

TOWARDS BETTER DIABETES THERAPEUTICS: DESIGNING A MORE STABLE  
INSULIN ANALOG

Oumoul Ghaniyya Faiza Sambou Oumarou

Submitted to the Faculty of the University Graduate School  
in Partial Fulfillment of the Requirements  
For The Degree  
Master of Science  
In The Department of Biochemistry and Molecular Biology  
Indiana University

March 2023

Accepted by the Graduate Faculty of Indiana University, in partial fulfillment of the requirements for the degree of Master of Science.

Master's Thesis Committee

---

Michael Weiss, MD, Ph.D., MBA, Chair

---

Millie M. Georgiadis, Ph.D.

---

Quyên Q. Hoang, Ph.D.

---

Emily Sims, MD, Ph.D.

© 2023

Oumoul Ghaniyya Faiza Sambou Oumarou

## DEDICATION

I am dedicating this thesis project to my dearest father, Sambou Oumarou, who has always been my biggest supporter, especially when it comes to pursuing my academic career. I still remember the days when I used to read books, solve math problems, and study for exams with you, and those are memories that I will cherish for a lifetime. You are the best father I could have asked for, and I am forever grateful for all you have done for me. You have always encouraged me to reach my academic goals and taught me that for everything I want to achieve, the sky is the limit. I hope this research thesis contributes to curing diabetes for patients all over the world, starting with you.

I am also dedicating this thesis to my dearest mother, Aissa Hama, whose strength, resilience, and good example have taught me to work hard in everything I want to achieve. Your patience and incredible personality make you the easiest person to live with. Your unconditional love, constant encouragement, advice, and prayers kept me going on a daily basis. Even with the thousands of miles that separate us, I always think about you because you are the woman I aspire to be one day. Thank you for all the values you have passed on to me and I cannot be prouder to call you my mother.

## ACKNOWLEDGMENT

First, I would like to express my deepest gratitude to my committee chair, Dr. Michael Weiss, for letting me be part of his laboratory, being a great mentor and guiding me through my thesis project with his immense knowledge. Thank you for all the valuable lessons you have taught me along the way that will certainly make a difference in my career and life. I would also like to thank my committee members, Drs. Millie Georgiadis, Quyen Q. Hoang, and Emily Sims, for believing in and supporting my work. I am extremely grateful to Dr. Georgiadis for her enthusiasm and for pushing me to excel in the graduate courses and my research. I cannot begin to express my thanks to Dr. Hoang for being the kindest professor and offering unwavering support, patience, and guidance through my thesis presentations and defense. I would also like to extend my deepest gratitude to Dr. Sims for providing exceptional clinical insights into my research thesis. I am sincerely grateful to my graduate advisor, Dr. Clark Wells, for all the much-needed guidance as a new graduate student trying to find my way in the department. Thank you for always making time out of your busy schedule and for your amazing advising skills that successfully helped me prepare for my committee meetings and thesis defense.

A debt of gratitude is owed to Nicolàs Varas for helping me conduct the experiments, teaching me many laboratory skills, and motivating me through the process. Thank you for your patience as I was learning to navigate the laboratory and asking a million questions. I would like to show appreciation to Yen-Shan Chen, Ph.D. and Yanwu Yang, Ph.D., for their valuable data which I used in my thesis. I would like to thank my lab mates, Joe Racca, Ph.D., Mark Jarosinski, Ph.D., Balamurugan Dhayalan,

Ph.D., Ian Burke, Adam Brabender, and Chun-Lun Ni, Ph.D. for making me feel welcome in the laboratory while offering support and helpful insights into my research. They also provided friendly advice in editing my thesis and made my graduate research experience incredible and unforgettable.

Last, but not least, I would like to thank my biggest inspiration and wonderful husband, Assoumane Abass Mallam, for his help with rehearsing my presentations and supporting my endeavors. You have always encouraged me to chase my dreams and reach even higher. To my boys Irfan and Mohammed, you have kept me entertained and made me a stronger and better version of myself. Thanks to my caring family for their continuous support, encouragement, and prayers, which I am forever grateful for. I would not have made it without all their support, unforgettable moments, and entertainment. Special thanks to my brilliant twin brother, Khidrou Fadhloullah, for not only being my confidant but also always being available to offer the best advice. Also, congratulations on successfully defending your thesis and I am sure your patients would be lucky to have such a caring doctor. I am also grateful to my friend Josette Damiba for being the best friend ever. Thank you for all the silly jokes, laughter, entertainment, friendly ear, and support you offered during this time.

Oumoul Ghaniyya Faiza Sambou Oumarou

TOWARDS BETTER DIABETES THERAPEUTICS: DESIGNING A MORE STABLE  
INSULIN ANALOG

Insulin is a hormone that plays a central role in the regulation of human metabolism, and as a drug, is used in the treatment of diabetes mellitus. Hyperglycemia characterizes this condition due to a range of reasons from impaired insulin production by pancreatic beta cells to abnormalities resulting in resistance to insulin action. Depending on time and mechanism of action, the main types of insulin analogs are basal and prandial. Basal insulin analogs are slow-acting insulins that maintain a continuous basal level of insulin in the bloodstream. Prandial insulin analogs are fast-acting and their therapeutic goal is to avoid immediate and late post-prandial hyperglycemia. Most analogs face the problem of chemical degradation and amyloid-like fibril formation (fibrillation) in delivery devices. Thus, many modifications have been made to insulin in the effort to make it more stable and faster-acting. This thesis aims to study the effects of modifications that could be used to design an insulin analog with improved chemical and physical properties, while maintaining biological activity.

We studied six amino-acid substitutions to native human insulin in different combinations: desB1<sup>1</sup>, AB2<sup>2</sup>, EB3, EA8<sup>3</sup>, EA14, and EB29. Analogs of the protein were chemically synthesized. Then, fibrillation and circular dichroism assays were performed using purified proteins. The results suggested that EB3 and EA14 are stabilizing modifications that prevent fibril formation, whereas EA8 and EA14 increase the

---

<sup>1</sup> DesB1 = taking out the first residue of the B chain of insulin (in this case phenylalanine).

<sup>2</sup> AB2 = replacing the native valine of the B chain of insulin at position 2 with an alanine residue.

<sup>3</sup> EA8 = replacing the native threonine of the A chain of insulin at position 8 with a glutamic acid residue.

structural stability of an analog. Our findings also suggested that certain modifications in isolation may not impact overall stability, but when combined with others, may show detectable effects, which is why EA8 and EA14 became the focus of further experiments. Cell-based activity assays indicated that all the analogs had similar biological activities. Future work will assess chemical degradation, solubility, amide proton exchange (as monitored by NMR), and mitogenicity.

Michael Weiss, MD, Ph.D., MBA, Chair

Millie Georgiadis, Ph.D.

Quyen Q. Hoang, Ph.D.

Emily Sims, MD, Ph.D.

## TABLE OF CONTENTS

List of Tables .....	x
List of Figures .....	xi
List of Abbreviations .....	xii
Chapter 1. Introduction .....	1
1.1 Diabetes Mellitus and Insulin .....	1
1.2 Insulin Processing and Mechanism of Action .....	2
1.3 Insulin Types.....	4
1.4 Prandial Insulin Therapy Challenges and Background on Previous Efforts.....	7
1.5 Current Work on Prandial Insulin Design and Rationale .....	8
Chapter 2. Materials and Methods .....	15
2.1 Peptide Synthesis .....	15
2.1.1 Amino Acid Loading .....	16
2.1.2 Peptide Cleavage and Deprotection .....	17
2.1.3 Peptide Folding .....	18
2.1.4 HPLC Purification and Analytical LC-MS Identification .....	18
2.1.5 Trypsinization .....	19
2.1.6 Trypsin-mediated Semi-Synthesis .....	20
2.2 Fibrillation Assay.....	21
2.3 CD Guanidine Titration Assay.....	22
2.4 CD Wavelength Scan Assay- Phosphorylation (CHO-IRB) .....	23
2.5 In vitro Insulin Receptor Binding Assay .....	23
Chapter 3. Results .....	25
3.1 Initial Fibrillation Assay Results for Insulin-PET and KP Analogs .....	25
3.2 Initial CD Assay Results for Insulin-PET and KP Analogs .....	28
3.3 Final CD and Fibrillation Assay Results for Selected Analogs.....	31
3.4 Phosphorylation Assay Results.....	35
Chapter 4. Discussion and Conclusions.....	37
4.1 DesB1 and AB2 Effects .....	37
4.2 EB3 for Potential Clinical Use.....	38
4.3 EA8 and EA14 for Potential Clinical Use .....	38
4.4 Biological Potency .....	40
4.5 Fibrillation and CD Conceptual Correlation .....	40
4.6 Maximum Fluorescence Correlation with Amount of Fibrils.....	41
4.7 Dimer Complication in Fibrillation and CD experiments.....	42
4.8 Future Directions .....	43
Appendices.....	45
Appendix A.....	45
Appendix B .....	46
References.....	47
Curriculum Vitae	

## LIST OF TABLES

Table 1: Rationale of Design Elements .....	10
Table 2: Design and Synthesis of Insulin Analogs .....	16
Table 3: Fibrillation Lag Times .....	28
Table 4: CD-based Guanidine Titrations ( $\Delta G_u$ in kcal/mol.) .....	30
Table 5: Insulin-PET and KP Analogs: EA8 and EA14 .....	31
Table 6: CD Guanidine Titrations ( $\Delta G_u$ ) of Selected Analogs .....	33

## LIST OF FIGURES

Figure 1: Insulin Processing.....	3
Figure 2: Insulin Hexamer .....	4
Figure 3: Design Elements of Interest for A. PET and B. KP analogs .....	11
Figure 4: Structure of a General Insulin Analog.....	14
Figure 5: Fibrillation of “PET” Analogs (37°C, 60 µM protein, PBS [pH 7.4] .....	26
Figure 6: Fibrillation of “KP” Analogs (37°C, 60 µM protein, PBS [pH 7.4] .....	27
Figure 7: CD Studies for Insulin-PET Analogs .....	29
Figure 8: CD Studies for Insulin-KP Analogs .....	30
Figure 9: CD Studies for Insulin-PET and KP analogs, focusing on EA8 and EA14 .....	32
Figure 10: Fibrillation Properties of EA8 and EA14.....	35
Figure 11: Biological Activities of Insulin Analogs .....	36
Figure 12: Biological Activities of Selected Insulin Analogs .....	36

## LIST OF ABBREVIATIONS

CD = Circular Dichroism  
CHO = Chinese Hamster Ovary cells  
Desdi = 49-amino acid insulin precursor, lacking two residues (B29 and B30)  
DOI = des-octapeptide [B23-B30]-insulin  
DOSY = Diffusion Ordered Spectroscopy  
DNA = Deoxyribonucleic acid  
ER = Endoplasmic Reticulum  
HPLC = High-Performance Liquid Chromatography  
IRB = Insulin Receptor, type B  
IR = Insulin Receptor  
IR-A = Insulin Receptor, type A  
IGF-I and IGF-II = Insulin-like Growth Factor I and Insulin-like Growth Factor II  
IGF-IR = Insulin-like Growth Factor I Receptor  
LDL = Low-Density Lipoprotein  
MD = Molecular Dynamics  
MODY = Maturity-Onset Diabetes of the Young  
MIDY = Mutant *INS* gene-induced Diabetes of Youth  
NMR = Nuclear Magnetic Resonance  
PK = Pharmacokinetics  
T1D = Type I Diabetes  
T2D = Type II Diabetes  
ThT = Thioflavin T  
 $\Delta G_u$  = Change in free energy of unfolding

# CHAPTER 1

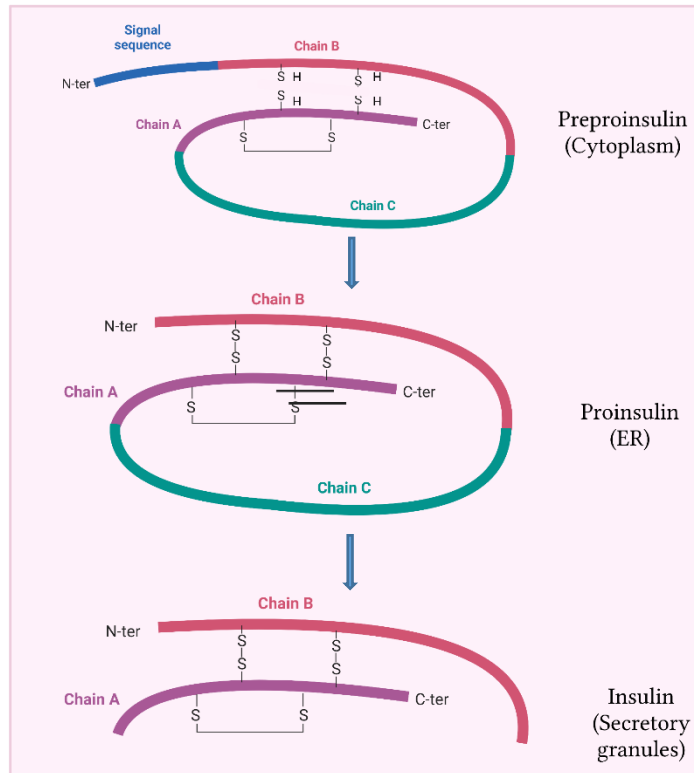
## INTRODUCTION

### 1.1 Diabetes Mellitus and Insulin

Diabetes mellitus is a chronic disease affecting more than five hundred million people worldwide with more than 11% living in the US [1]. The International Diabetes Federation (IDF) estimated an overall prevalence of diabetes mellitus of 366 million in 2011, and this is expected to rise to 552 million by 2030 [2, 3]. Classification of diabetes depends on many criteria, such as genetic classification (MIDY/MODY), age of onset and abruptness of hyperglycemia, but there are mainly two types: types I and II [2]. In healthy individuals, beta cells of the pancreas produce insulin that acts on target cells to induce glucose uptake, to keep blood-glucose levels within the normal range. In type I diabetes (T1D), autoimmune destruction of beta cells leads to failure to produce insulin and, therefore an increase in blood-glucose levels, termed hyperglycemia [2, 4]. In the most common type of diabetes, which is type II (T2D), beta cells can still produce insulin, but target cells become resistant to it, therefore leading to hyperglycemia when beta cell defect is present and compensation is incomplete [2]. Treatment in both cases can involve insulin therapy, which is essential in T1D, but common only in severe or long-standing T2D with beta-cell exhaustion. Other types exist, such as gestational diabetes, which occurs during pregnancy with elevated blood-glucose levels below those otherwise diagnostic of diabetes [5]. Untreated diabetes can lead to complications, such as loss of consciousness, ketoacidosis, kidney failure, heart disease, stroke, neuropathy, blindness, and eventually death [2, 6].

