. We can rotate points in a 3D space around the x-axis, y-axis, or z-axis using standard affine transformation rotation matrices.

For example, to rotate the atom representation,  $a$ , about the x axis by  $\theta$  degrees counterclockwise we multiply it by the following matrix,

$$R_x(\theta) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\theta) & -\sin(\theta) & 0 \\ 0 & \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (2.5)$$

Similarly, to rotate  $a$  about the y-axis by  $\alpha$  degrees counterclockwise we multiply it by the following matrix,

$$R_y(\alpha) = \begin{bmatrix} \cos(\alpha) & 0 & \sin(\alpha) & 0 \\ 0 & 1 & 0 & 0 \\ -\sin(\alpha) & 0 & \cos(\alpha) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (2.6)$$

Also, to rotate  $a$  about the z-axis by  $\beta$  degrees counterclockwise we multiply it by the following matrix,

$$R_z(\beta) = \begin{bmatrix} \cos(\beta) & -\sin(\beta) & 0 & 0 \\ \sin(\beta) & \cos(\beta) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (2.7)$$

We are not constrained to rotations about the x, y, and z axes only. Correctly combining the three rotation matrices above allows to rotate about any arbitrary axis. The power of Affine Transformations lies in the fact that we can combine and chain many transformation operations on a single point, or atom in this case.

The amyloids that make up fibrils have been reported to aggregate together and rotate slightly with every monomer addition about a fibril axis. The translation and rotation matrices applied together enable us to model this fibril aggregation-rotation phenomenon. For our computational framework, we meticulously derive the fibril axis and transformation matrices that build our fibrils next.

#### 2.4.6 CreateFibril: Fibril Axis Matrix

CreateFibril rotates aggregates around a fibril axis by  $\theta$  degrees and moves them along the axis by  $d$  angstroms by computing a transformation matrix  $F_{\theta,d}$  that applies rotation and translation matrices on the amyloid protein models.

The general idea behind constructing  $F_{\theta,d}$  is to first move the protein a distance of  $d$  angstroms along the fibril axis by applying a translation matrix  $T(d)$ . To rotate the protein around the fibril axis, we move the fibril axis to the origin, rotate the

axis so that it aligns with either the x, y, or z axis, apply a standard rotation transformation to it by  $\theta$  degrees with respect to the chosen axis, and then return the fibril axis to its original position in space. This procedure involves a combination of translation and rotation matrices that we believe is worthy of explaining. Without loss of generality, given a point  $p$  lying in the -x, -z, +y direction for which the fibril axis passes through, we translate the system that contains our fibril axis and shifted protein atoms by  $T(-d - p)$ . This operation translates all the objects in our system towards the origin. The fibril axis now passes through the origin and all distances between the atoms and the fibril axis are preserved.

We will choose to align the fibril axis with the x-axis. To do so, we first need to rotate the fibril axis by  $\phi$  degrees to intersect the x-y plane, and then rotate it by  $\psi$  degrees to align it with the x-axis. Therefore, we need to apply the rotation matrix  $R_y(\phi)$  to the fibril axis to intersect the x-y plane followed by the rotation matrix  $R_z(\psi)$  to rotate the fibril axis along the z-axis and align it with the x-axis.

We are now in a position to rotate our protein atoms around the fibril axis by  $\theta$  degrees. We use the standard x-axis rotation transformation since the x-axis and the fibril axis are aligned at this point. We apply the rotation matrix  $R_x(\theta)$  to the fibril axis and the protein. The fibril axis rotates around itself, while the protein atoms rotate around the fibril axis.

To conclude, we need to move our fibril and protein back to their initial positions in space. To undo all axis positioning rotations and translations we rotate the axis back by  $\psi$  degrees with  $R_z(-\psi)$  followed by  $-\phi$  degrees around the y-axis with  $R_y(-\phi)$ . We finally translate the axis and the protein back by  $T(d + p)$ . This puts

the fibril axis exactly back where it was and achieves our desired rotation of the protein atoms by  $\theta$  degrees around the fibril axis. Hence our fibril axis rotation matrix is,

$$F_{\theta,d} = O^{-1}(d, p, \phi, \psi) * R_x(\theta) * O(d, p, \phi, \psi) * T(d) \quad (2.8)$$

where  $O^{-1}$  and  $O$  are defined as,

$$O^{-1}(d, p, \phi, \psi) = T(d + p) * R_y(-\phi) * R_z(-\psi) \quad (2.9)$$

$$O(d, p, \phi, \psi) = R_z(\psi) * R_y(\phi) * T(-d - p). \quad (2.10)$$

The last piece of the puzzle is to explain how to obtain the correct values for  $\phi$  and  $\psi$ . The  $\phi$  angle is the angle needed to rotate the fibril axis to intersect the x-y plane, i.e., the angle between the fibril axis and the x-y plane. To find  $\phi$ , we set  $v$  to be the projection of the fibril axis on to the x-z plane ( $v$  is the fibril axis with the y component set to 0) and we set  $w$  to be a vector on the x axis ( $w$  is the fibril axis with the y and z component set to 0) and make use of the following dot product property to find  $\phi$ ,

$$\phi = \cos^{-1} \frac{v \cdot w}{\|v\| \|w\|}. \quad (2.11)$$

With this  $\phi$  value, we can rotate our fibril axis to fall into the x-y plane and denote it as  $f$ . If we set vector  $l$  to be the projection of the fibril axis in the x-y plane ( $l$  is  $f$  with the y component set to 0) then we can find  $\psi$  by,

$$\psi = \cos^{-1} \frac{l \cdot f}{\|l\| \|f\|}. \quad (2.12)$$

$F_{\theta,d}$  is dynamically calculated in `CreateFibril` whenever a rotation by  $\theta$  and translation by  $d$  is needed. Supplying `CreateFibril` with values for  $d$  and  $\theta$  creates fibrils that elongate at a distance  $d$  apart and rotate an angle of  $\theta$  degrees about the fibril axis.

## 2.5 Results & Discussion

Amyloid fibrils are known to grow by monomer addition [113]. Monomers aggregating at fibril ends usually create a helical structure of a single fibril filament but are capable of creating higher-order fibrils composed of multiple single filaments packed closely by hydrogen bond interactions on  $\beta$ -sheets. These higher-order fibrils can be composed of 2-filaments, 3-filaments or even a larger number of  $n$ -filaments. In designing the different possible geometric forms an  $n$ -filament fibril can assemble in, we observed three main configurations that characterize the possible interactions of fibril filaments. We classify these various shapes in the categories of: Ring Family, Polygon Family, and the Stack Family. Ring fibrils comprise of filaments that pack together at their  $\beta$ -sheet turns, creating a hollow ring shape in the main axis of a fibril. Polygon fibrils comprise of filaments that pack together at their  $\beta$ -strands, creating a hollow  $n$ -polygon shape through the main axis of a fibril. Finally, Stack fibrils comprise of filaments that pack together at their  $\beta$ -strands and pack laterally creating planar fibril sheets. All these kinds of fibrils have been recorded experimentally throughout the years, but a formal classification for fibrils has never been formalized yet. This nomenclature will be used throughout the paper to discuss the

various polymorphic fibrils we observe. A visual representation of this proposed novel classification is provided in Fig. 2–1. This section describes polymorphic preferences predicted by CreateFibril.

### 2.5.1 Stability landscape

We developed a tool, called CreateFibril (see Methods), to build energetically stable fibrils out of single experimentally validated amyloid monomers. We were interested in obtaining structural values of distances and rotations that characterize the architecture of stable fibrils. These structural values constitute the numerous fibril degrees of freedom and would be given as input “parameter” values to our tool to assemble monomers together and construct stable conformations (see Fig. 2–3 for an illustration of these “parameters”). CreateFibril intends to find suitable parameters that characterize stable structures (i.e. structures that would not diverge during MD simulations).

One brute force approach to find suitable “parameter” values is to run Molecular Dynamics (MD) simulations on a starting configuration, attempt to remove any atomic clashes, minimize the energy of the structure and reshape the fibril into a more stable conformation. However, such simulations are computationally expensive in resources and time, prone to numerical imprecisions, and do not guarantee a final stable conformation. For this reason, we introduce a new strategy that searches for suitable “parameter” values by means of generating fibril stability landscapes. Starting with an accurate crystal or NMR amyloid monomer, we first define a range

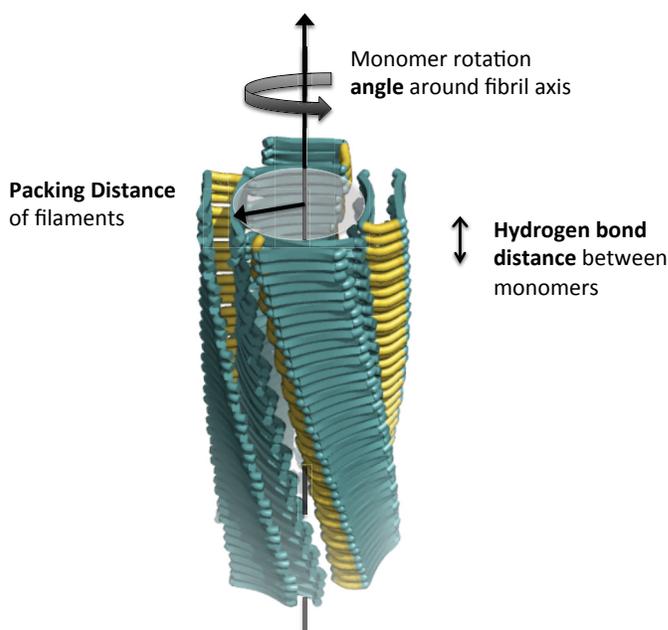


Figure 2–3: The different parameters used by CreateFibril to build structures. Parameters drawn are fibril axis location and direction, rotation angle of amyloid monomers along the fibril axis, hydrogen bond distance of  $\beta$ -sheets along the fibril axis, and packing distance of filaments perpendicular to fibril axis. Parameters not drawn include: protein PDB structure file, fibril class type, and length of fibril.

of naturally possible values for the various fibril degrees of freedom characterized by rotation angles, packing distances and beta strand proximities, as shown in Fig. 2–3. Second, CreateFibril utilizes these range of values to construct all possible fibril structures using rigid affine transformations (see SI Materials and Methods). Third, we perform light runs of Energy Minimization (a few hundred steps) on each generated structure to assess its initial stability sensitivity by calculating any enthalpy drift between final and initial conformations. This step creates the fibril stability landscape by exhausting all suitable “parameter” values. We then search the landscape for values that construct the most stable initial conformation. These

“parameter” values would create structures with lowest enthalpy drift and lowest initial Lennard-Jones and Coulomb terms. Structures with low enthalpy drifts allude to stable conformations (local minima on the structural energy landscape of fibrils), and structures with high energy drifts suggest parameters that produce unstable conformations. We applied this approach to explore the structures of HET-s, Abeta and Amylin fibrils by exhaustively searching the fibril degrees of freedom for values that produce maximum stability. Parameter mining of the three proteins are summarized in Table 2–1. Fig. 2–4 provides an enthalpy drift plot describing the rotation angle vs. beta strand distance landscape for the Single HET-s fibril and packing distance for the 3-Polygon HET-s.

### 2.5.2 HET-s

HET-s (PDB Id 2RNM) is a fungal prion involved in the programmed cell death of filamentous fungi *Podospora anserina*. HET-s fibrils have been studied extensively as a result of their available high resolution atomic structures [33]. CreateFibril explored the stability landscape of HET-s and simulated an ensemble of polymorphic fibril models that confirm the structural properties observed in experimental data.

**Capturing 3-Ring and Single HET-s fibril structures.** Seven different HET-s fibrils were modelled up to 80 nm in length. This large scale modelling of fibrils is the first of its kind providing models on the scale of structures observed experimentally (5 - 10 nm). The enthalpy drift calculations shown in Fig. 2–4 suggest that a left handed swirling orientation of the HET-s amyloids around their fibril axis is more favorable than a right handed twist. The Single and 3-Ring structures are

	Type	Rotation angle (deg)	H bond (Å)	Packing Distance (Å)	Max-length model (Å)	Helical Pitch 360° (Å)		Experimentally observed
						model	literature	
HET-s	Single	22	4.7	NA	800	402	410	Yes [37]
	2-Stack	16	4.7	8	406	656	-	-
	2-Ring	27	4.7	3	412	451	-	-
	3-Ring	17	4.7	7	238	668	1005 - 1083	Yes [37]
	4-Ring	14	4.7	8	199	738	-	-
	3-Polygon	16	4.7	9	290	637	-	-
	4-Polygon	13	4.7	16	203	773	-	-
Abeta	Single	6	4.7	NA	694	925	-	-
	2-Stack	4	4.7	9	373	1268	1140 - 1760	Yes [35]
	2-Ring	4	4.7	7	379	1390	1620 - 2980	Yes [35]
	3-Ring	4	4.7	7	248	1334	-	-
	4-Ring	4	4.7	9	183	1236	-	-
	3-Polygon	2	4.7	16	240	2748	2000 - 2800	Yes [48]
	4-Polygon	3	4.7	21	179	1684	-	-
Amylin	Single	8	5	NA	789	664	242 - 833	†Yes [104]
	2-Stack	9	5	3	396	591	-	* Yes [34, 120]
	2-StackE	-	-	-	244	-	486	* Yes [34]
	2-Ring	-8	5	5	419	624	-	* Yes [52]
	3-Ring	-7	5	8	269	870	-	* Yes [42]
	4-Ring	-3	5	12	195	1090	-	* Yes [42]
	3-Polygon	6	5	16	262	998	-	* Yes [42]
4-Polygon	4	5	28	190	1324	-	* Yes [42]	

\* It is ambiguous as to whether the Ring or Stack structures really form.

† A Single fibril model was built out of partial EPR distance measurements.

Table 2–1: Predicted structural parameters for HET-s, Abeta, and Amylin fibrils produced by CreateFibril. Note that the parameters produce structures that have very similar helical pitches to fibrils seen in nature, hence, the H bond lengths, angles, and packing distances are accurate predictions. Highlighted rows are the polymorphs CreateFibril predicted to form in greatest abundance and possessed the lowest total energies.

well known in the literature as the predominant forms of HET-s [37]. The Single fibrils come together and pack to form 3-Ring structures at pH values lower than 3. From CreateFibril’s structural findings for HET-s in Table 1, we verify that the

HET-s Single fibril helical pitch of 410-Å, beta sheet aggregation of 4.8-Å, and left-handed twist of the fibril reported by [37] all fall in the range of the most energetically favorable structural parameters of HET-s. Furthermore, the 3-Ring structures form with a packing radius of 7 Å and an axial repeat of 66.8 nm. Due to the large size of the HET-s protein and the many steric clashes that form during aggregation, it was expensive to run full EM to relax fibrils in preparation for calculating the total energies of the different polymorphs by our dipolar solvent model; a task not intended for the purposes of this fast method. Fig. 2–6 (a) and (b) show the predominant HET-s fibrils in nature. The Single fibril hides the hydrophobic regions in its core, while the 3-Ring fibril uses the branching residues of each fibril filament to further cover hydrophobic areas.

### 2.5.3 Abeta

$\beta$ -amyloid peptide (Abeta), found excessively in patients with Alzheimer’s disease, is believed to lead to neurodegeneration in humans [44]. This protein aggregates into various fibril shapes that form neuritic plaques and neurofibrillary tangles [45, 46, 47]. Abeta molecules are known to form into the 3-Polygon [48], 2-Stack and 2-Ring polymorphic shapes [49, 35]. CreateFibril structurally characterized Abeta’s polymorphic fibrils and computationally assessed their stability in solution.

**CreateFibril matched the helical pitches of Abeta polymorphs in nature.** Using rigid affine transformations and enthalpy drift measurements, CreateFibril built several potential polymorphs for Abeta. In Table 1, we reported these best structures along with their structural parameters. The helical pitches of the

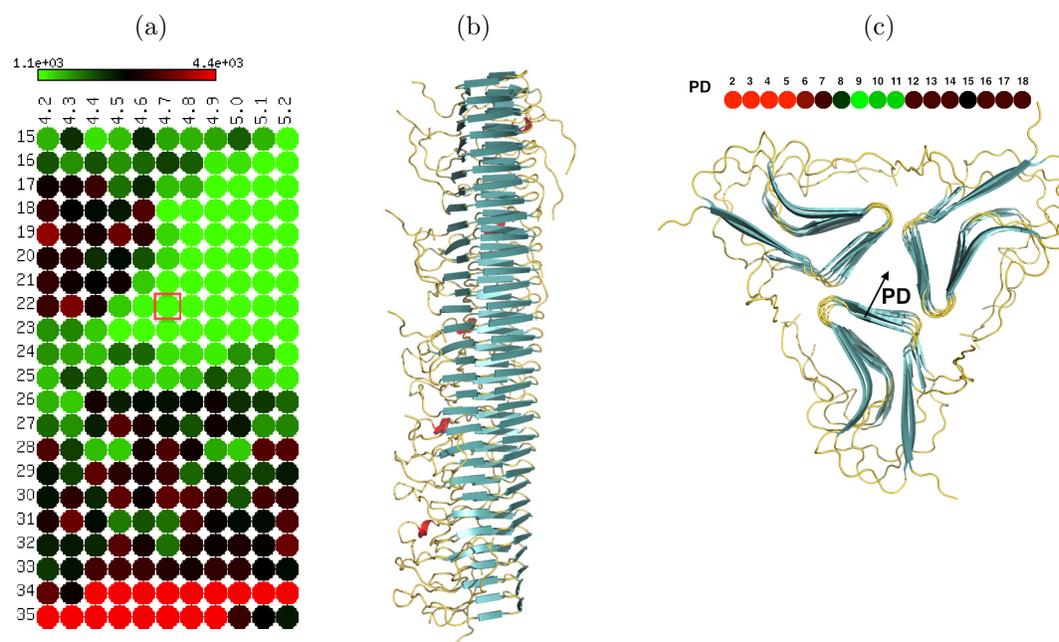


Figure 2-4: HET-s Single fibril parameter findings. (a) Heat map representation of the stability landscape of Single HET-s structures exploring rotational angle  $\theta$  (y-axis) and beta sheet bonding distance  $d$  (x-axis). Green circles indicate stable structures with low enthalpy drift, red circles indicate unstable structures with high enthalpy drift, and black circles are intermediately stable structures. (b) Single HET-s fibril built by CreateFibril with the values of  $\theta = 22$  degrees and  $d = 4.7$  Å taken from the best result in (a) enclosed by the red square. ?? Stability landscape of the Packing Distance (PD) of HET-s 3-Polygon. Energy values are in KJ/mol.

fibrils we modeled were inline with those stated in the literature [35]. Although the structure of Abeta (PDB Id 2BEG) was missing residues 1 - 16, CreateFibril was still able to choose the right rotation angles  $\theta$ , fibril packing distances, and beta sheet aggregation distances,  $d$ , for all polymorphs to reproduce the structures that have been observed in nature (see Table 2-1).

### 2-Stack, 3-Polygon, and Single Abeta polymorphs predicted to form.

Fig. 2-6 (a) describes the effect of water solvent on the aggregation of the Abeta

polymorphs modeled with our dipolar water formalism. Up to length 27 monomers, water favors the formation of the Single Abeta fibril. At length of 27 monomers, the 2-Stack and the 3-Polygon start outperforming the Single in aggregation. This could be explained by the packing of Single fibrils to produce 2-Stack and 3-Polygon structures. In Fig. 2-7 (b), the energies of all polymorphs grow negatively implying favorable aggregation with regards to solvation and enthalpy. At around 32 monomers of length, the figure suggests that the 2-Stack structure is the most stable polymorph in the set, followed by the 3-Polygon, and the Single fibrils, all of which have been observed experimentally. Fig. 2-6 (c) - (e) shows the hydration shell effect on the monomers that make up these fibrils and how aggregation attempts to hide hydrophobic regions. Fig. 2-5 (a) and (b) present the results of the implicit water model in calculating solvation energy and total free energy of Abeta polymorphs. This model proposes the emergence of the 4-Polygon and the 4-Ring fibrils which were never experimentally observed.

**CreateFibril characterized the stability landscape of wrapped Abeta fibrils. Wrapped structures can further stabilize 2-Stack fibrils.** Stroud *et al.* [49] used X-ray powder diffraction to observe that some Abeta fibrils are likely composed of laterally associated fibril filaments that twist around internal helical axes. These internal axes wrap around a common superhelical axis in a geometry that they term wrapping. When a filament is wrapped around a helical axis in this manner, it obtains a twist that is in phase with the fibril helix. Stroud *et al.* showed that higher crossing angles are related to greater curvature and increasingly large holes in fibrils, suggesting that Abeta fibril toxicity may be related to their potential

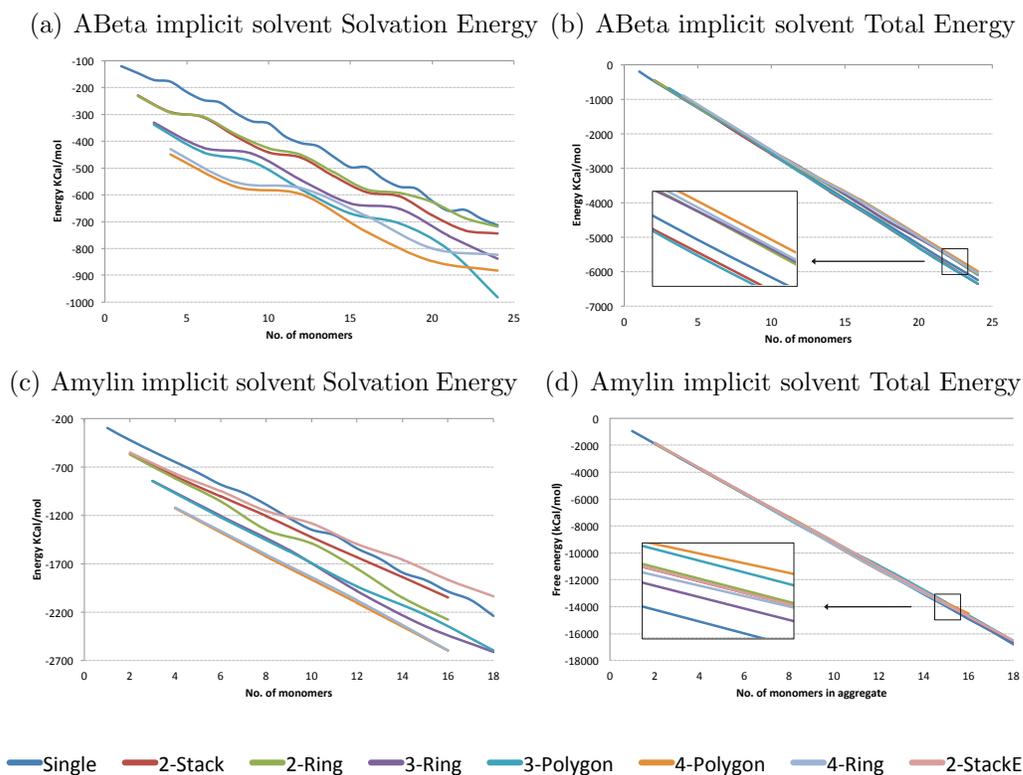


Figure 2-5: Energies of ABeta and Amylin fibrils as they aggregate. (a) and (c) Solvation energy by implicit solvent, (b) and (d) Free energy by implicit solvent.

for forming pores. In Fig. 2-9 (c), we constructed the stability landscape plot for Abeta wrapped fibrils with crossing angles between 0 and 88 degrees and rotation angles (with respect to main fibril axis) between  $-13$  and  $13$  degrees. Wrapped structures we modeled did not contain runaway domain swapping. Two stable wrapped structures suggested by the stability landscape had a crossing angle of 8 and 59, and a rotation angle of  $-3$  and  $-11$ , respectively. In Fig. 2-9 (d) the first wrapped structure obtained a more stable conformation than the 2-Stack model, validating that some 2-Stack Abeta fibrils are indeed wrapped. Fig. 2-8 expands the stability

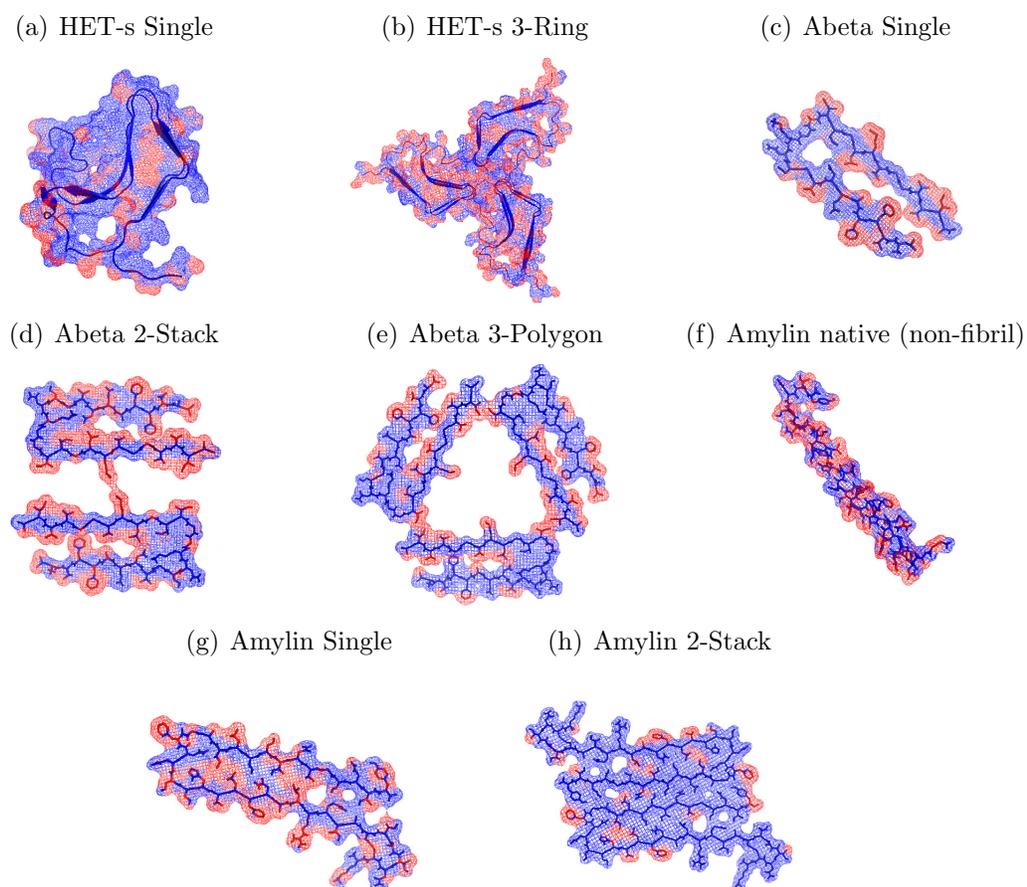


Figure 2-6: Cross-sectional view of the hydration shell effect on the hydrophobicity of the predominant HET-s, Abeta, and Amylin fibrils produced by AQUASOL. Blue regions represent hydrophilic residues while red regions represent hydrophobic residues. HET-s Single fibrils possess a hydrophobic core which they aggregate around (a), and the branching residues of the Single fibrils help hide hydrophobic residues of their neighboring fibril when packed in the 3-Ring structure (b). Abeta 2-Stack and 3-Polygon fibrils aggregate creating a hydrophobic core (d) and (e). The native amylin contains many hydrophobic residues (f) yet possess a lower energy than its amyloid counterpart (g). To hide their hydrophobic residues, amylin amyloids aggregate in the 2-Stack polymorph (h).

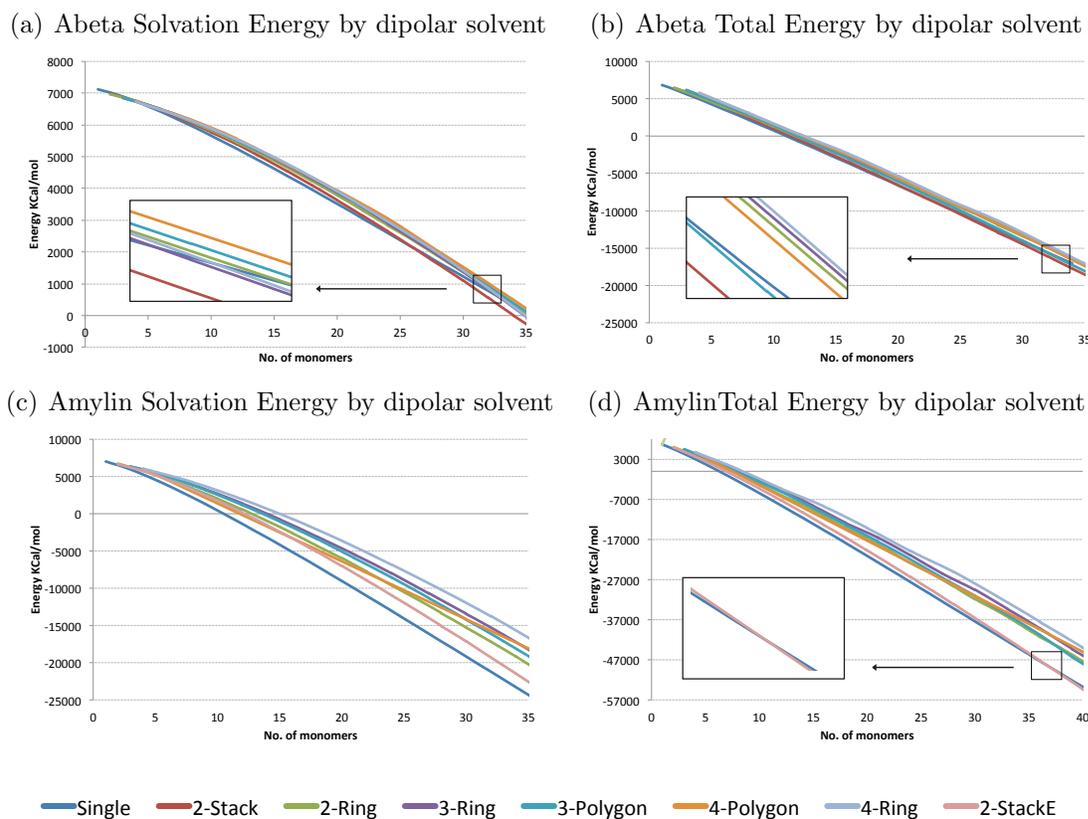


Figure 2-7: Energies of Abeta and Amylin fibrils as they aggregate. (a) and (c) Solvation energy by dipolar solvent, (b) and (d) Free energy by dipolar solvent.

landscape results by showing the enthalpy energies before and after minimization runs.

### 2.5.4 Amylin

Deposits of Islet amyloid polypeptide (Amylin) in the pancreas are toxic and believed to be a contributing factor to Type II diabetes [50, 51]. Amylin fibrils

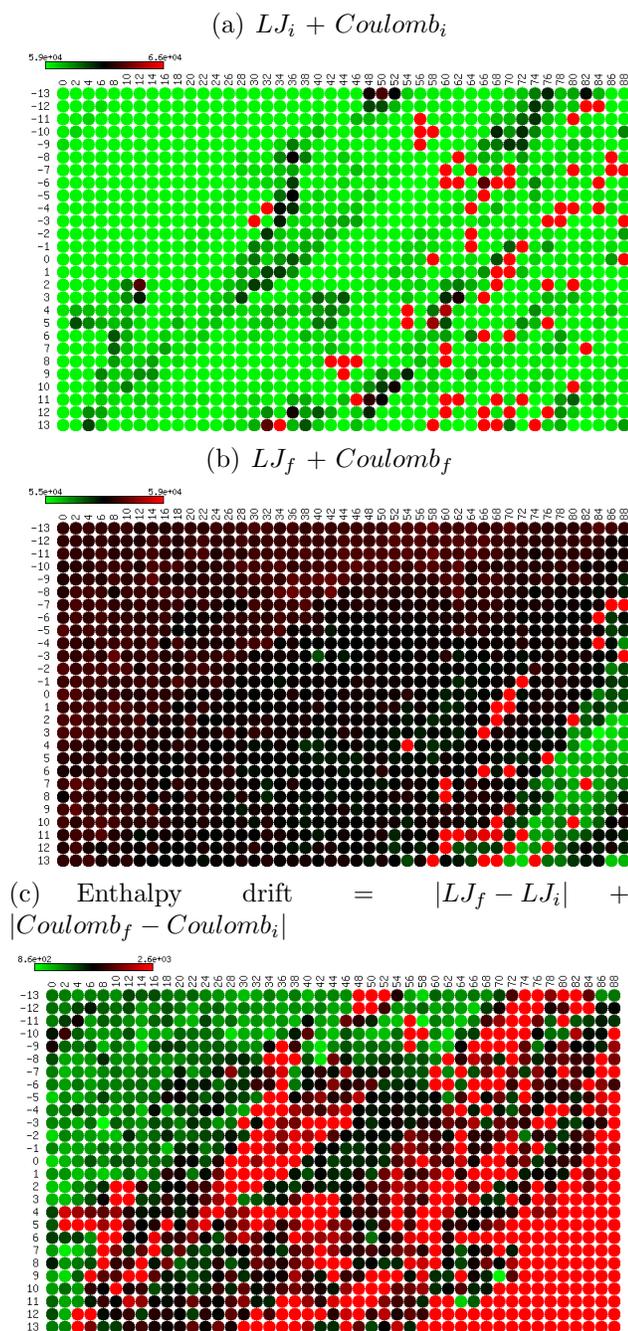


Figure 2–8: Heatmap representation of the stability landscape and the enthalpy drift of Abeta wrapped 2-Stack fibrils. Exploration covered crossing angles between 0 and 88 degrees and fibril rotation angles (main axis) between -13 and 13 degrees. (a) The energies of structures initially built by CreateFibril assembled through rigid affine transformations. (b) The energy plot of each structure after a run of Energy minimization. (c) The energy difference, or enthalpy drift, determines the initial structural stability of fibrils built by CreateFibril in (a). Structures that represent low drifts are considered fixed points and believed to live in local minimum neighbourhood. Energies in KJ/mol.

were successfully polymerized *in vitro* and showed a diverse ensemble of polymorphic shapes [52, 42]. To reconstruct the polymorphs computationally, CreateFibril required the atomic structure of the amyloid form of Amylin. Unfortunately to date, no one has been able to crystallize full-length human Amylin. Instead, many models for the monomeric form of amyloid Amylin have been proposed and among the most prominent structures are ones proposed by Wiltzius *et al.* [34], Luca *et al.* [120], and Bedrood:2012nx *et al.* [104]. The model of Luca *et al.* consists of a full atomic model of a single striated ribbon amylin polymorph based on constraints from solid-state NMR. This work in 2007 opened the doors for further Amylin experimental model development. Interestingly, we have observed that the Amylin structures proposed by Bedrood:2012nx *et al.* resemble the wrapped Abeta fibrils containing swapping runaway domains recently discovered by Stroud *et al.* [49]. Although the model of Bedrood:2012nx *et al.* gives new insight into a possibly novel Amylin conformation depicted by EPR distance measurements, the model was built by only considering a single stack of peptides and ignored additional restrictions that would rise from the packing of multiple fibril filaments around each other. Hence, the model Bedrood:2012nx *et al.* present might be a special form an Amylin monomer can take, and not the predominant conformation. The final prominent atomic model of Amylin proposed by Wiltzius *et al.* [34] was built using biochemical and structural data along with the fibril NNFGAIL and SSTNVG crystallized regions of Amylin to formulate a structure with atomistic details for the protein in the form of a 2-Stack. It is key to start CreateFibril with a very realistic amyloid monomer as this is the

basis for determining polymorphic shapes and predicting accurate likelihoods of formation. Perturbations at the monomer level can alter the fibril pitch and packing distances which might introduce artifacts when calculating likelihoods of formation. It is important to start with an amyloid crystal structure, when possible, or a precise model that is not very different from the actual amyloid ( $\approx$  RMSD 2 Å see Fig. 2–10 for convergence analysis). An RMSD difference of 2 Å is quite large for small monomers such as Abeta and Amylin. MD simulations and minimization techniques should be used to improve model quality when necessary. It is our understanding that this Wiltzius’ model is the best model in the literature regarding Amylin in its monomeric amyloid form and we refer to it as 2-StackE in Fig. 2–7 and in Table 2–1. We extracted one Amylin protein from this fibril model and used it as a starting template to build the other polymorphic fibrils. Having verified that our approach generated valid structural results for Abeta and HET-s and valid solvation results for Abeta, we endeavored to find novel findings for Amylin polymorphs.