## **1.2 Insulin Processing and Mechanism of Action**

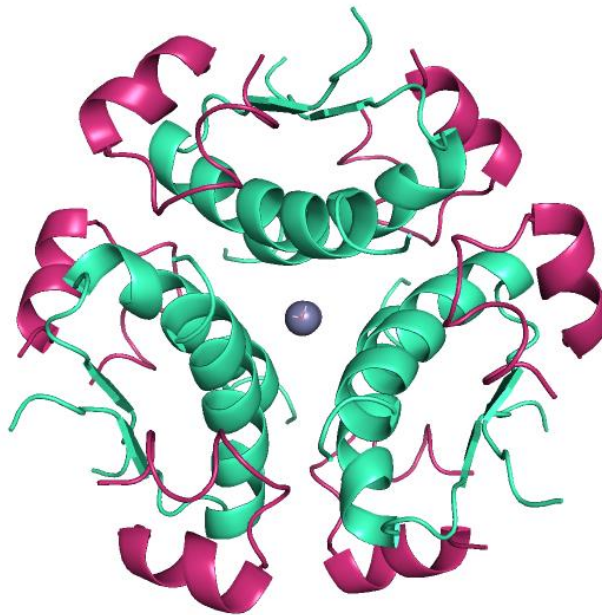
Insulin, first isolated in 1922 by Banting and Best, is an anabolic protein hormone that plays a central role in the regulation of vertebrate metabolism and, as a drug, is mainly used in the treatment of diabetes mellitus [7-9]. It is produced by pancreatic beta cells in the islet of Langerhans and secreted as a single-chain precursor designated proinsulin [10-13]. Proinsulin is composed of a 24-amino acid N-terminal hydrophobic signal peptide, followed by the 30-amino acid insulin B chain, an Arg-Arg sequence, 31 amino acids of the connecting peptide (C-peptide), a Lys-Arg sequence, and then the 21-amino acid insulin A chain [12]. The signal sequence allows for the translocation of proinsulin from the cytoplasm to the ER, whereas the C-domain is thought to facilitate structural alignment and folding of insulin for the proper formation of disulfide bridges [12]. In the ER nascent proinsulin undergoes oxidative folding coupled to the formation of three disulfide bridges to become folded proinsulin, which is then transferred to the Golgi [9]. There, proinsulin convertases PC1 and PC2 cleave proinsulin to remove the connecting domain (C-peptide), liberating the mature form of insulin-containing A and B chains, linked by two inter-chain disulfide bridges (Figure 1) [9, 14]. Insulin is then stored in beta-cell granules as zinc-stabilized hexamers until needed for release (Figure 2) [9, 15]. The monomeric protein's final structure is composed of three alpha helices (residues A1–A8, A12–A18 and B9–B19) constrained by one intra- and two interchain disulfide bonds, one beta-strand (B24-B28), and less well-ordered segments [16]. The mature hormone contains 51 amino-acid residues [17, 18].



**Figure 1: Insulin Processing.** Preproinsulin at the top is cleaved and undergoes oxidative phosphorylation to produce proinsulin (middle). Proinsulin is then cleaved to produce mature insulin and the C-peptide that is released.

When blood-glucose levels are high, typically after a meal, insulin is released from a special class of  $\beta$ -cell storage granules into the bloodstream, where the hormone binds to its receptor (IR) on target cells to induce glucose uptake from the blood to the cells [19]. Insulin uptake primarily affects carbohydrate, lipid, and protein metabolism with major target cells primarily located in the liver, skeletal muscle, and adipose tissue. In the liver the hormone promotes glycogenesis, LDL formation, cholesterol and protein synthesis whereas insulin signaling inhibits glycogenolysis, gluconeogenesis, and fatty acid oxidation [20]. In the muscle it increases glucose transport across the membrane, glycolysis, glycogenesis, triglyceride uptake, and protein synthesis, while decreasing

fatty acid oxidation and protein degradation. In adipose tissue insulin signaling promotes glucose transport across the membrane, glycolysis, glycogenesis, lipolysis, triglyceride uptake and protein synthesis [21]. In other cells insulin stimulates the uptake and breakdown of glucose to produce energy. Ultimately, all these processes promote the anabolism of carbohydrates, lipids, and proteins, while reducing their catabolism.



**Figure 2: Insulin Hexamer.** The A chain is in warmpink and B chain in greencyan (Pymol colors). Hexamer showing three dimers linked by zinc ions in the middle in grey color. PDB ID 1AI0.

### 1.3 Insulin Types

Because insulin therapy remains essential for the treatment of T1D and late-stage T2D, insulin analogs were sought to enhance therapeutic properties via the optimization of pharmacokinetics (PK). The use of analogs has increased over the past 25 years. In general, insulin analogs are genetically altered, or chemically modified forms of human insulin made to modify insulin's absorption properties or bioavailability. They are designed in a laboratory, using various techniques such as chemical synthesis or

recombinant DNA technology in bacteria and yeast [22]. Endogenous human insulin in  $\beta$ -cell granules self-assembles to form hexamers: six monomers linked by zinc ions and extensive self-association surfaces to form hexamers [18]. When released into the bloodstream, the hexamers rapidly dissociate into zinc-free monomers [4]. Insulin binds as a monomer to the receptor on target cells to induce effects. It has been reported that the monomeric form of insulin exhibits markedly higher receptor-binding activity than the dimeric form under standard conditions [23]. When insulin is injected subcutaneously, hexamers must dissociate into monomeric or dimeric forms before absorption into the bloodstream as hexamers are too big to enter capillaries [4, 22]. This dissociation step imposes a delay in the onset of action in lowering blood-glucose levels for injected insulin, compared to the portal secretion of endogenous insulin [4]. There are many types of insulin analogs, which may be classified based on their PK properties and respective mechanisms of absorption and formulation concentration: rapid, ultra-rapid, intermediate, and long-acting insulin analogs. These analogs can be separated into two major groups: basal and prandial insulin analogs.

Basal insulin analogs are formulated to be very stable once injected in the subcutaneous space and slowly released into the bloodstream, therefore taking longer before being cleared by the body. This is achieved in one subclass by isoelectric precipitation and the addition of zinc, making it insoluble at tissue pH, thus promoting the formation of stable aggregates of zinc hexamers in the subcutaneous depot [24]. These analogs are meant to always maintain a continuous basal level of insulin in the bloodstream (i.e., a “peakless” PK profile). Basal insulin analogs include intermediate- and long-acting insulin analogs. A distinct subclass is acetylated to enable albumin

binding as exemplified by formulations of long-acting insulin analogs such as Levemir (insulin detemir, typically given twice a day); and Tresiba (insulin degludec); the latter can confer at steady-state, a near-peakless 24-hour profile with once-a-day dosing [25]. Other less common methods such as pegylation of insulin at Phe<sup>B1</sup> or Lys<sup>B29</sup>, not only result in a reduction of immunogenicity, allergenicity, and antigenicity but also confer a longer half-life in systemic circulation relative to HI [26]. Another less exploited method is addition of a fourth disulfide bond (A10-B4), which produced fibrillation-resistant analogs that bind to the IR with picomolar affinities [27]. Intermediate-acting insulin analogs, such as neutral protamine Hagedorn (NPH), only have a duration of action of 10 to 16 hours [28]. Because its glucose-lowering effect is highly variable, NPH is mostly used by T2D patients as a basal insulin analog, especially as a lower-cost treatment option [28].

Prandial insulin analogs include rapid-, ultra-rapid, and short-acting insulin analogs or second-generation analog formulations. The central therapeutic goal of these analogs is to avoid immediate post-prandial hyperglycemia and late post-prandial hypoglycemia [24]. Rapid-acting insulin analogs were formulated to form less stable insulin hexamers after subcutaneous injection and thus be rapidly absorbed into the bloodstream [4]. Insulin Lispro, Aspart, and Glulisine are the three main rapid-acting insulin analogs available for clinical use [4, 24]. Such insulin analogs mitigate upward post-prandial excursions of blood-glucose concentrations. Combinations of different insulin analogs also act to mimic physiological profiles of endogenous insulin, depending on the body's needs. There are also disadvantages to such analog products, including

cost, weight gain and potential hypoglycemia (even unconsciousness) when excessive doses are administered [29, 30].

#### **1.4 Prandial Insulin Therapy Challenges and Background on Previous Efforts**

Although both groups of insulin analogs present with challenges, this thesis focuses on prandial insulin analogs. Despite the design goals of prandial analogs, immediate post-prandial hyperglycemia and late hypoglycemia remain a concern. In addition, there are problems linked to thermodynamic, chemical, and physical stability: the three analogs in current use are less stable than human insulin. A key challenge is posed by fibrillation, the predominant mode of thermal degradation, which refers to the formation of amyloid-like fibrils in insulin formulation vials [31]. (This phenomenon is also responsible for the very rare syndrome of insulin-derived amyloidosis (IDA) in repeated subcutaneous injections and overall poor glycemic control [32-34].) Fibrillation occurs when enough monomers undergo conformational changes to nucleate and form beta-sheets that grow into mature fibrils. Fibrillation's main driving force is usually the shielding of hydrophobic surfaces exposed in a partially unfolded state; the formation of intermolecular  $\beta$ -sheet may further stabilize the fibrillar structure [35]. Insulin inherently tends to form unwanted amyloid-like fibrils in storage or in the reservoirs or tubings of delivery devices, such as in pumps before injection [31]. Fibrils not only impair proper dosage by clogging delivery devices, but also are immunogenic and may limit the concentration of feasible insulin formulations (also limited by native aggregation delaying absorption) [36]. Insulin fibrillation limits shelf-life above room temperature and imposes a requirement for a global cold chain, which is problematic for the developing world.

Insulin also has some amino-acid positions that make it prone to chemical and physical degradation, such as solvent-exposed nonpolar side chains. According to a previous study, chemical deterioration of insulin during storage of pharmaceutical preparations is mainly due to two categories of chemical reactions, hydrolysis and intermolecular transformation reactions [37]. The authors further stated that the most common form of hydrolysis reaction is asparagine deamidation, which happens in insulin at residue A21 in acidic solutions and at B3 in neutral or basic solutions [37].

To address these challenges, many efforts have been made to modify insulin over the past half-century. Pharmaceutical chemistry, biotechnology and delivery technologies have each been exploited to make insulin formulations that would better meet the needs of patients with diabetes [22]. One such example is by introducing deletions and/or substitutions in the amino-acid sequence of the protein: by doing so, the properties of the insulin analog are changed. Some modifications lead to an increase in stability or reduction of insulin degradation, either chemical or physical. For example, a previous study of insulin Glulisine demonstrated how EB29 (Glutamic acid at position 29 of the B chain) reduces hexamer formation while allowing self-assembly into protective dimers, thereby promoting rapid absorption in the blood [38].

### **1.5 Current Work on Prandial Insulin Design and Rationale**

Since it has been first shown that the size of insulin self-assembled oligomers plays a critical role in insulin absorption, monomeric insulin design has been a priority [39]. For our purposes herein, six amino acid substitutions to insulin were made and studied; each intended to increase stability of insulin. Physical stability for our purposes is evaluated by fibrillation assays, while thermodynamic stability in insulin is being

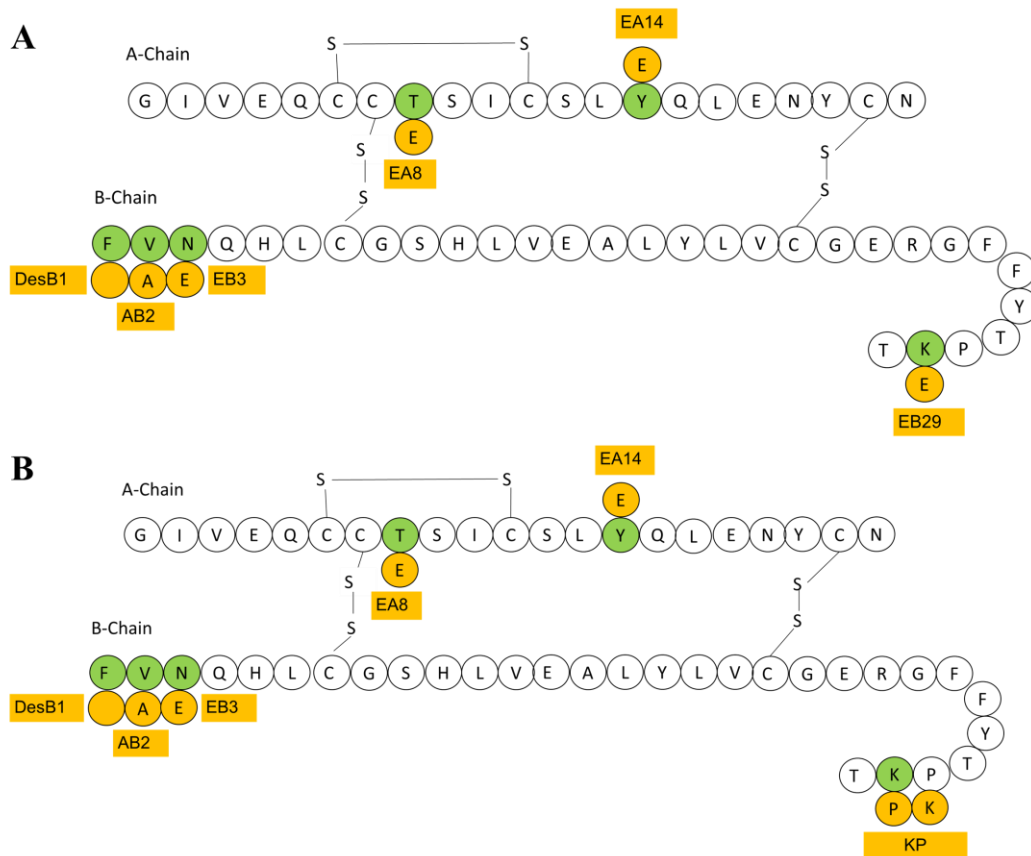
measured by CD experiments. Those modifications are DesB1, AB2, EB3, EA8, EA14, EB29 (PET), KB28 and PB29 (KP). A schematic outline of such design elements is shown in Figure 3, in relation to a model in Figure 4. Each of the above modifications was chosen for a specific reason and Table 1 gives a summary of the rationale behind each modification. DesB1, removing phenylalanine at position 1 of the B chain, was made to prevent fibrillation, based on a previous study where it was found that removal of B1 and B2 delayed fibril formation [40]. It was assumed that this delay arises from the destabilization of the proposed antiparallel  $\beta$ -sheet due to less favorable hydrophobic interactions [40]. As a highly flexible region [27], the constant free movement of phenylalanine at B1 of the N-terminus-end due to exposure of the non-polar ring could destabilize the molecule or mediate aberrant protein-protein interactions, possibly leading to fibrillation nucleation. In another study the lag time to visual fibrillation went from 7 hours in HI to 10 hours in DesB1-B2 insulin at 50°C and pH 1.6 [40]. With fibrillation assays, the longer the lag time, the better. This means that the lag time, defined as the time it takes for the protein to first form detectable fibrils, increased by 3 hours in assays of DesB1-B2 insulin [40].

Design Elements	Rationale for Modification
DesB1	Prevents Fibrillation
AB2	Increases helix propensity and decreases fibrillation
EB3	Increase solubility, prevent Asn deamidation (chemical degradation)
EA8	Increase solubility, increases stability by optimization of A1-A8 helical C-cap
EA14	Increases solubility and stability by replacing a solvent exposed Tyr
EB29	Known to stabilize insulin (insulin glulisine)
KB28 & PB29	Completely monomeric insulin

**Table 1: Rationale of Design Elements.** A summary of the rationale behind each modification is explained.

AB2, replacing valine at position 2 of the B chain with alanine, was intended to increase helix propensity and also decrease exposure of a nonpolar side chain, in turn delaying onset of fibrillation. Because valine has a  $\beta$ -branching side chain, it was hypothesized that it contributes to beta-sheet formation and therefore promotes fibrillation. Alanine has a shorter and less hydrophobic side chain, and has a higher helical propensity relative to valine, therefore promoting  $\alpha$ -helix formation, which in turn might reduce beta-sheet formation and fibrillation. EB3, replacing asparagine at position 3 of the B chain with glutamic acid, was made to increase solubility and prevent asparagine deamidation, a form of chemical degradation discussed above. Both asparagine and glutamic acid are hydrophilic. EA8 and EA14 were each hypothesized to increase insulin solubility and stability but by different mechanisms: (a) optimizing the A1-A8 helix cap with EA8 and (b) replacing a solvent-exposed nonpolar side chain

(tyrosine), thought to destabilize insulin, with EA14. EA8 replaces threonine at position 8 of the A chain with glutamic acid, both being hydrophilic amino acid residues. Threonine in this case is at the end of the A1-A8  $\alpha$ -helix, therefore making it a C-cap residue. Helix stability largely depends on helix propensity and helix capping interactions of residues involved [41]. Replacing the native threonine at A8 with glutamic acid, which has a higher  $\alpha$ -helical propensity and better preference to be a C-cap residue, might in principle increase insulin stability and solubility.



**Figure 3: Design Elements of Interest for A. PET and B. KP analogs.** Green circles contain native sequence amino acids found, while orange circles contain the substituted amino acid at that specific position, along with the designated name in the orange rectangles. One letter code of amino acids is used. PET analogs refer to all insulin analogs synthesized with the substitution EB29 and KP analogs refer to all insulin analogs synthesized with the substitutions KB28 and PB29.

EA14 designates the substitution of tyrosine at position 14 of the A chain with glutamic acid. It was found in 1992 that the A14 position tolerates a considerable structural diversity without affecting biological potency [42]. We now appreciate that this is because the A14 side chain does not directly contact the receptor [42, 43]. In the 1992 study same relative activities of EA14 insulin were investigated: although receptor binding was slightly decreased, lipogenesis was increased relative to unmodified bovine or porcine insulin [42]. Further, the native tyrosine at A14 has a large hydrophobic side chain that is hyper-exposed to the solvent, and therefore might incur a reverse-hydrophobic effect, where hydrophobic moieties are mobile surface elements instead of being sequestered within the folded protein core [44, 45]. Because the solvation of nonpolar residues destabilizes the protein, replacing the solvent-exposed tyrosine with a charged residue (like glutamic acid) has been shown to enhance overall stability by mitigating the reverse-hydrophobic effect [45].

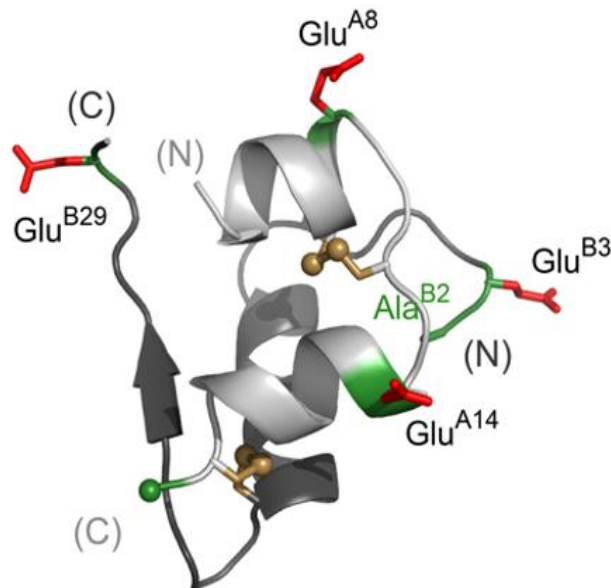
The last two modifications are in the B26-30 region of insulin, which is not particularly critical for receptor binding, but contributes to dimerization [43, 46-48]. This makes it a favorable target for prandial insulin analog design. Taking advantage of these properties, the current prandial analogs (insulin aspart, glulisine, and lispro: see above) contain substitutions in this segment. EB29 is a well-studied modification in insulin analog design, known to stabilize insulin and found in glulisine. In this analog, the native component of Apidra, the native lysine at position 29 of the B chain is replaced by glutamic acid resulting in a faster action profile due to the reduction of the extent of self-association with a rapid exchange between monomers, dimers and zinc-free hexamers [49]. Thus, EB29 reduces hexamer formation but allows self-assembly into short-lived

dimers, which promotes rapid absorption from the subcutaneous space into the blood [38]. This segment of the B chain may also contribute to fibrillation nucleation [43].

Insulin Lispro, a fast-acting insulin analog in clinical use, is identical to HI except for the transposition of proline and lysine at positions 28 and 29 in the C- terminus of the B chain [50]. This sequence change reduces dimerization by a factor of at least  $10^3$  and mimics insulin-like growth factors (IGF-I and II) which do not dimerize or hexamerize [43, 51]. Native insulin monomer interactions involve the C-terminal end of the B chain, leading to the formation of insulin dimers, termed dimerization. This in turn delays insulin absorption after subcutaneous injection since dimers' absorption is slower than monomers, leading to a slower insulin response for a prandial analog. IGF-I and HI share a sequence homology of almost 50% and in the absence of IGF-binding proteins, can elicit nearly the same hypoglycemic response. This modification eliminates a favorable ProB29-related dimer contact and also leads to a conformational change in the C-terminal segment of the B chain, which may sterically hinder dimer formation [49, 52]. Although phenol-stabilized zinc-insulin can nevertheless form, their rapid disassembly in the subcutaneous space in turn allows insulin monomers to enter the blood rapidly, thereby providing a faster onset of action to lower blood-glucose levels [50, 53].

It is important to note that the receptors IR and IGF-I receptor (IGF-IR) are structurally similar; thus each can bind the other's ligands but with different affinities and resulting functions [49]. On the one hand, IGF-IR mediates prenatal growth by the action of both IGF-I and IGF-II but only IGF-I acts on postnatal growth [54]. On the other hand, the IR mediates prenatal growth in response to IGF-II and postnatal metabolism upon binding to insulin [54]. Thus, IGF-II can bind both the insulin and IGF-I receptors, but

for different functions. While the insulin receptor primarily mediates metabolic signals, the IGF receptors are involved in growth and differentiation, which is why IGFs are more associated with mitogenic properties [55, 56]. Whether insulin is also mitogenic through its own receptor is controversial, but a study found that this can occur in the proper physiological context [56]. With that in mind, clinical insulin analogs were intended to either have similar or reduced mitogenicity levels as HI. Table 1 shows the synthesis design of all insulin-PET and -KP analogs with their final amino acid sequences.



**Figure 4: Structure of a General Insulin Analog.** Ala<sup>B2</sup> = AB2; Glu<sup>B3</sup> = EB3; Glu<sup>A8</sup> = EA8; Glu<sup>A14</sup> = EA14; and Glu<sup>B29</sup> = EB29 (PET). The structure shows three alpha-helices (50% of the insulin structure) and one beta-sheet (10%), along with random coils. The image was made by Yanwu Yang, Ph.D.

## CHAPTER 2

### MATERIALS AND METHODS

#### **2.1 Peptide Synthesis**

Analogs, shown in Table 2, were chemically synthesized and tested for stability by fibrillation and CD assays. The rationale was to add one substitution at a time, to ultimately have all the modifications added to the insulin analogs and study their combined effects. Because EB29 and KP are well-known modifications in the literature to make rapid-acting insulin analogs, all our analogs either contain EB29 (PET) or KP at the end of the B-chain. This allowed us to separate them into two groups, referred to as insulin-PET and insulin-KP analogs in this manuscript. PET analogs refer to all insulin analogs synthesized with the substitution EB29 and KP analogs refer to all insulin analogs synthesized with the substitutions KB28 and PB29.

The efficient synthesis of long peptides (> 35 residues) like insulin, requires optimized coupling, deprotection, and cleavage protocols, and the incorporation of sensitive monitoring techniques [57]. Because the chemical synthesis of insulin is challenging, all analogs were synthesized as DesDi precursors, which are versatile synthetic intermediates first described by DiMarchi [58]. The recent study used an optimized synthetic protocol to identify a 49 amino acid peptide named DesDi, with increased folding efficiency because of its optimized structure and potential to proteolytically be converted into a bioactive two-chain insulin [58]. Appendix A contains the peptide analogs with their identifying ions, both expected and observed as measured by LC-MS.

PET Analogs	Final Sequence	KP Analogs	Final Sequence
PET	FVNQHLCGSHLVEALYLVCGERGFFYTPET-GIVEQCCTSICSLYQLENYCN	KP	FVNQHLCGSHLVEALYLVCGERGFFYTKPT-GIVEQCCTSICSLYQLENYCN
DesB1-PET	VNQHLCGSHLVEALYLVCGERGFFYTPET-GIVEQCCTSICSLYQLENYCN	DesB1-KP	VNQHLCGSHLVEALYLVCGERGFFYTKPT-GIVEQCCTSICSLYQLENYCN
DesB1-AB2	ANQHLCGSHLVEALYLVCGERGFFYTPET-GIVEQCCTSICSLYQLENYCN	DesB1-AB2-KP	ANQHLCGSHLVEALYLVCGERGFFYTKPT-GIVEQCCTSICSLYQLENYCN
DesB1-EB3	VEQHLCGSHLVEALYLVCGERGFFYTPET-GIVEQCCTSICSLYQLENYCN	DesB1-EB3-KP	VEQHLCGSHLVEALYLVCGERGFFYTKPT-GIVEQCCTSICSLYQLENYCN
DesB1-AB2-EB3-PET	AEQHLCGSHLVEALYLVCGERGFFYTPET-GIVEQCCTSICSLYQLENYCN	DesB1-AB2-EB3-KP	AEQHLCGSHLVEALYLVCGERGFFYTKPT-GIVEQCCTSICSLYQLENYCN
DesB1-AB2-EB3-EA8-PET	AEQHLCGSHLVEALYLVCGERGFFYTPET-GIVEQCCE <sup>S</sup> ICSLYQLENYCN	DesB1-AB2-EB3-EA8-KP	AEQHLCGSHLVEALYLVCGERGFFYTKPT-GIVEQCCE <sup>S</sup> ICSLYQLENYCN
DesB1-AB2-EB3-EA14-PET	AEQHLCGSHLVEALYLVCGERGFFYTPET-GIVEQCCTSICSL <sup>E</sup> QLENYCN	DesB1-AB2-EB3-EA14-KP	AEQHLCGSHLVEALYLVCGERGFFYTKPT-GIVEQCCTSICSL <sup>E</sup> QLENYCN
DesB1-AB2-EB3-EA8-EA14-PET	AEQHLCGSHLVEALYLVCGERGFFYTPET-GIVEQCCE <sup>S</sup> ICSL <sup>E</sup> QLENYCN	DesB1-AB2-EB3-EA8-EA14-KP	AEQHLCGSHLVEALYLVCGERGFFYTKPT-GIVEQCCE <sup>S</sup> ICSL <sup>E</sup> QLENYCN

**Table 2: Design and Synthesis of Insulin Analogs.** PET analogs on the left and KP analogs on the right. One letter code of amino acids used for the final sequences of the analogs. The highlighted amino acids in red indicate the amino acid substitutions specific to the corresponding analog.

### 2.1.1 Amino Acid Loading

All peptide analogs used for experiments were chemically synthesized using the Tribute-UV/IR peptide synthesizer. In a 40 mL reaction vessel (Protein Technologies, Inc.), was loaded 0.1 mmol/g of ChemMatrix® resin H-ASN(TRT)-HMPB. Amino acids (Protein Technologies, Inc.) were weighed out in vials and loaded onto the instrument. Each vial or cartridge held 1 mmol of each amino acid. The resin and all amino acids, including their vials, were purchased from Protein Technologies. Solvents used were 20% Piperidine, 0.5 M Oxymapure solution (Protein Technologies, Inc.), 7.7% of N,N-Diisopropylcarbodiimide or DIC (Protein Technologies, Inc.), N,N-Dimethylformamide or DMF (Fisher), and Dichloromethane or DCM (Fisher). Piperidine was used to remove protective groups on amino acids like Fmoc (9-Fluorenylmethoxycarbonyl) while

Oxymapure and DIC were used as activators of reactive groups [59]. DMF was used as a polar protic solvent for amino acids to enhance solubility and peptide-resin swelling, and to improve coupling kinetics for difficult sequences such as insulin [57]. DCM's purpose was to improve the swelling of the resin, thus allowing improved diffusion of reagents to active sites [57, 59]. After peptide synthesis was completed, reaction vessels were removed, and contents were transferred into scintillation vials. Then, peptide cleavage was performed to remove peptides from the resin.

### ***2.1.2 Peptide Cleavage and Deprotection***

The goal of this step is to separate or cleave the peptides from resin beads while removing the protecting groups from the side chains [60]. Resins from the scintillation vials were mixed with a cleavage cocktail consisting of 90% TFA (CHEM-IMPEX INT'L INC.), and 2.5% of each of the 4 scavengers: Milli-Q Water, Anisol, 2,2-(Ethylene dioxy)diethanethiol (DODT), and Triisopropyl silane (TIPS) (SIGMA-ALDRICH). TFA is the main component of this cocktail since it cleaves the peptides from the resin and removes the protecting groups of the side chains. The scavengers were used to prevent any modification or destruction of amino acids by potentially reactive side chains. The choice of scavengers was mainly dependent upon the amino acid sequence of the peptide to be cleaved [59]. A ratio of 20 mL of cleavage cocktail per gram of resin was used. The reaction was allowed to stir for 2-3 hours in the scintillation vials, before being filtered in a 50 mL Falcon tube to remove the resin. Cold Diethyl Ether (Fisher) was added and mixed to the resulting supernatant to wash and precipitate the peptides and extract the scavengers [59]. The tubes were then centrifuged at 3000 rpm for 5 minutes. Ether was then removed, leaving the precipitated pellet at the bottom of the

tube. This washing process was repeated two other times for a total of three washes with cold Diethyl Ether, breaking out the pellet each time and mixing the tube contents thoroughly before centrifugation. On the last wash, the supernatant was removed, and pellets were placed into lyophilizing bottles that were connected to a vacuum hose for 24 hours to dry the peptides on the Labconco 700401000 FreeZone Lyophilizer. After 24 hours, the crude peptides were ready for folding.