**2-Stack and Single Amylin polymorphs have lowest total energies.**

Fig. 2–7 (d) graphs the total energy of the Amylin polymorphs and suggests that Amylin Single and 2-Stack polymorphs compete in solution. The Single structure dominates in lowest total energy until a 35 amyloid fibril is reached, after-which, the 2-Stack polymorph becomes more abundant. This figure also shows that the other polymorphs cluster together and are higher in energy. Assuming that the model we started with is the actual model of Amylin in amyloid form we can suggest that the 2-Ring, 3-Ring, 3-Polygon, 4-Ring and 4-Polygon are unlikely to form. Although Fig. 2–7 (c) suggests that solvation generally prefers the emergence of the Single structure

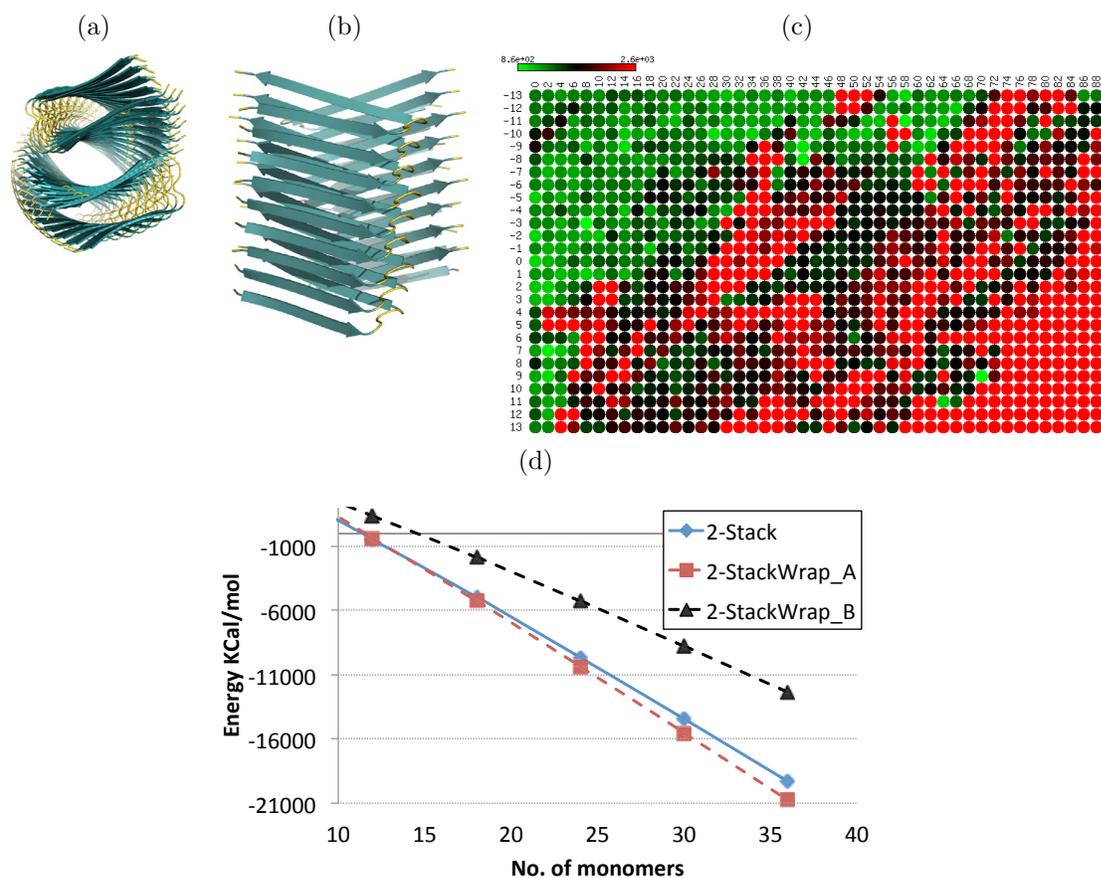


Figure 2-9: Abeta wrapped fibrils. (a) Cross-sectional top view and (b) side view of the wrapped Abeta fibril. (c) Heat map representation of the stability landscape of Abeta wrapped fibrils characterized by rotational angle  $\theta$  (y-axis) and crossing angle  $\omega$  (x-axis). Green circles indicate stable structures with low energy, red circles indicate unstable structures with high energy, and black circles are intermediately stable structures. Energy values are in KJ/mol. (d) 2-StackWrap\_A structure with crossing angle 8 and rotation -3 is more stable than a non-wrapped 2-Stack. 2-StackWrap\_B has a crossing angle 59 and rotation angle -11.

in solution over the 2-Stack, total stability and dominance of structure is determined by a combination of enthalpy and solvation. The dipolar water model in Fig. 2-7 (d) suggests that a 2-Stack, supported in the literature by [34, 120] should exist in

greater abundance and possess the greatest stability among polymorphs. Because Single fibrils initially are more energetically favorable than the 2-Stack and might contribute to the emergence of the 2-Stack, we propose that finding a way to limit the aggregation of the Single polymorph could greatly diminish all the fibrils that form in solution. This would be of critical importance for attempting to find therapeutics to combat the growth of Amylin fibrils in the Beta Cells of the pancreas [51]. Fig. 2-6 (f) - (h) presents the hydration shell effect on the hydrophobicity of Amylin when aggregated in the Single and 2-Stack fibril form and when left in native form. We observe that the 2-Stack model structure hides most hydrophobic regions, perhaps one explanation to its abundance in solution. The implicit water model suggests in Fig. 2-5 (c) that the 2-Stack Amylin structures are high in energy and are not likely to exist for a long period of time, a result that contradicts the experimental finding of Goldsbury *et al.* that claims the dominance of the 2-polymorph Amylin structures *in vitro* [52, 42]. On the other hand, the dipolar water model confirms these findings and furthermore suggests that the dominant 2-polymorph fibrils would take on a 2-Stack conformation and any emerging 3-polymorph fibril would take on a 3-Polygon conformation similar to the 3-Polygon of Abeta.

**Native Amylin predicted to be lowest in energy.** It has been unclear whether amyloid monomers form because they are lower in energy compared to their native counterparts, or because their aggregation produces lower energy structures compared to an accumulation of unbound native proteins. Using our dipolar solvent model, we found that the total energy of one Amylin monomer in native form (PDB ID 2KB8) was 6504.7 KCal/mol and the total energy of one Amylin monomer in

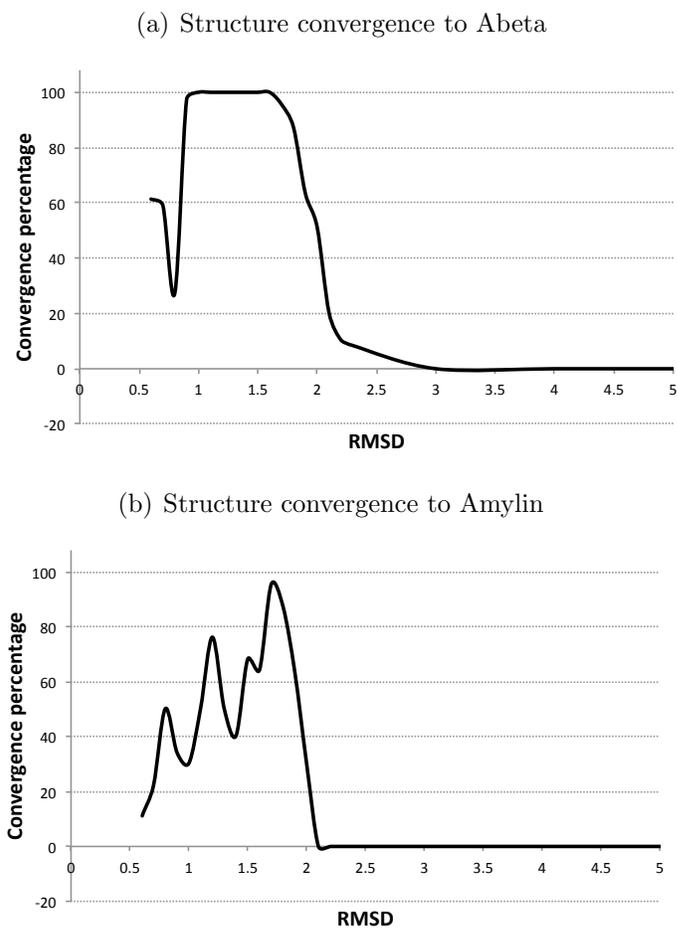


Figure 2–10: Convergence rate of Abeta and Amylin tweaked structures. Thousands of structures were created and tweaked for every RMSD decimal of convergence. MD simulations were performed on tweaked structures with the following parameters: CHARMM force field, SPC model, TIP3P box with a minimum of 15 Å from any edge of the box to any atom, and 20000 integration steps per minimization run following a steepest gradient decent algorithm.

amyloid form was 6660.7 KCal/mol. Assuming we started with an accurate Amylin model, this result reveals that the native form of Amylin is initially favored in nature over the amyloid form. As amyloids find each other and aggregate, the energy of an aggregate structure of  $k$  monomers becomes lower than the energy of  $k$  native molecules floating in solution for all  $k > 2$ . This could explain the rapid aggregation phenomenon of Amylin fibrils.

### 2.5.5 Online tool & Fibril database

In this paper, we have shown how to build amyloid fibrils with various geometries by using rigid affine transformations, starting from the amyloid structure of a single prion form of the protein. The free energy of these structures in water was computed by adding together the Lennard-Jones, Coulomb and solvation energies. The latter is a crucial component in the stability of the fibrils, and has been computed by using the AQUASOL program, a software which yields free energies in good agreement with those computed from long runs of MD. These computed free energies in turn allow to assess the stability of the various proposed structures and to classify their abundance in amyloid solutions. Our results are in very good agreement with current experimental findings, and in some cases predict the existence of stable forms of aggregates which have not yet been observed. Table 1 summarizes the stability landscape exploration parameters used by CreateFibril to build the most stable fibril structures for HET-s, A $\beta$ , and Amylin proteins. We collect all the generated polymorphic fibril structures in an open database at <http://amyloid.cs.mcgill.ca> and

provide CreateFibril as a free online application on the same site.

## **2.6 Acknowledgments**

MRS is supported by a Fellowship of the Canadian Institutes of Health Research System Biology Training program at McGill University. JW is supported by a Discovery grant from the Natural Science and Engineering Research Council of Canada. We thank the French Embassy in Canada for awarding a research travel fund between Montreal and Paris.

## CHAPTER 3

### Computational re-engineering of Amylin sequence with reduced amyloidogenic potential

#### 3.1 Preface

A significant effort in the field of amyloid inhibition has been spent in researching ways to limit the growth of amyloid fibrils, slow their production, and impede their formation. Molecules have been designed to target early oligomer aggregates to prevent amyloid fibril formation [121, 122, 123]. Others have been designed to enhance fibril formation to reduce the build up of oligomers in the cases where oligomers were toxic [124]. Other approaches included capping fibrils with docking proteins [125], inhibiting fibrils by methods of Lysine-specific molecular tweezers [126], and using sulfonated triphenyl methane derivatives as potent inhibitors [127]. Although some of these attempts have demonstrated fibril inhibition and slower reaction rates *in vitro*, the effect of introducing these various molecules on cellular processes, reaction pathways and other proteins is not clear.

We established in Chapter 1 that amino acid mutations affect the stability and initiation processes of amyloids. Single-point mutations are associated with the onset and progression of several diseases, including diabetes. The right mutations can

stabilize, or destabilize, fibril structures affecting their rates of formation, insolubility properties, sizes, and polymorphic assemblies.

The second objective of this thesis is to explore ways to destabilize or inhibit amyloid and fibril formation. Tweaking environment conditions such as pH and temperature around proteins to lower their amyloidogenicity was one of our prime interests. However, we found that 1) changing environment conditions is theoretically possible *in silico* but difficult *in vivo*, 2) tampering with environment conditions in the cell could affect other processes and organelles, 3) modeling environment conditions at an atomic level is computationally difficult and infeasible, and 4) certain environment conditions could be critical for the native structure of the protein to fold. All four factors motivated our idea to destabilize amyloids with sequence mutations that could be engineered into cells.

The tool from the previous chapter was the first of its kind in successfully creating fibril models from a single amyloid PDB structure. One of the main results obtained from the previous work is the prediction of amylin's polymorphic structure and the precise parameter values needed to make up its morphology. This has now enabled us, for the first time, to start exploring fibril dynamics and expound how different amino acid regions contribute to fibril stability.

In this chapter, we describe a protocol to analyze the structure of an amyloid protein and search for regions that could be exploited with sequence mutations to destabilize fibril structures. We apply this protocol to suggest mutations that potentially weaken the diabetes-related amylin 2-Stack fibril. In particular, we identify six key regions that contribute to amylin's fibril stability and use the tools from the

previous chapter to study the effect of single-point mutations on amylin nucleation and aggregation. We packaged the algorithm that assess the stability of regions into a tool called FibrilMutant. The goal behind this work is to explore impeding aggregation without stressing the cellular environment.

The remaining content of this chapter has been taken from a manuscript that was submitted for publication:

- M. R. Smaoui and J. Waldispühl. “Computational re-engineering of Amylin sequence with reduced amyloidogenic potential”. *Submitted for review in June 2014*.

### **3.2 Abstract**

The aggregation of amyloid proteins into fibrils is associated with neurodegenerative diseases such as Alzheimer’s and Type II Diabetes. Different methods have explored ways to impede and inhibit amyloid aggregation. Most attempts in the literature involve applying stress to the environment around amyloids. Varying pH levels, modifying temperature, applying pressure through protein crowding and ligand docking are classical examples of these methods. However, environmental stress usually affects molecular pathways and protein functions in the cell and is challenging to construct in vivo. In this paper, we explore destabilizing amyloid proteins through the manipulation of genetic code to create beneficial substitute molecules for patients with certain deficiencies. To unravel sequence mutations that destabilize amyloid fibrils yet simultaneously conserve native fold, we analyze the structural landscape of amyloid proteins and search for potential areas that could be exploited to weaken

aggregation. Our tool, FibrilMutant, analyzes these regions and studies the effect of amino acid point mutations on nucleation and aggregation. This multiple objective approach impedes aggregation without stressing the cellular environment. We identified six main regions in amyloid proteins that contribute to structural stability and generated amino acid mutations to destabilize those regions. Full length fibrils were built from the mutated amyloid monomers and a dipolar-solvent model capturing the effect of dipole-dipole interactions between water and very large molecular systems to assess their aqueous stability was used to generate energy plots. Our results are in agreement with experimental studies and suggest novel targeted single point mutations in Amylin, potentially creating a better therapeutic agent than the currently administered Pramlintide drug for Type I Diabetes patients.

### **3.3 Introduction**

Protein misfolding has been regarded as one of the most important events triggering a wide variety of neurodegenerative and systemic diseases including Alzheimer's, Parkinson's, Prion disease, and Type II Diabetes [128, 27, 30]. The misfolding of certain critical soluble proteins introduces conformational changes that favor aggregation and the creation of highly ordered beta sheet rich insoluble polymers [129, 130]. These structures, often referred to as amyloids in their monomeric form, or amyloid fibrils in their long aggregated form, have been observed to accumulate in the brain, heart, pancreas, and other organs. They are believed to contribute to many health problems including memory loss, brain lesions, senile plaques, synaptic spine loss, neurotic dystrophy, and cell death [22, 131].

Considerable amount of work has been spent into researching ways to limit the growth of amyloid fibrils, slow their production, and inhibit their formation. Molecules have been designed to target early oligomer aggregates to prevent amyloid fibril formation [121, 122, 123]. Interestingly, others have been designed to enhance fibril formation to reduce the build up of oligomers in the cases where oligomers were toxic [124]. Other approaches included capping fibrils with docking proteins [125], inhibiting fibrils by methods of Lysine-specific molecular tweezers [126], and using sulfonated triphenyl methane derivatives as potent inhibitors [127]. Although some of these attempts have demonstrated fibril inhibition and slower reaction rates in vitro, the effect of introducing these various molecules on cellular processes, reaction pathways and other proteins is not clear and can be unfavorable.

Amyloids have been observed to undergo mutations that change their amyloido-genicity and rate formation. Several cases of Parkinson's disease are associated with amino acid mutations of the alpha-synuclein ( $\alpha$ S) protein [58, 59, 60]. The A30P mutation in  $\alpha$ S decreases the overall rate of fibril formation [61, 62], while the H50Q, H50A, and G51D mutants aggregate more quickly than the wild type but more slowly than A53T and E46K mutants [63]. One point mutations have been observed to be sufficient to affect the landscape of the  $A\beta_{42}$  protein in Alzheimer's and change the internal dynamics between microstates [64]. R5A mutation studies showed a decrease both in the tendency towards  $A\beta$  aggregate formation and a reduced toxicity in Alzheimer's [65]. A mutation in amino acid position 25 of  $A\beta$ , the loop area connecting two beta strands has been show to destabilize  $A\beta$  fibrils [66]. Furthermore, a single mutation of serine-to-glycine at position 20 in Amylin in Chinese and

Japanese populations [67] is associated with early onset of Type II Diabetes [68, 69] and amplified amyloid formation [70, 71, 72]. Moreover, more than 120 single point mutations have been associated with the systemic disorder FAP [77].

As a result of the heavily observed natural mutations, research into deliberately mutating amino acids of amyloids has been proposed as an explorative method to destabilize fibrils and reduce toxicity. Computational methods coupled and steered with Molecular Dynamics runs have proved to be a viable strategy to study the impact of mutations, however, unscalable and expensive in resources and time [78]. Advancements in understanding the effects of amyloid sequence mutations on fibril toxicity and formation rate has paved the way for the development of therapeutic agents to replace highly amyloidogenic species. Re-engineering the genetic code of proteins and administering them as substitute agents for patients is a promising strategy for drug development. Pramlintide, a mutated protein version of Amylin, is used as a drug replacement in Type I Diabetes and has been shown to produce less fibrils and cause less beta cell death in the pancreas [79]. It is still unclear how point mutations alter the pathway of oligomerization and the kinetics of fibril conformational transitions [64], however, it is clear that proteins aggregate through nucleation-dependent polymerization [22, 63]. Hence, exploring mutations that affect the nucleation of amyloid monomers has the potential to introduce more therapeutic agents to inhibit oligomer formation and reduce the effect of disease.

In this paper we describe a protocol to analyze the structure of an amyloid protein and search for regions and residues that could be exploited to weaken aggregation. We focus our study on the diabetes-related protein, Amylin, and explore

six key regions potentially contributing to amyloid oligomerization and fibril production. We developed a tool, FibrilMutant, that implements this protocol and suggests several amino acid mutations that weaken fibril structure. Subsequently, we address the multi-objective problem of discovering point mutations that destabilize fibrils yet conserve the native fold in an efficient manner by running short Molecular Dynamics simulations on single mutated native Amylin proteins to detect any initial structural turbulence and utilizing a dipolar solvent model to assess the fibril aqueous stability of much larger mutated amyloid aggregates. Our approach generated several destabilizing mutations that respect the multiple objectives. Oligomers and fibrils were built from each mutation and assessed for structural stability with an energy function that takes into account solvation energy, hydrophobicity, electrostatic interactions, and hydrogen bonding. We validate our method with results of mutations determined experimentally for Amylin and suggest new mutations that show stronger amyloid destabilizing potential than the current best therapeutic agent for diabetes.

### **3.4 Materials and Methods**

We created FibrilMutant to explore the effect of sequence mutations on destabilizing amyloid fibrils. FibrilMutant takes a PDB file of a single amyloid monomer and a PDB file of its native protein conformation as input and generates single point mutations to destabilize the protein's fibril structure. We apply the mutations to both the native and the amyloid form of the protein and calculate the stability effect on both forms. The following procedure is followed to generate and assess the mutations:

1. Load the PDB file into FibrilMutant and analyze key structural characteristics contributing to aggregation.
2. Explore a set of mutations that could weaken regions contributing to aggregation.
3. Generate fibril stability landscapes to find the most stable fibril polymorph of the amyloid PDB monomer.
4. Build mutated fibrils of the most stable polymorph.
5. Assess the stability of each mutated fibril in water with a dipolar water model.
6. Discard any mutation that stabilizes the amyloid fibrils, and verify with short full Molecular Dynamics (MD) simulations that the final mutation list does not introduce structural lumps or turbulences that destabilize the native protein.

FibrilMutant builds on our recent work that enabled the simulation of accurate fibril models [132]. Its core development includes predicting a set of effective mutations and building oligomers and large fibrils to test the effects of these mutations on structural stability. A detailed description of the procedure is described below.

### **Step 1: Analyzing amyloid structures**

Amyloids share key structural similarities. FibrilMutant extracts from a PDB file regions with Beta strands, screens regions at Beta turns, identifies salt bridges, and examines hydrophobic, polar and charged residues. This collected data is then utilized to generate amino acid mutations in the extracted key regions and residues.

### **Step 2: Generating mutations**

Structural data collected from an amyloid PDB file is used to predict destabilizing mutations. The inner core of amyloids is known to be a hydrophobic core. One

way to disrupt this core is by introducing a mutation of one of its amino acids into a charged residue. We have explored several ways mutations can weaken the structure stability of amyloids. Upon selection of potential mutations, we run TANGO [98] to quickly rank candidate mutations and provide preliminary data on the potential destabilizing effect of each mutation. We understand that TANGO uses a coarse-grained model to perform high-throughput screenings of results, hence we only use it as a guide to initially rank the numerous mutations. Table 3–1 summarizes the method behind choosing these mutations.

### **Step 3: Fibril stability landscape**

Amylin has recently been found to form into a fibril structure composed of two stacked protofibrils [34]. However, the experimental structural parameters, mainly fibril rotation angle and protofibril packing distance, have not been published yet. MD simulations could be used to find these parameters but the process is computationally expensive in resources and time, and we estimated these by our previous work on Stability Landscapes [132].

Starting with an accurate crystal or NMR amyloid monomer, we first define a range of naturally possible values for the various fibril degrees of freedom characterized by rotation angles, packing distances and beta strand proximities. Second, we utilize these range of values to construct all possible fibril structures using rigid affine transformations. Third, we perform light runs of Energy Minimization (a few hundred steps of relaxing the protein structure and removing close clashing atoms) on each generated structure to assess its initial stability sensitivity by calculating any enthalpy drift between final and initial conformations. This step creates the fibril

Structure Characteristic	Contribution to Amyloid Stability	Disruption Method
Hydrophobic Core	Hides core residues from water and generates a packed core	Mutate a hydrophobic residue in the core into a charged one
Hydrophilic Surface	Provides a stable contact surface to water	Mutate a polar residue on the surface into a hydrophobic one
Beta Sheets	Constitute the backbone of fibrils	Decrease the number of hydrogen bonds between Beta strands
Beta Turns	Provide needed torsional flexibility for Beta sheets to form	Mutate the center residue and any Glycine amino acid of a Beta turn region into a Proline to limit torsional flexibility
Salt Bridges	Produce an ionic bond between fibril monomers or the monomer itself	Search the amyloid structure for bonds less than 4.5Å apart bonding the following pair of amino acids: ASP - LYS, ASP - ARG, GLU - LYS, GLU - ARG, and mutate one amino acid into a non charged, non polar residue to break the ionic bond.
Polar Regions	Contribute hydrogen bonds	Mutate polar residues into non polar ones to weaken hydrogen bonds

Table 3–1: Effect of mutation choice on structural stability. This table summarizes our approach to choosing mutations to test for fibril destabilization. We identify six main features of amyloids that contribute to structure stability and outline the methods we used to weaken their contribution to the amyloidogenicity of proteins.

stability landscape by exhausting all suitable “parameter” values. We then search the landscape for values that construct the most stable initial conformation. These “parameter” values would create structures with lowest enthalpy drift and lowest initial Lennard-Jones and Coulomb terms. Structures with low enthalpy drifts allude to stable conformations (local minima on the structural energy landscape of fibrils), and structures with high energy drifts suggest parameters that produce unstable conformations.

#### Step 4: Building fibril models

A different amyloid monomer is generated for every mutation. To test the stability effect of the mutations on fibrils, these mutated amyloid monomers need to assemble into fibrils. Fibrils are polymorphic and since we don’t know the polymorph this specific amyloid protein will aggregate in, we first need to figure out which

polymorph is the most stable for the current protein. To do this, we resort to our previous work, CreateFibril, a tool that builds and explores the stability of fibrils. Once we know the specific polymorph the protein will aggregate in, we apply mutations to the amyloid monomer and construct a new mutated structure with SCWRL [133]. SCWRL is a tool that determines side-chain conformations to a backbone structure. We specified the original backbone of the protein, but gave SCWRL the mutated sequence of amino acid to fit onto the structure. We then perform energy minimization on the mutated structures to remove any steric clashes due to mutations. The structure is then built into the correct fibril polymorph by CreateFibril.

#### **Step 5: Assessing mutated fibril structural stability**

Applying point mutations to a structure can introduce steric clashes between amino acid side chains. To combat this issue, we perform EM to relax all mutated amyloid and native structures. After this process of relaxation, we use our previous work to quickly calculate the Free energy of proteins. We calculate the LJ and Coulomb energies and use a fast and detailed dipolar water model to compute the solvation energy of molecules by solving the dipolar nonlinear Poisson-Boltzmann-Langevin equation. Together, the three energy terms are used to describe the stability behavior of fibrils. Fibrils with higher energy than the natural control are termed as unstable, and the mutations that generate them are kept for further analysis on the native structure. We used the program AquaSol [89] with the following setup: atomic charges and radii assigned with PDB2PQR using CHARMM force field at neutral pH. A grid of 257 points per edge spaced by 1 Å, a temperature of 300K,

and a solvent accessible surface with an Rprobe of 1.4 Å. All hydrogen-bonds were optimized. We used a trilinear interpolation protocol for projection of fixed charges on the grid, a lattice grid size for the solvent:  $a = 2.8$  Å, solvent made of dipoles of moment  $p_0 = 3.00D$  at a concentration of  $C_{dip} = 55M$ . No salt was added to the solution and small ions were used to equilibrate the system when needed. The electrostatic potential was set to zero at the boundaries, and the stopping criteria for residual was set to:  $1.10^{-6}$  (when possible).

$$F_E = F_{solv} + F_{coulomb14} + F_{vdw} \quad (3.1)$$

where,

$$\begin{aligned} F_{solv} &= F_{(p_0, C_{dip})} - F_{(0,0)} - N_w \mu_w \\ \mu_w &= k_B T \frac{\ln(1 - N_A C_{dip} a^3)}{N_A C_{dip} a^3} \\ N_w &= \int_{solvent} d\mathbf{r} \rho_{dip}(\mathbf{r}) \end{aligned}$$

### Step 6: Structural deviations of the native protein

This step is crucial in addressing the multi-objective problem of discovering point mutations that destabilize fibrils yet conserve the native fold. Mutations that create unstable fibrils are applied to the native protein form to assess any structural stability effect they could produce. We are interested in mutations that destabilize the amyloid but not the native form. Such mutations theoretically preserve structure and protein function and are candidates for therapeutic engineering. We use SCWRL

to build the mutated native proteins and run EM to relax the structures. We then run a full MD simulation and plot RMSD and RMSF graphs to verify any structure deviations caused by the mutation. In particular, we calculate the perturbation in structural motion with

$$\delta_{rmsd} = RMSD(mutant) - RMSD(native)$$

where,  $RMSD$  measures the root mean-square deviations, in angstroms, of the  $C_\alpha$  atom positions in a protein's residues over a simulation run.

We also calculate the root mean square fluctuations (RMSF), a measure of the deviation between the position of a particle  $i$  over a simulation run,

$$RMSF = \frac{1}{T} \sum_{t_j=1}^T (x_i(t_j) - \tilde{x}_i)^2$$

where  $T$  is the total simulation time, and  $\tilde{x}_i$  is the reference position of particle  $i$ . Low  $RMSF$  at a particular mutation site suggests the absence of local residual instability.

### 3.4.1 Molecular Dynamics and Energy Minimization

We used the GROMACS 4.5 [90] molecular simulation package to run molecular dynamics (MD) and energy minimization (EM) simulations. Our mutated proteins were solvated in a cubic box (with a minimum distance of 35 Å from any edge of the box to any atom) and neutralized with chloride ions and modeled using the GRO-MOS96 53a6 force field along with the SPC water model. We used a cutoff of 10 Å for van der Waals and short range electrostatic interactions, and calculated long range electrostatic interactions using a particle mesh Ewald sum [117, 118]. Simulations

were prepared for a full MD run in both isothermal-isobaric (100 ps) and canonical equilibration (100 ps) ensembles. Temperature and pressure were controlled at 300 K and 1 bar using the velocity rescaling thermostat and the Parrinello-Rahman barostat, respectively. A linear constraint solver was used to keep all bonds at their equilibrium length. Approximately ten million time steps were used with an integration time step of 2 fs. The system's coordinates were saved every 10 ps for further analysis.

### 3.4.2 Analyzing Energy Results

To assess the effect of mutations on amyloid fibril stability, we generated fibril mutants up to 25 monomers in size and used Eq. 3.1 to calculate their energies. The solvation term was calculated by AquaSol while the LJ and Coulomb terms were calculated by GROMACS. The same formula was used in the initial assessment of the mutated native structures. We generated RMSD and RMSF plots from MD simulations to analyze structural changes and residue perturbations in native Amylin mutants.

## 3.5 Results and Discussion

In this section we apply our methods to the protein Amylin (PDBID 2KB8), a 37 residue peptide hormone that is secreted from the pancreas in response to intake of food. The 2KB8 is a micelle-stabilized NMR structure suited for diabetes protein-membrane aggregation studies. Amylin normally contributes to glycemic control and inhibits the appearance of specific nutrients in the plasma [134, 135]. In patients with Type II Diabetes, Amylin has been found to misfold into destructive amyloid

monomers that aggregate in pancreatic beta cells and disturb cellular activity, disrupt flow of ions through membranes, and force cells to apoptosis [54, 57]. Little is known about the mechanism or pathway behind the misfolding event, however, the structure of Amylin’s amyloid protein is known. Patients with Type I Diabetes are unable to produce Amylin in their pancreas and require Amylin injections. In 2005, Pramlintide, a version of Amylin with three point mutations that has a lower affinity to form amyloids and fibrils was introduced in the treatment of Type I and Type II Diabetes [136], and has been a better substitute for Amylin in patients with diabetes. Pramlintide, however, is not optimal as patients still experience the emergence of some fibrils that further destroy their  $\beta$ -cells. In this section, we present the results of applying our FibrilMutant protocol on analyzing Amylin’s conformational regions and stability landscape. Moreover, we show that our energy function and destabilizing criteria are in agreement with experimentally tested Amylin mutations and discover novel mutations with stronger destabilizing potential and lower fibril affinities than Pramlintide.