### ***2.1.3 Peptide Folding***

Insulin is a folded protein, therefore making this step necessary in insulin analog synthesis. This is where the insulin chains fold by oxidation to create two interchain disulfide bridges (A7-B7 and A20-B19) holding the two chains together, and one intrachain disulfide bond in the A chain (A6-A11) [27]. Crude peptides were dissolved to make 0.1 mM Desdi peptide solutions using 2 mM Cysteine, 2 mM Cystine, 20 mM Glycine, and Milli-Q water. Cysteine was needed to oxidize the cysteine residues in insulin to help the folding reaction, while cystine, the oxidized form of cysteine, was used to accelerate the reaction. Glycine was chosen for its good buffering capacity. Cysteine, cystine, and glycine were all purchased from SIGMA-ALDRICH. The pH was adjusted to 10.5 using a 10N Sodium Hydroxide (NaOH) to dissolve the peptide and facilitate the folding reaction. The reaction was left stirring at room temperature for 24 hours and then purified using HPLC.

### ***2.1.4 HPLC Purification and Analytical LC-MS Identification***

High-performance liquid chromatography (HPLC) has been firmly established as the premier technique for the analysis and purification of a wide range of molecules, including peptides and proteins [61]. HPLC was used to purify the peptides, while liquid

chromatography-mass spectrometry (LC-MS) was used to analyze and identify them. The peptide solutions were filtered after acidification to pH 2-3 using a 5N HCl acid solution. PROTO 300 C4 or C8 columns (Higgins Analytical, Inc.) were used for the purifications. The column was equilibrated for 10 minutes at 95% solvent A (0.1% TFA in water) and solvent B (0.1% TFA in Acetonitrile) at a flow rate of 20 mL/min. The peptide solutions were then loaded at 100% solvent A, at a rate of 20 mL/min and peptides were purified on the Waters 2489 UV/Visible Detector instrument. Analytical LC-MS was then performed on the Agilent 1100 and LCQ Advantage Thermo Finnigan instruments, to confirm sample identification, according to the existing protocol [62]. Small vials were prepared with 30 mL of water and 30 mL of the purified peptide sample and loaded onto the LC-MS instrument. Masses were obtained by online electrospray mass spectrometry and mass ions specific to proteins were calculated and identified on each peptide to confirm identification [62]. The resulting Desdi peptides were then trypsinized.

### ***2.1.5 Trypsinization***

This reaction allows for the cleavage of the insulin chains by trypsin, which recognizes and cleaves after arginine and lysine residues. This was necessary to remove the short peptide, AEAFK, that was added to the insulin chain for synthesis convenience, because phenylalanine at position 1 of the B chain (Phe<sup>B1</sup>) was thought to help the insulin chains fold better by providing a flexible N-terminal arm [63]. This technique allows us to still have phenylalanine through insulin folding, but then easily remove it afterward to make the DesB1 analogs. The trypsinization reaction also acts on the Desdi to remove the hexapeptide GFFYTK that is normally part of the B-chain of insulin, but because it resides between arginine and lysine residues which are recognized by trypsin, this portion

of the B-chain gets cleaved. This leads to the formation of a desoctapeptide (DOI) from the Desdi, which is essentially a Desdi with eight residues removed. The solutions used for trypsinization were 8M urea and 0.1M ammonium bicarbonate ( $\text{NH}_4\text{HCO}_3$ ). Urea (Fisher) was used to open the single chain peptide making it more accessible for trypsin to cleave at Lys and Arg residues, while bicarbonate was used as a buffer to adjust the pH. For each 100 mg of a lyophilized insulin analog, 35 mL of 0.1 ammonium bicarbonate and 5 mL of 8M urea were used for a final solution of 1M urea. Then, 20% of w/w TPCK-treated trypsin (Thermo Scientific) was added to remove both groups: AEAFK and the hexapeptide, GFFYTK. Solutions were allowed to gently rock for 24-48 hours before being purified again by HPLC and resulting DOIs were identified by LC-MS using the protocol previously described. After successful DOI confirmation, semi-synthesis of the peptide analogs followed.

#### ***2.1.6 Trypsin-mediated Semi-Synthesis***

In the previous section, the hexapeptide GFFYTK was removed by the trypsinization reaction, but since it is part of the normal B-chain of insulin, it was necessary to reattach it. Therefore, semi-synthesis or Trypsin-mediated ligation was used to bring back the octapeptide to finish the chemical synthesis of insulin analogs. Equivalent weights of DOIs and octapeptides containing either KP or PET modifications were obtained by chemical synthesis and combined with 20% w/w Trypsin. GFFYTK**KPT** and GFFYTP**PET** were the sequences of the KP and PET octapeptides, respectively. All those were dissolved in a synthesis buffer composed of 35% N,N-dimethylacetamide (NDMA), 35% 1,4-butanediol, and 30% 0.2M Tris Acetate (SIGMA-ALDRICH). The pH of the solution was adjusted using 4-methylmorpholine (SIGMA-ALDRICH) and

then incubated in a 10-15 degrees Celsius circulating water bath (model T-2 of Brinkmann MGW Lauda RM3) for 48 hours. After 48 hours, samples were then mixed with 2 mL of 6M guanidine hydrochloride (GuHCl) containing 0.1% TFA and HPLC purification and LC-MS identification followed as previously described, but this time using the HPLC instrument Waters 2481. Three peaks should be obtained on the spectrum, corresponding to the octapeptide, DOI, and fusion product. Appendix A summarizes the expected and actual ion fragments for each semi-synthesis peptide obtained. Peptide solutions were then frozen and lyophilized on the Labconco 700401000 FreeZone. This completed the whole process of peptide synthesis, and the next step was running fibrillation, CD wavelength scan and CD guanidine titration assays.

## **2.2 Fibrillation Assay**

Peptide solutions were prepared at 60  $\mu$ M each, in a phosphate-buffered saline (PBS) solution at 37°C and pH 7.4, containing 0.02% NaN<sub>3</sub>. PBS was purchased from Invitrogen. Absorbances were measured using the Nanodrop ND-1000 spectrophotometer and concentrations were determined using Lambert-Beer law. Additional calculations were made to determine samples and buffer volumes to prepare the final solutions of 60  $\mu$ M. Next, 2  $\mu$ M of the dye indicator Thioflavin T or ThT (SIGMA) was prepared by dissolving 1 mg of powdered ThT in 1 ml of PBS buffer. ThT binds to mature fibrils, and with excitation and emission wavelengths at 450 and 484 nm, respectively, fluorescence measurement translates to the lag time. In this assay, 2  $\mu$ L of ThT was added for each 250  $\mu$ L of sample. From each peptide solution, 250  $\mu$ L of sample was transferred to each well of a 96-well assay plate purchased from COSTAR. The plate was then covered with a film plastic to properly seal the wells before loading it onto the fibrillation assay

instrument SYNERGY H1 Microplate Reader from BIOTEK. The assay was run until all sample wells fibrillated, and in this case, a minimum of one week. The results were then processed and analyzed using Excel and GraphPad Prism.

### **2.3 CD Guanidine Titration Assays**

Each lyophilized protein was dissolved in Milli-Q water, 100  $\mu\text{g}$  for wavelength scan assays and 350  $\mu\text{g}$  for guanidine titration assays. Concentrations were measured and calculations were made to make sure there was enough protein for the assay. Two buffers were used. The first one was prepared using 50 mM KCl and 10 mM  $\text{H}_2\text{PO}_4$  and adjusting their pH to 7.4 with KOH. The other buffer contained the same salt concentrations as the previous one, but also had 8 M GuHCl. For each sample, 8 mL of 8M GuHCl buffer was used, so calculations were made accordingly to determine the amount of guanidine buffer needed for all samples. With the CD guanidine titration assay, guanidine was used as a denaturant to destabilize the structures of the analogs used. The lyophilized samples were dissolved in 1.2 mL of the CD buffer without guanidine to make approximately 50  $\mu\text{M}$  of stock solutions of peptides. Of those solutions, 800  $\mu\text{L}$  were transferred to 15 mL Falcon tubes to be frozen and lyophilized. The remaining 400  $\mu\text{L}$  were left in microtubes in a  $-4^\circ\text{C}$  fridge, until all samples were ready and, in this case, overnight. The next day refrigerated samples and buffers were taken out 30 minutes before the experiment. The lyophilized samples in 15 mL tubes were resuspended in 8 mL GuHCl buffer each, using a 10 mL serological pipette. The CD cuvette was loaded with 1620  $\mu\text{L}$  of CD buffer and 180  $\mu\text{L}$  of the stock solution, for a total volume of 1.8 mL at 5  $\mu\text{M}$ . Samples from the microtubes and 15 mL Falcon tubes were loaded onto the CD instrument, J-1500 CD Spectrometer from JASCO. Assays were

run at pH 7.4 for all the samples, using the respective buffers. Data were analyzed using Interval Data Analysis and Spectra Analysis, while Excel and GraphPad Prism were used to generate results and graphs. Replicates were measured but +/- errors in result tables come from fitting of the data to a two-state unfolding model.

#### **2.4 CD Wavelength Scan Assay-Phosphorylation (CHO-IRB)**

The 100 ug lyophilized protein samples were dissolved in CD buffer and exact concentrations were determined using the spectrophotometer GENESYS 150 from Thermoscientific. From there, 200  $\mu$ L of peptide solutions at 25  $\mu$ M were made and loaded on the CD instrument, J-1500 CD Spectrometer from JASCO. Assays were run at pH 7.4 for all samples, using the respective CD buffers. Data were analyzed using the Interval Data Analysis, Spectra Analysis, and results generated using Excel and GraphPad.

#### **2.5 In vitro Insulin Receptor Binding Assay**

This is an insulin phosphorylation assay for CHO-IR type B. The signaling assay protocol used in a paper published by Chen, Y.C., *et al*, was followed for the first two sets of analogs shown in Table 2, but a few changes were made for the last two sets of analogs shown in Table 5 [64]. CHO-IRB cells were used instead of HepG2 cells, cultured in an F-12 Glutamax media from Fisher. Serial analog dilutions were prepared in Dulbecco's PBS, with calcium and magnesium (900 nM with 3-fold dilutions). The blocking buffer added to fix the cells was 250  $\mu$ L instead of 100  $\mu$ L. The primary antibody was prepared at 10  $\mu$ L anti-pTyr 4G10 into 10 mL blocking buffer while the second antibody was prepared at 2  $\mu$ L of DRAQ, 100  $\mu$ L anti-mouse IgG-R-

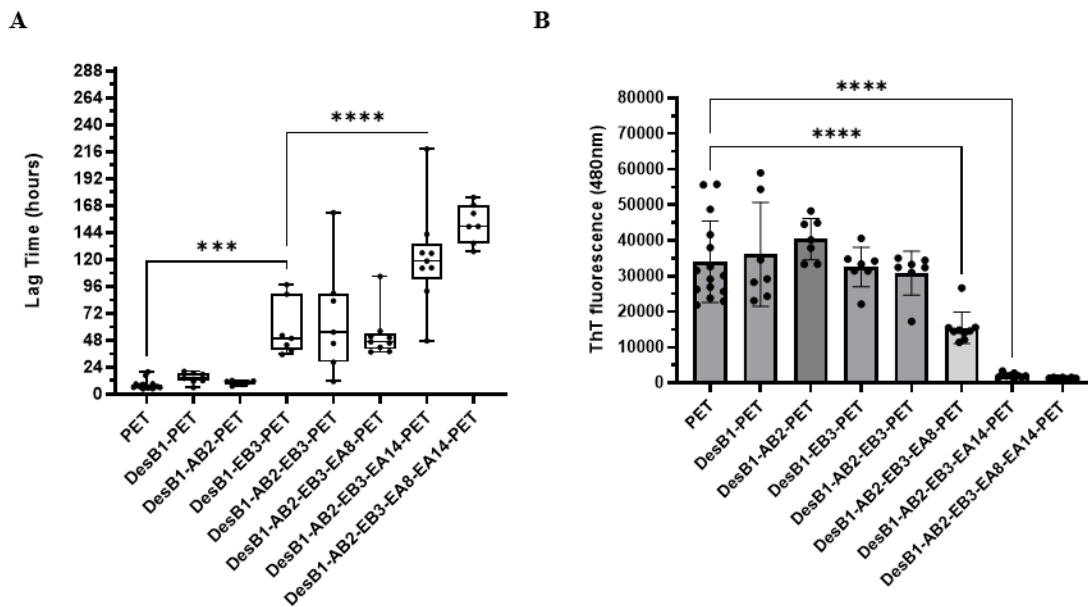
Phycoerythrin antibody into 10 mL blocking buffer. The fluorescence signal was detected at 700 nm and 578 nm using Biotek Neo2 multimode reader.

## CHAPTER 3

### RESULTS

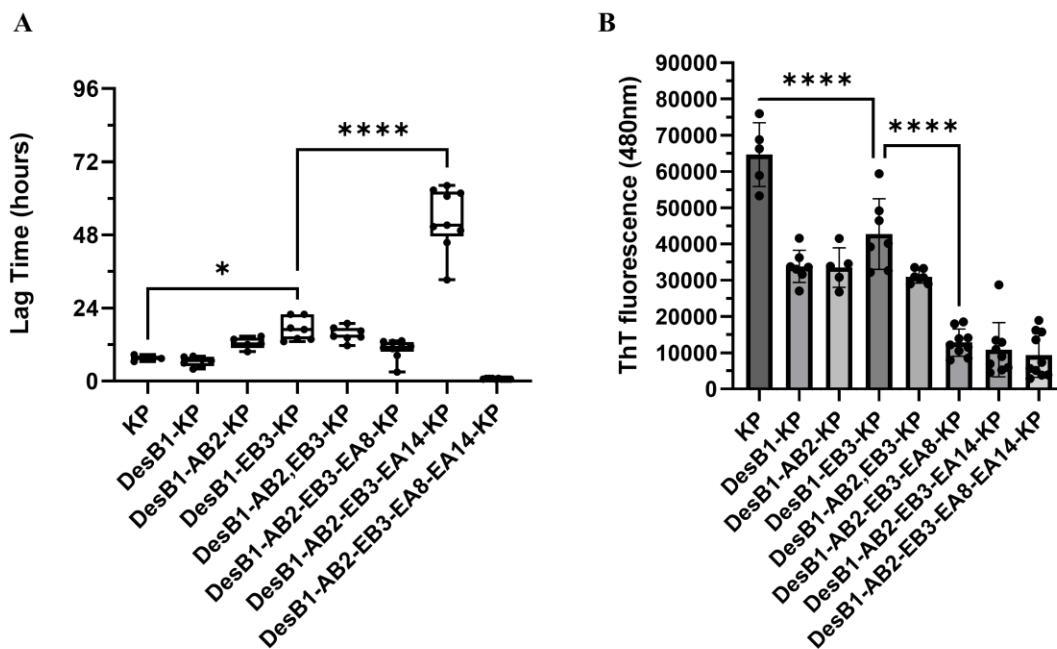
#### 3.1 Initial Fibrillation Assay Results for Insulin-PET and KP Analogs

Analogs in Table 2 were all tested for fibrillation, and results are shown in Figures 5 & 6 and Table 3. The control PET-insulin formed fibrils within 24 hours of the assay, along with analogs DesB1-PET and DesB1-AB2-PET (Figure 5A). Although DesB1-PET fibrillated within 24 hours, it noticeably delayed fibrillation by almost doubling the lag time, from 8.3 in PET-insulin to 14.2 hours in DesB1-PET insulin (Table 3A). The lag time started to increase with DesB1-EB3-PET and DesB1-AB2-EB3-PET, which also formed fibrils within 58 - 72 hours (Table 3A). Interestingly, adding EA8 did not delay fibrillation, but EA14 significantly increased the lag time from 52.4 to 121.6 hours (Table 3A). When EA8 and EA14 were combined in the same analog, the lag time increased to 152.1 hours (Table 3A).