### 3.5.1 Exploring key stability regions of Amylin

The protocol we implemented into FibrilMutant identified six key stability regions in Amylin that contribute to the emergence of amyloids and the growth of their fibrils. It generated twenty three single point mutations with potential to destabilize Amylin fibrils, possibly hindering their production or slowing down their aggregation. The mutations were initially ranked by a statistical mechanics algorithm used in TANGO [98] to help us prioritize simulation order. Table 3-2 displays these mutations. These suggested mutations imply that Amylin amyloid fibrils are stabilized

Original Amino Acid	Residue No.	Mutated Amino Acid	Disruption Method	TANGO Rank
A	13	R	making core charged	1
F	15	P	mutating an amino acid on a beta strand	2
F	15	D	making core charged	3
L	16	D	making core charged	4
A	25	R	making core charged	5
I	26	R	making core charged	6
G	24	P	mutating GLY at a turn	7
L	27	R	making an amino acid on a beta strand charged	8
F	23	E	making core charged	9
G	24	D	making core charged	10
V	17	E	making core charged	11
Q	10	H	making protein surface hydrophobic	12
N	21	P	mutating an amino acid at a turn	14
C	2	Q	making protein surface hydrophobic	15
T	6	M	making protein surface hydrophobic	16
T	4	S	making protein surface hydrophobic	17
V	32	K	making an amino acid on a beta strand charged	18
N	3	H	making protein surface hydrophobic	19
A	8	E	making an amino acid on a beta strand charged	20
T	9	N	making protein surface hydrophobic	21
L	12	E	making core charged	23
C	7	T	making protein surface hydrophobic	24
G	33	E	making an amino acid on a beta strand charged	25
S	20	G	discovered experimentally	13
S	20	K	discovered experimentally	22
N	21	L	discovered experimentally	26
N	14	L	discovered experimentally	27

Table 3–2: Amylin Mutations generated by FibrilMutant with destabilizing potential. Mutations above the horizontal line are destabilizing mutations proposed by FibrilMutant, and mutations below the line have been suggested and tested experimentally. Mutations are ranked by TANGO from lowest aggregation potential to highest aggregation potential.

by the following four main factors: a hydrophilic surface contact with water, a large hydrophobic core region, beta strands, and glycine amino acids at beta sheet turns as illustrated in Fig. 3–1. All proposed mutations attempt to destabilize these regions to weaken Amylin fibril structures.

### 3.5.2 Generating Amylin fibrils

To test the effectiveness of the mutations suggested by FibrilMutant, we apply the mutations to Amylin fibrils ranging in size from one to twenty five monomers as

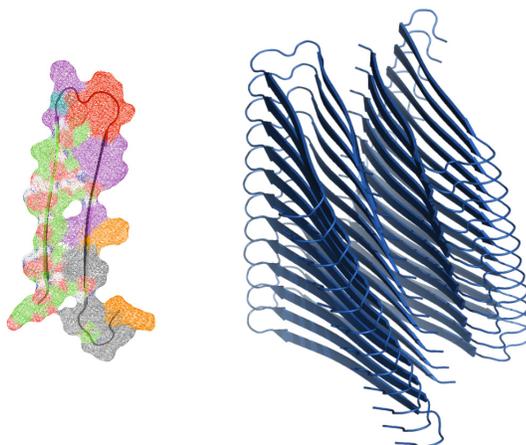


Figure 3–1: Amylin amyloids. Left: Identification of key stability regions of amyloid Amylin by FibrilMutant. Beta strands are colored green, beta turns red, charged residues orange, hydrophobic residues purple, polar residues grey, and glycine residues at turns blue. Right: Full Amylin fibrils of 25 monomers in size.

shown in Fig. 3–1. Amylin fibrils have been found to form into a dimer conformation recently classified as a 2-Stack. Using fibril stability landscapes from our previous work [132], we construct the mutated structures with a fibril packing distance of 3.0 Å, Hbond distance of 5.0 Å between monomers and a rotation angle of 9 degrees along the fibril axis. Each fibril contains one amino acid mutation on each of its amyloid monomers. The stability landscapes are efficient-exhaustive search heuristics for structural parameters that create energetically optimal fibril shapes.

### 3.5.3 Analyzing Amylin fibrils

Fig. 3–2 shows the free energies of the most significant fibrils at their nucleation phase and at their extended aggregation phase. Energy values higher than the control fibril correspond to a decrease in stability, while lower energies correspond to an

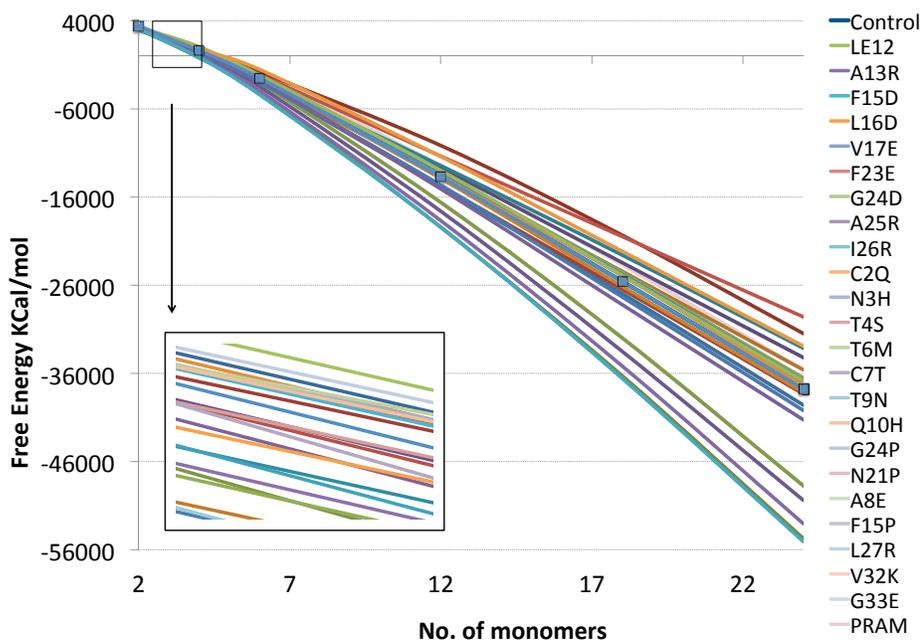


Figure 3-2: Free energies of Amylin mutated fibrils calculated with Eq. 3.1. The inner box is a close up showing energies at the nucleation phase. The control wild type fibril is marked with boxes along its curve. All fibrils with lines above the control are less stable than the control, and all fibrils below the control increase fibril stability.

increase in fibril stability. We are interested in mutations that create the most instability. To be able to sort out these better mutations, we developed three formulas to rank the mutations according to the metrics that measure a protein amyloidogenicity factor, a fibril nucleation factor, and a fibril aggregation extension potential factor. Together, these metrics are intended to measure stability deviations in the various stages of fibril development and growth. The amyloidogenicity factor is measured by

$$\Delta G^i = F_{a_1}^i - F_{n_1} \quad (3.2)$$

where  $\Delta G^i$  is the free energy resulting from transforming an Amylin protein from a mutated native to a mutated amyloid fold,  $F_{a_1}^i$  is the free energy of a single amyloid monomer and  $F_{n_1}$  is the free energy of a single native protein for all mutated fibrils  $i$ .

The second metric is

$$\Delta N^i = F_{a_4}^i - 4 F_{a_1}^i \quad (3.3)$$

where  $\Delta N^i$  is the free energy of nucleation resulting from joining four free amyloid monomers into a fibril structure,  $F_{a_4}^i$  is the free energy of an Amylin fibril composed of 4 monomers and  $F_{a_1}^i$  is the free energy of a single amyloid monomer for all mutated fibrils  $i$ .

The third metric,  $\Delta \tilde{F}^i$ , measures the difference in energy between the mutated fibril and the control averaged out over the length of the fibril,

$$\Delta \tilde{F}^i = \sum_j^n \frac{F_{a_j}^i - F_{a_j}^c}{j} \quad (3.4)$$

where  $F_{a_j}^i$  is the free energy of the mutated fibril  $i$  at length  $j$  and  $F_{a_j}^c$  is the free energy of the control wild type fibril at length  $j$ .

Together,  $\Delta G$ ,  $\Delta N$  and  $\Delta \tilde{F}$  provide insights into the stability perturbations caused by the mutations at the amyloid formation phase, fibril nucleation phase, and fibril elongation phase, respectively. Table 3–3 shows the energy values of all the three metrics applied to the mutated fibrils. A positive  $\Delta G$  value describes an endothermic reaction where native Amylin structures required an input of energy to form into amyloid monomers. The higher the  $\Delta G$ , the higher the gap in energy

between amyloids and native Amylin. Negative  $\Delta G$  suggest favorable exothermic reactions and possibly spontaneous formation of fibrils. Hence, mutations in Table 3–3 with negative  $\Delta G$  are eliminated in red.

$\Delta N$  is useful in comparing the strength of fibrils created at nucleation. The more negative  $\Delta N$  values produce stronger exothermic reactions, and hence more stable fibrils. For our study, we want to explore the mutations that produce a  $\Delta N^i$  greater than  $\Delta N^c$  (control). We rank the mutations in the middle column of Table 3–3 from weakest to highest and remove all mutations smaller than  $\Delta N^c$ . Finally,  $\Delta \tilde{F}^i$  estimates the stability deviation of the mutated fibril from the control wild type. Positive values suggest fibrils that are weaker than the control and negative values suggest fibrils more stable than the control. We observe that the energy gap widens between fibrils and the control as fibrils grow in size which suggests that unstable fibrils (high energy difference with respect to the control) are likely to create energetically preferred shorter structures, possibly a better chance for degrading enzymes and macrophages to destruct them [137, 138]. We ranked the filtered mutations in Table 3–3 and highlighted in green the unstable fibrils out of our set.

The sensitivity of amyloid formation to point mutations can be exploited to design slower Amylin aggregating variants that inhibit fibril formation. There has been no reported systematic analysis of all of the amino acid positions of IAPP or their amyloidogenicity potential, and mutation studies are sparse [139]. We validate our method and rankings by first considering the few mutations explored experimentally [140, 72]. Amylin N14L and N21L mutants did not form amyloids experimentally

Mutation ID	$\Delta G$		$\Delta N$		$\Delta F$
G33E	296.52	T6M	-15494.80	L12E	969.34
F23E	242.23	A8E	-15592.10	A8E	831.79
F15P	217.67	T9N	-15635.66	G33E	764.78
A25R	207.75	G24D	-15694.84	L16D	296.82
N3H	207.72	V17E	-15755.21	F15P	266.55
Control	205.79	F15P	-15771.38	F15D	249.14
C7T	205.12	G33E	-15794.61	C7T	227.84
L16D	200.64	V32K	-15802.45	N21P	171.83
A13R	197.74	G24P	-15866.30	T6M	136.14
L12E	190.56	L12E	-15915.03	Q10H	115.75
F15D	184.83	N21P	-16044.82	Control	0.00
I26R	181.08	C7T	-16098.65	G24P	-12.28
L27R	167.15	Q10H	-16104.24	C2Q	-74.43
N21P	161.64	F15D	-16144.06	N3H	-101.91
Q10H	153.91	L16D	-16145.39	V17E	-116.60
T4S	144.96	A25R	-16147.63	G24D	-345.73
G24P	125.15	Control	-16158.22	T4S	-464.05
V17E	94.30	T4S	-16177.21	F23E	-615.08
C2Q	80.36	N3H	-16249.20	T9N	-636.75
G24D	72.18	C2Q	-16312.04	A25R	-1338.74
A8E	56.52	I26R	-16424.93	I26R	-1644.14
T6M	36.38	F23E	-16429.73	L27R	-2056.31
T9N	-11.13	A13R	-16557.11	A13R	-2363.51
V32K	-13.97	L27R	-16623.31	V32K	-2487.62
S20G	226.36	PRAM	-15592.91	PRAM	436.84
N14L	223.82	N21L	-15768.26	N21L	473.79
N21L	193.17	S20K	-15865.49	N14L	316.25
S20K	146.17	N14L	-15967.28	S20G	106.47
PRAM	42.08	S20G	-16044.40	S20K	-1340.63

Table 3–3: Fibril stability results. Stability values of mutated fibrils. Mutations above the horizontal line were proposed by FibrilMutant. Mutations below the line were explored experimentally. The “Control” fibril is the non-mutated, naturally occurring fibril. Units are in KCal/mol.

while, the S20K mutant lengthened the lag phase by a factor of 18 and had a significant effect on amyloid formation and S20G was observed to form amyloids. The lower part of Table 3–3 shows the  $\Delta G$ ,  $\Delta N$ , and  $\Delta\tilde{F}$  of these 4 mutations. We observe that N14L and N21 have a higher  $\Delta N$  and  $\Delta\tilde{F}$  than the control, suggesting that these mutations destabilize their fibrils, and hence could explain why they do not form experimentally. S20K has a  $\Delta N$  that is also higher than the control, suggesting that the nucleation product is less stable than the control’s oligomer, also suggesting a longer nucleation phase as observed experimentally. The S20K  $\Delta\tilde{F}$  is quite small, suggesting that this mutation might form unstable fibrils, as observed experimentally. The S20G mutant was observed to form amyloids, and its corresponding  $\Delta\tilde{F}$  also suggests this finding. It is key to note that Pramlintide ranked as the highest unstable mutant explored experimentally with a  $\Delta\tilde{F} = 436.84$ , close in instability to the N21L mutant.

Our results indicate that the mutant L12E causes the most instability to fibrils and has a high  $\Delta G$ ,  $\Delta N$ , and  $\Delta\tilde{F}$ . In fact, its  $\Delta\tilde{F}$  is more than twice as large as the PRAM  $\Delta\tilde{F}$ , suggesting that it might inhibit fibrils altogether. The mutant A8E also has twice as large a  $\Delta\tilde{F}$  than PRAM, but also has a slower more unstable nucleation phase, indicating that it has a strong potential to inhibit fibril formation. In fact, the results of running AmyloidMutants [99] on the A8E mutation suggest that this point mutation destabilizes amylin fibrils, as shown in Fig. 3–3. The last competitive mutant, G33E, also exhibits a higher  $\Delta\tilde{F}$  than Pramlintide and shows a high  $\Delta N$  value, also suggesting high instability in the nucleation phase and fibril elongation phase. Together, these observations recommend a Glutamic acid mutation in Amylin

to stop it from forming fibrils. Since Amylin contains no acidic residues, the addition of this charged, acidic residue will enhance the formation of a quasi-infinite array that destabilized the fibrils with unfavorable electrostatic interactions created along the fibril length [72]. The other highlighted mutations in green in Table 3–3 also have the potential to destabilize and inhibit fibrils, and their effect might be similar or smaller than Pramlinitide. Fig. 3–2 confirms that the formation of fibrils is always favorable and that aggregation of amyloid monomers contributes to stability. The key thing to note is that some mutations produce fibrils that are less stable than others, and some mutations might produce fibrils that have high activation energies and will not form in physiological conditions. An excellent example is rat Amylin; it does not form any fibrils *in vivo*, however, under the right environment conditions, it has recently been reported to create long fibrils [141]. The results we report in this study are important to design stronger alternative variants to the Pramlintide antihyperglycemic drug with a minimalistic mutation approach for diabetes patients.

TANGO and AmyloidMutants are current computational tools exploring amyloid stability and analyzing the effect of secondary structure modifications on increasing amyloidogenicity and protein aggregation. Such tools use coarse-grained models that enable them to perform high-throughput screenings, but cannot achieve the accuracy of higher resolution models [78]. The use of AmyloidMutants and TANGO assisted in ranking candidate destabilizing mutations prior to running our computationally expensive dipolar solvent model to accurately assess the instability caused in the mutated fibrils. Although, the TANGO and AmyloidMutants results did not match well with our results, they did provide some valuable insight. TANGO results

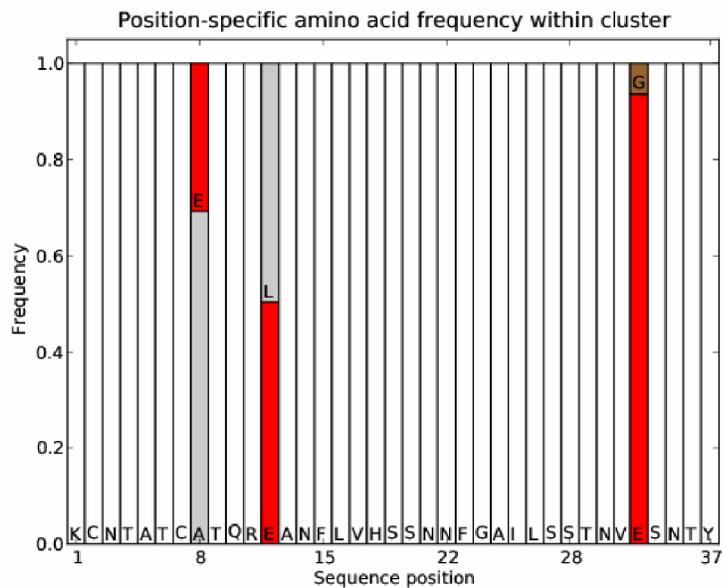


Figure 3–3: Top 3 mutation results from AmyloidMutants. Mutation A8E presents a high destabilization fibril aggregation frequency.

in Table 3–4 suggest that the A8E mutation doubles the potential of alpha helices compared to the Control and the L12E mutation lowers the amyloidogenicity by a factor of 42 compared to the Control. AmyloidMutants also suggested that the A8E mutation should destabilize fibrils.

TANGO Rank	Mutation	AGG	AMYLO	TURN	HELIX	BETA
1	A13R	17.4231	0.00229692	26.2032	46.2422	50.1558
2	F15P	17.4231	0.00483631	24.9158	50.5216	44.2985
3	F15D	18.6542	0.00494376	26.6371	18.5491	45.443
4	L16D	18.6542	5086.05	27.7912	26.473	55.8228
5	A25R	19.9643	11265.8	27.9887	38.6424	55.1474
6	I26R	19.9643	11265.8	27.6605	39.8376	43.7964
7	G24P	19.9643	11265.8	25.3545	38.6424	47.8267
8	L27R	19.9643	11265.8	27.6375	38.6424	47.5071
9	F23E	20.1387	11265.3	28.4565	42.7176	46.663
10	G24D	20.1387	11265.3	26.3146	40.1409	52.1159
11	V17E	27.6387	11303.4	27.8741	37.1066	42.0603
12	Q10H	36.4611	11348.7	27.7108	4.22761	52.4841
14	N21P	37.2516	11271.1	39.9629	41.3196	47.075
15	C2Q	37.3071	11266.1	27.5314	41.2988	46.1128
16	T6M	37.311	3600.44	26.6308	44.1818	41.6555
17	T4S	37.3528	12404.1	27.732	43.2968	47.3828
18	V32K	37.3697	11265.8	29.2945	41.3929	45.9617
19	Control	37.3874	11265.8	27.3615	41.4001	47.7778
20	N3H	37.4431	13070.4	27.169	37.9999	51.2714
21	A8E	37.4799	12261.2	26.8771	92.3333	53.2163
22	T9N	37.5127	12620.4	28.8211	33.0901	45.0847
24	L12E	37.6587	264.028	25.8306	23.7728	48.1199
25	C7T	37.6787	4923.45	26.176	26.4939	47.923
26	G33E	38.7927	11265.3	21.6298	41.3269	64.2312
13	S20G	36.959	11266.7	31.3738	41.3718	47.3916
23	S20K	37.5908	11268.8	24.84	39.3714	49.6538
27	N21L	42.4584	11266.8	20.5032	42.4313	61.4794
28	N14L	503.96	12778.5	24.8149	23.0706	51.5248

Table 3–4: TANGO results on amylin mutations. Mutations above the horizontal line are destabilizing mutations proposed by FibrilMutant, and mutations below the line have been suggested and tested experimentally. Mutations are ranked by TANGO from lowest AGG potential to highest AGG potential. The TANGO algorithm predicts cross-beta aggregation in proteins and populates structures according to a Boltzmann distribution to 4 structural stats: beta-sheet aggregation, beta-turn, alpha-helix, and alpha-helical aggregation. The results in this table describe the following in respective order: percentage of aggregation, amyloidogenicity potential, percentage of beta-turn conformation, percentage of alpha-helical conformation, and percentage of beta-strand conformation

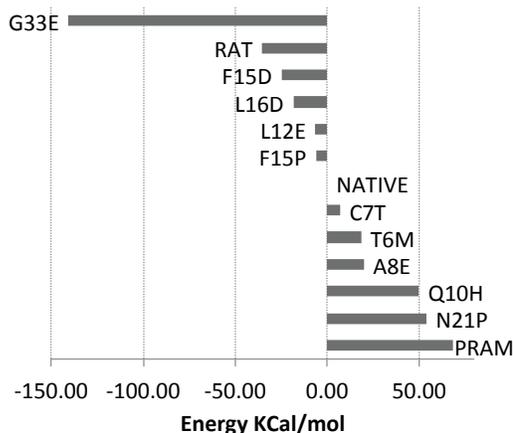


Figure 3–4: Stability of the Amylin mutants in their native fold. Each bar represents one mutated protein in native fold, where the energies are the free energy differences between the mutant and the non-mutated Amylin.

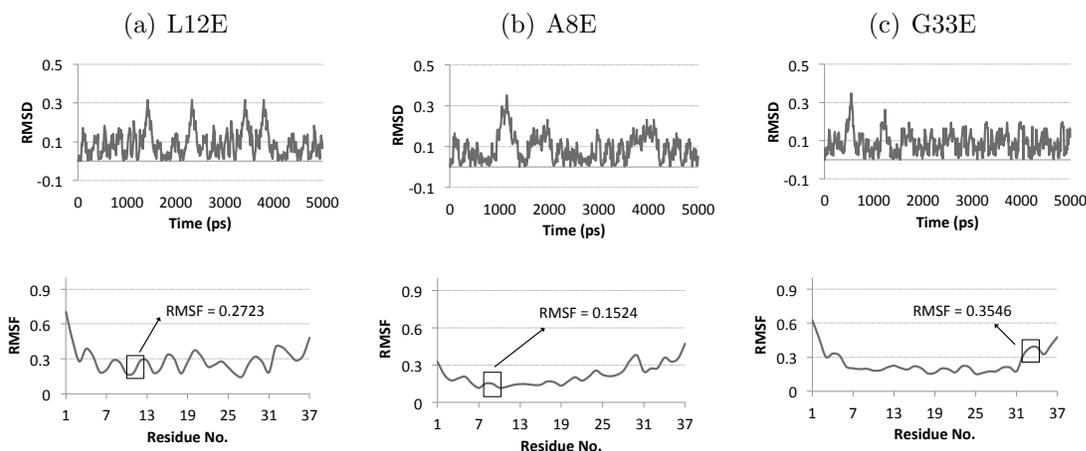


Figure 3–5:  $\delta_{rmsd}$  and RMSF plots for mutants L12E, A8E, and G33E over a 5ns simulation. The top figures display  $\delta_{rmsd}$  graphs between each mutant and the native Amylin, and the bottom plots are the RMSF plots.  $\delta_{rmsd}$  graphs show extremely small modifications in structure  $\ll 1 \text{ \AA}$ . RMSF plots also show very minimal structural variance at the amino acid mutation sites.

### 3.5.4 Maintaining native structure & function

Introducing a mutation to the sequence of Amylin might cause a change in its structure and force it to fold into a different shape, affecting its normal function. For this reason, we have tested the effect of the destabilizing mutations on the native structure of Amylin. The free energies of mutated Amylin in Fig. 3–4 show that the mutations result in energy states close to the native control, rat Amylin, and pramlintide. Fig. 3–4 suggests that these mutations do not make native Amylin unstable. To further verify this, we ran short MD simulations on the top 3 native mutants (L12E, A8E, and G33E) to test any structural turbulence or major energetic imbalances as a result of the point wise substitution. We observed very slight deviations in  $\delta_{rmsd}$ , stable RMSF in mutated regions and smooth energy trajectories, suggesting a preservation in shape and function. Fig. 3–5 shows the  $\delta_{rmsd}$  and RMSF results for best mutants, L12E, A8E, and G33E.

## 3.6 Conclusion

The process of amyloid protein formation and aggregation is sensitive to amino acid sequence point mutations. We discuss in this manuscript how altering the genetic code of these amyloid proteins has been shown to affect fibril propagation, dynamics, growth, stability, and infectivity. Certain regions in amyloid proteins contribute to fibril structural stability, compactness, and insolubility. Altering some amino acids that make up these regions, such as the amino acids that are involved in creating a hydrophobic core, can create energy perturbations and imbalances that weaken an individual amyloid protein monomer and subsequently carry on the effect

to weaken every amyloid monomer on the fibril, resulting in an accumulated fibril destabilization effect. Tackling the problem from this perspective has enabled us to computationally perform an amino acid mutation analysis on Amylin to unravel modifications that potentially destabilize fibrils, yet are restricted to conserve the native fold of the protein. Addressing this multi-objective problem can be generally useful in suggesting novel therapeutic agents or improving existing treatments for cases where drugs have to be administered to patients. In our case, addressing this problem has opened up discussions on the 3 potential efficacy improvements in the Pramlintide drug for Type I Diabetes.

### **3.7 Acknowledgements**

M.R.S. was supported by a fellowship from the Canadian Institutes of Health Research System Biology Training program at McGill University and a grant from the Fonds de recherche Nature et technologies Quebec. J.W. was supported by a Discovery grant from the Natural Science and Engineering Research Council of Canada.

## CHAPTER 4

### Complete characterization of the mutation landscape reveals the effect on amylin stability and amyloidogenicity

#### 4.1 Preface

We have presented in the previous two chapters the importance of amino acid mutations in affecting amyloid structure and toxicity. Researchers have been trying to deduce which mutations, out of the set of all possible mutations, affect the stability of amyloids. Since every residue position of a protein can be mutated into 19 different possibilities, this results in a mutation space,  $M$ , given by,

$$M = \binom{n}{k} * 19^k \quad (4.1)$$

where  $n$  is the number of residues in a protein and  $k$  is the number of residues that we wish to mutate. Even for low values of  $k$ , the mutation space is infeasible to explore experimentally.

In the previous chapter, we relied on properties of amyloid structures to deduce constructive mutations. We used a heuristic method to return a small set of mutations that potentially lower the amyloidogenic of amylin. We generated complete fibrils for the mutations and quantified their nucleation and elongation.

In this chapter, we explore the mutational landscape of amylin for  $k = 1$  and assess the effect of mutations on amylin's stability and potential to form amyloids. Computing the free energies of amylin mutant structures for  $k = 1$  was an expensive brute-force task and exploring mutational landscapes of  $k > 1$  is improvident. However, from the results of the mutational landscape for  $k = 1$ , we constructed a way to quickly estimate the stability and amyloidogenicity effect of multiple-point mutations for  $k > 1$ . In this work, we restrict our search to  $k = 3$  mutations to limit the compounding error on the estimation of stability and amyloidogenicity values. This opened up the possibility to efficiently explore larger mutational landscapes and significantly reduced the search space for experimentalists. According to our results, the leading Pramlintide drug in diabetes can be significantly improved with the selection of different amino acid mutations that engineer more stability and less amyloidogenicity.

The remaining content of this chapter has been taken from a manuscript that was submitted for publication:

- M. R. Smaoui and J. Waldispühl. "Complete characterization of the mutation landscape reveals the effect on amylin stability and amyloidogenicity". *Submitted for review in Sept 2014*.

## 4.2 Abstract

Type II diabetes is characterized by insulin resistance and a failure of the pancreas to supply enough insulin. This failure is believed to be partially aggravated by the emergence of toxic amylin protein deposits in the extracellular space of the

pancreas  $\beta$ -cells. Amylin in normal form is a regulatory hormone that is co-secreted with insulin, however, it has been observed to misfold into toxic structures. Pramlintide, an FDA approved injectable amylin analog mutated at positions 25, 28 and 29 was therefore developed to create a more stable, soluble, less-aggregating, and equipotent peptide that is used as an adjunctive therapy for diabetes. Although Pramlintide has shown some therapeutic benefits over amylin, it is far from ideal. Researchers have been attempting to optimize this drug by further chemically modifying its amino acids to generate better therapeutic amylin analogs. In this work, we assist the finding of optimal analogs by computationally revealing the mutational landscape of amylin. We computed the structure energies of all the possible single-point mutations on amylin and studied the effect they have on protein stability and amyloidogenicity. Each of the 37 amylin residues was mutated into the 19 canonical amino acids and an energy function computing the Lennard-Jones, Coulomb and Solvation energy was used to analyze changes in stability. The mutation landscape enabled us to identify amylin's conserved stable regions, residues that can be tweaked to further stabilize structure, regions that are susceptible to mutations, and mutations that are amyloidogenic. We used the data of the single-point mutational landscape to generate estimations for higher-order multiple-point mutational landscapes and discovered millions of 3-point mutations that are more stable and less amyloidogenic than Pramlintide. The landscapes provided an explanation for the effect of the S20G and Q10R mutations on the onset of diabetes of the Chinese and Maori populations, respectively.

### 4.3 Author Summary

Amylin is a crucial protein in balancing the blood glucose levels in the human body. When the pancreas undergoes significant, prolonged pressure to produce large quantities of this molecule, amylin starts to misfold into toxic structures that kill cells in the pancreas. Researchers have tried to stop this phenomenon by creating amylin analogs that have lower tendencies to misfold. So far, the FDA approved an injectable amylin analog substance, called Pramlintide, engineered with three sequence mutations that proved to lower the potential of misfolding. It is very likely that there exist different mutations that produce a better analog than Pramlintide. However, because there are far more combinations of possible mutations to experimentally test for than there are atoms in the universe, proper testing is not feasible. Instead, we can computationally explore the effect of the 740 point-mutations on amylin's stability and likelihood to misfold. Examining these mutations enabled us to identify the regions in amylin that are prone to misfold and regions that can be mutated to improve stability. From this analysis, we were able to efficiently search through all the 62 million combinations of 3-point mutations and find many analogs that are potentially more advantageous than Pramlintide.

### 4.4 Introduction

Type II diabetes affects approximately 333 million people around the world and is likely to increase by 50.7% over the next 20 years [142]. It is characterized by insulin resistance [7, 8] and the failure of the pancreatic beta cells to supply required levels of insulin [9]. The failure to secrete enough insulin is caused by beta cell

dysfunction and reduced beta cell mass [10, 11, 12]. Many studies have found an association between this failure to produce insulin and the emergence of islet amyloid protein deposits in humans [13, 14, 15], in non-human primates [16, 17], and cats [18]. Interest in studying the potential pathogenic role of islet amyloid deposits in type II diabetes has been incentivized by experiments reporting that mice transgenic for human islet amyloid develop hyperglycemia [19, 20, 21], a signature condition for type II diabetes.

Islet amyloid polypeptide, also widely known as amylin [143], is a 37 amino acid peptide that is co-stored and co-secreted with insulin in secretory granules in beta cells [144, 145]. Amylin's functions include modulating gastric emptying [146], inhibiting glucagon to prevent postprandial spikes in blood glucose levels [147], and inducing satiety leading to decreased food intake and weight loss [148, 149]. Together, amylin, insulin, and glucagon play a crucial role in maintaining glucose homeostasis [53]. However, amylin has been observed to counterbalance the stability of beta cells by misfolding into beta-sheet rich amyloid deposits. The peptide has been observed to form oligomer and fibril deposits in the intra and extra-cellular compartments of beta cells [54, 55]. The extreme insolubility of these deposits have hindered our understanding of how they form, what they are made of, and how they bind to different molecules. Although causative factors for the conversion of the normally soluble native amylin peptide into insoluble, aggregating amyloid fibrils remain largely obscure, these amyloid depositions are among the most common pathological features of type II diabetes arising in almost all diabetic subjects [56, 57].

Although the function and primary shape of native amylin have been well conserved [150], there exist some amino acid mutations across species that influence amyloid formation. As mentioned earlier, diabetes associated amyloid is observed in cats and in non-human primates, but not in rodents. Rodents and mice exhibit six point mutations in their amylin peptide that inhibit fibril formation [74, 75, 76]. Nonetheless, amyloid fibrils have been observed to form in transgenic mice producing human amylin [151] suggesting that point-mutations might alter the energy barrier required for amylin to misfold. The misfolding is believed to happen inside granules as they transport insulin and amylin from the cell cytoplasm to the membrane [152].

Point-wise mutations have also been observed in humans. The S20G mutation of the amylin gene in Japanese type II diabetes patients accounted for 4.1% of all cases and increased the risks of early and severe onset of type II diabetes [69]. Similarly, the S20G mutation has also been found to increase the diabetes risk of Chinese patients [153]. Moreover, the Q10R mutation of the amylin gene in the New Zealand Maori population is believed to explain the high susceptibility and prevalence of diabetes in Maori patients [73].

Extensive amyloid deposition has been correlated with reduced islet function, the loss of beta cell mass, and the need for insulin injection therapy in humans. Although the exact relationship of amylin to the onset and progression of type II diabetes is unknown, hindering the misfolding of amylin and reducing the formation of amyloids and fibrils might prolong the onset and severity of type II diabetes [152]. Pramlintide, an FDA approved injectable amylin analog administered subcutaneously at mealtimes, was therefore developed by performing three-point mutations

on amylin replacing amino acids 25, 28, and 29 with prolines to create a stable, soluble, non-aggregating, and equipotent peptide that is used as an adjunctive therapy for type I and type II diabetes [154, 155].

Although Pramlintide has shown some therapeutic benefits over amylin, its pharmacokinetic properties are not ideal. Because it precipitates above a pH of 5.5, patients need to inject it separately from insulin. Pramlintide has a half life of only 45 minutes, requiring a daily injection three times [156, 157]. For these reasons, researchers have been attempting to optimize this drug by further chemically modifying its amino acids and generating several other stable amylin analogs to offer a more appealing therapeutic agent [158].

In this work, we aim to reveal the mutational landscape of amylin to assist in the further development of analogs for type I and type II diabetes. We compute all the possible 740 single-point mutations on amylin and study the effect they have on protein stability and amyloidogenicity. Each of the 37 amylin residues is mutated into the 19 canonical amino acids and an energy function computing the Lennard-Jones, Coulomb and Solvation energy is used to analyze and detect changes in energy. The mutational landscape enabled us to identify amylin’s conserved stable regions, regions that are susceptible to mutations, and mutations that are amyloidogenic. We used the data of the single-point mutational landscape to generate estimations for higher-order multiple-point mutational landscapes and efficiently explored millions of 3-point mutated analogs that are more stable and less amyloidogenic than Pramlintide.

## 4.5 Methods

In this section we present, MAPOR, the tool we built to generate the mutational landscape of proteins. We will outline the algorithm it uses to measure the stability and amyloidogenicity of each structure along trajectories in the mutational landscape. MAPOR can be downloaded from <http://amyloid.cs.mcgill.ca/>

### 4.5.1 Computing the landscape

MAPOR (MutAtion landscaPe generatOR) is a tool that analyzes the mutational landscape of proteins by assessing changes in energy and stability. Its basic procedure is outlined in Algorithm 1 of Table 4-1. Given a protein's complete PDB structures in amyloid and non-amyloid form, MAPOR performs all possible mutations on the amino acid backbone of the protein. With each mutation, the Lennard-Jones, Coulomb, and Solvation energies are computed and compared to the non-mutant energies as shown in Eqs. 4.2-4.4.

---

```

Require:  $Structure_{Native}, Structure_{Amyloid}, SEQUENCE$ 
Ensure:  $Structure_{Native}.Length == Structure_{Amyloid}.Length ==$ 
 $SEQUENCE.Length$ 
 $MutationList \leftarrow \text{List}(\text{"ARNDCQEGHILKMFPSTWYV"})$ 
 $Results_{Native} \leftarrow []$ 
 $Results_{Amyloid} \leftarrow []$ 
for  $x$  in  $[1, 2, \dots, SEQUENCE.Length]$  do
  for  $m$  in  $MutationList$  do
     $MutatedSequence \leftarrow \text{mutateSequence}(\text{position} \leftarrow x, \text{mutation} \leftarrow m,$ 
 $SEQUENCE)$ 
     $Mutant_{Native} \leftarrow \text{createMutatedStructure}(MutatedSequence,$ 
 $Structure_{Native})$ 
     $Coulomb_E \leftarrow \text{calculateCoulombEnergy}(Mutant_{Native})$ 
     $LJ_E \leftarrow \text{calculateLJEnergy}(Mutant_{Native})$ 
     $Solvation_E \leftarrow \text{calculateSolvationEnergy}(Mutant_{Native})$ 
     $Results_{Native}.add(MutatedSequence, Coulomb_E, LJ_E, Solvation_E)$ 
     $Mutant_{Amyloid} \leftarrow \text{createMutatedStructure}(MutatedSequence,$ 
 $Structure_{Amyloid})$ 
     $Coulomb_E \leftarrow \text{calculateCoulombEnergy}(Mutant_{Amyloid})$ 
     $LJ_E \leftarrow \text{calculateLJEnergy}(Mutant_{Amyloid})$ 
     $Solvation_E \leftarrow \text{calculateSolvationEnergy}(Mutant_{Amyloid})$ 
     $Results_{Amyloid}.add(MutatedSequence, Coulomb_E, LJ_E, Solvation_E)$ 
  end for
end for
return  $Results_{Native}, Results_{Amyloid}$ 

```

---

Table 4–1: Generating the mutational and amyloidogenic landscapes of amylin

$$E = \textit{Solvation} + \textit{LJ} + \textit{Coulomb} \quad (4.2)$$

$$\Delta E = E - E_0 \quad (4.3)$$

where  $E_0$  is the energy of the native non-mutant protein and  $E$  is the energy of a mutant structure, and

$$\Delta E' = E' - E'_0 \quad (4.4)$$

where  $E'_0$  is the energy of the amyloid non-mutant protein and  $E'$  is the energy of a mutant amyloid structure. Together,  $\Delta E$  and  $\Delta E'$  measure the difference in energy caused by a mutation in both the native and amyloid forms, respectively. Changes in energy correlate to changes in stability. Positive  $\Delta E$  values correlate to a decrease in stability and negative  $\Delta E$  values correlate to an increase in stability. We measure the difference in energy between amyloid and native structures by computing  $\Delta G$  in Eq. 4.5.

$$\Delta G = E' - E. \quad (4.5)$$

$\Delta G$  is the minimum energy required to change the protein from its native to amyloid form. It is estimated that the larger the  $\Delta G$ , the larger the energy barrier between the native and amyloid structures. In this work, we correlate amyloidogenicity with  $\Delta G$  values. Mutations that significantly increase  $\Delta G$  values lower amyloidogenicity

and mutations that decrease  $\Delta G$  values make it potentially easier for amylin to misfold into its amyloid form.

The landscape returned by MAPOR is 3-dimensional. The 20 possible amino acid mutations make up 1 dimension, the amino acid length of the protein being studied makes up the second dimension, and the energy values of each mutated structure make up the third dimension.

#### 4.5.2 Mutant Structures

We performed 19 mutations for every amino acid in amylin. Mutations were constructed with SCWRL4 [133], a tool to determine the side-chain conformations to a backbone structure. For each generated mutant structure, SCWRL took as input the original native amylin protein (PDBID 2kb8) and the new mutated amino acid sequence to fit onto the structure. We then performed energy minimization on the mutated structures to remove any steric clashes due to mutations prior to calculating the LJ, Coulomb, and Solvation energies. We repeated the process for the amyloid mutant forms of amylin [34].

#### 4.5.3 Calculating structure energy

We perform Energy Minimization (EM) on all mutant structures to reduce any steric clashes that might occur in the amino acid mutation process. After this process of relaxation, we resort to our previous work to quickly calculate the Free energy of proteins [132] computed by Eq. 4.2. The Solvation energy term in Eq. 4.2 is produced using a fast and detailed dipolar water model that solves the dipolar nonlinear

Poisson-Boltzmann-Langevin equation. The three energy terms in Eq. 4.2 are used to describe the stability of molecules. Native amylin mutants that have a  $\Delta E$  that is negative are more stable than amylin itself, and amyloid mutants that have a  $\Delta E'$  that is positive possess greater instability than amylin's amyloid molecule. Solvation energy was precisely calculated using AquaSol [89] with the following setup: atomic charges and radii assigned with PDB2PQR using CHARMM force field at neutral pH. A grid of 257 points per edge spaced by 1 Å, a temperature of 300K, and a solvent accessible surface with an Rprobe of 1.4 Å. All hydrogen-bonds were optimized. We used a trilinear interpolation protocol for projection of fixed charges on the grid, a lattice grid size for the solvent:  $a = 2.8 \text{ Å}$ , solvent made of dipoles of moment  $p_0 = 3.00\text{D}$  at a concentration of  $C_{dip} = 55\text{M}$ . No salt was added to the solution and small ions were used to equilibrate the system when needed. The electrostatic potential was set to zero at the boundaries, and the stopping criteria for residual was set to:  $1.10^{-6}$  (when possible).

#### 4.5.4 Molecular Dynamics and Energy Minimization

We used the GROMACS 4.5 [90] molecular simulation package to run molecular dynamics (MD) and energy minimization (EM) simulations. Our mutant molecules were solvated in a cubic box (with a minimum distance of 35 Å from any edge of the box to any atom) and neutralized with chloride ions and modeled using the GROMOS96 53a6 force field along with the SPC water model. We used a cutoff of 10 Å for van der Waals and short range electrostatic interactions, and calculated long range electrostatic interactions using a particle mesh Ewald sum [117, 118]. Simulations

were prepared for a full MD run in both isothermal-isobaric (100 ps) and canonical equilibration (100 ps) ensembles. Temperature and pressure were controlled at 300 K and 1 bar using the velocity rescaling thermostat and the Parrinello-Rahman barostat, respectively. A linear constraint solver was used to keep all bonds at their equilibrium length. One million time steps were used with an integration time step of 2 fs to assess any potential turbulence introduced into the molecules by mutations. The system’s coordinates were saved every 10 ps for further analysis.

#### 4.5.5 Assessing structural deviations

After a mutant structure is built and relaxed with EM, we assess the structural effect of mutations by calculating perturbations in amino acid positions using *RMSD* and *RMSF* graphs. *RMSD* measures the root mean-square deviations, in angstroms, of the  $C_\alpha$  atom positions in a protein’s residues over a simulation run, whereas the *RMSF* measures the root mean square fluctuations, a measure of the deviation between the position of a particle  $i$  over a simulation run given by,

$$RMSF = \frac{1}{T} \sum_{t_j=1}^T (x_i(t_j) - \tilde{x}_i)^2$$

where  $T$  is the total simulation time, and  $\tilde{x}_i$  is the reference position of particle  $i$ . Low *RMSF* at a particular mutation site suggests the absence of local residual instability.

#### 4.5.6 Calculating energies for multiple-point mutations

Algorithm 2 of Table 4-2 explains in detail the procedure for calculating approximations of  $\Delta E$  and  $\Delta G$  given by  $\Delta\tilde{E}$  and  $\Delta\tilde{G}$ , respectively, for multiple-point

mutations. Given a sequence of length  $k$ , the possible combination of  $n$ -point amino acid mutations is

$$\binom{k}{n} * 20^n. \quad (4.6)$$

Starting with the  $\Delta E$  and  $\Delta G$  tables generated by MAPOR using Algorithm 1 from Table 4-1 for the single-point mutations, we compute the estimates  $\Delta\tilde{E}$  and  $\Delta\tilde{G}$  for  $n$ -point mutations by Eqs. 4.7-4.9,

$$\Delta\tilde{E} = \sum_{i=1}^n \Delta E(m_i, p_i) \quad (4.7)$$

$$\Delta\tilde{E}' = \sum_{i=1}^n \Delta E'(m_i, p_i) \quad (4.8)$$

$$\Delta\tilde{G} = \Delta G_0 - \Delta\tilde{E} + \Delta\tilde{E}' \quad (4.9)$$

where  $m_i$  is the mutation number  $i$ ,  $x_i$ , is the residue position that the mutation  $m_i$  should take effect on,  $n$  is the total number of desired mutations and the dimension of the landscape, and  $\Delta G_0$  is the  $\Delta G$  value for native with no mutations. The estimates are calculated directly by summing values from the  $\Delta E$ ,  $\Delta E'$ , and  $\Delta G$  tables and without running the MAPOR simulations. We found the estimates to be close to values returned by simulation runs. We hypothesize that the estimates deviate slightly from real values because they do not take into account the pairwise electrostatic and coulomb effects of the mutations with one another. Nevertheless, they provide good estimates and allow us to compute values for millions of mutations

efficiently.

---

```

Require:  $Results_{Native}, Results_{Amyloid}, E_0, E'_0, \Delta G_0, SEQUENCE$ 
 $MutationList \leftarrow \text{List}(\text{"ARNDCQEGHILKMFPSTWYV"})$ 
Table  $\Delta E \leftarrow Results_{Native} - E_0$  // for each element in Results
Table  $\Delta E' \leftarrow Results_{Amyloid} - E'_0$  // for each element in Results
 $Results \leftarrow []$ 
Tuples  $Mutations \leftarrow \text{CHOOSE}(n, SEQUENCE.Length)$ 
for all  $(x_1, x_2, \dots, x_n)$  tuples in  $Mutations$  do
  for  $m_1$  in  $MutationList$  do
    for  $m_2$  in  $MutationList$  do
       $\vdots$ 
    for  $\tilde{m}_n$  in  $MutationList$  do
       $MSeq \leftarrow \text{mutateSequence}(\text{positions} \leftarrow (x_1, x_2, \dots, x_n), \text{mutations} \leftarrow$ 
         $(m_1, m_2, \dots, m_n), SEQUENCE)$ 
       $\Delta \tilde{E} \leftarrow \sum_{i=1}^n \Delta E(m_i, x_i)$ 
       $\Delta \tilde{E}' \leftarrow \sum_{i=1}^n \Delta E'(m_i, x_i)$ 
       $\Delta \tilde{G} \leftarrow \Delta G - \Delta \tilde{E} + \Delta \tilde{E}'$ 
       $Results.add(MSeq, \Delta \tilde{E}, \Delta \tilde{G}, (x_1, x_2, \dots, x_n), (m_1, m_2, \dots, m_n))$ 
    end for
   $\vdots$ 
  end for
end for
end for
 $Results.SORT(Key \leftarrow 3)$  //Sorts on third element descending
return  $Results$ 

```

---

Table 4-2: Generating  $\Delta \tilde{E}$  and  $\Delta \tilde{G}$  for  $n$ -point mutations

## 4.6 Results & Discussion

We applied Algorithm 1 in Table 4-1 to the protein amylin (PDBID 2kb8) and generated its mutational landscape consisting of 740 single-point mutations as shown in Fig. 4-1(a). Each amino acid position in the 37 residues of amylin was mutated into the 19 different canonical amino acids. For each mutation, a separate amylin mutant structure was created and its Coulomb, Lennard-Jones, and Solvation energies were calculated as defined in the Methods section. The result of the simulation runs is the 3-dimensional plot in Fig. 4-1(a). The 20 possible amino acid mutations make up 1 dimension, the amino acid length of the protein being studied makes up the second dimension, and the energy values of each mutated structure make up the third dimension. The same procedure was repeated for the amyloid form of amylin. We used the structure given by Wiltzius *et al.* [34] as amylin's amyloid form. The difference in energies between the amyloid landscape and the native (non-amyloid) landscape gave us values for  $\Delta G$  and are plotted in Fig. 4-1(b).

The mutational landscape of amylin's native structure exhibits several features. First, it contains rugged regions that explain the sensitivity of amylin to point mutations. Second, the landscape reveals amylin's unstable residue positions and unstable regions. These unstable single residues and regions return high energy values compared to the non-mutant amylin. Third, it identifies amino acids mutations that have the potential to further stabilize amylin. Fourth, it pinpoints several mutations across the landscape that have a high potential to destabilize amylin. Finally, it details a residue segment in amylin that is energetically conserved against destabilizing

(a) Mutational stability landscape for native amylin

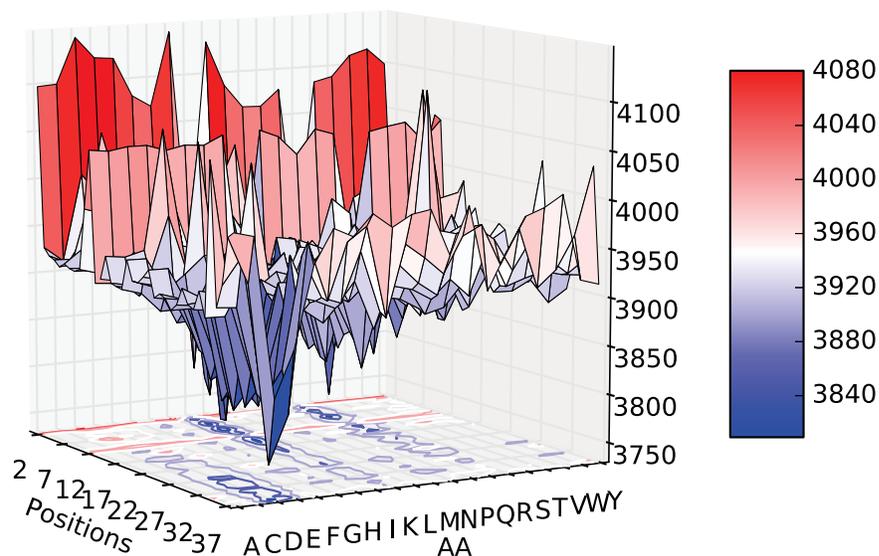
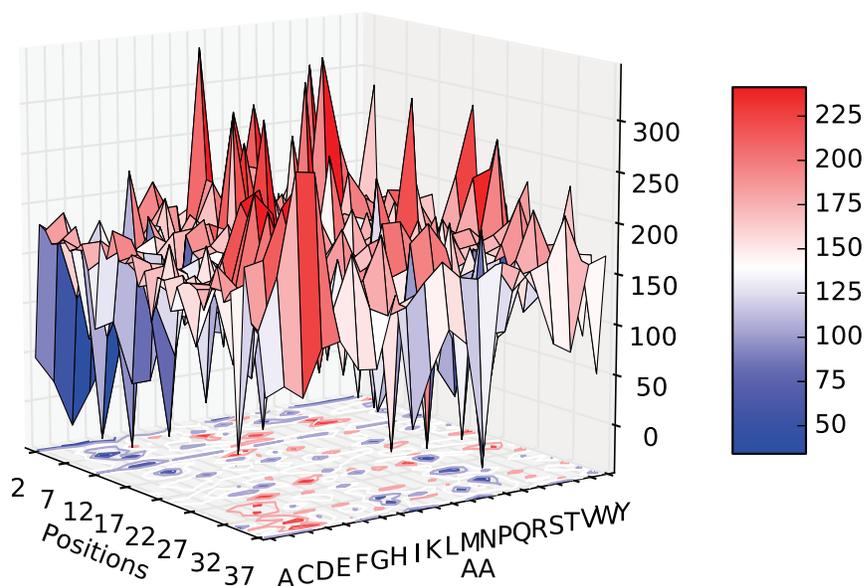
(b) Amyloidogenicity landscape  $\Delta G$ 

Figure 4–1: Stability and amyloidogenicity mutational landscapes. The energy values represent the free energy of structures given by Eq. 1 and are in kcal/mol. The energies of all 740 mutant native 2kb8 structures are shown in (a). The energy difference between native and amyloid structures corresponding to the same mutation are given by  $\Delta G$  and shown in (b).

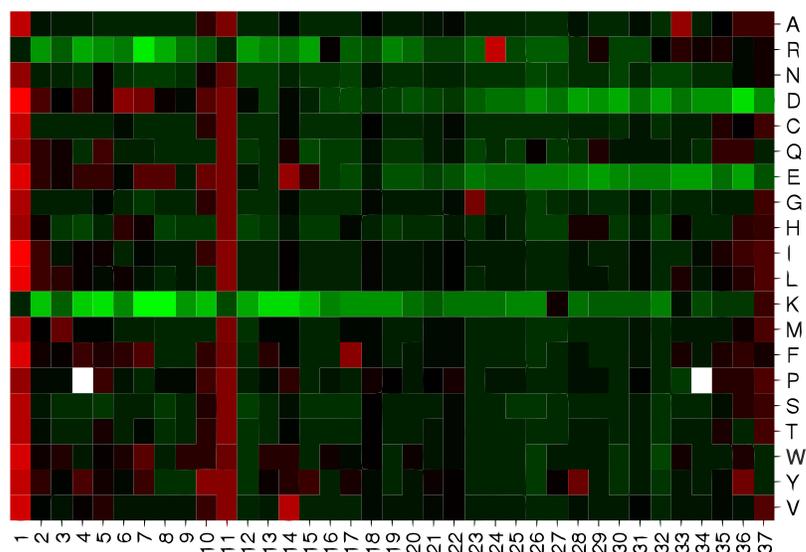
mutation effects.

#### 4.6.1 Amyloidogenicity mutational landscape analysis

The change in energy between mutant amyloid structures compared to the non-mutant amyloid structure is represented by  $\Delta E'$  as outlined in Eq. 4.4. Although  $\Delta E'$  values tell us how mutations affect the stability of amylin's amyloid form, it is more insightful to consider the change in energy between the native mutants and the amyloid mutant forms given by  $\Delta G$  from Eq. 4.5 and plotted in Fig. 4-1(b). Although  $\Delta G$  values are not correlated with the actual energy barriers between structures,  $\Delta G$  is the minimum energy required to transform a native structure into its amyloid form. The greater this lowerbound is, the bigger the gap between structures and the lower the amyloidogenicity.

Fig. 4-2(b) presents the projection of  $\Delta G$  values onto a 2-dimensional graph. Light green values represent large energy gaps between a mutant's amyloid and native structures, the intense red colored boxes represent very low  $\Delta G$  values, and the spectrum between green and red represents the energy values in between. Generating this amyloidogenicity mutational landscape allows us to easily identify point mutations that lower the gap in energy between native and amyloid form, and equally distinguishes point mutations that increase the gap in energy between native and amyloid form and decrease amyloidogenicity.

(a) 2D projection of mutational landscape stability for native amylin



(b) Amyloidogenicity landscape 2D projection

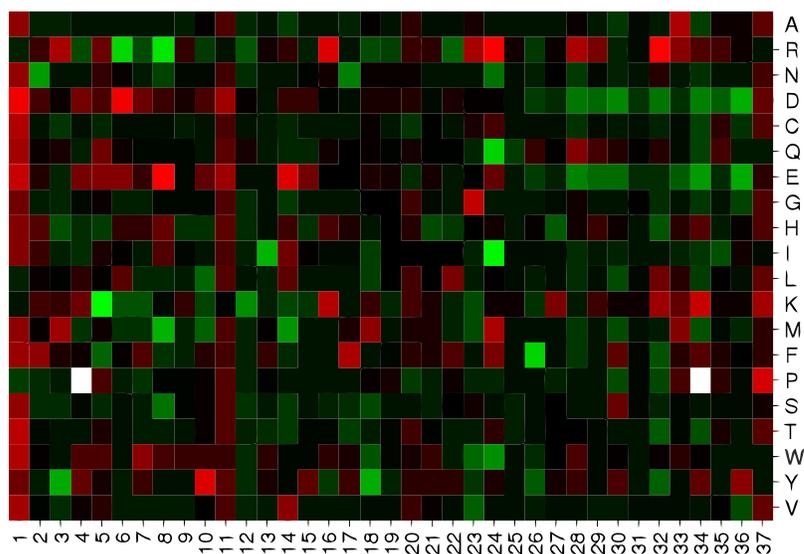


Figure 4-2: Projections of mutational landscapes. (a) The projection of Fig. 4-1(a) onto a 2-dimensional plane. The x-axis is the amino acid positions of amylin, the y-axis are the 20 possible amino acid mutations, and the color intensity represents the energy of the structure. Light green are the most stable mutants while bright red are the most unstable. (b) The projection of Fig. 4-1(b). List green are mutants with highest  $\Delta G$ , while bright red are mutants with the smallest  $\Delta G$ .

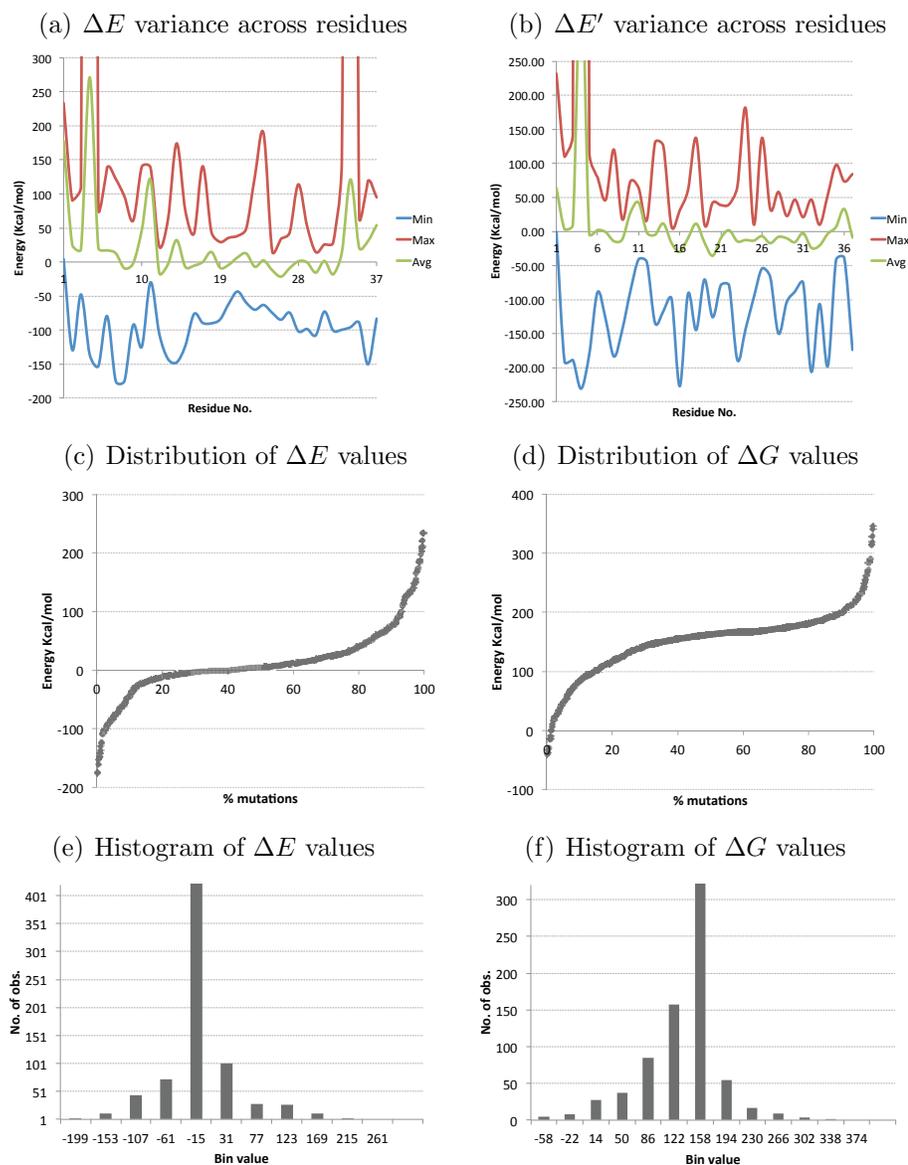


Figure 4-3: Stability analysis of mutations. Figs. (a) and (b) show the variance in energy across amylin residues for native and amyloid mutants, respectively. Close to 20 percent of mutations will stabilize amylin in (c) and most mutations show a  $\Delta G \approx 175$  Kcal/mol. Figs. (e) and (f) outline the distribution of the values in (c) and (d), respectively.

### 4.6.2 Stable and unstable regions in native amylin

To further dissect the mutational landscape and identify key mutations, Fig. 4-3(a) presents the maximum and minimum energy value perturbations caused by mutations for every residue position in amylin. The average plot is a measure of the stability susceptibility of every residue and is essentially the mean of energies produced by the 19 mutations on a single position. The stable residue positions that are immune to destabilizing mutational effects are those with an average  $\Delta E$  value close to 0 and min and max values relatively close to one another. We could see from Fig. 4-2(a) and Fig. 4-3(a) that these positions constitute the segment with positions 12-34. The local maxima on the max graph in Fig. 4-3(a) show the positions that can be altered to significantly destabilize amylin and they represent the following mutations: K1I, T4P, Q10Y, R11L, N14V, V17F, F23G, G24R, S28Y, S34P, T36Y and Y37V. Similarly, we can see that the following mutations drastically stabilize amylin: \*(5-25)K, I26D, L27E, S29E, T30D, N31E, G32D, and N35D. We highlight these residue positions on amylin's structure in Figs. 4-4(a) and 4-4(b).

Fig. 4-3(c) shows the percentage of mutations that create amylin mutant structures below a certain  $\Delta E$  value. We observe that approximately 60% of all single-point mutations destabilize amylin, while less than 10% of mutations show a drastic improvement in stability ( $\Delta E \leq -50$  Kcal/mol). Fig. 4-3(e) is a histogram of the mutational landscape showing that most mutations fall in the  $\Delta E$  bins between -15 and 77 Kcal/mol.

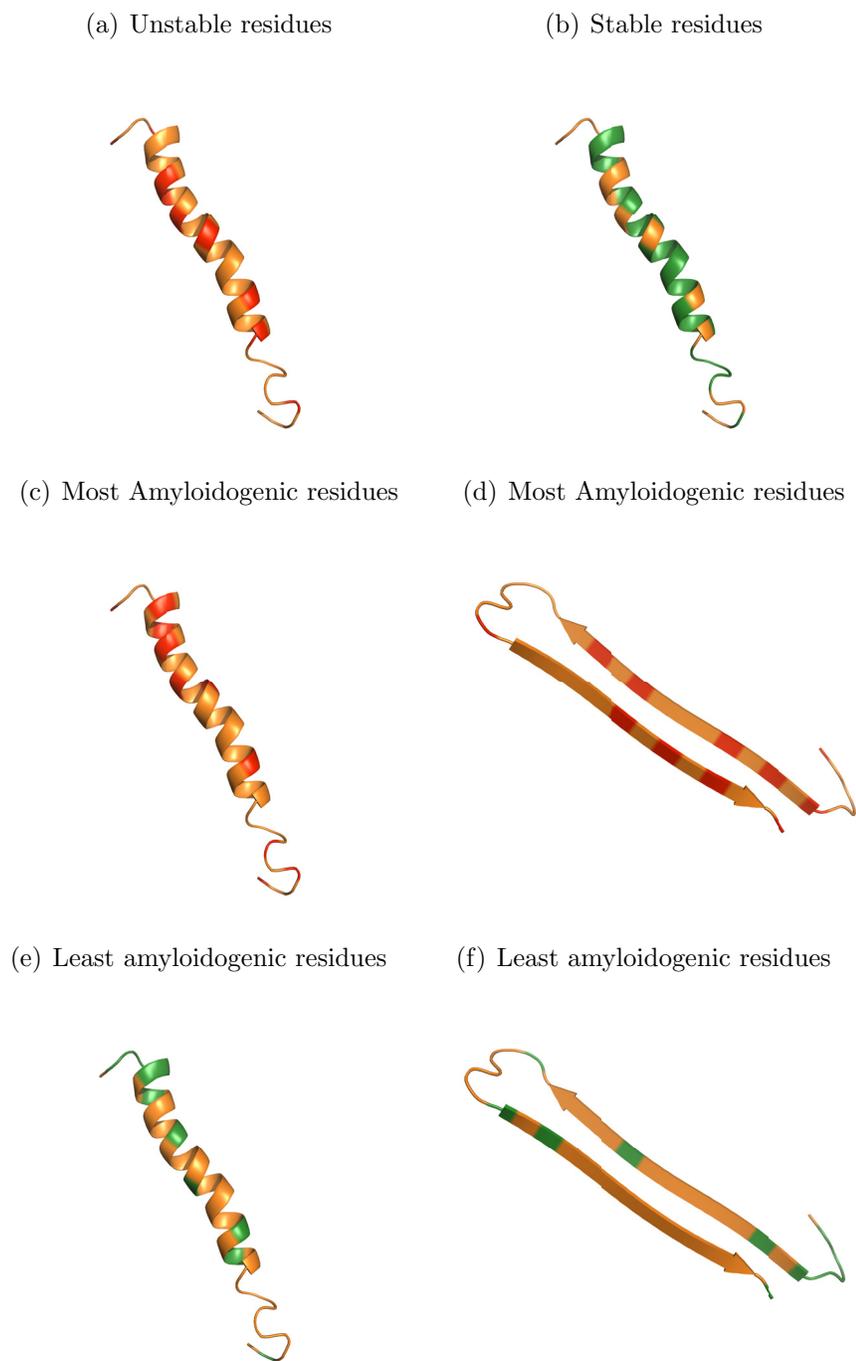


Figure 4–4: Residues that contribute to stability and amyloidogenicity. (a) The red amino acid positions (1, 4, 10, 11, 14, 17, 24, 28, and 34) can be mutated to cause the most instability. (b) The green positions that are most stable. (c) Residues (1, 6, 8, 10, 14, 16, 24, 32, 34, and 37) can be mutated to produce the smallest  $\Delta G$  values. (d) The same residues in (c) but colored on the amyloid form. (e) The green colored positions (2, 3, 4, 5, 6, 8, 13, 18, 24, 26, and 36) can be mutated to produce the highest  $\Delta G$  values and (f) shows the same residues but on the amyloid form of amylin.

### 4.6.3 Amyloidogenic regions in amylin

Similar to Fig. 4-3(a), Fig. 4-3(b) plots the energy value perturbations caused by mutations for every residue position in the amyloid form. We refer to these values as  $\Delta E'$  as defined in Eq. 4.8. For each residue position, the plot outlines the minimum, maximum, and average energy perturbations caused by the 19 mutations. Some of the positions that show resistance to destabilization include positions 10, 12, 31, 33, and 35. Fig. 4-3(d) shows the percentage of mutations that produce a  $\Delta G$  lower than a value  $y$  Kcal/mol. According to the histogram in Fig. 4-3(f), most mutations in the amyloidogenic landscape result in creating a minimum  $\Delta G$  energy barrier between 86 and 230 Kcal/mol. However, a few mutations result in negative  $\Delta G$ s, suggesting that the misfolding process can possibly undergo an exothermic reaction to produce amyloid oligomers and fibrils spontaneously.

The amyloidogenic landscape in Fig. 4-2(b) does not seem to be as smooth as the landscape in Fig. 4-2(a). It is very rugged and contains pockets of stable and unstable regions, suggesting a strong sensitivity to point mutations. A mutation into a D or an E in the region 26-36 seems to lower amyloidogenicity yet increase it in the region 1-11. Similarly, a mutation into a K in positions 5-16 lowers amyloidogenicity while increasing it in the region 27-36. Moreover, any mutation in positions 1, 11, or 37 increases amyloidogenicity and the probability of the native structure to misfold.