**Figure 5: Fibrillation of “PET” Analogs (37°C, 60  $\mu$ M protein, PBS [pH 7.4].** A. Lag time measure in hours for each PET-analog with B. the highest amount of fibrils formed measured by ThT fluorescence at 480 nm. PET-insulin is the control and the other ones are novel analogs. Significance is used to compare PET-insulin vs DesB1-EB3-PET and DesB1-AB2-EB3-PET vs DesB1-AB2-EB3-EA14-PET, with cutoff value  $p=0.05$ . GraphPad Prism was used to normalize and integrate the data to generate graphs.

Figure **5B** shows the highest amount of fluorescence detected in each well throughout the entire duration of the assay, which then correlates to the maximum amount of fibrils. Based on the trend seen on the figure, fluorescence kept decreasing as more substitutions were added onto the insulin analogs, suggesting that the total amount of fibrils was decreasing. Fluorescence went from about 35,000 in PET-insulin to less than 5,000 in the analog containing all the modifications (Figure **5B**). This suggests that EA14 was not only significantly delaying fibrillation but was also reducing the amount of fibrils formed.



**Figure 6: Fibrillation “KP” Analogs (37°C, 60 μM protein, PBS [pH 7.4].** A. Lag time measure in hours for each KP-analog with B. the highest amount of fibrils formed measured by ThT fluorescence at 480 nm. KP-insulin is the control and the others are novel analogs. Significance is used to compare KP-insulin vs DesB1-EB3-KP and DesB1-AB2-EB3-KP vs DesB1-AB2-EB3-EA14-KP, with cutoff value p=0.05. GraphPad Prism was used to normalize and integrate the data to generate graphs.

The same protocol described above was used for the fibrillation assay of the insulin-KP analogs, and the results are summarized in Figure 6 and Table 3B. KP-insulin was a control with a lag time of 7.6 hours, but unlike what is seen in PET analogs, DesB1-KP had about the same lag time as the control, while DesB1-AB2-KP almost doubled that time at 12.5 hours (Table 3B). EB3 and EA8 did not seem to affect the lag time, until EA14 was added, which increased the lag time to 53.3 hours. The last analog that had all the modifications, unexpectedly, had the worst lag time at 0.8 hours. Figure 6B showing the highest amount of fluorescence, suggests a similar trend as seen with PET analogs, but the amount of fluorescence went from 90,000 in KP-insulin to 10,000

in DesB1-AB2-EB3-EA8-EA14-KP. All in one, only EA14 significantly delayed fibrillation but results with KP analogs markedly differ with PET analogs.

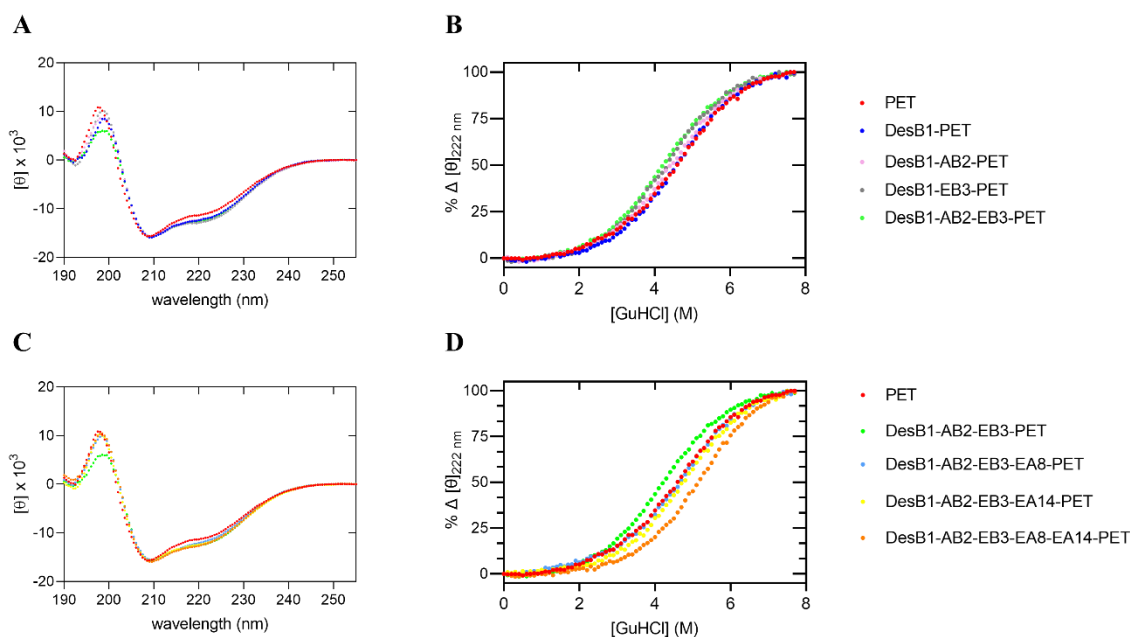
A		B	
Analog	Lag time $\pm$ SD (hours)	Analog	Lag time $\pm$ SD (hours)
PET	8.3 $\pm$ 4.2	KP	7.6 $\pm$ 0.8
DesB1-PET	14.2 $\pm$ 5.0	DesB1-KP	6.5 $\pm$ 1.6
DesB1-AB2-PET	10.2 $\pm$ 1.9	DesB1-AB2-KP	12.5 $\pm$ 1.9
DesB1-EB3-PET	58.0 $\pm$ 24.9	DesB1-EB3-KP	17.0 $\pm$ 3.8
DesB1-AB2-EB3-PET	67.8 $\pm$ 49.9	DesB1-AB2-EB3-KP	15.5 $\pm$ 2.4
DesB1-AB2-EB3-EA8-PET	52.4 $\pm$ 20.65	DesB1-AB2-EB3-EA8-KP	10.5 $\pm$ 3.2
DesB1-AB2-EB3-EA14-PET	121.6 $\pm$ 45.3	DesB1-AB2-EB3-EA14-KP	53.3 $\pm$ 10.1
DesB1-AB2-EB3-EA8-EA14-PET	152.1 $\pm$ 17.5	DesB1-AB2-EB3-EA8-EA14-KP	0.8 $\pm$ 0.1

**Table 3: Fibrillation Lag Times** with A. Insulin PET-analogs on the left and corresponding lag times in hours. B. insulin KP-analogs and corresponding lag times in hours, on the right. The red-highlighted rows show the modifications that increased the lag time.

### 3.2 Initial CD Assay Results for Insulin-PET and KP Analogs

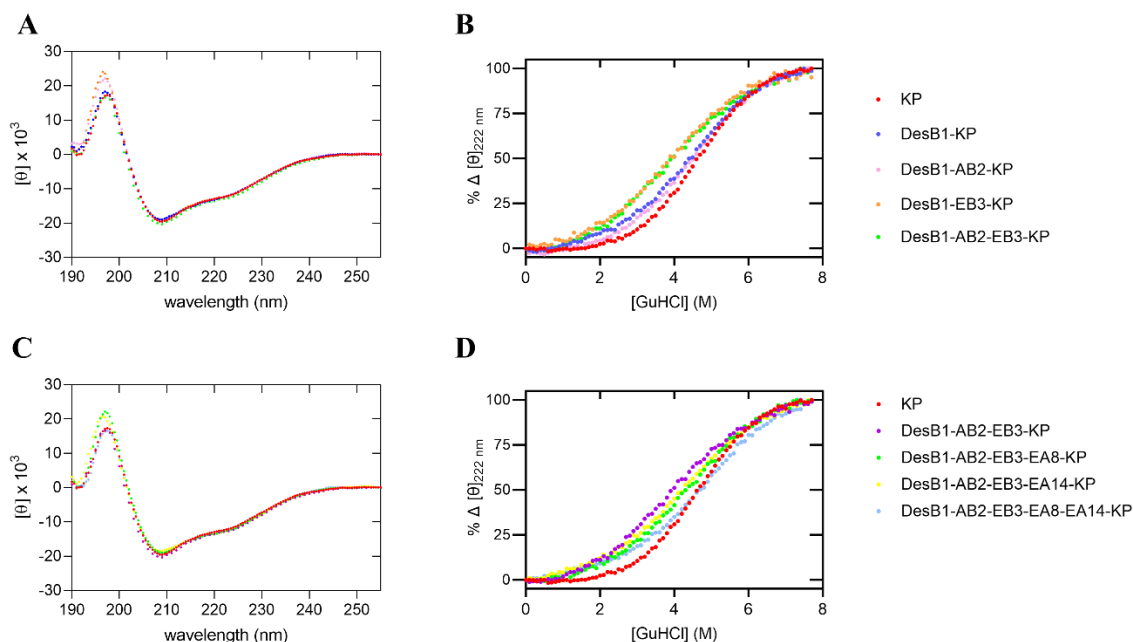
The CD wavelength and guanidine titration assays results are displayed in Figures 7 & 8 and the corresponding Table 4 shows each analog with their  $\Delta G_u$ . The CD wavelength data reveal that all PET and KP analogs show little to no major structural changes, especially at 222 nm (Figures 7A & C and 8A & C). PET-insulin was the control with a 3.0 Kcal/mol  $\Delta G_u$  (Table 4A). When curves are shifted to the right, it suggests that the analog is more stable because it is resisting guanidine denaturation, as opposed to a shift of the curve to the left. When the modifications DesB1, AB2, and EB3 were added to the PET analogs, the curves were shifted to the left, suggesting that they are destabilizing modifications (Figure 7B). Starting with DesB1-AB2-EB3-EA8-PET,

the  $\Delta G_u$  significantly increased to 3.5 kcal/mol and then decreased to 3.3 kcal/mol with DesB1-AB2-EB3-EA14-PET (Table 4A). The last analog, which bears all the modifications had the highest  $\Delta G_u$  at 3.7 kcal/mol (Table 4A), suggesting that it is the most stable of all the PET analogs (Figure 7D).



**Figure 7: CD Studies for Insulin-PET Analogs.** A & C. CD guanidine wavelength (left) and B & D. titration (right) results for the PET analogs with a legend on the far right to design the corresponding color-coded graphs. GraphPad Prism was used to normalize and integrate the data to generate graphs.

With KP analogs, KP-insulin was the control with a  $\Delta G_u$  of 3.0 kcal/mol, just like PET-insulin (Table 4B). DesB1-KP and DesB1-AB2-KP lowered the values to 2.8 and 2.6 kcal/mol, respectively, but adding EB3 gave the lowest  $\Delta G_u$  value at 1.7 kcal/mol (Table 4B) and so its CD curve was the most shifted to the left (Figure 8B). In Figure 8D, analogs containing EA8 and EA14 separately, had CD curves that were a slightly shifted to the right compared to the previous ones, but the analog combining both EA8 and EA14 had the curve that was the most shifted to the right, with a  $\Delta G_u$  of 3.1 kcal/mol (Table 4B).



**Figure 8: CD Studies for Insulin-KP Analogs.** A & C. CD guanine wavelength (left) and B & D. titration (right) results for the KP analogs with a legend on the far right to design the corresponding color-coded graphs. KP-insulin is the control and the other ones are novel analogs. GraphPad Prism was used to normalize and integrate the data to generate graphs.

A		B	
Analog	$\Delta G_U$ (Kcal/mol)	Analog	$\Delta G_U$ (Kcal/mol)
PET	$3.0 \pm 0.08$	KP	$3.0 \pm 0.06$
DesB1-PET	$3.1 \pm 0.08$	DesB1-KP	$2.8 \pm 0.11$
DesB1-AB2-PET	$2.9 \pm 0.09$	DesB1-AB2-KP	$2.6 \pm 0.09$
DesB1-EB3-PET	$2.8 \pm 0.09$	DesB1-EB3-KP	$2.0 \pm 0.19$
DesB1-AB2-EB3-PET	$2.7 \pm 0.07$	DesB1-AB2-EB3-KP	$1.7 \pm 0.17$
DesB1-AB2-EB3-EA8-PET	$3.5 \pm 0.10$	DesB1-AB2-EB3-EA8-KP	$2.1 \pm 0.13$
DesB1-AB2-EB3-EA14-PET	$3.3 \pm 0.10$	DesB1-AB2-EB3-EA14-KP	$2.6 \pm 0.12$
DesB1-AB2-EB3-EA8-EA14-PET	$3.7 \pm 0.07$	DesB1-AB2-EB3-EA8-EA14-KP	$3.1 \pm 0.12$

**Table 4: CD-based Guanidine Titrations ( $\Delta G_u$  in kcal/mol.)** A. Insulin-PET analogs with  $\Delta G_u$  in kcal/mol. B. Insulin-KP analogs with  $\Delta G_u$  in kcal/mol. PET- and KP-insulin are the controls and the other ones are novel analogs. The last row of each table is highlighted in red, showing the highest values of  $\Delta G_u$ .