### 4.6.4 Evolutionary mutations

The S20G mutation has been observed to increase the risk of early and severe onset of type II diabetes in Chinese and Japanese patients [69, 153]. According to

	1-----9-----18-----27-----37	$\Delta E$	$\Delta \tilde{E}$	$\Delta E'$	$\Delta \tilde{E}'$	$\Delta G$	$\Delta \tilde{G}$
Human	KCNTATCATQLRANFLVHSSNFGAILSSSTNVGSNTY					167.6	
Monkey	-----R-----T-----D--	-125.1	-121.8	-30.3	-7.5	262.5	281.9
Cat	-----IR----L----P-----	-12.0	-11.2	0.2	39.0	179.8	217.8
Dog	-----RT---L----P-----	-6.8	-8.6	-8.2	31.5	166.2	207.8
Rat	-----R----L-PV-PP-----	3.9	18.8	-144.0	68.1	19.8	216.9
Mouse	-----R----L-PV-PP-----	3.9	18.8	-144.0	68.1	19.8	216.9
Hamster	-----N-L-PV-P-----	40.7	41.3	180.0	-67.3	306.9	59.0
G-pig	-----T-----R--H-L-A-LP-D-----	-128.7	-129.3	11.1	54.7	307.4	351.5
Degu	-----T-----R--H-L-A-PP-K-----	-82.5	-57.6	-89.6	15.3	160.5	240.6
Rabbit	-----I-----F-PPS	29.0	33.2	174.3	85.7	312.9	220.1
Hare	-----I-----F-PP-	26.6	29.1	165.7	174.0	306.7	312.5
Cougar	-*-----*-----IRSS*-----*	-38.9	-27.5	12.3	10.2	218.8	205.2
Pig	-M-----H-----DR-R-L:-T-F-P-K---	40.3	27.0	-102.5	-163.1	24.9	-22.5
Pram	-----P--PP-----	39.5	51.2	35.0	46.7	163.1	163.1

Table 4–3: Evolutionary amylin mutations in species. Amylin sequence alignment across 13 species. Amino acids shared with the human sequence are indicated with a (-), (\*) are amino acids that are unknown, and (: ) in the Pig sequence is a missing residue.  $\Delta \tilde{G}$ ,  $\Delta \tilde{E}'$ , and  $\Delta \tilde{E}$  estimates are computed using Eqs. 6-8.

the mutation and amyloidogenic landscapes, the S20G mutation has a negative  $\Delta G$  value (-55.8 Kcal/mol) and a slight  $\Delta E$  instability of value (5.4 Kcal/mol). The negative  $\Delta G$  could explain the emergence of amyloids and early onset of diabetes for these patients. The Q10R mutation observed in the New Zealand Maori population [73] also exhibits a negative  $\Delta G$  value of -11.3 Kcal/mol, yet a more stable  $\Delta E$  value of -40.6 Kcal/mol. The S20K experimentally explored mutation lengthened the lag phase by a factor of 18 before misfolding into amyloids [72]. In our results, the S20K mutation improves the stability of native amylin with a  $\Delta E = -61.1$  Kcal/mol and significantly lowers the  $\Delta E'$  resulting in a negative, highly amyloidogenic  $\Delta G = -125.7$  Kcal/mol.

In Table 4–3, we calculate the  $\Delta E$ ,  $\Delta E'$ , and  $\Delta G$  values for the mutant amylin structures of twelve species and for the recently administered diabetes drug Pramlintide. We find that regardless of the number of point mutations, the  $\Delta E$  value

is small and between -128.7 and 40.7 Kcal/mol. In addition, according to our generated mutational landscape, each single-point mutation in these species does not individually destabilize the native structure. In fact, many of the mutations individually contribute to the increase of  $\Delta G$ , suggesting that the evolutionary mutations do not prefer destabilizing the native form of amylin. The  $\Delta G$  values range from 20 to 312 Kcal/mol. Although some of these species produce amyloids and some do not, it is difficult to explain the phenomenon of amyloid aggregation without accurately estimating the energy barrier and required activation energy between native and amyloid forms. This is specially evident for the  $\Delta G$ s of Pramlintide and the Rat and Mouse which supposedly do not form fibrils in normal physiological conditions.

It is intriguing to note that while computing the mutational landscape for multi-point mutations is extremely expensive in time and resources, we found that calculating multi-point mutations by using sums of single-point mutations provides a reasonable estimate for stability in the various species. We defined estimates of  $\Delta E$ ,  $\Delta E'$ , and  $\Delta G$  as  $\Delta\tilde{E}$ ,  $\Delta\tilde{E}'$ , and  $\Delta\tilde{G}$ , respectively, as shown in Eqs. 4.7-4.9. We computed the values of the estimates for the twelve species in Table 4-3 and observed very close results to the real values. We believe the small discrepancy in values are a result of not taking into account the long-range interactions caused by multiple-point mutations on the entire structure. Nevertheless, this method allowed us to explore the landscapes of multi-point mutations and estimate the energy results of millions of mutations efficiently.

#### 4.6.5 Efficient estimation of stability and amyloidogenicity in multiple-point mutation landscapes

Working with the assumption that  $\Delta\tilde{E}$  and  $\Delta\tilde{G}$  values are good estimates for stability and amyloidogenicity in multiple-point mutation landscapes, we embarked on exploring the 3-point mutational landscape of amylin and assessing whether a more stable and less amyloidogenic structure than Pramlintide can be found. Such structure can potentially be a better therapeutic agent than Pramlintide, producing less amyloid fibrils and affecting beta cells less. We restrict our search to 3-point mutations to limit the compounding error on the estimation of  $\Delta\tilde{E}$  and  $\Delta\tilde{G}$  values. Using Eq. 4.6 and Algorithm 2 of Table 4–2, we generated the  $\Delta\tilde{E}$  and  $\Delta\tilde{G}$  of all 62,160,000 distinct 3-point mutations in amylin and sorted them by descending  $\Delta\tilde{G}$  and  $\Delta\tilde{E}$  values. This returned the mutants with the most stability in native form and the highest minimum required activation energy for misfolding into amyloid form. Using this method Pramlintide ranked in the top 32% of all mutations in stability and amyloidogenicity. There were millions of other candidate analogs that showed better stability. We report the best 20 results in Table 4–4 and the top 1000 results in Table S1. Algorithm 2 of Table 4–2 can be used to explore the n-point mutational landscape of any protein. However, for the reason of finding the most parsimonious mutations in amylin and reducing the complexity of the search space (see Eq. 4.6), we capped our search at 3-point mutations.

1-----9-----18-----27-----37	$\Delta\tilde{E}$	$\Delta\tilde{G}$	Mutations
KCNTATCATQRLANFLVHSSNFGAILSSTNVGSNTY			
----K--R-----I-----	-254.5	681.2	A5K A8R G24I
----KR-----I-----	-215.7	670.8	A5K T6R G24I
----K-----I-F-----	-154.3	667.1	A5K G24I I26F
----KR-R-----	-333.5	658.6	A5K T6R A8R
----K--R-----F-----	-272.1	654.9	A5K A8R I26F
----K--R-----Q-----	-275.4	653.9	A5K A8R G24Q
----R-R-----I-----	-172.5	653.4	T6R A8R G24I
----R-----I-F-----	-111.2	649.7	A8R G24I I26F
----KR-----F-----	-233.3	644.5	A5K T6R I26F
----KR-----Q-----	-236.6	643.5	A5K T6R G24Q
----K-----I-----I-----	-137.4	643.4	A5K A13I G24I
----K--M-----I-----	-146.1	641.6	A5K A8M G24I
----K-----Q-F-----	-175.3	639.8	A5K G24Q I26F
----R-----I-F-----	-72.4	639.3	T6R G24I I26F
----K-----I-----D-----	-295.0	639.0	A5K G24I T36D
----K-----Y-----I-----	-125.7	638.0	A5K H18Y G24I
----K-----I-----E-----	-247.4	636.1	A5K G24I T36E
--Y-K-----I-----	-118.3	636.0	N3Y A5K G24I
----K--R----I-----	-255.3	631.3	A5K A8R A13I
----R-R-----F-----	-190.2	627.1	T6R A8R I26F

Table 4–4: Top 20 results for amylin 3-point mutations. Pramlintide (3 mutations of 25P, 28P, and 29P) ranked no. 19,918,936 out of all the 62,160,000 at the top 32% of all 3-point mutations for amylin stability and amyloidogenicity. The 3-point mutation landscape incorporates the 1-point and 2-point mutational landscapes in its results.

#### 4.6.6 Mutant (A5K, A8R, G24I) is stable and less amyloidogenic

The top result in the 3-point mutational landscape of amylin has a  $\Delta\tilde{E}$  of -254.5 and a  $\Delta\tilde{G}$  of 681.2 Kcal/mol, suggesting a super stable native structure and an extremely large gap between native and amyloid forms. The large gap could possibly indicate the high instability of the amyloid form. We performed a complete Molecular Dynamics simulation on both native and amyloid structures of this mutant and found that indeed the native structure is very stable while the amyloid structure is unstable to a point that it unfolds out of its beta-structure and starts to form helices again. Figs. 4-5(a) and 4-5(b) show the *RMSD* results for the native and amyloid forms. The native *RMSD* is low while the amyloid is extremely high. Figs. 4-5(c) and 4-5(d) show the *RMSF* plots for the native and amyloid and Figs. 4-5(e) and 4-5(f) show the final shape of the native and amyloid structures after 18n seconds of MD simulation. It is clear from the graphs that the amyloid form has misfolded and is highly unstable and unfavorable, while the native maintained its structure.

#### 4.7 Conclusion

The misfolding of amylin molecules in diabetes patients contributes to the  $\beta$ -cell death and deterioration of the pancreas. The sensitivity of amylin to point mutations has been exploited to design the drug Pramlintide, an amylin analog that is more stable and less amyloidogenic than amylin. However, because amylin mutation studies are sparse and because no reported systemic analysis of all the amino acid positions or their amyloidogenicity potentials exists [139], it is highly likely that there are many other amylin analogs that could perform better than Pramlintide. Testing

(a) Native mutant after 18ns MD run (b) Amyloid mutant after 18ns MD run

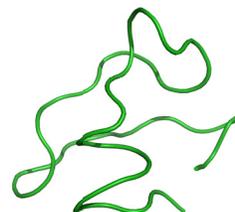
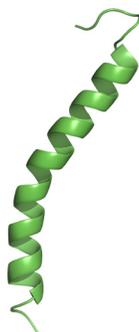
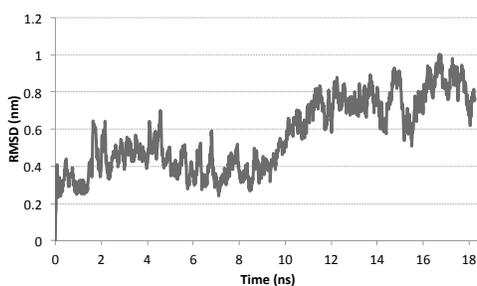
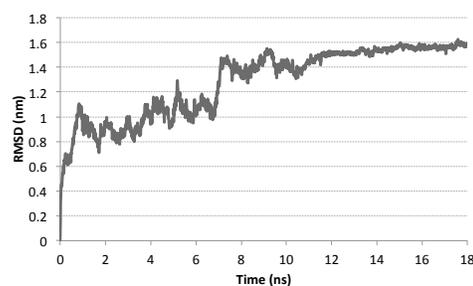
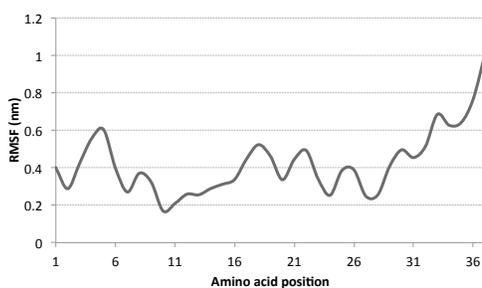
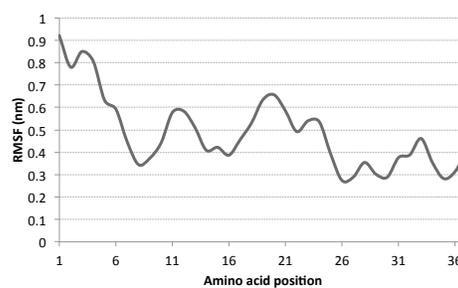
(c) *RMSD* graph for native mutant(d) *RMSD* graph for amyloid mutant(e) *RMSF* graph for native mutant(f) *RMSF* graph for amyloid mutant

Figure 4–5: MD results of mutation (A5K, A8R, G24I). (a) The final mutant native structure after 18ns of MD simulation. The structure looks stable and according to its low *RMSD* values in (c) and low *RMSF* values in (e), the structure is believed to be conserved. (b) The final mutant amyloid structure after 18ns of MD simulation. The structure has collapsed on itself and broken its two beta-strands to create multiple alpha helices. (d) The *RMSD* values of the mutant amyloid and (f) the *RMSF* values of the mutant amyloid are high and suggest instability and severe deviation in structure.

experimentally for these mutations is not feasible as the number of possible tests are innumerable. Using the dipolar water model that calculates precisely the free energies of proteins, we constructed the single-point mutational and amyloidogenic landscape of amylin and reported the effects of any and all mutations on amylin's stability and amyloidogenicity potential. We found that increasing the number of point mutations can decrease amyloidogenicity. This motivated us to generate millions of 3-point mutations and explore mutants that show a higher stability than Pramlintide and lower amyloidogenicity potential. With the use of Molecular Dynamics packages, promising mutants can be further tested for stability prior to lab experimentation. The tool we developed to produce the landscape values, MAPOR, is freely available at <http://amyloid.cs.mcgill.ca>. It can be used to generate the landscapes of other proteins and explore their n-point mutations.

#### 4.8 Acknowledgments

M.R.S. and J.W. designed the study and analyzed results. M.R.S. carried out the experiments, developed code and drafted the manuscript. M.R.S. and J.W. read and approved the final manuscript. *Funding:* M.R.S. was supported by a fellowship from the Canadian Institutes of Health Research System Biology Training program at McGill University and the from the Fonds de recherche du Qubec (FQRNT). J.W. was supported by a Discovery grant from the Natural Science and Engineering Research Council of Canada.

#### **4.9 Supplementary material**

Table S1 from the supplementary material is too long to include in this thesis.

To view the Table online, please visit: <http://amyloid.cs.mcgill.ca/Suppl3>

## CHAPTER 5

### Probing the binding affinity of amyloids to reduce toxicity of oligomers in diabetes

#### 5.1 Preface

Obesity, unhealthy diet, and low physical activity can eventually lead to type II diabetes, which affects around 350 million people worldwide. It is a disease that results in the failure of the pancreas to produce enough insulin and amylin proteins in the human body. This failure of meeting the body's insulin and amylin demands results in high glucose concentration in the blood and low glucose absorption into the heart, muscles and adipose tissue. The pancreas fails to produce enough insulin due to the death of its overworked and overstressed  $\beta$ -cells. As such, patients with this disease are required to inject insulin and amylin replacements daily to survive.

Because amylin has a tendency to misfold into amyloid aggregates that further infect  $\beta$ -cells, leading to an accelerated failure of the pancreas, we explored in the previous chapter the mutational landscape of amylin to understand the effect of mutations on its stability and amyloidogenicity. The landscapes revealed regions in the amylin protein that were conserved and regions prone to instability. We suggested improving the efficiency of the Pramlintide drug by engineering a novel amylin analog

with stronger stabilizing mutations. The increase in stability will further reduce the rate of misfolding into amyloid structures.

In this chapter, we explore the reverse problem. In diabetes, short amyloid aggregates (oligomers) in the pancreas are toxic. These oligomers inject themselves into membranes of  $\beta$ -cells in the pancreas and create pores that disrupt the electric potential across the membranes. However, long aggregates (fibrils) do not permeate cells and are not toxic. The idea behind the work in this chapter is to design a molecule (amylin mutant analog) that forces the short oligomers to extend their length. We engineer a molecule that binds very strongly to the oligomer extension surface and induces aggregation. The extension lowers the toxicity created by oligomer pores by slowly restoring the disrupted electric potential that exists across the membrane to its equilibrium. The uncontrolled flow of ions is hindered by lengthening the oligomer structures. We design the molecule to be administered with Pramlintide injections. By design, the molecule is engineered to repel from Pramlintide and not interact with it, eliminating any potential for Pramlintide to misfold. Upon injection into the body, the molecule performs its amylin normal functions. When the molecule reaches the pancreas, it should misfold into amyloid shapes and bind to the oligomers.

We extend the insulin-glucose mathematical model to present a novel system of integral equations that model the effect of amyloid production on the insulin-glucose relationship. The model is the first in incorporating the effect of amyloids in diabetes and shows how reducing the toxicity of oligomers keeps the pancreas in a healthier state.

The remaining content of this chapter has been taken from a manuscript that was submitted for publication:

- M. R. Smaoui, H. Orland, and J. Waldispühl. “Probing the binding affinity of amyloids to reduce toxicity of oligomers in diabetes”. *Submitted for review in Sept 2014*.

## 5.2 Abstract

**Motivation:** Amyloids play a role in the degradation of  $\beta$ -cells in diabetes patients. In particular, short amyloid oligomers inject themselves into the membranes of these cells and create pores that disrupt the strictly controlled flow of ions through the membranes. This leads to cell death. Getting rid of the short oligomers either by a deconstruction process or by elongating them into longer fibrils will reduce this toxicity and allow the  $\beta$ -cells to live longer.

**Results:** We develop a computational method to probe the binding affinity of amyloid structures and produce an amylin analog that binds to oligomers and extends their length. The binding and extension lower toxicity and  $\beta$ -cell death. The amylin analog is designed through a parsimonious selection of mutations and is to be administered with the Pramlintide drug, but not to interact with it. The mutations (T9K L12K S28H T30K) produce a stable native structure, strong binding affinity to oligomers, and long fibrils. We present an extended mathematical model for the insulin-glucose relationship and demonstrate how affecting the concentration of oligomers with such analog is strictly coupled with insulin release and  $\beta$ -cell fitness.

**Availability:** SEMBA, the tool to probe the binding affinity of amyloid proteins and generate the binding affinity scoring matrices and  $R$ -scores is available at: <http://amyloid.cs.mcgill.ca>

**Contact:** [jeromew@cs.mcgill.ca](mailto:jeromew@cs.mcgill.ca)

### 5.3 Introduction

Insulin and amylin molecules are co-secreted from  $\beta$ -cells in the pancreas to lower the concentration of glucose in the blood. Some of the amylin misfold into amyloids and do not continue to carry on their normal functions in the body. In type II diabetes, these amylin amyloids are observed to form in the extracellular space surrounding  $\beta$ -cells and build up into different sizes, some small (oligomers) and others large (fibrils) [152]. The exact effect of the very large fibril deposits on the progression of diabetes and the inflammation of  $\beta$ -cells is not really known. However, amyloid oligomers have been clearly observed to be toxic, while the longer fibrils are non-toxic [159]. These aggregates inject themselves into cell membranes and create weakly-selective pores that introduce an uncontrolled influx of ions into and out of the cell [160, 161]. Their short length allows for ions to pass through them easily, unlike fibrils. The influx of ions is mainly due to the high concentration gradient of calcium molecules separating the cytoplasm and extracellular space. This influx disrupts cell coupling, impairs insulin secretion, depletes ATP, depolarizes the membrane, and weakens cells inducing apoptosis [162]. Upon cell breakdown, the oligomers find other live cells to target [159]. Fig. 5-1 shows an illustration of amylin oligomers permeating a cell membrane.

When too many amylin molecules misfold into amyloids, due to insulin resistance and extreme pressure on the pancreas to secrete insulin, the required levels of amylin in the body to modulate gastric emptying and prevent postprandial spikes in glucose levels are not met. This further complicates the conditions of type II diabetes patients. Several injectable amylin analogs have been created in an effort to find a suitable replacement for this lack of the necessary levels of amylin. The most promising one, Pramlintide, is an analog containing 3 proline substitutions at positions 25, 28 and 29, and reveals a weak tendency to aggregate [163]. Pramlintide is an effective amylin replacement agent that acts as a synergetic partner to insulin [164]. Pramlintide does not enter the cells of the pancreas, however, it moves through the circulatory system to reach necessary organs and compensates for the lack of sufficient amylin levels.

Although Pramlintide has a weaker tendency to aggregate, and compensates for amylin's lost functions, it does not solve the problem of amyloid oligomer toxicity. The inhibition of the amyloid production has been studied as a way to elongate cell life in diabetes. Molecules like rifampicin have been designed to prevent fibril formation [165], but failed to remove the short toxic oligomers that speed up the death of cells [159]. Other attempts at designing inhibiting molecules include the red wine compound resveratrol [166], Vanadium complexes [167], Polyphenols [168, 169], selenium-containing phycocyanin molecules [170], and N-methylation of amylin [171]. Many of these molecules presented strong inhibiting results that lower the rate of oligomerization *in vitro* and are currently undergoing more experimentation. However, none of them are at a stage of being produced as a therapeutic agent.

In this work we explore a different technique to reduce the toxicity of oligomers. Instead of inhibiting fibril extension, we promote it. At a higher rate of amyloid aggregation, longer structures are created. The higher the rate, the lower the concentration of oligomers and the higher the concentration of fibrils. We aim to achieve this high aggregation rate by constructing a molecule that when injected into the human body, binds to these short oligomers and extends their length, promoting aggregation and forcing them into long non-toxic fibrils. A circumspect approach to create such molecule that does not perform unknown and undesirable interactions in the body is to engineer the molecule as an amylin analog with a parsimonious selection of mutations to perform the task. We construct this molecule as an analog of high amyloidogenicity, to have very strong binding affinity to the amyloids, and to have an extremely low dissociation rate. The low dissociation rate is essential in preventing the emergence of more short oligomers through fibrils breaking up. This analog is designed to be administered with the Pramlintide drug in a native, non-amyloid form. The analog is engineered to have a low binding affinity to Pramlintide to prevent inducing Pramlintide to misfold. To conserve structure, the analog is designed through a parsimonious selection of mutations.

#### **5.4 Approach**

Ridding the pancreas from toxic, short amyloid oligomers will reduce  $\beta$ -cell death. In theory, we could remove these oligomers in three ways:

1. Altering the environment around amyloids to reduce the rate of nucleation and inhibit oligomerization.

2. Designing molecules that bind to the ends of oligomers to stop the flow of ions through pores and to inhibit amyloid extension.
3. Promoting amyloid extension and fibril growth to reduce oligomer toxicity.

In this work, we develop a novel method to assist researchers with solutions 2 and 3. Although we do not focus on exploring molecules to inhibit amyloid growth, our work will show how the method could be used to assess the binding preferences and affinity of amyloids to other molecules. In particular, we use the method to improve the binding of amylin amyloids to other amyloids, stimulating growth and extension. Promoting amyloid extension forces oligomers to grow into fibrils and decreases their toxicity.

Amyloids inherently bind to each other. They have a strong affinity to aggregate. Introducing the right amino acid mutations into amyloids can increase their binding affinity and likelihood to aggregate. This increase results in longer structures, forcing short oligomers to grow into fibrils. The Binding Affinity Scoring Matrix (BASM) method we introduce in this paper explores the affinity of amyloid residues to all possible amino acid bindings. We explore the effect of pairing a single amino acid with an amyloid structure at a specific residue position. For each of the residue positions of an amyloid, we generate 20 binding affinity readings, one for every position and amino acid pair, as illustrated in Fig. 5-2. The affinity measures the Lennard-Jones potential between the amino acid and the amyloid structure, the Coulomb interactions, and the Solvation energy of the system. We organize the readings into a  $n \times 20$  matrix. The first dimension of the matrix comprises the 20 amino acids, the second dimension is the protein residue positions, and each record is what we call an

$R$ -score. The lower the  $R$ -scores, the stronger the binding affinity between an amino acid and the amyloid structure at the specific residue position.

The goal behind the BASM method is to identify bindings that increase affinity. The best structure bindings could be used to engineer an amylin analog that produces amyloids that bind better to the ones in the pancreas. If designed with the proper amyloidogenicity, the engineered analog will not only bind well to oligomers, but will also promote amyloid extension. This extension reduces oligomer toxicity and  $\beta$ -cell death.

The analog can be administered in native, non-amyloid form with Pramlintide injections. In its native form, it will carry out the normal functions of amylin and Pramlintide. Once it reaches the pancreas (through the bloodstream) and makes contact with the amyloids, it should misfold and aggregate with the surrounding amyloids and oligomers due to its high amyloidogenicity. To prevent any potential aggregation with Pramlintide, the analog is designed with lowest affinity to Pramlintide. In the Methods section we explain in detail how we develop this analog by generating the BASM  $R$ -scores for amylin and Pramlintide amyloids.

## 5.5 Methods

### BASM $R$ -scores

The free energy of a system,  $F_E$ , that includes an amyloid protein and an amino acid separated far apart from each other can be given by,

$$F_E = (LJ_p + Coul_p) + (LJ_a + Coul_a) + S \quad (5.1)$$

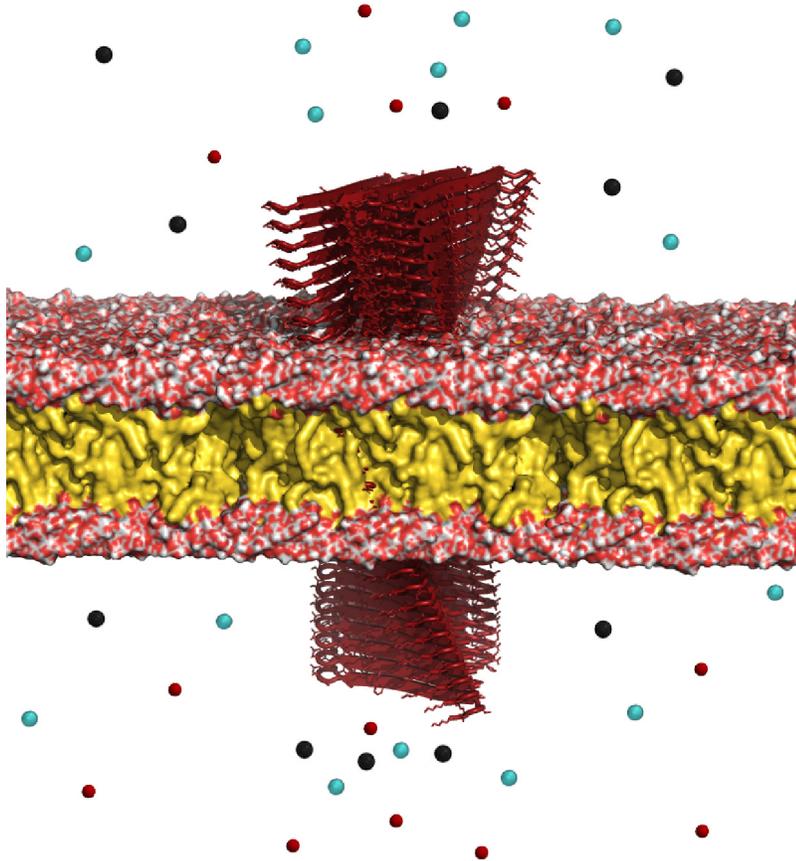


Figure 5-1: Oligomer toxicity. Amylin amyloid oligomers (red) permeating the cell membrane and disrupting the tightly controlled flow of ions through the membrane. Figure created with PyMol [172]

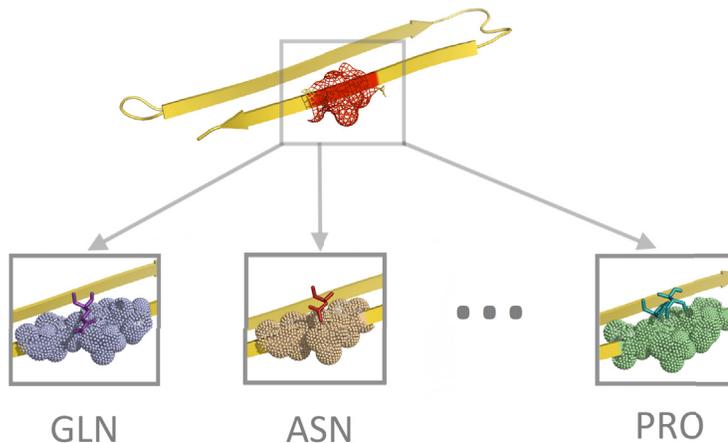


Figure 5–2: Binding affinity method. Twenty amino acids are placed on top of a specific residue position on the amylin protein, one at a time, to compute the  $R$ -scores in Eq. 5.4

where  $LJ$  is the Lennard-Jones potential,  $Coul$  is the Coulomb energy, and  $S$  is the solvation energy of the entire system. The terms with a  $p$  subscript denote the energies of the amyloid protein and the terms with an  $a$  subscript denote energies of a separate amino acid. If we alter the system such that we place the amino acid in close proximity to the protein to interact with it, the free energy becomes

$$\tilde{F}_E = (LJ_p + Coul_p) + (LJ_a + Coul_a) + (LJ_{ap} + Coul_{ap}) + S_{ap} \quad (5.2)$$

where  $LJ_{ap}$  and  $Coul_{ap}$  are the Lennard-Jones and Coulomb terms resulting from the interaction of the atoms of the amino acid with the atoms of the protein, respectively.  $S_{ap}$  is the solvation energy of the coupled system. Subtracting Eq. 5.2 from 5.1 yields the residue binding affinity,  $R$ , of the amino acid to the protein at the position it

was coupled at and gives,

$$R = LJ_{ap} + Coul_{ap} + S_{ap} - S \quad (5.3)$$

or in more formal terms,

$$R(N_1, N_2) = \sum_{i=1}^{N_1} \sum_{j>i}^{N_2} 4\epsilon \left[ \left( \frac{\sigma}{r_{ij}} \right)^{12} - \left( \frac{\sigma}{r_{ij}} \right)^6 \right] + \sum_{i=1}^{N_1} \sum_{j>i}^{N_2} \left[ f \frac{q_i q_j}{\epsilon_r r_{ij}} \right] + S_{ap} - S \quad (5.4)$$

where  $N_1$  are the atoms of the amino acid and  $N_2$  are the atoms in the protein.  $\epsilon$  is the depth of the Lennard-Jones potential well,  $\sigma$  is the finite distance at which the inter-particle potential is zero,  $r_{ij}$  is the distance between the protein atoms  $i$  and  $j$ , and  $q_i$  and  $q_j$  represent the charges of the two atoms. The solvation terms are given by

$$S = F_{(p_0, C_{dip})} - F_{(0,0)} - \left( k_B T \frac{\ln(1 - N_A C_{dip} a^3)}{N_A C_{dip} a^3} \right) \int_{solvent} d\mathbf{r} \rho_{dip}(\mathbf{r}). \quad (5.5)$$

where  $F_{(p_0, C_{dip})}$  defines the free energy of the system defined at dipoles of moment values  $p_0$  and concentration  $C_{dip}$ ,  $F_{(0,0)}$  the free energy of the system with solvent concentration set to zero,  $a^3$  is the lattice grid size volume of the solvent,  $k_B$  is the Boltzmann constant,  $T$  temperature in Kelvin, and  $r$  is the surface definition, solvent-accessible surface probe. The lower the  $R$ -scores, the stronger the binding affinity of an amino acid to a particular position within the protein.

### 5.5.1 BASM matrix

Let  $\mathbf{M}$  be the set of canonical amino acids given by,

$$\mathbf{M} = \{A, R, N, D, C, Q, E, G, H, I, L, K, M, F, P, S, T, W, Y, V\} \quad (5.6)$$

we construct a protein’s residue-Binding Affinity Scoring Matrix (BASM) using the  $R$  scores in Eq. 5.4. In particular, for every residue position  $i$  of a protein  $p$ , we compute twenty  $R$  scores, one for each  $m \in \mathbf{M}$  computed with

$$BASM[i, m] = R(N_p, N_m), \quad 1 \leq i \leq n, m \in \mathbf{M} \quad (5.7)$$

To physically model this system where we have an amino acid interacting with the amyloid, we start with a PDB of dimer amyloid structure. We mutate an amino acid in one of the amyloid structures using SCWRL [133], and remove all the non-mutated amino acids from that amyloid. This leaves an amino acid placed on top of an amyloid monomer at the right binding distance.

### Binding affinity function

Let  $S$  be the sequence of amino acids representing a protein with length  $n$  such that  $S \in \mathbf{M}^n$  and  $B$  the residue binding affinity scoring matrix (BASM) of sequence  $S$ . We compute,  $F(R)$ , the global-sequence binding affinity function of  $B$  to a sequence of amino acids  $R \in \mathbf{M}^n$  by,

$$F(R) = \sum_{k=1}^n B[k, R_k] \quad (5.8)$$

where  $R_k$  is an amino acid at position  $k$ . We can also compute partial-sequence binding affinities of  $B$  to  $R$  and point-binding affinity functions by,

$$F(R)_{i,j} = \sum_{k=i}^j B[k, R_k] \quad (5.9)$$

$$F(R)_k = B[k, R_k] \quad (5.10)$$

respectively.  $i$  and  $j$  are the starting and ending amino acid positions of subsequences  $S$  and  $R$ .

To construct a sequence  $S^*$  with optimal BASM binding affinity to  $S$ , we must find  $F(S^*)$  such that

$$F(S^*) = \min F(R), \quad \forall R \in \mathbf{M}^n \quad (5.11)$$

Because the values in a BASM matrix are independent of one another,  $F(S^*)$  can be computed by

$$F(S^*) = \sum_{i=1}^n \left[ \min_{j \in \mathbf{M}} B[i, j] \right] \quad (5.12)$$

and the sequence  $S^*$  can be constructed by,

$$S_i^* = S_{i-1}^* + J_i, \quad 1 \leq i \leq n \quad (5.13)$$

where  $J_i \in \mathbf{M}$  and  $J_i$  satisfies

$$B[i, J_i] = \min_{j \in \mathbf{M}} B[i, j], \quad 1 \leq i \leq n \quad (5.14)$$

### Finding best fit to amylin and worst to Pramlintide

Let  $B^1$  be the BASM of the amylin sequence  $S^1$  and  $B^2$  the BASM of the Pramlintide sequence  $S^2$ . To construct a sequence with strong binding affinity to the sequence of amylin amyloids  $S^1$  and simultaneously weak binding affinity to Pramlintide amyloid sequence  $S^2$ , we generate a BASM  $\bar{B}$  that satisfies this constraint

and combines  $B^1$  and  $B^2$  in the following way,

$$\bar{B} = \alpha * B^1 - \beta * B^2 \quad (5.15)$$

where  $\alpha$  and  $\beta$  can be considered as amyloid and pramlintide affinity coefficients, respectively. Because we aim to find a binding sequence that produces very stable bonds with amyloids and simultaneously, very unstable bonds with pramlintide, setting  $\frac{\alpha}{\beta} = 1$  generates a  $\bar{B}$  with amino acid affinities that prefer amylin amyloids as much as they repel Pramlintide. Similarly, choosing  $\frac{\alpha}{\beta} > 1$  results in a sequence that binds better to amylin amyloids and repels Pramlintide less.

If we let  $\bar{F}$  be the sequence binding affinity function for  $\bar{B}$ , we can construct  $S^*$  for this system by using Eqs. 5.12 - 5.14.  $S^*$  might contain 1 to  $n$  differences in amino acids compared to amylin's sequence,  $S^1$ . However, too many differences in amino acids might create a structure that differs from amylin and does not bind to the amyloids. In order to guarantee that we have a sequence that represents a structurally conserved, folded amylin analog, we will restrict the number of possible differences between  $S^*$  and  $S^1$ . Using  $S^*$ , we will construct a parsimonious sequence  $S^P$ .

If we allow for only 1 difference in sequence between  $S^*$  and  $S^1$ , we can generate a 1-residue parsimonious sequence  $S^{P_1}$  such that,

$$\bar{F}(S^{P_1}) = \min_{1 \leq i \leq n} \{F(S^1)_{1,i-1} + \bar{F}(S^*)_i + F(S^1)_{i+1,n}\} \quad (5.16)$$

where  $F$  is the sequence binding affinity function for  $B^1$  and

$$S^{P_1} = S_1^1 S_2^1 \cdots S_{l-1}^1 S_l^* S_{l+1}^1 \cdots S_n^1 \quad (5.17)$$

where  $l$  is the  $i$  value that returns the minimum in Eq. 5.16. We can generalize Eq. 5.16 to allow for  $k$  differences in sequence between  $S^*$  and  $S^1$  and generate a  $k$ -residue parsimonious sequence  $S_k^P$  by solving

$$\min_{i_1, i_2, \dots, i_k} \{F(S^1)_{1, i_1-1} + \sum_{j=1}^n \bar{F}(S^*)_{i_j} + \sum_{j=1}^n F(S^1)_{i_{j+1}, i_{(j+1)}-1} + F(S^1)_{i_{k+1}, n}\} \quad (5.18)$$

where  $1 \leq i_1 < i_2 < \dots < i_k \leq n$ .  $S^{P_k}$  can be constructed by

$$S^{P_k} = S_1^1 \cdots S_{l_1-1}^1 S_{l_1}^* S_{l_1+1}^1 \cdots S_{l_2-1}^1 S_{l_2}^* S_{l_2+1}^1 \cdots S_{l_k-1}^1 S_{l_k}^* S_{l_k+1}^1 \cdots S_n^1 \quad (5.19)$$

where  $l_1, l_2, \dots, l_k$  are the  $i$  values that solve for the minimum in Eq. 5.18. Applying the BASM technique on amylin will generate an optimal binding sequence for chosen  $\alpha$  and  $\beta$  values. Ensuring a parsimonious selection of mutations increases the complexity of the problem and reduces the optimality of the binding affinity for amyloids. Nonetheless, this reduces the risk of losing structure and function.

### 5.5.2 Amyloidogenicity

In solving for the minimum binding score in Eq. 5.18 for amylin, we pick the top 10 results (Table 5-2). Out of the 10 results, we pick the one that has the highest potential to misfold into amyloids and has the lowest dissociation potential. The higher the amyloidogenicity, the more likely the analog will misfold in the presence of amyloids and aggregate. The lower the dissociation, the less oligomers will form out of the analog. We estimate a lower bound on amyloidogenicity by calculating  $\Delta G$ , the difference in energy (LJ, Coulomb and Solvation) between the amyloid and native forms as follows,

$$\Delta G = E_{amyloid} - E_{native} \quad (5.20)$$

Although  $\Delta G$  does not reflect the true energy barrier between native and amyloid form, we use it to estimate a lower bound on the energy needed for a structure to misfold.

### 5.5.3 Oligomer concentrations

Out of the set of solutions for Eq. 5.18 we choose the parsimonious solution that has the lowest dissociation potential, producing the lowest number of oligomers. For each solution, we measure the concentration of oligomers they might create. We can quantitatively express these concentrations by computing the free energies of the oligomers. The expression for the concentration of these structures in terms of the chemical potential of the system  $\mu$ , the free energies of oligomers  $F_k$ , and the number of monomers  $k$  at each state can be computed by the law of mass action.

Let's denote by  $N$  the initial number of proteins in the solution,  $V$  the volume of the system,  $N_k$  the number of oligomers of type  $k$ . Let  $a^3$  be the volume of a protein. The initial volume fraction  $\phi$  of the proteins is defined by  $\phi = Na^3/V = ca^3$  where  $c = N/V$  is the initial concentration of proteins and the volume fraction of the oligomers of type  $k$  by  $\phi_k = N_k k a^3 / V$ . Note that all  $\phi_k$  and  $\phi$  must be positive and smaller than 1.

The partition function of the system is written as

$$\begin{aligned} Z &= e^{-\beta F} \\ &= \sum_{N_1 \dots N_M \dots = 0}^{\infty} e^{\Lambda + \Gamma} \end{aligned} \quad (5.21)$$

where  $\beta = 1/(k_B T)$  is the inverse temperature and  $F$  is the total free energy of the system and  $\Lambda$  and  $\Gamma$  are

$$\Lambda = \beta \mu \sum_k k N_k - \sum_k N_k \log \frac{N_k}{e} \quad (5.22)$$

$$\Gamma = \sum_k N_k \log \frac{V}{k a^3} - \beta \sum_k N_k F_k \quad (5.23)$$

The expression for the concentrations in Eq. 5.21 is obtained by minimizing the exponent with respect to the  $N_k$

$$\beta \mu k - \log N_k + \log \frac{V}{k a^3} - \beta F_k = 0 \quad (5.24)$$

or equivalently

$$\phi_k = \frac{N_k k a^3}{V} = e^{-\beta(F_k - \mu k)} \quad (5.25)$$

The chemical potential  $\mu$  of the system is determined by solving the equation

$$\frac{1}{\beta} \frac{\partial}{\partial \mu} \log Z = \sum_k k N_k = N \quad (5.26)$$

or in term of the volume fractions

$$\phi = \sum_k e^{-\beta(F_k - \mu k)} \quad (5.27)$$

#### 5.5.4 Molecular Dynamics and Energy Minimization

We used the GROMACS 4.5 [90] molecular simulation package to run molecular dynamics (MD) and energy minimization (EM) simulations. Our mutant molecules were solvated in a cubic box (with a minimum distance of 35 Å from any edge of the box to any atom) and neutralized with chloride ions and modeled using the GRO-MOS96 53a6 force field along with the SPC water model. We used a cutoff of 10 Å for van der Waals and short range electrostatic interactions, and calculated long range electrostatic interactions using a particle mesh Ewald sum [117, 118]. Simulations were prepared for a full MD run in both isothermal-isobaric (100 ps) and canonical equilibration (100 ps) ensembles. Temperature and pressure were controlled at 300 K and 1 bar using the velocity rescaling thermostat and the Parrinello-Rahman barostat, respectively. A linear constraint solver was used to keep all bonds at their equilibrium length. One million time steps were used with an integration time step of 2 fs. The system’s coordinates were saved every 10 ps for further analysis.

#### 5.5.5 Dipolar water solvent

We resort to our previous work in [132] to calculate the Free energy of protein aggregates and plot Fig. 5–3. In particular, we compute a Lennard-Jones, Coulomb and Solvation energy terms. The Solvation term is computed using a fast and detailed dipolar water model that solves the dipolar nonlinear Poisson-Boltzmann-Langevin equation. The three energy terms are used to describe the stability of molecular forces in molecules. Solvation energy was precisely calculated using the AQUASOL routine [89] with the following setup: atomic charges and radii assigned with PDB2PQR

using CHARMM force field at neutral pH. A grid of 257 points per edge spaced by 1 Å, a temperature of 300K, and a solvent accessible surface with an Rprobe of 1.4 Å. All hydrogen-bonds were optimized. We used a trilinear interpolation protocol for projection of fixed charges on the grid, a lattice grid size for the solvent:  $a = 2.8$  Å, solvent made of dipoles of moment  $p_0 = 3.00\text{D}$  at a concentration of  $C_{dip} = 55\text{M}$ . No salt was added to the solution and small ions were used to equilibrate the system when needed. The electrostatic potential was set to zero at the boundaries, and the stopping criteria for residual was set to:  $1.10^{-6}$  (when possible).

### 5.5.6 Building oligomer structures

The 3D structures (PDB files) are required in the process of assessing the concentration of oligomers produced by a particular analog. To create the 3D oligomers for amylin, we start with the amylin PDB structure solved by Wiltzius *et. al* [34] and use SCWRL [133] to perform the necessary mutations introduced in each analog. The mutated structure is then used by the CreateFibril tool [132] to construct the oligomers with the following parameters: 8 deg rotation angle, 5.0 Å between monomers, and a fibril packing distance of 3 units.

### 5.5.7 The effect of amyloid oligomers on the Insulin-Glucose system

The process of response to glucose intake in the human body is complex. Many factors determine how the pancreas responds to such stimulus. The natural hereditary reaction would be to release enough insulin into the bloodstream to signal other organs to consume the glucose. The following system is the basic insulin-glucose

model proposed in 1961 by Bolie [173]

$$\frac{dG}{dt} = q - \gamma I - \delta G \quad (5.28)$$

$$\frac{dI}{dt} = p - \alpha I + \beta G \quad (5.29)$$

where  $G$  is the level of glucose in the blood,  $I$  is the level of insulin released from the pancreas and the rest of the variables are coefficients that were determined experimentally. This model is a high level representation of the normal production of insulin in healthy patients.

To capture the effect of amyloids on the insulin-glucose relationship and to incorporate the effect of oligomers on  $\beta$ -cell death, we extend Eqs. 5.28 and 5.29 and

propose the following system of integral equations,

$$\bar{G}(t) = \bar{G}(0) + \int_0^t q - \gamma \bar{I} - \delta \bar{G} dt \quad (5.30)$$

$$\bar{I}(t) = \bar{I}(0) + \theta(R) * \left(1 - \frac{R}{C}\right) * \left[ \int_0^t p - \alpha \bar{I} + \beta \bar{G} dt \right] \quad (5.31)$$

$$R(t) = \left[ \frac{1}{j} \int_0^t O(t) dt \right] \left[ \left(1 + \frac{i}{100}\right)^{i*t} - 1 \right] \quad (5.32)$$

$$O(t) = \begin{cases} \frac{\dot{A}(t)}{k_O} * \sigma_1 - M(t) * \sigma_2 & \text{for } \frac{\dot{A}(t)}{k_O} \sigma_1 \geq M(t) \sigma_2 \\ 0 & \text{otherwise} \end{cases} \quad (5.33)$$

$$\dot{A}(t) = F(A(t)) \quad (5.34)$$

$$A(t) = \frac{\bar{I}(t)}{100} \quad (5.35)$$

$$F(A(t)) = \begin{cases} \epsilon & \text{for } A(t) < h \\ (1 - \epsilon') * A(t) & \text{for } A(t) \geq h \end{cases} \quad (5.36)$$

$$\theta(R) = \begin{cases} \frac{1}{1 - \frac{R}{C}} & \text{for } R(t) < d \\ \frac{1}{1 - \frac{d}{C}} & \text{for } R(t) \geq d \end{cases} \quad (5.37)$$

where  $\theta$  is the over burden term expressing how cells produce more insulin to compensate for their dead neighbors,  $C$  is the total number of  $\beta$ -cells in a pancreas,  $R$  is the number of dead  $\beta$ -cells in the pancreas at time  $t$ ,  $A$  is the level of amylin molecules released with insulin,  $\dot{A}$  is the number of amyloid proteins generated from amylin,  $O$  is the number of oligomers in the pancreas,  $k_O$  is the average length of oligomers,  $\sigma_1$  is the percentage of amyloid structures that form into oligomers,  $M$  is the number of engineered molecules injected into the system to reduce the toxicity

---

$\alpha$	$\beta$	Sequence $S^*$	$F(S^*)$
1	0	KKKKMKKQPKKKKMKNNYHKKYFHNKPLKKNMKNKQK	-5484
0	1	ETIWDGESDDYRPLTTTQKTFTTFWDFEESKRPEYET	-3944
1	1	KTAKKKKKKKKKKKKIKKKKFTKPNKHESKKVKNKHK	-3162
2	1	KKKKMKKQKAKKKMKNEYYKKYFHNKHKKKVKQKFK	-7913
4	1	KKKKMKKQPKKKKMKNNYYKKYFHNKHKKNWKNKFK	-18737
10	1	KKKKMKKQPKKKKMKNNYYKKYFHNKPLKKNMKNKFK	-51585
1	2	ETIWDGEKDIYRWLTTTQKTFTKRFDHESKRPEHET	-5720

Table 5–1: Sequences  $S^*$  for various  $\alpha$  and  $\beta$  binding affinity values.  $\frac{\alpha}{\beta} = 1$  results in a sequence  $S^*$  that has a binding affinity to amylin amyloids that is equal to its repelling affinity towards Pramlintide. A  $\frac{\alpha}{\beta} > 1$  results in a sequence that binds better to amylin amyloids and repels Pramlintide less.  $F(S^*)$  values are in Kcal/mol.

of oligomers,  $\sigma_2 \in [0, 1]$  is the binding affinity of molecules  $M$  to the oligomers,  $i$  is the infection rate of oligomers in killing cells (how fast do they kill cells),  $j$  the number of oligomers required to kill a cell,  $\epsilon$  a small value equal to the amount of amyloids being produced before reaching the threshold  $h$ ,  $\epsilon'$  is a small term to correct the amount of amylin molecules that do not misfold after reaching the threshold  $h$ , and  $d$  is the threshold of dead  $\beta$ -cells before the pancreas can stop compensating for dead cells.

The model is analyzed further in the Discussion section and provides a foundation for exploring the relationship between oligomers and  $\beta$ -cell fitness.

## 5.6 Results

We introduced the method to compute  $R$ -scores of BASM in the Methods section and described how they can be used to engineer an injectable amylin analog that binds well to amyloids in the pancreas while preventing Pramlintide from misfolding.

Rank	1-----9-----18-----27-----37	$F(S^{P_4})$	$\Delta G$	Mutations
	KCNTATCATQRLANFLVHSSNNFGAILSSTNVGSNTY	-1650.0	167.6	NA
1	-----K-----K-H-K-----	-3013.5	125.8	L12K I26K S28H T30K
2	-----K-----N--H-K-----	-3007.5	178.0	L12K A25N S28H T30K
3	-----K-----K-----H-K-----	-3005.4	185.9	L12K S20K S28H T30K
4	---K-----K-----H-K-----	-2993.3	246.9	T4K L12K S28H T30K
5	-----K-----H--H-K-----	-2992.3	210.5	L12K G24H S28H T30K
6	-----K-----NK--K-----	-2965.3	115.0	L12K A25N I26K T30K
7	-----Q--K-----H-K-----	-2957.7	191.3	A8Q L12K S28H T30K
8	-----K-----K----N---K-----	-2957.2	122.4	L12K S20K A25N T30K
9	-----K-K-----H-K-----	-2954.9	167.7	T9K L12K S28H T30K
10	-----K-----H-K--K-----	-2950.1	129.2	L12K G24H I26K T30K

Table 5–2: Top results solving the 4-residue parsimonious sequences for  $S^{P_4}$  with  $\alpha = 2$  and  $\beta = 1$ .  $F(S^{P_4})$  values are in Kcal/mol. The complete list of all 4-residue parsimonious sequences is given in Table S3.

Table 5–3 shows the  $R$ -scores that we generated for all the residue positions of amylin amyloid, and Table 5–4 shows the  $R$ -scores for Pramlintide’s amyloid residues. In Table 5–1, we solve for the sequence  $S^*$  with optimal binding affinity to BASM  $\bar{B}$  for various  $\alpha$  and  $\beta$  values. Setting a fixed  $\beta = 1$  value and increasing  $\alpha$  doesn’t drastically change the sequence  $S^*$ , suggesting that the residues of  $S^*$  possess a strong binding affinity to amylin. We proceed with sequence  $S^*$  generated by  $\alpha = 2$  and  $\beta = 1$ . This sequence encompasses a binding affinity to amylin that is twice as large as the repelling forces to Pramlintide.

### Constructing optimal binding analogs

As explained in the Methods section,  $S^*$  will naturally have a low sequence similarity to amylin. To construct an amylin analog that is structurally conserved and binds well to amylin amyloids and extends the length of oligomers, we’ll introduce some of the amino acid variations from  $S^*$  into the analog. To introduce a parsimonious selection of variations to the amylin sequence, we solve Eqs. 5.18 and 5.19 for a

Res.	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
1	100.0	7.2	55.9	183.5	113.7	150.7	196.1	94.1	116.2	101.1	49.3	-4.2	81.0	60.6	32.6	115.8	121.5	42.8	82.6	126.6
2	6.2	-26.8	-0.6	116.3	5.9	61.0	124.6	-8.1	111.2	48.4	38.5	-111.8	25.5	68.3	50.1	163.7	141.8	53.7	55.2	46.0
3	-115.7	93.2	12.9	89.3	50.3	-11.1	101.8	-123.9	-93.8	226.1	-104.6	-160.2	-0.7	-69.3	-150.7	-108.7	111.2	-108.2	-80.1	50.4
4	91.7	28.1	-10.9	85.8	81.5	87.7	82.0	43.2	84.3	56.0	-19.7	-61.4	64.6	-55.8	49.8	87.6	131.3	128.7	48.5	110.0
5	38.1	10.1	55.5	184.5	107.0	66.9	92.8	90.2	28.3	58.0	62.9	36.8	-21.0	-1.1	-3.4	100.1	59.4	36.1	68.2	99.2
6	-53.8	-69.2	-4.9	-96.5	-71.6	-135.9	-127.0	51.9	-169.2	-142.9	12.2	-197.0	-70.8	44.4	39.5	7.3	-99.3	-108.7	14.4	-124.9
7	-79.8	-92.8	-128.1	-19.4	-87.1	-84.4	-4.1	-75.7	-92.0	-67.2	-68.2	-173.4	-93.5	-83.3	-114.0	-70.9	-52.7	-75.7	-72.0	-56.1
8	46.9	-113.9	-28.3	-122.5	-24.6	-138.7	36.2	44.8	47.3	36.9	-90.0	-3.8	-123.1	-11.8	-7.3	79.8	-105.2	12.6	26.6	-49.1
9	13.1	-7.5	-1.2	128.5	18.4	30.1	76.6	20.6	3.7	24.7	19.1	-113.1	-122.4	13.2	-130.3	23.1	46.8	77.2	41.3	30.5
10	-125.0	-131.1	-27.8	78.9	-128.9	-112.6	-60.3	-125.9	-49.7	25.1	-26.7	-161.3	-148.1	-104.8	-105.1	-132.1	-104.9	-97.7	-102.9	-106.1
11	-106.7	-147.3	-165.8	-46.5	-115.0	-110.2	-68.3	-120.4	-132.6	-114.4	-119.1	-212.2	-143.8	-121.7	-139.1	-86.1	-64.3	-138.5	11.0	-99.1
12	16.6	144.1	-173.6	-92.5	25.4	-4.2	51.6	1.8	-22.5	-103.6	79.6	-205.1	-78.1	-27.9	13.7	33.1	46.6	16.0	115.4	50.5
13	-166.7	-188.2	-238.5	-171.9	-164.8	-165.9	-123.9	-166.1	-194.1	-182.0	-186.0	-238.7	-209.6	-181.4	47.0	-156.3	-155.6	20.7	-175.9	-175.5
14	-82.7	-17.3	-218.9	-95.5	-63.4	-167.7	-57.0	-60.5	-192.0	-38.2	50.9	-22.2	-220.4	-175.0	-62.1	-60.4	-20.6	-166.9	-160.5	-162.6
15	-132.6	-136.4	-179.7	-122.3	-131.5	-105.8	-129.6	-138.9	-153.3	-21.6	-133.4	-194.2	-166.3	-142.3	-153.8	-99.5	-70.9	-186.6	-123.6	-109.1
16	11.2	19.6	-171.0	-78.3	66.8	-147.1	-5.5	14.5	-49.8	-42.0	-147.3	-50.9	-143.8	-131.8	57.5	-7.9	139.9	21.9	-122.8	-94.8
17	10.0	34.9	-181.5	2.4	-3.3	102.5	-151.2	11.8	-31.1	-26.8	-99.9	-95.5	-65.1	27.4	-118.0	25.6	55.9	13.6	100.5	-9.6
18	-134.6	20.2	-3.8	63.0	-131.6	148.1	18.2	-127.8	-111.0	-113.8	-22.5	-25.6	7.1	-112.4	-71.2	-116.6	-85.4	-133.4	-154.1	70.4
19	-78.0	-50.4	-102.6	-67.8	-82.6	-85.5	-75.4	-86.7	-121.6	-77.9	-89.1	-36.1	19.8	-109.9	20.3	-67.4	74.0	-83.1	-121.2	-52.9
20	116.0	109.6	3.3	69.2	22.1	63.7	56.1	-55.2	42.5	-76.4	-65.3	-82.9	-22.4	48.9	6.2	67.9	154.5	33.7	-6.3	111.2
21	28.8	90.1	-48.1	41.4	21.2	31.4	36.4	95.7	37.7	-43.0	30.8	-51.6	38.3	60.6	137.2	54.7	122.1	42.2	-37.2	87.5
22	-1.2	67.0	-85.8	-55.6	-1.7	-6.0	10.8	-7.5	-15.5	-80.7	-100.6	-6.5	56.1	-56.3	-99.4	11.7	36.0	-7.4	-102.9	26.3
23	28.8	28.0	-142.5	12.4	46.9	15.7	-48.6	-77.5	-119.2	-151.3	-156.7	40.0	-130.2	-174.1	-117.8	-59.8	138.0	-51.0	-146.7	-28.9
24	-107.9	69.3	-147.6	-30.2	36.5	-112.6	15.3	32.7	-189.8	-62.6	17.0	-52.6	-155.4	102.6	-29.9	-114.5	36.2	25.7	54.1	99.8
25	-75.5	-101.7	-221.3	-178.7	-108.1	-166.7	-183.7	-126.1	-169.4	-189.6	-123.3	-221.3	-195.6	-100.5	-177.2	-126.2	-89.9	-14.9	-117.1	-154.9
26	14.0	9.2	-13.0	38.3	-3.6	-120.7	-46.5	17.9	-119.9	87.6	-107.0	-129.4	-104.6	-79.3	11.2	17.2	-23.0	-78.9	-53.5	-20.8
27	-57.1	-74.7	-96.7	-58.4	-51.5	-78.7	-83.9	-34.8	-93.1	-50.0	-81.7	-139.6	-108.4	54.0	-151.8	-54.3	16.1	-92.0	-77.0	-53.6
28	74.2	1.4	56.7	121.9	53.3	14.7	85.8	101.0	-93.6	-4.1	-129.7	-29.5	34.5	-9.1	86.4	56.0	213.1	77.1	-26.5	27.7
29	-19.6	-89.5	-67.4	-55.5	-38.3	-62.9	-40.6	-32.6	-65.3	-47.2	-90.3	-121.0	-120.4	-81.4	-55.5	-74.2	-53.5	-70.3	-58.3	-47.4
30	89.4	-137.3	-118.8	91.8	78.0	-79.6	99.9	77.3	68.2	-98.7	-98.2	-153.3	33.3	-6.1	66.2	73.0	117.0	73.9	88.0	47.1
31	-133.4	-85.3	-172.3	-59.5	-122.1	-90.2	-144.1	-126.6	-116.5	-138.6	-156.7	-124.8	-164.7	-133.7	-67.0	-98.0	-116.6	-92.7	-142.4	-155.8
32	-55.5	124.3	76.1	110.7	-69.3	-47.5	181.6	-0.4	-2.8	-65.2	6.9	-31.1	-91.2	-66.1	-8.8	-27.7	-14.4	-90.4	-73.1	-69.3
33	-143.4	-142.1	-141.9	-112.0	-126.8	-171.8	-92.3	-108.1	-108.4	-131.5	-142.7	-198.9	-4.4	-129.6	8.8	-139.5	-100.7	-90.8	-103.3	-94.2
34	-5.7	96.2	-141.1	-10.4	-9.4	-125.4	72.8	-15.8	101.0	-97.6	-20.7	-28.1	-98.5	23.4	-4.1	-5.8	-37.0	-35.3	9.1	-10.9
35	-129.9	-104.1	-169.9	-100.9	-151.5	-118.3	-102.0	-140.3	-146.1	-162.0	-106.5	-208.0	-182.4	-70.4	-140.8	-118.2	-98.0	-127.5	-54.3	-114.0
36	52.0	42.6	-4.4	74.9	51.3	-86.5	98.1	44.0	-45.7	49.2	-9.7	6.7	24.4	-86.3	34.9	51.5	67.4	-65.0	-25.2	55.1
37	-156.2	-141.3	-160.9	-157.4	-156.3	-167.6	-154.3	-23.0	-156.6	-170.1	-5.2	-219.5	-165.4	-157.0	8.8	-159.6	14.2	-22.2	-61.5	-151.6

Table 5–3: The Binding Affinity Scoring Matrix (BASM) R score values of the amylin amyloid. The first column represents the amino acid residue positions 1-37 of the amylin amyloid. The rest of the columns show the R-scores of the 20 amino acids interacting with every residue position of amylin.

4-residue parsimonious sequence. The assumption we make here is that the binding affinity effect of separate positions are additive. This should hold for residues that are not adjacent. We rank all the possible  $S^{P_4}$  mutation combinations according to their global binding affinity values and organize them in Table S4. The top 10 results are shown in Table 5–2. For each of the top 10 results, we calculate their  $\Delta G$  values using Eq. 5.20 to obtain a lower-bound on their amyloidogenicity potential. All of the  $\Delta G$  values are positive, indicating that the analog will not spontaneously misfold. Their stability is expected to be similar to amylin because of the similarity

## CHAPTER 5

Res.	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
1	132.5	25.4	83.9	213.6	128.2	154.3	224.0	126.2	130.9	135.0	76.9	67.1	113.4	82.2	68.0	-26.8	122.1	65.3	100.4	149.8
2	52.1	34.5	43.6	165.4	39.9	104.6	170.5	37.9	163.1	96.1	85.2	-49.6	72.4	117.5	81.1	164.5	217.2	99.4	100.4	90.4
3	-85.1	109.2	36.3	119.7	42.0	25.7	109.6	-95.1	-67.0	247.3	-107.4	-86.9	27.0	-43.6	-128.6	-77.9	125.4	-100.9	-71.0	96.9
4	93.5	43.3	38.9	145.7	100.9	43.0	81.4	62.7	103.5	107.6	8.0	11.3	56.5	-22.0	69.4	102.4	151.8	190.7	81.2	134.8
5	85.8	52.4	99.2	231.3	147.4	112.8	122.3	139.5	59.2	105.7	109.2	132.0	22.5	45.7	41.4	145.8	107.0	82.5	114.2	147.5
6	-21.5	-50.4	8.7	-60.5	-41.9	-108.4	-94.5	93.4	-139.7	-111.1	21.7	-115.2	-38.2	32.7	78.2	17.8	-65.4	-89.0	4.3	-86.6
7	-47.2	-60.9	-93.9	11.1	-54.1	-49.1	24.8	-49.0	-63.9	-20.7	-34.7	-89.9	-59.0	-51.4	-82.2	-41.3	-13.7	-43.9	-40.7	-22.1
8	97.0	-97.8	-3.4	-82.7	19.9	-94.5	52.9	94.1	96.3	47.5	-52.1	101.2	-94.3	27.4	41.2	102.5	-59.1	18.3	55.2	-14.8
9	31.9	14.4	14.2	150.6	37.9	51.2	92.6	39.0	24.6	44.3	39.3	-45.9	-84.7	35.9	-81.2	43.3	68.3	69.0	58.2	48.6
10	-89.8	-94.2	18.5	65.9	-92.5	-77.0	-33.4	-93.5	-14.2	65.5	-9.7	-79.3	-120.9	-66.8	-61.8	-98.8	-69.1	-60.6	-70.9	-78.5
11	-77.7	-124.4	-144.9	-18.3	-79.1	-85.7	-21.8	-94.1	-110.6	-76.6	-99.9	-139.7	-111.3	-114.8	-109.3	-54.9	-26.3	-114.5	25.4	-64.2
12	52.5	139.5	-149.1	-76.5	59.3	4.4	39.6	40.3	43.9	-82.8	66.7	-97.6	-61.2	5.4	62.2	66.9	81.7	57.4	110.2	90.8
13	-139.7	-157.0	-211.4	-145.1	-146.6	-135.0	-84.3	-134.1	-161.0	-140.0	-150.2	-167.7	-174.8	-150.2	7.9	-114.0	-121.8	1.9	-140.7	-139.6
14	-28.7	10.1	-186.3	-56.0	-44.7	-135.3	11.2	-14.8	-148.2	20.7	69.7	31.9	-177.0	-135.0	-31.8	-56.4	18.8	-115.4	-123.7	-116.2
15	-108.3	-111.2	-154.7	-99.6	-107.1	-96.2	-100.9	-117.4	-124.6	-16.2	-107.9	-111.7	-136.8	-108.6	-121.4	-66.4	8.5	-149.5	-94.3	-91.3
16	27.7	42.4	-149.4	-130.2	60.0	-128.4	25.6	34.0	-20.4	63.9	-111.6	18.8	-125.1	-103.7	79.2	2.5	162.7	-35.3	-92.2	-59.8
17	26.0	51.0	-155.8	54.9	32.8	57.0	-93.1	33.4	-5.5	9.8	-80.0	-23.0	-35.4	22.4	-110.1	58.3	88.3	62.1	86.8	-0.3
18	-106.2	65.5	-4.0	74.2	-112.5	136.2	1.8	-90.1	-84.0	-81.4	0.8	29.2	20.8	-79.6	-37.8	-89.9	-42.4	-86.5	-124.8	69.1
19	-60.1	-39.6	-88.6	-45.3	-64.9	-77.3	-51.9	-75.4	-95.6	-64.0	-58.1	52.0	46.9	-86.6	35.0	-30.5	20.8	-46.6	-84.2	-43.7
20	100.8	112.2	-0.8	96.5	36.1	106.7	90.9	-32.1	48.6	-31.6	-91.4	43.9	-5.6	79.3	9.8	63.3	180.2	55.3	-37.7	128.1
21	58.8	74.8	6.8	59.3	53.6	57.8	73.2	100.9	13.5	-39.6	89.1	7.3	67.2	124.9	81.6	85.5	77.5	68.1	-18.1	99.1
22	20.1	97.4	-44.6	-38.5	32.5	36.3	87.4	32.1	1.3	-44.9	-71.5	53.1	96.4	-7.0	-73.9	36.2	133.7	11.9	-75.1	65.3
23	36.1	80.4	-123.9	18.6	19.7	36.5	-32.8	-38.7	-117.8	-109.1	-132.6	118.1	-87.4	-126.9	-72.6	-39.8	129.0	-79.6	-93.2	13.1
24	-94.4	106.7	-161.0	0.5	27.0	-84.7	11.9	32.9	-147.8	-48.3	0.5	-10.7	-138.3	108.5	21.1	-76.9	76.0	47.1	52.6	66.3
25	-51.2	-77.3	-87.5	-127.1	-55.6	-123.7	-155.3	-96.9	-138.9	-162.9	-138.5	-149.6	-187.2	5.1	-160.6	-108.9	-61.7	36.6	-66.3	-134.6
26	41.8	-0.3	42.8	74.0	41.0	-110.1	-11.7	47.6	-77.5	24.1	-118.0	-60.7	-68.2	-54.8	10.1	62.2	-10.1	-55.4	-26.5	-7.6
27	-54.8	-52.6	-78.9	-29.2	-17.9	-44.4	-54.6	-9.6	25.0	-32.1	-50.0	-61.7	-73.7	69.2	-113.4	-31.6	10.3	-72.0	-56.7	-48.8
28	122.1	129.6	92.4	139.4	97.0	97.9	279.7	162.3	84.6	18.0	-104.1	62.9	72.1	1.8	113.9	98.6	259.6	64.8	-22.2	61.0
29	-26.4	-69.2	-57.6	-39.4	-50.3	-40.0	-19.9	-25.5	-70.9	-39.2	-76.5	-83.3	-87.2	-66.7	-45.2	-36.0	-35.9	-68.9	-43.6	-35.4
30	118.2	-111.7	-124.3	131.5	100.2	-73.5	81.1	121.1	-97.7	-70.6	-106.2	-65.9	88.0	49.6	116.5	133.8	114.6	99.5	132.7	53.5
31	-56.7	-64.3	-136.2	-70.8	-109.1	-91.7	-100.6	-87.5	-94.2	-89.3	-92.1	-21.1	-106.1	-104.4	-72.7	-57.7	-109.4	-37.1	-94.1	-107.6
32	-23.1	175.9	74.9	105.0	-42.3	-25.7	163.7	27.2	-3.9	-12.6	46.4	32.4	-58.1	-11.8	14.5	-12.6	19.9	-51.8	-20.2	5.3
33	-81.9	-102.2	-61.5	-69.1	-97.1	-116.1	-68.5	-91.7	-121.9	-95.6	-106.4	-119.5	15.4	-66.9	25.7	-105.9	-71.2	-56.2	-111.5	-60.6
34	30.2	117.4	-113.7	20.2	37.3	-75.9	131.4	17.9	94.3	-62.8	11.3	45.6	-68.2	74.7	20.5	30.1	3.6	1.4	46.0	13.3
35	-102.5	-78.1	-145.7	-72.3	-125.1	-91.7	-76.3	-112.8	-63.2	-124.2	-81.2	-133.9	-149.0	-104.4	-124.9	-101.3	-77.0	-121.7	-29.3	-92.1
36	82.0	65.1	21.0	82.8	62.0	-78.1	134.1	50.7	-5.9	76.9	16.8	75.2	53.0	-37.4	56.8	81.9	86.6	-23.5	8.5	83.2
37	-120.5	-106.7	-124.3	-111.8	-119.6	-124.3	-107.3	-12.8	-111.1	-142.7	-19.3	-130.1	-132.4	-130.4	18.6	-122.2	39.6	8.3	-58.9	-112.9

Table 5–4: The Binding Affinity Scoring Matrix (BASM) R score values of the Pramlintide amyloid. The first column represents the amino acid residue positions 1-37 of the Pramlintide amyloid. The rest of the columns show the R-scores of the 20 amino acids interacting with every residue position of Pramlintide.

in  $\Delta G$  values. The global binding affinity score for the top 10 results are almost twice that of an amylin amyloid, suggesting that the binding between an amyloid and one of these analogs will be twice as strong as the binding between two amylin amyloids. We used TANGO [98] to explore the amyloidogenicity potential of these 10 sequences and found that all sequences were predicted to create beta-strands, were predicted to not form any alpha-helices, and expressed an extremely high propensity for aggregation (see Table 5–5). This validation by TANGO supports the high values

of  $\Delta G$  in the top 10 parsimonious sequences.