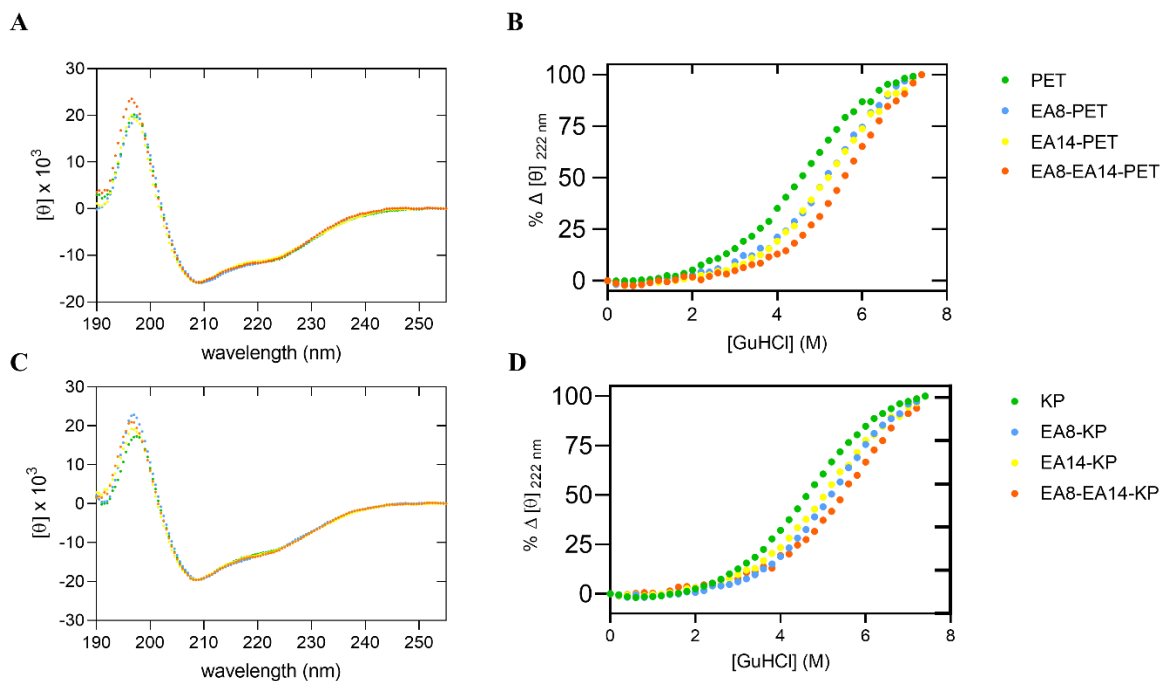
### 3.3 Final CD and Fibrillation Assays Results for Selected Analogs

All results combined suggest that the substitutions that provided stability the most were EB3, EA8, and EA14 by delaying fibrillation, and EA8 and EA14 by providing physical stability with CD guanidine assays. This is the reason why EA8 and EA14 became the focus and highlight of further investigations by the same assays. Analogs containing only EA8 and EA14 were chemically synthesized in both PET and KP forms to be tested via fibrillation and CD assays (Table 5). CD wavelength results for both PET and KP analogs show no structural variation for all analogs (Figure 9A & C).

Analog	Final Sequence
PET	FVNQHLCGSHLVEALYLVCGERGFFYTPET-GIVEQCCTSICSLYQLENYCN
EA8-PET	FVNQHLCGSHLVEALYLVCGERGFFYTPET-GIVEQCCESICSLYQLENYCN
EA14-PET	FVNQHLCGSHLVEALYLVCGERGFFYTPET-GIVEQCCTSICSLYQLENYCN
EA8-EA14-PET	FVNQHLCGSHLVEALYLVCGERGFFYTPET-GIVEQCCESICSLYQLENYCN
KP	FVNQHLCGSHLVEALYLVCGERGFFYTKPT-GIVEQCCESICSLYQLENYCN
EA8-KP	FVNQHLCGSHLVEALYLVCGERGFFYTKPT-GIVEQCCESICSLYQLENYCN
EA14-KP	FVNQHLCGSHLVEALYLVCGERGFFYTKPT-GIVEQCCTSICSLYQLENYCN
EA8-EA14-KP	FVNQHLCGSHLVEALYLVCGERGFFYTKPT-GIVEQCCESICSLYQLENYCN

**Table 5: Insulin-PET and KP Analogs: EA8 and EA14.** The highlighted amino acids are the ones that are replacing the native amino acids of insulin.

Comparing PET analogs,  $\Delta G_u$  increased from 3.0 in the control PET-insulin to 3.9 Kcal/mol in EA8-PET and 3.6 Kcal/mol in EA14-PET (Table 6A). As expected though,  $\Delta G_u$  of the last analog EA8-EA14-PET was the highest of all, at 4.3 kcal/mol (Table 6A), with the curve that is the most-shifted to the right (Figure 9B). A similar trend is also seen with KP analogs, with EA8-KP being more stable than EA14-KP, but EA8-EA14-KP was the most stable of all KP-insulin analogs, with a  $\Delta G_u$  of 4.1 kcal/mol (Table 6B). Accordingly, the CD guanidine titration graph also shows that the curve for EA8-EA14-KP was the most shifted to the right, and thus the most stable (Figure 9D).



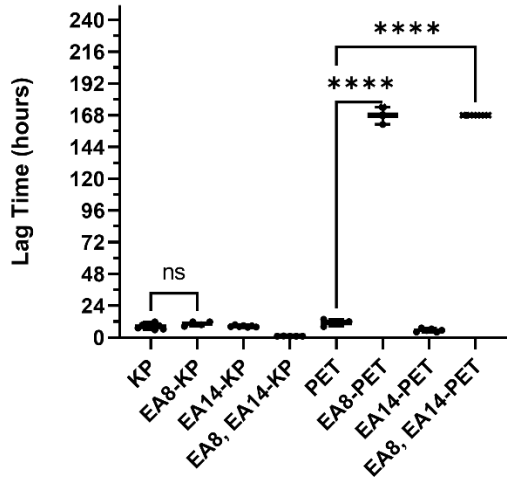
**Figure 9:** CD Studies for Insulin-PET and KP analogs, focusing on EA8 & EA14. A & C. CD guanidine wavelength (left) B & D. CD titration (right) results for the KP analogs with a legend on the far right to design the color-coded graphs. GraphPad Prism was used to normalize and integrate the data to generate graphs.

<b>A</b>		<b>B</b>	
<b>Analog</b>	<b><math>\Delta G_U</math> (Kcal/mol)</b>	<b>Analog</b>	<b><math>\Delta G_U</math> (Kcal/mol)</b>
PET	$3.0 \pm 0.08$	KP	$2.7 \pm 0.06$
EA8-PET	$3.9 \pm 0.15$	EA8-KP	$3.6 \pm 0.10$
EA14-PET	$3.6 \pm 0.16$	EA14-KP	$3.2 \pm 0.13$
EA8-EA14-PET	$4.3 \pm 0.28$	EA8-EA14-KP	$4.1 \pm 0.18$

**Table 6: CD Guanidine Titrations ( $\Delta G_U$ ) of Selected Analogs.** A. Insulin-PET analogs. B. Insulin-KP analogs. The last row of each table shows that combining both EA8 and EA14 in the same analog gives the highest  $\Delta G_U$ .

The fibrillation assay was repeated twice and both results are shown in Figure 10. The same trend is seen in both assays, where only EA8 delayed fibrillation. All insulin-KP analogs fibrillated within 10 hours, with lag times that vary between 1 and 10 hours in Figure 10B and 8 and 44 hours in Figure 10D. The insulin-PET analogs showed an increase in the lag time, especially with EA8-PET and EA8-EA14-PET, that did not fibrillate after 168 hours (7 days) for both (Figures 10A & B). In Figures 10E & 10F, the highest amount of fluorescence suggests that both EA8 and EA14 are reducing the amount of fibrils formed, but the effect was more pronounced with EA8 than EA14.

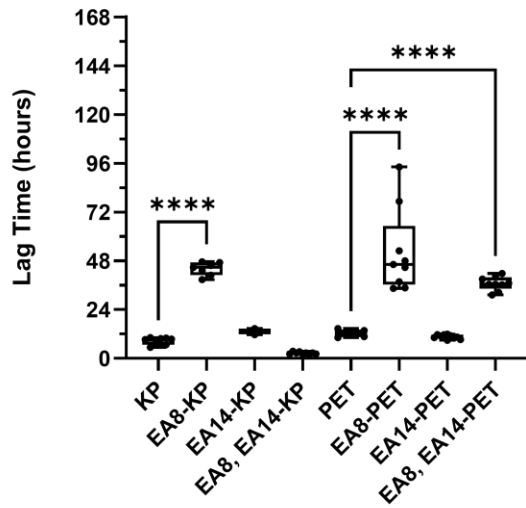
A



B

Analog	Lag time $\pm$ SD (hours)
KP	8.3 $\pm$ 1.9
EA8-KP	10.4 $\pm$ 1.6
EA14-KP	8.5 $\pm$ 0.7
EA8-EA14-KP	1.04 $\pm$ 0.1
PET	11.5 $\pm$ 2.3
EA8-PET	167.7 $\pm$ 6.6
EA14-PET	5.5 $\pm$ 1.3
EA8-EA14-PET	> 168.0

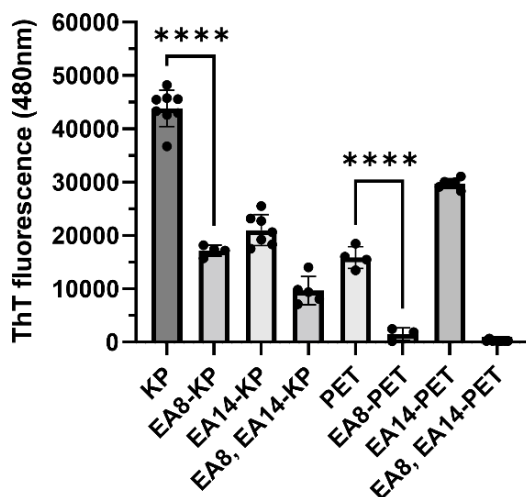
C



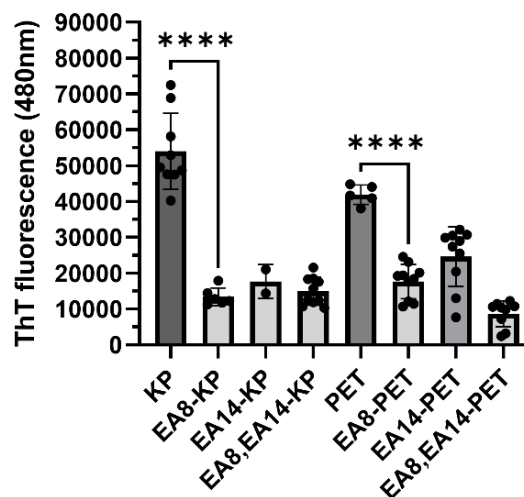
D

Analog	Lag time $\pm$ SD (hours)
KP	8.3 $\pm$ 1.7
EA8-KP	44.0 $\pm$ 3.3
EA14-KP	13.2 $\pm$ 2.1
EA8-EA14-KP	2.06 $\pm$ 0.5
PET	12.3 $\pm$ 1.9
EA8-PET	52.3 $\pm$ 20.4
EA14-PET	10.7 $\pm$ 1.1
EA8-EA14-PET	36.8 $\pm$ 3.4

E



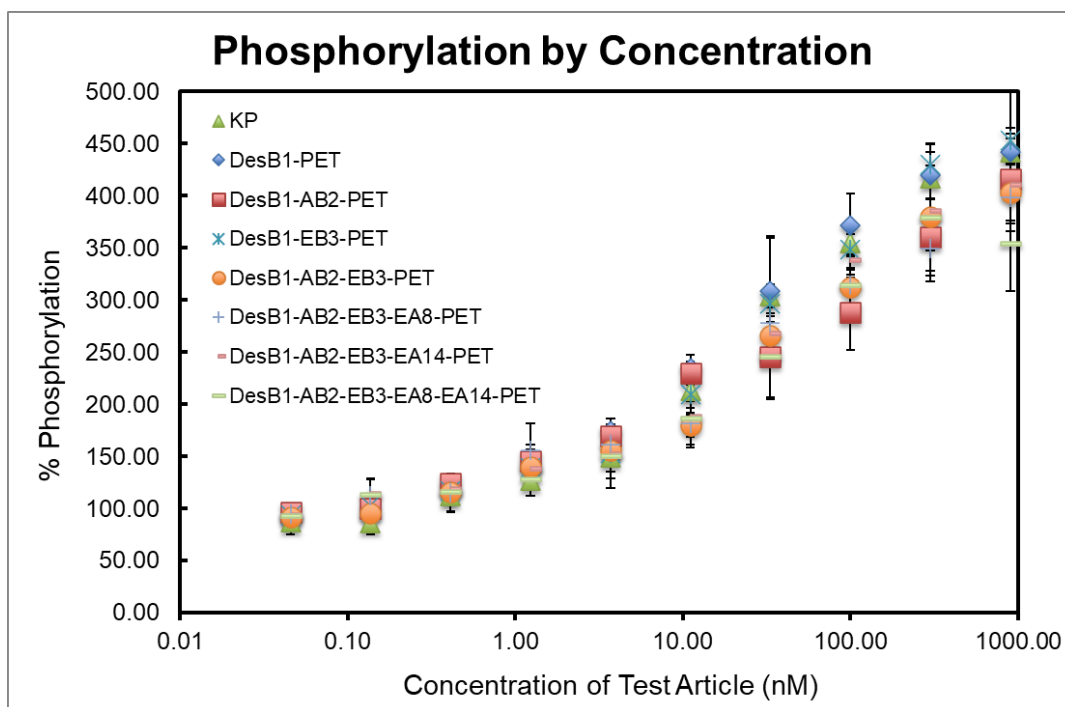
F



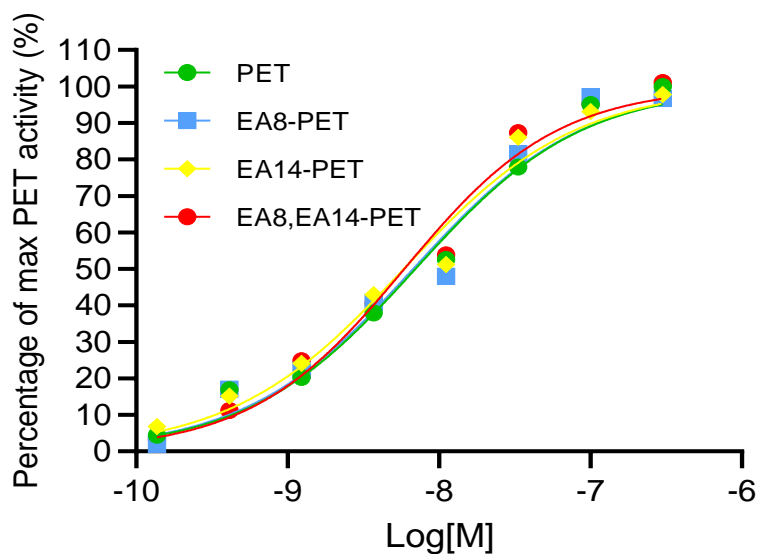
**Figure 10: Fibrillation Properties of EA8 and EA14.** Fibrillation assay was repeated twice and both results are shown, with the graphs showing **lag Time** in A & C and **quantification of lag time** in B & D for selected analogs. E & F show the highest amount of fluorescence attained during the assay. The highlighted rows show the analogs with the highest lag time and significance is shown with asterisks with  $p < 0.05$ .