Mutant	AGG	AMYLO	TURN	HELIX	HELAGG	BETA
M7	20.366	0.0755151	31.0521	16.326	0	52.5555
M4	20.3613	0.000988828	31.1604	20.7997	0	39.2829
M9	20.3598	0.0213931	31.3389	29.4205	0	39.2994
M3	19.6618	0.0699662	28.7727	21.8182	0	42.3278
M8	1.85245	0.141604	36.4188	19.4214	0	43.4389
M2	1.84561	0.0589466	39.0022	18.7469	0	39.0683
M1	0.94131	0.0589388	31.2728	26.4465	0	39.4491
M5	0.94131	0.059078	28.3981	18.7469	0	47.8042
M6	0.94131	0.0589388	37.8203	19.8651	0	38.7392
M10	0.94131	0.0589369	29.2805	21.189	0	45.6678

Table 5–5: Exploring aggregation properties of top 10 results with TANGO. The AGG column ranks the beta-sheet aggregation propensity of each sequence. TANGO suggests that the top 6 sequences are highly amyloidogenic. All the sequences possess a HELAGG (helix-aggregation propensity) of 0. Although TANGO predicts the aggregation propensity of proteins from sequence alone, the dipolar solvent model provides much more accurate results since it calculates precisely the solvation and enthalpy terms based on the full 3D atomic structure of proteins. It is supportive however to find that the results from TANGO are in-line with our predictions.

### Oligomer concentrations

The  $F(S^{P_4})$  values in Table 5–2 give good binding affinities to the amyloids. In addition to a strong binding affinity, the analog must form long fibril structures. We built the aggregate structures for all ten results and computed their free energies (Eq. 5.1) in Fig. 5–3. We observe that all the ten sequences in Fig. 5–3 produce more stable oligomers than the amylin amyloid, suggesting that they would form longer aggregates since it is energetically favorable to do so. To determine the  $S^{P_4}$  that generates the highest potential to form long aggregates, we compute molecule concentrations (Eq. 5.27) by the law of mass action and report the volume fractions of

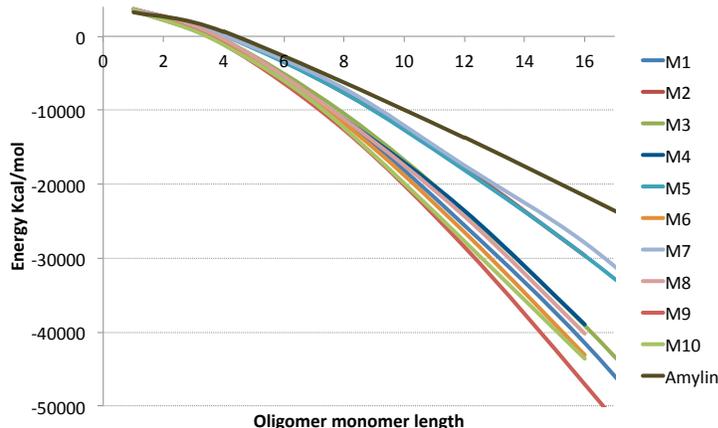


Figure 5-3: Energies for top ten  $S^{P_4}$  amylin analog oligomers.

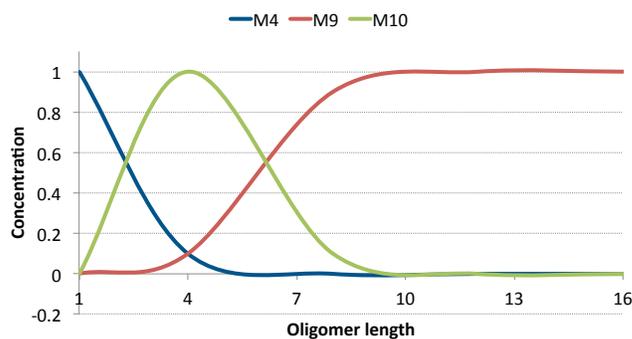


Figure 5-4: Concentration of oligomers in solution.

oligomers at various lengths in Fig. 5-4. The sequence that ranked 9th in Table 5-2 expressed the most stability in Fig. 5-3 and dominated the volume fractions in Fig. 5-4 at oligomers of size greater than 7. This suggests that this sequence creates the longest aggregates out of the 10 results. The longer it is, the more stable it gets. In contrast, we observe that the  $S^{P_4}$  ranked 10th in Table 5-2 would create many short oligomers of length 4 compared to the ten sequences. It is also worth mentioning that the most stable amyloid  $S^{P_4}$  monomer is produced by the sequence ranked 1.

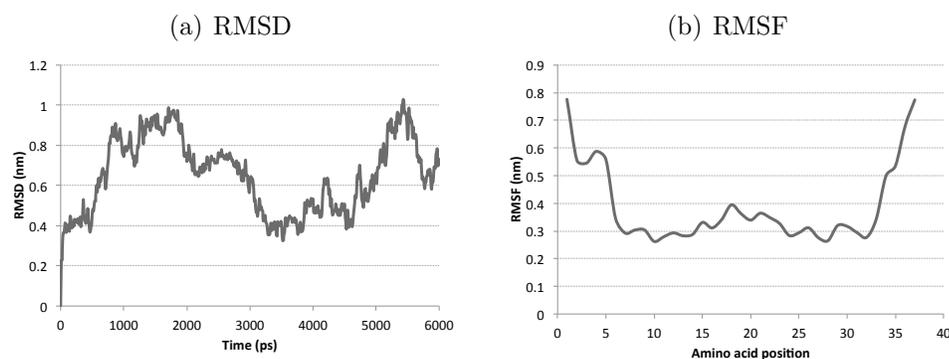


Figure 5–5: RMSD and RMSF plots for the  $S^{P4}$  sequence with the T9K, L12K, S28H and T30K mutations over a 6ns simulation.

In Fig. 5–4 we see that this sequence dominates the concentrations at length 1 but does not have a strong affinity to aggregate when compared to the other sequences.

### Maintaining native form

Introducing four mutations to the sequence of amylin might destabilize it and force it to misfold. To ensure that the (T9K L12K S28H T30K) mutation preserves amylin structure and does not induce Pramlintide to misfold, we ran an MD simulation of 6 ns on this amylin mutant. We report in Fig 5–5 (a) and (b) the RMSD and RMSF graphs, respectively. The RMSD graph shows low values ( $< 1$ ), indicating that the mutant structure is very stable. From the RMSF graph, we see that the positions 9, 12, 28, and 30 that correspond to the mutations all have low RMSF values, indicating that the mutations do not cause structural turbulence and are not disadvantageous to amylin. This preservation of structure should preserve amylin’s functions.

## Modeling the effect of oligomers on the insulin-glucose relationship

Upon having a meal, glucose levels increase in the bloodstream. The increase in levels signals the  $\beta$ -cells in the pancreas to produce insulin to instruct body organs such as the heart and muscles to absorb and metabolize glucose. This process lowers the glucose level in the blood. The amylin polypeptide is also secreted with insulin in this process at a ratio of 1:100 [51]. Amylin's functions include modulating gastric emptying, inhibiting glucagon to prevent postprandial spikes in blood glucose levels, and inducing satiety leading to decreased food intake and weight loss [148, 149]. Together, amylin and insulin play a crucial role in maintaining glucose homeostasis.

### 5.7 Discussion

#### Binding to oligomers

It is a challenging and computationally intensive task to find a molecule  $A$  that binds well to a molecule  $B$  or subregions of  $B$ . One can perform docking simulations with millions of molecules to narrow the search of finding a good fit. In this work, we introduced the BASM method of calculating binding affinities of optimal sequences to cut down on this search process and analyze the binding preferences of  $B$ , or subregions of  $B$ . This procedure allowed us to explore engineering analogs that bind strongly to oligomers and promote their aggregation.

Applying the BASM method to engineer an amylin analog that can be administered in native, non-amyloid form with Pramlintide injections resulted in discovering a mutant (T9K L12K S28H T30K) that is predicted to bind twice as strong to oligomers as amylin does to oligomers. This has been achieved by computing

$R$ -scores that bind well to oligomer surfaces and repel from Pramlintide's amyloid surface (Table 5–1). Using the optimal sequence constructed by these  $R$ -scores, we searched through 66 thousand combinations of parsimonious mutations to find the mutant with optimal binding and aggregation potential. This aggregation potential was both assessed by the dipolar solvent model from [132] and by TANGO [98]. Furthermore, we validated the stability of the mutant in native form by performing an MD production of 6 ns. The results showed that the mutant was stable in native form, and hence could be a potential candidate for therapeutic injections.

The function of the mutant is intended to reduce oligomer toxicity and decrease the rate of  $\beta$ -cell death. The mutant is stable in native form, and forms long fibrils in amyloid form. Since its native structure is conserved and is analogous to amylin, it will carry out the normal functions of amylin and Pramlintide. By design, the analog has low affinity to Pramlintide and should not interfere with its function. It will act as a replacement to amylin molecules and modulate gastric emptying, inhibit glucagon, and induce satiety [148, 149].

Once the mutant analog reaches the pancreas and makes contact with the amyloids, some of it should misfold and aggregate with the surrounding amyloids and oligomers, due to its relatively high  $\Delta G$  (amyloidogenicity potential). Although long fibrils have not been observed to be toxic in the pancreas, the pharmacological effect of elongating the oligomers still needs to be further studied. If very long fibrils are observed to be unfavorable, controlling aggregation and capping the growth of oligomers at a certain length will be required and can be achieved by choosing mutants that exhibit lower oligomer concentration potentials or regulating the dosage

of the injectable analogs in the body. The contribution of the BASM approach and R-scores to limit the toxicity of oligomers by favoring aggregation could be extended to explore creating potential therapeutic analog candidates for other amyloid related diseases such as Huntington and Parkinson's.

### **Modeling the effect of oligomers on the insulin-glucose relationship**

Upon having a meal, glucose levels increase in the bloodstream. The increase in levels signals the  $\beta$ -cells in the pancreas to produce insulin to instruct body organs such as the heart and muscles to absorb and metabolize glucose. This process lowers the glucose level in the blood. The amylin polypeptide is also secreted with insulin in this process at a ratio of 1:100 [51]. Together, amylin and insulin play a crucial role in maintaining glucose homeostasis.

Several mathematical models have been proposed to represent the complex process of response to glucose intake [174]. However, to date, no model has incorporated the production of amyloids and the effect of oligomers on  $\beta$ -cell death. We think that by taking into account the observed effect of oligomers on cell death the analytical solution of insulin in Eq. 5.29 introduced by [173] will be the one presented in Eq. 5.31.

We introduce a  $\theta$  and  $R$  function into Eq. 5.31, where  $\theta$  is the over burden term expressing how cells produce more insulin to compensate for their dead neighbors and  $R$  is the number of dead  $\beta$ -cells in the pancreas at time  $t$ . The  $(1 - \frac{R}{C})$  term lowers the expected production of insulin proportional to the number of dead  $\beta$ -cells in  $R$ . The more  $\beta$ -cells die, the less insulin should be produced. However,

this relationship is not linear. Upon death of  $\beta$ -cells, the pancreas still attempts to generate the required insulin by signalling the remaining live  $\beta$ -cells to secrete higher levels [175]. This overburdening of cells is modelled with  $\theta$ . The live cells will produce higher levels of insulin and compensate for the dead cells until a threshold  $d$  is reached, after-which the cells cannot produce any higher levels and the required level of insulin is not met. When this happens, glucose levels in Eq. 5.30 are not returned to normal, and the pancreas is required to continue pumping insulin. More insulin produces more amylin (Eq. 5.35) at a ratio of 1:100 [135], and more amylin creates higher levels of amyloids (Eq. 5.34). Again, this relationship is not linear. When amylin concentration is lower than a threshold  $h$ , minimal amyloids form, and when the threshold is exceeded, amyloid production is initiated in high concentration. Eq. 5.33 models the percentage of amyloids in the pancreas system that form into oligomers of average length  $k_O$ . This percentage is positive or 0, depending on the efficacy of any introduced oligomer inhibiting molecules. Because the death rate of  $\beta$ -cells due to oligomers resembles a very slow exponential-like curve [176], we use the compound interest formula to model this phenomenon in Eq. 5.32 .

The key to suppress the deterioration of  $\beta$ -cells is to minimize the number of oligomers infecting cells, modelled in Eq. 5.33. Producing  $M$  molecules that bind well to the oligomers achieves this result. In return, this affects Eq. 5.32, lowering the number of dead cells and stabilizing Eq. 5.31 to produce the right quantity of insulin. If the opposite occurs, Eq. 5.33 produces more oligomers, increasing the dead cells in Eq. 5.32, and in turn lowering the production of insulin in Eq. 5.31. Glucose levels will remain high in Eq. 5.30, prompting more insulin to be produced,

triggering an increase in amylin and amyloid production in Eqs. 5.34 and 5.35. This in turn creates more oligomers, and the cycle repeats.

Although the behavior of the subsystems of the theoretical model presented in Eqs. 5.30 - 5.37 are known biologically, we do not provide a simulation at different time points because there are way too many parameters in the system of integral equations that we dont have experimental values for. The aim of the model is to provide the theoretical framework that incorporates amyloids, oligomers, and toxicity of the beta cells, and to show at a high level the general effect of lowering oligomer concentration on the fitness of  $\beta$ -cells. It will become useful to extract data from simulating the model once many of the parameters are discovered experimentally.

### **Online tool**

The BASM method and  $R$ -scores procedure have been packaged into a tool called Single-rEsidue Mutational based Binding Affinity (SEMBA) to probe the mutation landscape of amyloid proteins. The tool can be downloaded as a standalone software from <http://amyloids.cs.mcgill.ca>

### **Acknowledgement**

M.R.S. designed the study. M.R.S. and J.W. analyzed results. H.O. designed the concentration formulas. M.R.S. carried out the experiments, developed code and drafted the manuscript. M.R.S., H.O., and J.W. read and approved the final manuscript.

*Funding:* M.R.S. was supported by a fellowship from the Canadian Institutes of Health Research System Biology Training program at McGill University and a grant from the Fonds de recherche Nature et technologies Quebec. J.W. was supported by a Discovery grant from the Natural Science and Engineering Research Council of Canada.

### 5.8 Supplementary material

Table 5-6 (S4) was too long to include in this thesis. It can be found at the following link: <http://amyloid.cs.mcgill.ca/SEMBA>

At the same link is an executable file called “GetRank.py” that can be used to find the specific ranking and affinity of any of the 4-point mutations.

## CHAPTER 6

### Conclusion

#### 6.1 Summary

In this thesis we developed a framework to simulate the atomic structures of amyloid fibrils and explore ways to limit their aggregation and toxicity. Our work has focused in particular on amyloids in diabetes, but can certainly be applied to study amyloids in Alzheimer's, Huntington, and Parkinson's diseases.

In Chapter 2, we developed the stability landscape method to find suitable structural parameters to build polymorphic fibril aggregates with our CreateFibril tool. The CreateFibril tool uses the structure of one amyloid protein to simulate and build the possible polymorphic assemblies using replication and rigid affine transformation matrices. This method replicated the experimental results for the  $A\beta$  and HET-s amyloids, and predicted the correct amylin fibril structures, which were validated during the time of our study.

In Chapter 3, we used the CreateFibril tool and results from Chapter 2 to construct the structures of amyloid fibrils that develop in diabetes patients. We developed a fast method to analyze an amyloid structure and predict destabilizing

mutations. The method studied the effect of the destabilizing mutations on nucleation, amyloidogenicity and fibril extension. Using this method, we found a set of mutations that destabilize amylin amyloids and potentially lower aggregation rates and amyloid formation. The mutations can be further studied and validated experimentally to engineer a replacement drug for amylin amyloids.

In Chapter 4, we explore and characterize the entire single-point mutation landscape for the amylin protein that misfolds in diabetes. We reveal the effect of single-point mutations on amylin's stability and potential to misfold into amyloids. Using the data of single-point mutations, we construct a way to estimate efficiently the effect of  $n$ -point mutations. This allowed us to explore millions of potential mutation combinations that stabilize amylin and inhibit fibrils more than the leading Pramlintide drug. Because experimental testing of all these mutation combinations is not feasible, this computational work can be used as a guide to shortlist the candidate mutations for potential drug design.

In Chapter 5, we attempted to tackle the problem of toxic oligomers (short amyloid aggregates) killing  $\beta$ -cells and aggravating the conditions of diabetes. We developed the Binding Affinity Scoring Matrix (BASM) method to probe the binding affinity of amylin amyloids and engineered an amylin mutant analog that elongates oligomer structures to reduce their toxicity. We designed the analog to be administered with Pramlintide shots for type I and type II diabetes patients.

## 6.2 Contribution to amyloid research

The results of this thesis are of high importance to the fields of drug design, development, and therapy. The computational tools to build amyloid fibrils and assess their stability under the application of mutations can facilitate the exploration, development, and enhancement of therapeutics that hinder or inhibit fibril formation. In the case of the Pramlintide drug in diabetes, our tools showed that various improvements to the sequence of the drug can be made to offer a more stable replacement, an insight that can't be made through experimental efforts alone. The ability to screen billions of possible mutation combinations to find a few good candidates is becoming crucial in advancing therapeutics.

The methods in this thesis could be applied to explore the mutation landscape of the  $A\beta$  protein in Alzheimer's and the  $\alpha$ -s protein in Parkinson's, among other diseases. The landscapes can determine the mutations required to create more stable analogs and restore lost biological function. The BASM method in Chapter 5 could be used to probe the binding affinity of  $A\beta$  and  $\alpha$ -s to construct molecules that bind strongly to their amyloids and either inhibit or promote their extensions.

## 6.3 Future direction

We have emphasized in this work that amino acid mutations alter the stability of proteins and their potential to misfold into amyloid structures. Researchers have been engineering  $A\beta$  and amylin analogs with mutations that affect their amyloidogenicity rate and their membrane permeability. However, computationally quantifying the effect of a mutation on the energy barrier between a native protein structure and its

amyloid form is an unexplored challenge. Understanding the effect of mutations on energy barriers will help us design more efficient analogs for disease. In this thesis, we quantified the effect of mutations on the stability of native proteins, and estimated a lower bound on the energy barrier between native and amyloid forms. Determining the full energy barrier will enable us to predict the effect of mutations on altering the probability of misfolding into amyloids. With this information, we can confidently select candidates for drug design that not only introduce stability to the native form, but also increase the energy barrier to decrease the chance of misfolding.

One way to determine this barrier is to construct possible folding pathway trajectories between a native and amyloid shape. Molecular dynamics and affine transformation matrices can be used to assist a structure in misfolding along the pathways and assessing structural energy at specific points along a trajectory. For each predicted folding point in a trajectory that we make, we let the molecule “relax” by running energy minimization and MD, upon which we predict the pathways that can be taken from that point to reach to the amyloid form. We can repeat this procedure for several points along the trajectory, until the structure converges to an amyloid. This generates a rough estimate of the energy barrier along the trajectory. Solving for the optimal trajectory returns an estimate for the energy barrier between the native and amyloid form.

## References

- [1] V. N. Uversky and A. L. Fink, “Conformational constraints for amyloid fibrillation: the importance of being unfolded,” *Biochim Biophys Acta*, vol. 1698, pp. 131–53, May 2004.
- [2] P. J. Landrigan, B. Sonawane, R. N. Butler, L. Trasande, R. Callan, and D. Droller, “Early environmental origins of neurodegenerative disease in later life,” *Environ Health Perspect*, vol. 113, pp. 1230–3, Sep 2005.
- [3] S. C. Kirkwood, J. L. Su, P. Conneally, and T. Foroud, “Progression of symptoms in the early and middle stages of huntington disease,” *Arch Neurol*, vol. 58, pp. 273–8, Feb 2001.
- [4] S. H. Golden, J. E. Williams, D. E. Ford, H.-C. Yeh, C. Paton Sanford, F. J. Nieto, F. L. Brancati, and Atherosclerosis Risk in Communities study, “Depressive symptoms and the risk of type 2 diabetes: the atherosclerosis risk in communities study,” *Diabetes Care*, vol. 27, pp. 429–35, Feb 2004.
- [5] L. G. Goldfarb, P. Brown, E. Mitrovà, L. Cervenàková, L. Goldin, A. D. Karczyn, J. Chapman, S. Gálvez, L. Cartier, and R. Rubenstein, “Creutzfeldt-jacob disease associated with the prnp codon 200lys mutation: an analysis of 45 families,” *Eur J Epidemiol*, vol. 7, pp. 477–86, Sep 1991.
- [6] M. F. Mendez, A. Selwood, A. R. Mastri, and W. H. Frey, 2nd, “Pick’s disease versus alzheimer’s disease: a comparison of clinical characteristics,” *Neurology*, vol. 43, pp. 289–92, Feb 1993.
- [7] S. Lillioja, D. M. Mott, B. V. Howard, P. H. Bennett, H. Yki-Järvinen, D. Freymond, B. L. Nyomba, F. Zurlo, B. Swinburn, and C. Bogardus, “Impaired glucose tolerance as a disorder of insulin action. longitudinal and cross-sectional studies in pima indians,” *N Engl J Med*, vol. 318, pp. 1217–25, May 1988.
- [8] J. H. Warram, B. C. Martin, A. S. Krolewski, J. S. Soeldner, and C. R. Kahn, “Slow glucose removal rate and hyperinsulinemia precede the development of

- type ii diabetes in the offspring of diabetic parents,” *Ann Intern Med*, vol. 113, pp. 909–15, Dec 1990.
- [9] G. M. Reaven, “Banting lecture 1988. role of insulin resistance in human disease,” *Diabetes*, vol. 37, pp. 1595–607, Dec 1988.
- [10] V. Poitout and R. P. Robertson, “Minireview: Secondary beta-cell failure in type 2 diabetes—a convergence of glucotoxicity and lipotoxicity,” *Endocrinology*, vol. 143, pp. 339–42, Feb 2002.
- [11] M. Prentki, E. Joly, W. El-Assaad, and R. Roduit, “Malonyl-coa signaling, lipid partitioning, and glucolipotoxicity: role in beta-cell adaptation and failure in the etiology of diabetes,” *Diabetes*, vol. 51 Suppl 3, pp. S405–13, Dec 2002.
- [12] N. Kaiser, G. Leibowitz, and R. Nesher, “Glucotoxicity and beta-cell failure in type 2 diabetes mellitus,” *J Pediatr Endocrinol Metab*, vol. 16, pp. 5–22, Jan 2003.
- [13] A. Clark, S. B. Chargé, M. K. Badman, and E. J. de Koning, “Islet amyloid in type 2 (non-insulin-dependent) diabetes,” *APMIS*, vol. 104, pp. 12–8, Jan 1996.
- [14] C. Röcken, R. P. Linke, and W. Saeger, “Immunohistology of islet amyloid polypeptide in diabetes mellitus: semi-quantitative studies in a post-mortem series,” *Virchows Arch A Pathol Anat Histopathol*, vol. 421, no. 4, pp. 339–44, 1992.
- [15] A. L. Maloy, D. S. Longnecker, and E. R. Greenberg, “The relation of islet amyloid to the clinical type of diabetes,” *Hum Pathol*, vol. 12, pp. 917–22, Oct 1981.
- [16] C. F. Howard, Jr, “Longitudinal studies on the development of diabetes in individual macaca nigra,” *Diabetologia*, vol. 29, pp. 301–6, May 1986.
- [17] E. J. de Koning, N. L. Bodkin, B. C. Hansen, and A. Clark, “Diabetes mellitus in macaca mulatta monkeys is characterised by islet amyloidosis and reduction in beta-cell population,” *Diabetologia*, vol. 36, pp. 378–84, May 1993.
- [18] T. D. O’Brien, D. W. Hayden, K. H. Johnson, and T. F. Fletcher, “Immunohistochemical morphometry of pancreatic endocrine cells in diabetic, normoglycaemic glucose-intolerant and normal cats,” *J Comp Pathol*, vol. 96, pp. 357–69, Jul 1986.

- [19] C. B. Verchere, D. A. D'Alessio, R. D. Palmiter, G. C. Weir, S. Bonner-Weir, D. G. Baskin, and S. E. Kahn, "Islet amyloid formation associated with hyperglycemia in transgenic mice with pancreatic beta cell expression of human islet amyloid polypeptide," *Proc Natl Acad Sci U S A*, vol. 93, pp. 3492–6, Apr 1996.
- [20] J. Janson, W. C. Soeller, P. C. Roche, R. T. Nelson, A. J. Torchia, D. K. Kreutter, and P. C. Butler, "Spontaneous diabetes mellitus in transgenic mice expressing human islet amyloid polypeptide," *Proc Natl Acad Sci U S A*, vol. 93, pp. 7283–8, Jul 1996.
- [21] M. Couce, L. A. Kane, T. D. O'Brien, J. Charlesworth, W. Soeller, J. McNeish, D. Kreutter, P. Roche, and P. C. Butler, "Treatment with growth hormone and dexamethasone in mice transgenic for human islet amyloid polypeptide causes islet amyloidosis and beta-cell dysfunction," *Diabetes*, vol. 45, pp. 1094–101, Aug 1996.
- [22] J. Hardy and D. J. Selkoe, "The amyloid hypothesis of alzheimer's disease: progress and problems on the road to therapeutics," *Science*, vol. 297, pp. 353–6, Jul 2002.
- [23] M. Sunde, L. C. Serpell, M. Bartlam, P. E. Fraser, M. B. Pepys, and C. C. Blake, "Common core structure of amyloid fibrils by synchrotron x-ray diffraction," *J Mol Biol*, vol. 273, pp. 729–39, Oct 1997.
- [24] R. Nelson, M. R. Sawaya, M. Balbirnie, A. Ø. Madsen, C. Riek, R. Grothe, and D. Eisenberg, "Structure of the cross-beta spine of amyloid-like fibrils," *Nature*, vol. 435, pp. 773–8, Jun 2005.
- [25] B. T. Hyman, G. W. Van Hoesen, A. R. Damasio, and C. L. Barnes, "Alzheimer's disease: cell-specific pathology isolates the hippocampal formation," *Science*, vol. 225, pp. 1168–70, Sep 1984.
- [26] P. J. Whitehouse, D. L. Price, R. G. Struble, A. W. Clark, J. T. Coyle, and M. R. Delon, "Alzheimer's disease and senile dementia: loss of neurons in the basal forebrain," *Science*, vol. 215, pp. 1237–9, Mar 1982.
- [27] C. M. Dobson, "Protein misfolding, evolution and disease," *Trends Biochem Sci*, vol. 24, pp. 329–32, Sep 1999.

- [28] S. Tzotzos and A. J. Doig, “Amyloidogenic sequences in native protein structures,” *Protein Sci*, vol. 19, pp. 327–48, Feb 2010.
- [29] C. Haass and D. J. Selkoe, “Soluble protein oligomers in neurodegeneration: lessons from the alzheimer’s amyloid beta-peptide,” *Nat Rev Mol Cell Biol*, vol. 8, pp. 101–12, Feb 2007.
- [30] F. Chiti and C. M. Dobson, “Protein misfolding, functional amyloid, and human disease,” *Annu Rev Biochem*, vol. 75, pp. 333–66, 2006.
- [31] R. Tycko, “Progress towards a molecular-level structural understanding of amyloid fibrils,” *Curr Opin Struct Biol*, vol. 14, pp. 96–103, Feb 2004.
- [32] B. H. Toyama and J. S. Weissman, “Amyloid structure: conformational diversity and consequences,” *Annu Rev Biochem*, vol. 80, pp. 557–85, 2011.
- [33] C. Wasmer, A. Lange, H. Van Melckebeke, A. B. Siemer, R. Riek, and B. H. Meier, “Amyloid fibrils of the het-s(218-289) prion form a beta solenoid with a triangular hydrophobic core,” *Science*, vol. 319, pp. 1523–6, Mar 2008.
- [34] J. J. W. Wiltzius, S. A. Sievers, M. R. Sawaya, D. Cascio, D. Popov, C. Riek, and D. Eisenberg, “Atomic structure of the cross-beta spine of islet amyloid polypeptide (amylin),” *Protein Sci*, vol. 17, pp. 1467–74, Sep 2008.
- [35] J. Meinhardt, C. Sachse, P. Hortschansky, N. Grigorieff, and M. Fändrich, “Abeta(1-40) fibril polymorphism implies diverse interaction patterns in amyloid fibrils,” *J Mol Biol*, vol. 386, pp. 869–77, Feb 2009.
- [36] A. K. Chamberlain, C. E. MacPhee, J. Zurdo, L. A. Morozova-Roche, H. A. Hill, C. M. Dobson, and J. J. Davis, “Ultrastructural organization of amyloid fibrils by atomic force microscopy,” *Biophys J*, vol. 79, pp. 3282–93, Dec 2000.
- [37] N. Mizuno, U. Baxa, and A. C. Steven, “Structural dependence of het-s amyloid fibril infectivity assessed by cryoelectron microscopy,” *Proc Natl Acad Sci U S A*, vol. 108, pp. 3252–7, Feb 2011.
- [38] B. Seilheimer, B. Bohrmann, L. Bondolfi, F. Müller, D. Stüber, and H. Döbeli, “The toxicity of the alzheimer’s beta-amyloid peptide correlates with a distinct fiber morphology,” *J Struct Biol*, vol. 119, pp. 59–71, Jun 1997.

- [39] R. Diaz-Avalos, C.-Y. King, J. Wall, M. Simon, and D. L. D. Caspar, "Strain-specific morphologies of yeast prion amyloid fibrils," *Proc Natl Acad Sci U S A*, vol. 102, pp. 10165–70, Jul 2005.
- [40] K.-I. Yamaguchi, S. Takahashi, T. Kawai, H. Naiki, and Y. Goto, "Seeding-dependent propagation and maturation of amyloid fibril conformation," *J Mol Biol*, vol. 352, pp. 952–60, Sep 2005.
- [41] J. L. Jiménez, E. J. Nettleton, M. Bouchard, C. V. Robinson, C. M. Dobson, and H. R. Saibil, "The protofilament structure of insulin amyloid fibrils," *Proc Natl Acad Sci U S A*, vol. 99, pp. 9196–201, Jul 2002.
- [42] C. Goldsbury, K. Goldie, J. Pellaud, J. Seelig, P. Frey, S. A. Müller, J. Kistler, G. J. Cooper, and U. Aepli, "Amyloid fibril formation from full-length and fragments of amylin," *J Struct Biol*, vol. 130, pp. 352–62, Jun 2000.
- [43] C. S. Goldsbury, S. Wirtz, S. A. Müller, S. Sunderji, P. Wicki, U. Aepli, and P. Frey, "Studies on the in vitro assembly of a beta 1-40: implications for the search for a beta fibril formation inhibitors," *J Struct Biol*, vol. 130, pp. 217–31, Jun 2000.
- [44] J. A. Hardy and G. A. Higgins, "Alzheimer's disease: the amyloid cascade hypothesis," *Science*, vol. 256, pp. 184–5, Apr 1992.
- [45] Z. S. Khachaturian, "Diagnosis of alzheimer's disease," *Arch Neurol*, vol. 42, pp. 1097–105, Nov 1985.
- [46] S. S. Mirra, A. Heyman, D. McKeel, S. M. Sumi, B. J. Crain, L. M. Brownlee, F. S. Vogel, J. P. Hughes, G. van Belle, and L. Berg, "The consortium to establish a registry for alzheimer's disease (cerad). part ii. standardization of the neuropathologic assessment of alzheimer's disease," *Neurology*, vol. 41, pp. 479–86, Apr 1991.
- [47] M. M. Esiri, R. C. Pearson, J. E. Steele, D. M. Bowen, and T. P. Powell, "A quantitative study of the neurofibrillary tangles and the choline acetyltransferase activity in the cerebral cortex and the amygdala in alzheimer's disease," *J Neurol Neurosurg Psychiatry*, vol. 53, pp. 161–5, Feb 1990.
- [48] A. K. Paravastu, R. D. Leapman, W.-M. Yau, and R. Tycko, "Molecular structural basis for polymorphism in alzheimer's beta-amyloid fibrils," *Proc Natl Acad Sci U S A*, vol. 105, pp. 18349–54, Nov 2008.