### 3.4 Phosphorylation Assay Results

Phosphorylation assay results shown in Figure 11 for the PET-analogs were performed by Yen-Shan Chen, PhD. The graph is plotting the percent phosphorylation as a function of the concentration of each analog. Percent phosphorylation increased as the concentration of each analog was increasing. All the analogs are clustered around the same point on the graph, suggesting that they all have the similar activity levels. Figure 12 also shows the same phosphorylation assay results for the selected analogs containing only EA8 and EA14.



**Figure 11: Biological Activities of Insulin Analogs.** The percent phosphorylation of the insulin receptor is plotted as a function of increasing concentrations (nM) of different analogs. All points are clustered around the same point at each concentration, suggesting that all analogs have similar biological activity. Assay and graph were prepared by Yen-Shen Chen, Ph.D., n = 1.



**Figure 12: Biological Activities of Selected Insulin Analogs.** All points are clustered around the same point at each concentration, suggesting that all analogs have similar biological activity relative to PET-insulin, which was the control. n = 1.

## CHAPTER 4

### DISCUSSIONS AND CONCLUSIONS

#### 4.1 DesB1 and AB2 Effects

Insulin oligomerization starts with monomer-monomer interactions to form dimers mostly via the residues present in the C-terminal end of the B-chain (B23-29), but also B8, B9, B12, B16, B21, and A21 [65]. In a process called hexamerization, dimers would interact to form hexamers via residues A13-14, A17, B1, B2, B4, B10, B13, B14, and B17-19 [65]. Because the constant movement of B1 around the insulin monomer could lead to fibrillation nucleation in addition to its role as a surface contact in hexamerization, its importance in fibrillation is well defined [65]. Therefore, removing it in DesB1-analogs was thought to delay fibrillation. Based on the fibrillation assay results, it is noticeable that DesB1 delayed fibril formation by almost doubling the lag time relative to PET-insulin. Nielsen *et al* found that the removal of B1 and B2 resulted in a longer lag time compared to HI, indicating that B1 and B2 were somewhat important for the stabilization of the fibril nucleus [66]. This is understandable because both B1 and B2 are at the dimer interface of insulin, and so their removal is expected to delay fibrillation by destabilizing the fibril nuclei [65]. Based on these results, combining both DesB1 with DesB2 instead of AB2 could have had more effect in delaying fibrillation. Given that DesB1-AB2-PET had an even lower lag time compared to DesB1-PET, one would argue that AB2 did not fulfill its purpose of decreasing fibrillation by increasing helical propensity. This can be seen on the CD assays with the reduced  $\Delta G_u$  of both DesB1-AB2-PET and DesB1-AB2-KP analogs (Table 4). Eventually, this should be further

investigated with structural biology means, such as MD simulations, to analyze how AB2 affects the A1-A8 helix.

#### **4.2 EB3 for Potential Clinical Use**

The EB3 substitution was primarily done to prevent chemical degradation, but results show that it contributed to delaying fibrillation in the insulin analogs, only when combined with EB29 (PET). However, this effect is lost with insulin-KP analogs (Table **3B**). This was certainly not the purpose of this substitution, but the outcomes are both surprising and promising. This could be due to the acidic residue glutamic acid, being both polar and charged, that adds to the overall negative charge of the individual insulin molecules. This would then translate to repulsion between negatively charged molecules and lead to a decrease in monomer association. Four out of the six substitutions made involve glutamic acid, so each insulin monomer analog has a net negative charge of four, and monomers may be compelled to repulse each other, decreasing chances of fibril formation. This might explain why fibrillation was significantly delayed when substitutions involving glutamic acid residues were added to the analogs, especially in the last analog, DesB1-AB2-EB3-EA8-EA14 where they were all present (Figure 5).

#### **4.3 EA8 and EA14 Potential Clinical Use**

In Table **3A**, there is an interesting effect with EA8 and EA14. Taken separately, EA8 does not seem to delay fibrillation, but combining it with EA14 shows that they are working together to increase the lag time. Therefore, the last analog containing all the modifications had the highest lag time (Table **3A**). Because both EA8 and EA14 are glutamic acid residues, the charge repulsion phenomenon explained above for EB3 might be at play to slow down fibril formation, therefore showing a longer lag time when

combined in the same analog. It is important to note that this trend is only seen with insulin-PET analogs, and not insulin-KP analogs. A14 has been shown to be involved in dimer-dimer interactions, so changing the native residue to glutamic acid would prevent those interactions to impair seed nucleation, therefore reducing fibrillation [65]. A previous paper that evaluated fibrillation of different insulin mutants where they only had one or two mutations at a time, showed that all mutants fibrillated faster than human insulin [66]. In this present thesis, the results suggest that combining many specific modifications rather than individual ones might show a greater effect.

The exception to that is seen with insulin-KP analogs in Table **3B**, where the last analog, shockingly had the lowest lag time of all. Unlike insulin-PET analogs, EB3 and EA8 did not really delay fibrillation in insulin-KP analogs; only EA14 did. This phenomenon is not only seen in the first sets of insulin-KP analogs bearing all the modifications but is also seen in the second set of analogs only bearing EA8 and EA14. The reasons behind this are unknown, but further investigations would determine the causes.

In the last set of fibrillation results shown in Figure **10**, some unexpected results are seen with EA8 and EA14 of insulin-PET analogs. EA8, which previously didn't seem to delay fibrillation in combination with the other modifications (DesB1-AB2-EB3-EA8-PET) with a lag time of 52 hours, significantly delayed fibrillation when by itself with a lag time of 167.7 hours (EA8-PET). Overall, both EA8-PET and EA8-EA14-PET did not fibrillate after 168 hours, so one can easily conclude that only EA8 is really delaying fibrillation in these analogs (Figure **10**). However, the analog containing only EA14 (EA14-PET) did not delay fibrillation at all with a lag time of 5 hours, relative to the

corresponding analog bearing the other modifications (DesB1-AB2-EB3-EA14-PET) that had the highest lag time of 152 hours. The initial set of analogs show that EA14 only delayed fibrillation when combined with the other modifications, but not by itself. Since EB3 delayed fibrillation in the initial set of analogs, combining EB3 with EA14 could be more promising. These results are very unexpected, but they suggest that focusing on analogs bearing only EA8 and EA14 might just be a tradeoff with the first sets of analogs bearing all modifications. The choice seems to come down to either EA8-insulin analogs or EA14-insulin analogs that possibly carry other modifications such as EB3, to effectively delay fibrillation. This needs to be further investigated to uncover the mechanisms involved in either case.

#### **4.4 Biological Potency**

Based on the phosphorylation assay in Figures **11** & **12**, all analogs show a similar level of activity in CHO-IRB cells, which was expected. This is because the goal of the modifications was to increase stability, while still retaining potency, thus in principle, none of the substitutions made to the analogs involved residues that interact with the insulin receptor. According to a previous study, residues A1, A5, A19, A21, B10, B16, B23, B24, and B25 interact with the IR on target cells [65, 67].

#### **4.5 Fibrillation and CD Conceptual Correlation**

On one hand, fibrillation assays give a measure of lag time to show how long a particular protein takes to form detectable fibrils. A higher lag time translates to a delay in fibrillation, thus, a more stable protein. On the other hand, the most widespread use of CD spectroscopy is to identify a protein's secondary structure content. In theory, CD assays measure  $\Delta G_u$  to predict stability, so the higher the value, the more stable the

protein is. By combining these two concepts, one can conclude that a higher value of  $\Delta G_u$  should correlate with less fibrillation, but these two are not always correlated. The reason being the fact that fibrillation measures physical stability, while CD measures thermodynamic stability of a protein. This is particularly seen with EB3, which delayed fibrillation but was shown to be a destabilizing modification by the CD assays. This explains the tradeoff that is usually involved in amino acid substitutions to most proteins like insulin, because doing so may confer an advantage while losing another with the native residue. This often complicates protein mutations, but ultimately, may lead to discovery of novel analogs.

The spectra of human insulin with and without zinc exhibit two minima at 208 and 222 nm, characteristic of the  $\alpha$ -helical conformation [66]. Almost all CD wavelength data show little variation between the curves at 222 nm, indicating that the insulin secondary structure is not dramatically changed by the modifications done to the analogs, which correlates with previous data [66].

#### **4.6 Maximum Fluorescence Correlation with Amount of Fibrils**

In the fibrillation assay, we are correlating the maximum fluorescence levels to the amount of fibrils formed during the assay because ThT binds to mature fibrils. Thus, a higher fluorescence would mean that more fibrils formed in the sample while a lower fluorescence would mean that less fibrils are formed during the assay. Many reasons could also explain a decrease in fluorescence levels, such as sample degradation, which can be physical or chemical, insulin chain separation, and different types of fibril formation although not yet described for insulin. To explore some of these possibilities, samples were recovered and purified using UPLC and LC-MS to evaluate possible

sample degradation. Results indicate that additional peaks were seen on the UV spectra of some analogs that could indicate possible sample degradation.

#### **4.7 Dimer Complication in Fibrillation and CD experiments**

In previous sections, it was mentioned that insulin-KP is completely monomeric, while insulin-PET is mostly monomeric with formation of short-lived dimers. This could be reflected in the fibrillation and CD experiments in different ways. Because fibril formation requires at least partial unfolding of the insulin monomer, dimers would need to be dissociated into monomers before starting the fibrillation process. This would impose a delay in the onset of fibrillation, seen with the insulin-PET analogs due to the presence of dimers. In contrast, the insulin-KP analogs that were completely monomeric were more prone to form fibrils because of the absence of the delay present in the PET analogs. This could explain why insulin-KP analogs fibrillated faster than PET, and why the data collected from the insulin-KP analogs had less variability in the fibrillation assay than that of the insulin-PET analogs.

In the CD wavelength scan assay, it could be speculated that hexamers would have a deeper minimum at 222 nm due to their higher helical content relative to the dimers and monomers. Therefore, monomers would have the weakest signal at 222 nm, which could explain the less pronounced minimum on the wavelength scan data. Because of the low concentration of peptide solutions used in the CD guanidine titration assay (5  $\mu\text{M}$ ), insulin analogs were expected to be mostly monomeric in solution. If dimers were present, they would most likely be more resistant to guanidine denaturation compared to monomers. Limitations of the assay do not give information on the monomer/dimer contents of the samples, but other assays such as size exclusion chromatography or

DOSY could be helpful in analyzing the diffusion coefficient of the oligomeric state of the insulin analogs.

#### **4.8 Future Directions**

From the discovery of insulin 100 years ago, the pharmaceutical industry has been developing new ways to make insulin analogs that would mimic the physiological action profile of insulin by introducing both basal and prandial insulin analogs. Recombinant DNA technology remains one of the most powerful techniques to discover new ways to improve upon existing insulin analogs, getting us closer to finding a cure for DM. This is even more important for T1D patients that must rely on insulin therapy. EB3, EA8, and EA14 were modifications that were done to increase solubility. Because they are all glutamic acid residues, it is expected to show promising results, but solubility assays were not yet performed. It is important to test solubility because there is an urgent need to produce more concentrated formulations to not only reduce the amount of insulin injected, but also the number of injections needed by patients, all while preserving the rapid onset of action of insulin. This will also revolutionize the prospect of miniaturizing portable insulin pumps with hybrid-closed-loop or – ultimately – closed-loop systems, as higher concentrations of insulin between U-200 and U-1000, are needed for those devices [22]. Insulin formulations are very limited in concentration because of hexamer formation, which would impair the action profile of rapid-acting insulin analogs, so having solubility data can bring us one step closer to that goal.

In this study, EB3, EA8, and EA14 really proved to solve many problems of insulin storage such as physical (fibrillation) and thermodynamic stability. Investigating other aspects of those three substitutions is fundamental in providing a complete picture

of the modifications that would be most beneficial for a potential prandial insulin analog. Chemical degradation and solubility assays were not performed in this work, so it would be interesting to see how those modifications would affect insulin degradation and solubility. These could then be incorporated into new formulations of prandial insulin analogs to facilitate current prandial insulin therapy and eventually improve patients' quality of life.

## APPENDICES

### Appendix A

#### LCMS Data for Synthesized Analogs

Analog	Molecular Weight (g/mol)	Expected Ions (m/z)		Actual Ions (m/z)	
		+3	+4	+3	+4
<b>PET</b>	5808.61	1937.2033	1453.1525	1936.67	1453.07
<b>DesB1-PET</b>	5661.43	1888.1433	1416.3575	1887.6	1416.4
<b>DesB1-AB2-PET</b>	5633.38	1878.7933	1409.345	1878.27	1409.2
<b>DesB1-EB3-PET</b>	5676.45	1893.15	1420.1125	1892.6	1420.07
<b>DesB1-AB2-EB3-PET</b>	5648.39	1883.7967	1413.0975	1883.4	1413
<b>DesB1-AB2-EB3-EA8-PET</b>	5676.40	1893.1333	1420.1	1892.73	1420.13
<b>DesB1-AB2-EB3-EA14-PET</b>	5614.33	1872.13	1404.53	1872.13	1404.53
<b>DesB1-AB2-EB3-EA8-EA14-PET</b>	5614.33	1881.78	1411.585	1881.4	1411.2
<b>KP</b>	5807.67	1936.89	1452.9175	1936.33	1452.60
<b>DesB1-KP</b>	5660.49	1887.83	1416.1225	1887.2	1416
<b>DesB1-AB2-KP</b>	5632.44	1878.48	1409.11	1878	1409.07
<b>DesB1-EB3-KP</b>	5675.50	1892.8333	1419.875	1892.67	1419.6
<b>DesB1-AB2-EB3-KP</b>	5647.45	1883.4833	1412.8625	1882.87	1412.8
<b>DesB1-AB2-EB3-EA8-KP</b>	5675.46	1892.82	1419.865	1892.12	1419.87
<b>DesB1-AB2-EB3-EA14-KP</b>	5613.39	1872.13	1404.3475	1871.6	1404.13
<b>DesB1-AB2-EB3-EA8-EA14-KP</b>	5641.40	1881.4667	1411.35	1880.87	1411.07
<b>EA8-PET</b>	5836.62	1946.54	1460.155	1946.07	1483.47
<b>EA14-PET</b>	5774.55	1925.85	1444.6375	1925.67	1444.67
<b>EA8-EA14-PET</b>	5802.56	1935.1867	1451.64	1935.07	1451.27
<b>EA8-KP</b>	5835.32	1946.1067	1459.83	1946.07	1459.8
<b>EA14-KP</b>	5773.25	1925.4167	1444.3125	1925.33	1444.07
<b>EA8-EA14-KP</b>	5801.26	1934.7533	1451.315	1934.4	1450.93