- [49] J. C. Stroud, C. Liu, P. K. Teng, and D. Eisenberg, "Toxic fibrillar oligomers of amyloid- have cross- structure," *Proc Natl Acad Sci U S A*, vol. 109, pp. 7717–22, May 2012.
- [50] L. Marzban, K. Park, and C. B. Verchere, "Islet amyloid polypeptide and type 2 diabetes," *Exp Gerontol*, vol. 38, pp. 347–51, Apr 2003.
- [51] P. Westermark, A. Andersson, and G. T. Westermark, "Islet amyloid polypeptide, islet amyloid, and diabetes mellitus," *Physiol Rev*, vol. 91, pp. 795–826, Jul 2011.
- [52] C. S. Goldsbury, G. J. Cooper, K. N. Goldie, S. A. Müller, E. L. Saafi, W. T. Gruijters, M. P. Misur, A. Engel, U. Aebi, and J. Kistler, "Polymorphic fibrillar assembly of human amylin," *J Struct Biol*, vol. 119, pp. 17–27, Jun 1997.
- [53] G. Jiang and B. B. Zhang, "Glucagon and regulation of glucose metabolism," *Am J Physiol Endocrinol Metab*, vol. 284, pp. E671–8, Apr 2003.
- [54] A. Clark, C. A. Wells, I. D. Buley, J. K. Cruickshank, R. I. Vanhegan, D. R. Matthews, G. J. Cooper, R. R. Holman, and R. C. Turner, "Islet amyloid, increased a-cells, reduced b-cells and exocrine fibrosis: quantitative changes in the pancreas in type 2 diabetes," *Diabetes Res*, vol. 9, pp. 151–9, Dec 1988.
- [55] C. Sempoux, Y. Guiot, D. Dubois, P. Moulin, and J. Rahier, "Human type 2 diabetes: morphological evidence for abnormal beta-cell function," *Diabetes*, vol. 50 Suppl 1, pp. S172–7, Feb 2001.
- [56] H.-M. Schneider, S. Störkel, and W. Will, "Das amyloid der langerhansschen inseln und seine beziehung zum diabetes mellitus," *DMW-Deutsche Medizinische Wochenschrift*, vol. 105, no. 33, pp. 1143–1147, 1980.
- [57] A. Lorenzo, B. Razzaboni, G. C. Weir, and B. A. Yankner, "Pancreatic islet cell toxicity of amylin associated with type-2 diabetes mellitus," *Nature*, vol. 368, pp. 756–60, Apr 1994.
- [58] M. Polymeropoulos, C. Lavedan, E. Leroy, S. Ide, A. Dehejia, A. Dutra, B. Pike, H. Root, J. Rubenstein, R. Boyer, E. Stenroos, S. Chandrasekharappa, A. Athanassiadou, T. Papapetropoulos, W. Johnson, A. Lazzarini, R. Duvoisin, G. DiIorio, L. Golbe, and R. Nussbaum, "Mutation in the alpha-synuclein gene identified in families with parkinson's disease," *Science*, vol. 276, pp. 2045–2047, 1997.

- [59] J. J. Zarranz, J. Alegre, J. C. Gómez-Esteban, E. Lezcano, R. Ros, I. Ampuero, L. Vidal, J. Hoenicka, O. Rodriguez, B. Atarés, V. Llorens, E. Gomez Tortosa, T. del Ser, D. G. Muñoz, and J. G. de Yebenes, “The new mutation, e46k, of alpha-synuclein causes parkinson and lewy body dementia,” *Ann Neurol*, vol. 55, pp. 164–73, Feb 2004.
- [60] R. Krüger, W. Kuhn, T. Müller, D. Voitalla, M. Graeber, S. Kösel, H. Przun-tek, J. T. Epplen, L. Schöls, and O. Riess, “Ala30pro mutation in the gene encoding alpha-synuclein in parkinson’s disease,” *Nat Genet*, vol. 18, pp. 106–8, Feb 1998.
- [61] H. A. Lashuel, C. R. Overk, A. Oueslati, and E. Masliah, “The many faces of -synuclein: from structure and toxicity to therapeutic target,” *Nat Rev Neurosci*, vol. 14, pp. 38–48, Jan 2013.
- [62] K. Conway, S. Lee, J. Rochet, T. Ding, R. Williamson, and P. Lansbury, “Acceleration of oligomerization, not fibrillization, is a shared property of both alpha-synuclein mutations linked to early-onset parkinson’s disease: Implications for pathogenesis and therapy,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, pp. 571–576, 2000.
- [63] D. Ghosh, M. Mondal, G. M. Mohite, P. K. Singh, P. Ranjan, A. Anoop, S. Ghosh, N. N. Jha, A. Kumar, and S. K. Maji, “The parkinson’s disease-associated h50q mutation accelerates -synuclein aggregation in vitro,” *Biochemistry*, Sep 2013.
- [64] L. Xu, S. Shan, and X. Wang, “Single point mutation alters the microstate dynamics of amyloid -protein a42 as revealed by dihedral dynamics analyses,” *J Phys Chem B*, vol. 117, pp. 6206–16, May 2013.
- [65] Y. Yoshiike, T. Akagi, and A. Takashima, “Surface structure of amyloid-beta fibrils contributes to cytotoxicity,” *Biochemistry*, vol. 46, pp. 9805–12, Aug 2007.
- [66] A. D. Williams, S. Shivaprasad, and R. Wetzel, “Alanine scanning mutagenesis of abeta(1-40) amyloid fibril stability,” *J Mol Biol*, vol. 357, pp. 1283–94, Apr 2006.
- [67] S. Morita, S. Sakagashira, M. Ueyama, Y. Shimajiri, M. Furuta, and T. Sanke, “Progressive deterioration of insulin secretion in japanese type 2 diabetic patients in comparison with those who carry the s20g mutation of the islet amyloid

- polypeptide gene: A long-term follow-up study,” *Journal of Diabetes Investigation*, vol. 2, pp. 287–292, 2011.
- [68] S. Lee, Y. Hashim, J. Li, G. Ko, J. Critchley, C. Cockram, and J. Chan, “The islet amyloid polypeptide (amylin) gene s20g mutation in chinese subjects: Evidence for associations with type 2 diabetes and cholesterol levels,” *Clinical Endocrinology*, vol. 54, pp. 541–546, 2001.
- [69] S. Sakagashira, T. Sanke, T. Hanabusa, H. Shimomura, S. Ohagi, K. Y. Kumagaye, K. Nakajima, and K. Nanjo, “Missense mutation of amylin gene (s20g) in japanese niddm patients,” *Diabetes*, vol. 45, pp. 1279–81, Sep 1996.
- [70] Z. Ma, G. T. Westermark, S. Sakagashira, T. Sanke, A. Gustavsson, H. Sakamoto, U. Engström, K. Nanjo, and P. Westermark, “Enhanced in vitro production of amyloid-like fibrils from mutant (s20g) islet amyloid polypeptide,” *Amyloid*, vol. 8, pp. 242–9, Dec 2001.
- [71] S. Sakagashira, H. J. Hiddinga, K. Tateishi, T. Sanke, T. Hanabusa, K. Nanjo, and N. L. Eberhardt, “S20g mutant amylin exhibits increased in vitro amyloidogenicity and increased intracellular cytotoxicity compared to wild-type amylin,” *Am J Pathol*, vol. 157, pp. 2101–9, Dec 2000.
- [72] P. Cao, L.-H. Tu, A. Abedini, O. Levsh, R. Akter, V. Patsalo, A. M. Schmidt, and D. P. Raleigh, “Sensitivity of amyloid formation by human islet amyloid polypeptide to mutations at residue 20,” *J Mol Biol*, vol. 421, pp. 282–95, Aug 2012.
- [73] N. R. Poa, G. J. S. Cooper, and P. F. Edgar, “Amylin gene promoter mutations predispose to type 2 diabetes in new zealand maori,” *Diabetologia*, vol. 46, pp. 574–8, Apr 2003.
- [74] C. Betsholtz, L. Christmansson, U. Engström, F. Rorsman, V. Svensson, K. H. Johnson, and P. Westermark, “Sequence divergence in a specific region of islet amyloid polypeptide (iapp) explains differences in islet amyloid formation between species,” *FEBS Lett*, vol. 251, pp. 261–4, Jul 1989.
- [75] P. Westermark, U. Engström, K. H. Johnson, G. T. Westermark, and C. Betsholtz, “Islet amyloid polypeptide: pinpointing amino acid residues linked to amyloid fibril formation,” *Proc Natl Acad Sci U S A*, vol. 87, pp. 5036–40, Jul 1990.

- [76] D. F. Moriarty and D. P. Raleigh, "Effects of sequential proline substitutions on amyloid formation by human amylin<sub>20-29</sub>," *Biochemistry*, vol. 38, pp. 1811–8, Feb 1999.
- [77] V. Planté-Bordeneuve and G. Said, "Familial amyloid polyneuropathy," *Lancet Neurol*, vol. 10, pp. 1086–97, Dec 2011.
- [78] H. Ndlovu, A. E. Ashcroft, S. E. Radford, and S. A. Harris, "Effect of sequence variation on the mechanical response of amyloid fibrils probed by steered molecular dynamics simulation," *Biophys J*, vol. 102, pp. 587–96, Feb 2012.
- [79] C. Weyer, D. G. Maggs, A. A. Young, and O. G. Kolterman, "Amylin replacement with pramlintide as an adjunct to insulin therapy in type 1 and type 2 diabetes mellitus: a physiological approach toward improved metabolic control," *Curr Pharm Des*, vol. 7, pp. 1353–73, Sep 2001.
- [80] T. Cellmer, D. Bratko, J. M. Prausnitz, and H. W. Blanch, "Protein aggregation in silico," *Trends Biotechnol*, vol. 25, pp. 254–61, Jun 2007.
- [81] B. Urbanc, L. Cruz, S. Yun, S. V. Buldyrev, G. Bitan, D. B. Teplow, and H. E. Stanley, "In silico study of amyloid beta-protein folding and oligomerization," *Proc Natl Acad Sci U S A*, vol. 101, pp. 17345–50, Dec 2004.
- [82] S. Santini, N. Mousseau, and P. Derreumaux, "In silico assembly of alzheimer's abeta<sub>16-22</sub> peptide into beta-sheets," *J Am Chem Soc*, vol. 126, pp. 11509–16, Sep 2004.
- [83] J. M. Borreguero, B. Urbanc, N. D. Lazo, S. V. Buldyrev, D. B. Teplow, and H. E. Stanley, "Folding events in the 21-30 region of amyloid beta-protein (abeta) studied in silico," *Proc Natl Acad Sci U S A*, vol. 102, pp. 6015–20, Apr 2005.
- [84] P. Gupta, C. K. Hall, and A. C. Voegler, "Effect of denaturant and protein concentrations upon protein refolding and aggregation: a simple lattice model," *Protein Sci*, vol. 7, pp. 2642–52, Dec 1998.
- [85] H. D. Nguyen and C. K. Hall, "Effect of rate of chemical or thermal renaturation on refolding and aggregation of a simple lattice protein," *Biotechnol Bioeng*, vol. 80, pp. 823–34, Dec 2002.

- [86] T. Cellmer, D. Bratko, J. M. Prausnitz, and H. Blanch, “The competition between protein folding and aggregation: off-lattice minimalist model studies,” *Biotechnol Bioeng*, vol. 89, pp. 78–87, Jan 2005.
- [87] A. Cooper, C. M. Johnson, J. H. Lakey, and M. Nöllmann, “Heat does not come in different colours: entropy-enthalpy compensation, free energy windows, quantum confinement, pressure perturbation calorimetry, solvation and the multiple causes of heat capacity effects in biomolecular interactions,” *Biophys Chem*, vol. 93, pp. 215–30, Nov 2001.
- [88] R. LoBrutto, A. Jones, Y. V. Kazakevich, and H. M. McNair, “Effect of the eluent pH and acidic modifiers in high-performance liquid chromatography retention of basic analytes,” *J Chromatogr A*, vol. 913, pp. 173–87, Apr 2001.
- [89] P. Koehl and M. Delarue, “Aqvasol: An efficient solver for the dipolar poisson-boltzmann-langevin equation,” *J Chem Phys*, vol. 132, p. 064101, Feb 2010.
- [90] S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess, and E. Lindahl, “Gromacs 4.5: a high-throughput and highly parallel open source molecular simulation toolkit,” *Bioinformatics*, vol. 29, pp. 845–54, Apr 2013.
- [91] B. Hess, C. Kutzner, D. Van der Spoel, and E. Lindahl, “Gromacs 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation.,” *Journal of Chemical Theory and Computation*, vol. 4, no. 3, pp. 435–447, 2008.
- [92] J.-T. Guo and Y. Xu, “Amyloid fibril structure modeling using protein threading and molecular dynamics simulations,” *Methods Enzymol*, vol. 412, pp. 300–14, 2006.
- [93] M. Källberg, H. Wang, S. Wang, J. Peng, Z. Wang, H. Lu, and J. Xu, “Template-based protein structure modeling using the raptorx web server,” *Nat Protoc*, vol. 7, pp. 1511–22, Aug 2012.
- [94] J. Peng and J. Xu, “Raptorx: exploiting structure information for protein alignment by statistical inference,” *Proteins*, vol. 79 Suppl 10, pp. 161–71, 2011.
- [95] M. Källberg, G. Margaryan, S. Wang, J. Ma, and J. Xu, “Raptorx server: a resource for template-based protein structure modeling,” *Methods Mol Biol*, vol. 1137, pp. 17–27, 2014.

- [96] J. Waldispühl, C. W. O'Donnell, S. Devadas, P. Clote, and B. Berger, "Modeling ensembles of transmembrane beta-barrel proteins," *Proteins*, vol. 71, pp. 1097–112, May 2008.
- [97] S. Shenker, C. W. O'Donnell, S. Devadas, B. Berger, and J. Waldispühl, "Efficient traversal of beta-sheet protein folding pathways using ensemble models," *J Comput Biol*, vol. 18, pp. 1635–47, Nov 2011.
- [98] A.-M. Fernandez-Escamilla, F. Rousseau, J. Schymkowitz, and L. Serrano, "Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins," *Nat Biotechnol*, vol. 22, pp. 1302–6, Oct 2004.
- [99] C. W. O'Donnell, J. Waldispühl, M. Lis, R. Halfmann, S. Devadas, S. Lindquist, and B. Berger, "A method for probing the mutational landscape of amyloid structure," *Bioinformatics*, vol. 27, pp. i34–42, Jul 2011.
- [100] S. J. Russell, P. Norvig, and E. Davis, *Artificial intelligence: a modern approach*. Prentice Hall Upper Saddle River, 3rd ed edition, 2010.
- [101] R. Zhang, X. Hu, H. Khant, S. J. Ludtke, W. Chiu, M. F. Schmid, C. Frieden, and J.-M. Lee, "Interprotofilament interactions between alzheimer's abeta1-42 peptides in amyloid fibrils revealed by cryoem," *Proc Natl Acad Sci U S A*, vol. 106, pp. 4653–8, Mar 2009.
- [102] A. T. Petkova, W.-M. Yau, and R. Tycko, "Experimental constraints on quaternary structure in alzheimer's beta-amyloid fibrils," *Biochemistry*, vol. 45, pp. 498–512, Jan 2006.
- [103] C. Wasmer, A. Soragni, R. Sabaté, A. Lange, R. Riek, and B. H. Meier, "Infectious and noninfectious amyloids of the het-s(218-289) prion have different nmr spectra," *Angew Chem Int Ed Engl*, vol. 47, no. 31, pp. 5839–41, 2008.
- [104] S. Bedrood, Y. Li, J. M. Isas, B. G. Hegde, U. Baxa, I. S. Haworth, and R. Langen, "Fibril structure of human islet amyloid polypeptide," *J Biol Chem*, vol. 287, pp. 5235–41, Feb 2012.
- [105] M. Fändrich, "On the structural definition of amyloid fibrils and other polypeptide aggregates," *Cell Mol Life Sci*, vol. 64, pp. 2066–78, Aug 2007.
- [106] C. M. Dobson, "Protein folding and misfolding," *Nature*, vol. 426, pp. 884–90, Dec 2003.

- [107] R. D. Hills, Jr and C. L. Brooks, 3rd, "Hydrophobic cooperativity as a mechanism for amyloid nucleation," *J Mol Biol*, vol. 368, pp. 894–901, May 2007.
- [108] N. L. Fawzi, E.-H. Yap, Y. Okabe, K. L. Kohlstedt, S. P. Brown, and T. Head-Gordon, "Contrasting disease and nondisease protein aggregation by molecular simulation," *Acc Chem Res*, vol. 41, pp. 1037–47, Aug 2008.
- [109] R. Halfmann, S. Alberti, R. Krishnan, N. Lyle, C. W. O'Donnell, O. D. King, B. Berger, R. V. Pappu, and S. Lindquist, "Opposing effects of glutamine and asparagine govern prion formation by intrinsically disordered proteins," *Mol Cell*, vol. 43, pp. 72–84, Jul 2011.
- [110] C. Azuara, H. Orland, M. Bon, P. Koehl, and M. Delarue, "Incorporating dipolar solvents with variable density in poisson-boltzmann electrostatics," *Biophys J*, vol. 95, pp. 5587–605, Dec 2008.
- [111] P. Shirley, M. Ashikhmin, and S. Marschner, *Fundamentals of Computer Graphics*. A. K. Peters, 2009.
- [112] H. M. Berman, T. Battistuz, T. N. Bhat, W. F. Bluhm, P. E. Bourne, K. Burkhardt, Z. Feng, G. L. Gilliland, L. Iype, S. Jain, P. Fagan, J. Marvin, D. Padilla, V. Ravichandran, B. Schneider, N. Thanki, H. Weissig, J. D. Westbrook, and C. Zardecki, "The protein data bank," *Acta Crystallogr D Biol Crystallogr*, vol. 58, pp. 899–907, Jun 2002.
- [113] S. R. Collins, A. Douglass, R. D. Vale, and J. S. Weissman, "Mechanism of prion propagation: amyloid growth occurs by monomer addition," *PLoS Biol*, vol. 2, p. e321, Oct 2004.
- [114] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans, "Interaction models for water in relation to protein hydration," *Intermolecular Forces*, 1981.
- [115] W. Jorgensen, J. Chandrasekhar, J. Madura, R. Impey, and M. Klein, "Comparison of simple potential functions for simulating liquid water," *ournal of Chemical Physics*, vol. 79, no. 2, pp. 926–935, 1983.
- [116] M. Mahoney and W. Jorgensen, "A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions," *Journal of Chemical Physics*, vol. 112, no. 20, pp. 8910–8922, 2000.

- [117] T. Darden, D. York, and L. Pedersen, "Particle mesh ewald - an  $n \cdot \log(n)$  method for ewald sums in large systems," *Journal of Chemical Physics*, vol. 98, pp. 10089–10092, 1993.
- [118] U. Essmann, L. Perera, M. Berkowitz, T. Darden, H. Lee, and L. Pedersen, "A smooth particle mesh ewald method," *Journal of Chemical Physics*, vol. 103, pp. 8577–8593, 1995.
- [119] Y. N. Vorobjev, J. A. Vila, and H. A. Scheraga, "Fambe-ph: a fast and accurate method to compute the total solvation free energies of proteins," *J Phys Chem B*, vol. 112, pp. 11122–36, Sep 2008.
- [120] S. Luca, W.-M. Yau, R. Leapman, and R. Tycko, "Peptide conformation and supramolecular organization in amylin fibrils: constraints from solid-state nmr," *Biochemistry*, vol. 46, pp. 13505–22, Nov 2007.
- [121] M. Citron, "Strategies for disease modification in alzheimer's disease," *Nat Rev Neurosci*, vol. 5, pp. 677–85, Sep 2004.
- [122] B. Caughey and P. Lansbury, "Protofibrils, pores, fibrils, and neurodegeneration: Separating the responsible protein aggregates from the innocent bystanders," *Annual Review of Neuroscience*, vol. 26, pp. 267–298, 2003.
- [123] Y. Chebaro and P. Derreumaux, "Targeting the early steps of abeta16-22 protofibril disassembly by n-methylated inhibitors: a numerical study," *Proteins*, vol. 75, pp. 442–52, May 2009.
- [124] J. Bieschke, M. Herbst, T. Wiglenda, R. P. Friedrich, A. Boeddrich, F. Schiele, D. Kleckers, J. M. Lopez del Amo, B. A. Grüning, Q. Wang, M. R. Schmidt, R. Lurz, R. Anwyl, S. Schnoegl, M. Fändrich, R. F. Frank, B. Reif, S. Günther, D. M. Walsh, and E. E. Wanker, "Small-molecule conversion of toxic oligomers to nontoxic -sheet-rich amyloid fibrils," *Nat Chem Biol*, vol. 8, pp. 93–101, Jan 2012.
- [125] R. Gupta, N. Kapoor, D. P. Raleigh, and T. P. Sakmar, "Nucleobindin 1 caps human islet amyloid polypeptide protofibrils to prevent amyloid fibril formation," *J Mol Biol*, vol. 421, pp. 378–89, Aug 2012.
- [126] S. Sinha, D. H. J. Lopes, Z. Du, E. S. Pang, A. Shanmugam, A. Lomakin, P. Talbiersky, A. Tennstaedt, K. McDaniel, R. Bakshi, P.-Y. Kuo, M. Ehrmann, G. B. Benedek, J. A. Loo, F.-G. Klärner, T. Schrader, C. Wang, and G. Bitan,

- “Lysine-specific molecular tweezers are broad-spectrum inhibitors of assembly and toxicity of amyloid proteins,” *J Am Chem Soc*, vol. 133, pp. 16958–69, Oct 2011.
- [127] F. Meng, A. Abedini, A. Plesner, C. T. Middleton, K. J. Potter, M. T. Zanni, C. B. Verchere, and D. P. Raleigh, “The sulfated triphenyl methane derivative acid fuchsin is a potent inhibitor of amyloid formation by human islet amyloid polypeptide and protects against the toxic effects of amyloid formation,” *J Mol Biol*, vol. 400, pp. 555–66, Jul 2010.
- [128] P. T. Lansbury, Jr, “Evolution of amyloid: what normal protein folding may tell us about fibrillogenesis and disease,” *Proc Natl Acad Sci U S A*, vol. 96, pp. 3342–4, Mar 1999.
- [129] C. A. Ross and M. A. Poirier, “Protein aggregation and neurodegenerative disease,” *Nat Med*, vol. 10 Suppl, pp. S10–7, Jul 2004.
- [130] D. J. Selkoe, “Folding proteins in fatal ways,” *Nature*, vol. 426, pp. 900–4, Dec 2003.
- [131] A. Alonso, T. Zaidi, M. Novak, I. Grundke-Iqbal, and K. Iqbal, “Hyperphosphorylation induces self-assembly of tau into tangles of paired helical filaments/straight filaments,” *Proc Natl Acad Sci U S A*, vol. 98, pp. 6923–8, Jun 2001.
- [132] M. R. Smaoui, F. Poitevin, M. Delarue, P. Koehl, H. Orland, and J. Waldispühl, “Computational assembly of polymorphic amyloid fibrils reveals stable aggregates,” *Biophys J*, vol. 104, pp. 683–93, Feb 2013.
- [133] Q. Wang, A. A. Canutescu, and R. L. Dunbrack, Jr, “Scwrl and molide: computer programs for side-chain conformation prediction and homology modeling,” *Nat Protoc*, vol. 3, no. 12, pp. 1832–47, 2008.
- [134] A. Clark, G. J. Cooper, C. E. Lewis, J. F. Morris, A. C. Willis, K. B. Reid, and R. C. Turner, “Islet amyloid formed from diabetes-associated peptide may be pathogenic in type-2 diabetes,” *Lancet*, vol. 2, pp. 231–4, Aug 1987.
- [135] S. E. Kahn, D. A. D’Alessio, M. W. Schwartz, W. Y. Fujimoto, J. W. Ensink, G. J. Taborsky, Jr, and D. Porte, Jr, “Evidence of cosecretion of islet amyloid polypeptide and insulin by beta-cells,” *Diabetes*, vol. 39, pp. 634–8, May 1990.

- [136] G. J. Ryan, L. J. Jobe, and R. Martin, "Pramlintide in the treatment of type 1 and type 2 diabetes mellitus," *Clin Ther*, vol. 27, pp. 1500–12, Oct 2005.
- [137] J. Pacheco-Quinto and E. A. Eckman, "Endothelin-converting enzymes degrade intracellular  $\alpha$ -amyloid produced within the endosomal/lysosomal pathway and autophagosomes," *J Biol Chem*, vol. 288, pp. 5606–15, Feb 2013.
- [138] M. K. Badman, R. A. Pryce, S. B. Chargé, J. F. Morris, and A. Clark, "Fibrillar islet amyloid polypeptide (amylin) is internalised by macrophages but resists proteolytic degradation," *Cell Tissue Res*, vol. 291, pp. 285–94, Feb 1998.
- [139] P. Cao, P. Marek, H. Noor, V. Patsalo, L.-H. Tu, H. Wang, A. Abedini, and D. P. Raleigh, "Islet amyloid: from fundamental biophysics to mechanisms of cytotoxicity," *FEBS Lett*, vol. 587, pp. 1106–18, Apr 2013.
- [140] B. W. Koo, J. A. Hebda, and A. D. Miranker, "Amide inequivalence in the fibrillar assembly of islet amyloid polypeptide," *Protein Eng Des Sel*, vol. 21, pp. 147–54, Mar 2008.
- [141] N. G. N. Milton and J. R. Harris, "Fibril formation and toxicity of the non-amyloidogenic rat amylin peptide," *Micron*, vol. 44, pp. 246–53, Jan 2013.
- [142] D. R. Whiting, L. Guariguata, C. Weil, and J. Shaw, "Idf diabetes atlas: global estimates of the prevalence of diabetes for 2011 and 2030," *Diabetes Res Clin Pract*, vol. 94, pp. 311–21, Dec 2011.
- [143] G. J. Cooper, B. Leighton, G. D. Dimitriadis, M. Parry-Billings, J. M. Kowalchuk, K. Howland, J. B. Rothbard, A. C. Willis, and K. B. Reid, "Amylin found in amyloid deposits in human type 2 diabetes mellitus may be a hormone that regulates glycogen metabolism in skeletal muscle," *Proc Natl Acad Sci U S A*, vol. 85, pp. 7763–6, Oct 1988.
- [144] S. G. Straub and G. W. G. Sharp, "Glucose-stimulated signaling pathways in biphasic insulin secretion," *Diabetes Metab Res Rev*, vol. 18, no. 6, pp. 451–63, 2002.
- [145] T. K. Bratanova-Tochkova, H. Cheng, S. Daniel, S. Gunawardana, Y.-J. Liu, J. Mulvaney-Musa, T. Schermerhorn, S. G. Straub, H. Yajima, and G. W. G. Sharp, "Triggering and augmentation mechanisms, granule pools, and biphasic insulin secretion," *Diabetes*, vol. 51 Suppl 1, pp. S83–90, Feb 2002.

- [146] A. A. Young, B. Gedulin, W. Vine, A. Percy, and T. J. Rink, "Gastric emptying is accelerated in diabetic bb rats and is slowed by subcutaneous injections of amylin," *Diabetologia*, vol. 38, pp. 642–8, Jun 1995.
- [147] B. R. Gedulin, T. J. Rink, and A. A. Young, "Dose-response for glucagonostatic effect of amylin in rats," *Metabolism*, vol. 46, pp. 67–70, Jan 1997.
- [148] P. A. Rushing, M. M. Hagan, R. J. Seeley, T. A. Lutz, and S. C. Woods, "Amylin: a novel action in the brain to reduce body weight," *Endocrinology*, vol. 141, pp. 850–3, Feb 2000.
- [149] T. A. Lutz, N. Geary, M. M. Szabady, E. Del Prete, and E. Scharrer, "Amylin decreases meal size in rats," *Physiol Behav*, vol. 58, pp. 1197–202, Dec 1995.
- [150] E. T. Jaikaran and A. Clark, "Islet amyloid and type 2 diabetes: from molecular misfolding to islet pathophysiology," *Biochim Biophys Acta*, vol. 1537, pp. 179–203, Nov 2001.
- [151] J. W. Höppener, J. S. Verbeek, E. J. de Koning, C. Oosterwijk, K. L. van Hulst, H. J. Visser-Vernooy, F. M. Hofhuis, S. van Gaalen, M. J. Berends, and W. H. Hackeng, "Chronic overproduction of islet amyloid polypeptide/amylin in transgenic mice: lysosomal localization of human islet amyloid polypeptide and lack of marked hyperglycaemia or hyperinsulinaemia," *Diabetologia*, vol. 36, pp. 1258–65, Dec 1993.
- [152] D. L. MacArthur, E. J. de Koning, J. S. Verbeek, J. F. Morris, and A. Clark, "Amyloid fibril formation is progressive and correlates with beta-cell secretion in transgenic mouse isolated islets," *Diabetologia*, vol. 42, pp. 1219–27, Oct 1999.
- [153] S. C. Lee, Y. Hashim, J. K. Li, G. T. Ko, J. A. Critchley, C. S. Cockram, and J. C. Chan, "The islet amyloid polypeptide (amylin) gene s20g mutation in chinese subjects: evidence for associations with type 2 diabetes and cholesterol levels," *Clin Endocrinol (Oxf)*, vol. 54, pp. 541–6, Apr 2001.
- [154] G. Grunberger, "Novel therapies for the management of type 2 diabetes mellitus: part 1. pramlintide and bromocriptine-qr," *J Diabetes*, vol. 5, pp. 110–7, Jun 2013.

- [155] L. M. Younk, M. Mikeladze, and S. N. Davis, "Pramlintide and the treatment of diabetes: a review of the data since its introduction," *Expert Opin Pharmacother*, vol. 12, pp. 1439–51, Jun 2011.
- [156] O. Schmitz, B. Brock, and J. Rungby, "Amylin agonists: a novel approach in the treatment of diabetes," *Diabetes*, vol. 53 Suppl 3, pp. S233–8, Dec 2004.
- [157] D. Singh-Franco, G. Robles, and D. Gazze, "Pramlintide acetate injection for the treatment of type 1 and type 2 diabetes mellitus," *Clin Ther*, vol. 29, pp. 535–62, Apr 2007.
- [158] J. S. Freeman, "Insulin analog therapy: improving the match with physiologic insulin secretion," *J Am Osteopath Assoc*, vol. 109, pp. 26–36, Jan 2009.
- [159] J. J. Meier, R. Kaye, C.-Y. Lin, T. Gurlo, L. Haataja, S. Jayasinghe, R. Langen, C. G. Glabe, and P. C. Butler, "Inhibition of human iapp fibril formation does not prevent beta-cell death: evidence for distinct actions of oligomers and fibrils of human iapp," *Am J Physiol Endocrinol Metab*, vol. 291, pp. E1317–24, Dec 2006.
- [160] T. A. Mirzabekov, M. C. Lin, and B. L. Kagan, "Pore formation by the cytotoxic islet amyloid peptide amylin," *J Biol Chem*, vol. 271, pp. 1988–92, Jan 1996.
- [161] H. A. Lashuel, D. Hartley, B. M. Petre, T. Walz, and P. T. Lansbury, Jr, "Neurodegenerative disease: amyloid pores from pathogenic mutations," *Nature*, vol. 418, p. 291, Jul 2002.
- [162] R. A. Ritzel, J. J. Meier, C.-Y. Lin, J. D. Veldhuis, and P. C. Butler, "Human islet amyloid polypeptide oligomers disrupt cell coupling, induce apoptosis, and impair insulin secretion in isolated human islets," *Diabetes*, vol. 56, pp. 65–71, Jan 2007.
- [163] J. R. Cort, Z. Liu, G. M. Lee, K. N. L. Huggins, S. Janes, K. Prickett, and N. H. Andersen, "Solution state structures of human pancreatic amylin and pramlintide," *Protein Eng Des Sel*, vol. 22, pp. 497–513, Aug 2009.
- [164] M. Riddle, J. Frias, B. Zhang, H. Maier, C. Brown, K. Lutz, and O. Kolterman, "Pramlintide improved glycemic control and reduced weight in patients with type 2 diabetes using basal insulin," *Diabetes Care*, vol. 30, pp. 2794–9, Nov 2007.

- [165] T. Tomiyama, H. Kaneko, K. i. Kataoka, S. Asano, and N. Endo, "Rifampicin inhibits the toxicity of pre-aggregated amyloid peptides by binding to peptide fibrils and preventing amyloid-cell interaction," *Biochem J*, vol. 322 ( Pt 3), pp. 859–65, Mar 1997.
- [166] R. Mishra, D. Sellin, D. Radovan, A. Gohlke, and R. Winter, "Inhibiting islet amyloid polypeptide fibril formation by the red wine compound resveratrol," *Chembiochem*, vol. 10, pp. 445–9, Feb 2009.
- [167] L. He, X. Wang, C. Zhao, D. Zhu, and W. Du, "Inhibition of human amylin fibril formation by insulin-mimetic vanadium complexes," *Metallomics*, vol. 6, pp. 1087–96, May 2014.
- [168] Y. Porat, A. Abramowitz, and E. Gazit, "Inhibition of amyloid fibril formation by polyphenols: structural similarity and aromatic interactions as a common inhibition mechanism," *Chem Biol Drug Des*, vol. 67, pp. 27–37, Jan 2006.
- [169] Y. Porat, Y. Mazor, S. Efrat, and E. Gazit, "Inhibition of islet amyloid polypeptide fibril formation: a potential role for heteroaromatic interactions," *Biochemistry*, vol. 43, pp. 14454–62, Nov 2004.
- [170] X. Li, L. Ma, W. Zheng, and T. Chen, "Inhibition of islet amyloid polypeptide fibril formation by selenium-containing phycocyanin and prevention of beta cell apoptosis," *Biomaterials*, vol. 35, pp. 8596–604, Oct 2014.
- [171] M. Tatarek-Nossol, L.-M. Yan, A. Schmauder, K. Tenidis, G. Westermark, and A. Kapurniotu, "Inhibition of hiapp amyloid-fibril formation and apoptotic cell death by a designed hiapp amyloid- core-containing hexapeptide," *Chem Biol*, vol. 12, pp. 797–809, Jul 2005.
- [172] L. Schrodinger, "The PyMOL molecular graphics system, version 1.3r1." Aug 2010.
- [173] V. W. Bolie, "Coefficients of normal blood glucose regulation," *J Appl Physiol*, vol. 16, pp. 783–8, Sep 1961.
- [174] A. Boutayeb and A. Chetouani, "A critical review of mathematical models and data used in diabetology," *Biomed Eng Online*, vol. 5, p. 43, 2006.
- [175] A. Pick, J. Clark, C. Kubstrup, M. Levisetti, W. Pugh, S. Bonner-Weir, and K. S. Polonsky, "Role of apoptosis in failure of beta-cell mass compensation for

insulin resistance and beta-cell defects in the male zucker diabetic fatty rat,” *Diabetes*, vol. 47, pp. 358–64, Mar 1998.

- [176] G. C. Weir and S. Bonner-Weir, “Five stages of evolving beta-cell dysfunction during progression to diabetes,” *Diabetes*, vol. 53 Suppl 3, pp. S16–21, Dec 2004.