## **Appendix B**

### **After Fibrillation Concentrations**

<b>Analogs</b>	<b>Absorbance (280 nm)</b>	<b>Concentrations (uM)</b>
<b>PET</b>	0.071	11.21
<b>DesB1-PET</b>	0.143	22.57
<b>DesB1-AB2-PET</b>	0.107	16.89
<b>DesB1-EB3-PET</b>	0.271	42.78
<b>DesB1-AB2-EB3-PET</b>	0.155	24.48
<b>DesB1-AB2-EB3-EA8-PET</b>	0.340	53.67
<b>DesB1-AB2-EB3-EA14-PET</b>	0.426	87.93
<b>DesB1-AB2-EB3-EA8-EA14-PET</b>	0.427	88.13
<b>KP</b>	0.048	7.58
<b>DesB1-KP</b>	0.075	11.84
<b>DesB1-AB2-KP</b>	0.093	14.68
<b>DesB1-EB3-KP</b>	0.205	32.36
<b>DesB1-AB2-EB3-KP</b>	0.108	17.05
<b>DesB1-AB2-EB3-EA8-KP</b>	0.073	11.52
<b>DesB1-AB2-EB3-EA14-KP</b>	0.122	25.18
<b>DesB1-AB2-EB3-EA8-EA14-KP</b>	0.133	27.45
<b>EA8-PET (1)<sup>4</sup></b>	0.528	83.35
<b>EA8-PET (2)<sup>5</sup></b>	0.066	10.42
<b>EA14-PET (1)</b>	0.082	16.92
<b>EA14-PET (2)</b>	0.102	21.05
<b>EA8-EA14-PET (1)</b>	0.394	81.32
<b>EA8-EA14-PET (2)</b>	0.205	42.31
<b>EA8-KP (1)</b>	0.076	11.997
<b>EA8-KP (2)</b>	0.065	10.26
<b>EA14-KP (1)</b>	0.073	15.07
<b>EA14-KP (2)</b>	0.071	14.65
<b>EA8-EA14-KP (1)</b>	0.081	16.72
<b>EA8-EA14-KP (2)</b>	0.069	14.24

---

<sup>4</sup> First set of co-dialysis samples

<sup>5</sup> Second set of co-dialysis samples

## BIBLIOGRAPHY

1. Federation, I., *IDF Diabetes Atlas, tenth*. 2021, International Diabetes.
2. Alam, U., et al., *General aspects of diabetes mellitus*. Handbook of clinical neurology, 2014. **126**: p. 211-222.
3. Whiting, D.R., et al., *IDF diabetes atlas: global estimates of the prevalence of diabetes for 2011 and 2030*. Diabetes research and clinical practice, 2011. **94**(3): p. 311-321.
4. Mathieu, C., P. Gillard, and K. Benhalima, *Insulin analogues in type 1 diabetes mellitus: getting better all the time*. Nature Reviews Endocrinology, 2017. **13**(7): p. 385-399.
5. Organization, W.H. *Diabetes*. 2021 [cited 2021; Available from: <https://www.who.int/news-room/fact-sheets/detail/diabetes>].
6. Clinic, C. *Diabetes: An Overview*. 2021; Available from: <https://my.clevelandclinic.org/health/diseases/7104-diabetes-mellitus-an-overview>.
7. Weiss, M., D.F. Steiner, and L.H. Philipson, *Insulin Biosynthesis, Secretion, Structure, and Structure-Activity Relationships*, in *Endotext*, K.R. Feingold, et al., Editors. 2000: South Dartmouth (MA).
8. Banting, F.G. and C.H. Best, *The internal secretion of the pancreas*. Indian Journal of Medical Research, 2007. **125**(3): p. L251.
9. Weiss, M.A., *Proinsulin and the Genetics of Diabetes Mellitus\**. Journal of Biological Chemistry, 2009. **284**(29): p. 19159-19163.

10. Bell, G.I., et al., *Sequence of the human insulin gene*. Nature, 1980. **284**(5751): p. 26-32.
11. Chan, S.J., P. Keim, and D.F. Steiner, *Cell-free synthesis of rat preproinsulins: characterization and partial amino acid sequence determination*. Proceedings of the National Academy of Sciences, 1976. **73**(6): p. 1964-1968.
12. Rhodes, C.J., *Processing of the insulin molecule*. Diabetes Mellitus: a fundamental and clinical text, 2000: p. 27-50.
13. Tsuchiya, Y., et al., *IRE1–XBPI pathway regulates oxidative proinsulin folding in pancreatic  $\beta$  cells*. Journal of Cell Biology, 2018. **217**(4): p. 1287-1301.
14. Hutton, J.C., *Insulin secretory granule biogenesis and the proinsulin-processing endopeptidases*. Diabetologia, 1994. **37**: p. S48-S56.
15. Guest, P.C., et al., *Insulin secretory granule biogenesis. Co-ordinate regulation of the biosynthesis of the majority of constituent proteins*. Biochemical Journal, 1991. **274**(1): p. 73-78.
16. Adams, M.J., et al., *Structure of rhombohedral 2 zinc insulin crystals*. Nature, 1969. **224**(5218): p. 491-495.
17. Menting, J.G., et al., *How insulin engages its primary binding site on the insulin receptor*. Nature, 2013. **493**(7431): p. 241-245.
18. Liu, M., et al., *Biosynthesis, structure, and folding of the insulin precursor protein*. Diabetes, Obesity and Metabolism, 2018. **20**: p. 28-50.
19. Cuatrecasas, P., *Membrane receptors*. Annual review of biochemistry, 1974. **43**(1): p. 169-214.

20. Newsholme, E.A. and G. Dimitriadis, *Integration of biochemical and physiologic effects of insulin on glucose metabolism*. *Exp Clin Endocrinol Diabetes*, 2001. **109 Suppl 2**: p. S122-34.
21. Thakrar, N., *Insulin*. TeachMePhysiology, 2022.
22. Kurtzhals, P., et al., *Commemorating insulin's centennial: engineering insulin pharmacology towards physiology*. *Trends in Pharmacological Sciences*, 2021. **42**(8): p. 620-639.
23. Fujita-Yamaguchi, Y. and J.T. Harmon, *A monomer-dimer model explains the results of radiation inactivation: binding characteristics of insulin receptor purified from human placenta*. *Biochemistry*, 1988. **27**(9): p. 3252-3260.
24. Jarosinski, M.A., et al., *Structural principles of insulin formulation and analog design: A century of innovation*. *Mol Metab*, 2021. **52**: p. 101325.
25. Porcellati, F., et al., *Pharmacokinetics, Pharmacodynamics, and Modulation of Hepatic Glucose Production With Insulin Glargine U300 and Glargine U100 at Steady State With Individualized Clinical Doses in Type 1 Diabetes*. *Diabetes Care*, 2019. **42**(1): p. 85-92.
26. Hinds, K.D. and S.W. Kim, *Effects of PEG conjugation on insulin properties*. *Advanced Drug Delivery Reviews*, 2002. **54**(4): p. 505-530.
27. Vinther, T.N., et al., *Additional disulfide bonds in insulin: Prediction, recombinant expression, receptor binding affinity, and stability*. *Protein Science*, 2015. **24**(5): p. 779-788.

28. Kramer, C.K., R. Retnakaran, and B. Zinman, *Insulin and insulin analogs as antidiabetic therapy: A perspective from clinical trials*. *Cell Metabolism*, 2021. **33**(4): p. 740-747.
29. Grunberger, G., *The need for better insulin therapy*. *Diabetes, Obesity and Metabolism*, 2013. **15**(s1): p. 1-5.
30. Handelsman, Y., et al., *American Association of Clinical Endocrinologists and American College of Endocrinology—clinical practice guidelines for developing a diabetes mellitus comprehensive care plan—2015—executive summary*. *Endocrine Practice*, 2015. **21**(4): p. 413-437.
31. Yang, Y., et al., *An Achilles' heel in an amyloidogenic protein and its repair: insulin fibrillation and therapeutic design*. *Journal of Biological Chemistry*, 2010. **285**(14): p. 10806-10821.
32. Ohno, Y., et al., *Investigation of factors that cause insulin precipitation and/or amyloid formation in insulin formulations*. *Journal of Pharmaceutical Health Care and Sciences*, 2019. **5**(1): p. 22.
33. Nagase, T., et al., *The insulin ball*. *The Lancet*, 2009. **373**(9658): p. 184.
34. Sie, M., et al., *Human recombinant insulin and amyloidosis: An unexpected association*. *The Netherlands journal of medicine*, 2010. **68**: p. 138-40.
35. Brange, J., et al., *Toward Understanding Insulin Fibrillation*. *Journal of Pharmaceutical Sciences*, 1997. **86**(5): p. 517-525.
36. Loughheed, W.D., et al., *Insulin aggregation in artificial delivery systems*. *Diabetologia*, 1980. **19**(1): p. 1-9.

37. Brange, J. and L. Langkjoer, *Insulin structure and stability*. Pharm Biotechnol, 1993. **5**: p. 315-50.
38. Yamada, S., *Insulin glulisine in the management of diabetes*. Diabetes, metabolic syndrome and obesity : targets and therapy, 2009. **2**: p. 111-115.
39. Brange, J., et al., *Monomeric insulins obtained by protein engineering and their medical implications*. Nature, 1988. **333**(6174): p. 679-682.
40. Garriques, L.N., et al., *The effect of mutations on the structure of insulin fibrils studied by Fourier transform infrared (FTIR) spectroscopy and electron microscopy*. Journal of pharmaceutical sciences, 2002. **91**(12): p. 2473-2480.
41. Chakrabartty, A., T. Kortemme, and R.L. Baldwin, *Helix propensities of the amino acids measured in alanine-based peptides without helix-stabilizing side-chain interactions*. Protein Science, 1994. **3**(5): p. 843-852.
42. Chu, Y.C., et al., *The A14 position of insulin tolerates considerable structural alterations with modest effects on the biological behavior of the hormone*. J Protein Chem, 1992. **11**(5): p. 571-7.
43. Shoelson, S.E., et al., *Mutations at the dimer, hexamer, and receptor-binding surfaces of insulin independently affect insulin-insulin and insulin-receptor interactions*. Biochemistry, 1992. **31**(6): p. 1757-1767.
44. Camilloni, C., et al., *Towards a structural biology of the hydrophobic effect in protein folding*. Scientific reports, 2016. **6**(1): p. 1-9.
45. Pakula, A.A. and R.T. Sauer, *Reverse hydrophobic effects relieved by amino-acid substitutions at a protein surface*. Nature, 1990. **344**(6264): p. 363-364.

46. Slieker, L., et al., *Modifications in the B10 and B26–30 regions of the B chain of human insulin alter affinity for the human IGF-I receptor more than for the insulin receptor*. *Diabetologia*, 1997. **40**(2): p. S54-S61.
47. Bi, R.C., et al., *Insulin's structure as a modified and monomeric molecule*. *Biopolymers: Original Research on Biomolecules*, 1984. **23**(3): p. 391-395.
48. Baker, E.N., et al., *The structure of 2Zn pig insulin crystals at 1.5 Å resolution*. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 1988. **319**(1195): p. 369-456.
49. Vigneri, R., S. Squatrito, and L. Sciacca, *Insulin and its analogs: actions via insulin and IGF receptors*. *Acta diabetologica*, 2010. **47**(4): p. 271-278.
50. Wilde, M.I. and D. McTavish, *Insulin Lispro*. *Drugs*, 1997. **54**(4): p. 597-614.
51. De Meyts, P., *The structural basis of insulin and insulin-like growth factor-I receptor binding and negative co-operativity, and its relevance to mitogenic versus metabolic signalling*. *Diabetologia*, 1994. **37**(Suppl 2): p. S135-S148.
52. Holleman, F. and J.B. Hoekstra, *Insulin lispro*. *New England Journal of Medicine*, 1997. **337**(3): p. 176-183.
53. Simpson, D., et al., *Insulin Lispro*. *Drugs*, 2007. **67**(3): p. 407-434.
54. Nakae, J., Y. Kido, and D. Accili, *Distinct and Overlapping Functions of Insulin and IGF-I Receptors*. *Endocrine Reviews*, 2001. **22**(6): p. 818-835.
55. Blakesley, V.A., et al., *Signaling via the insulin-like growth factor-I receptor: Does it differ from insulin receptor signaling?* *Cytokine & Growth Factor Reviews*, 1996. **7**(2): p. 153-159.

56. Ish-Shalom, D., et al., *Mitogenic properties of insulin and insulin analogues mediated by the insulin receptor*. *Diabetologia*, 1997. **40**(2): p. S25-S31.
57. FIELDS, G.B. and R.L. NOBLE, *Solid phase peptide synthesis utilizing 9-fluorenylmethoxycarbonyl amino acids*. *International Journal of Peptide and Protein Research*, 1990. **35**(3): p. 161-214.
58. Zaykov, A.N., et al., *Chemical synthesis of insulin analogs through a novel precursor*. *ACS Chem Biol*, 2014. **9**(3): p. 683-91.
59. Wellings, D.A.A., Eric, *Standard Fmoc Protocols*, in *Methods in ENZYMOLOGY*, G.B. Fields, Editor. 1997, ACADEMIC PRESS. p. 44-67.
60. Lloyd-Williams, P., F. Albericio, and E. Giralt, *Chemical approaches to the synthesis of peptides and proteins*. 2020: CRC Press.
61. Aguilar, M.-I., *HPLC of Peptides and Proteins*, in *HPLC of Peptides and Proteins*. 2004, Springer. p. 3-8.
62. Dhayalan, B., et al., *Efficient Total Chemical Synthesis of (13) C=(18) O Isotopomers of Human Insulin for Isotope-Edited FTIR*. *Chembiochem*, 2016. **17**(5): p. 415-20.
63. Liu, M., et al., *Deciphering the hidden informational content of protein sequences: foldability of proinsulin hinges on a flexible arm that is dispensable in the mature hormone*. *J Biol Chem*, 2010. **285**(40): p. 30989-1001.

64. Chen, Y.-S., et al., *Insertion of a synthetic switch into insulin provides metabolite-dependent regulation of hormone–receptor activation*. Proceedings of the National Academy of Sciences, 2021. **118**(30): p. e2103518118.
- . Owens, D.R., et al., *Insulin Centennial: Milestones influencing the development of insulin preparations since 1922*. Diabetes, Obesity and Metabolism, 2022. **24**: p. 27-42.
66. Nielsen, L., et al., *Probing the Mechanism of Insulin Fibril Formation with Insulin Mutants*. Biochemistry, 2001. **40**(28): p. 8397-8409.
67. Brange, J., et al., *Monomeric insulins and their experimental and clinical implications*. Diabetes Care, 1990. **13**(9): p. 923-954.

## **CURRICULUM VITAE**

**Oumoul Ghaniyya Faiza Sambou Oumarou**

### **Education**

#### **Indiana University School of Medicine, Indianapolis, IN**

*Master of Science, Biochemistry and Molecular Biology* **March 2023**

*Bachelor of Science, Clinical Laboratory Science with Summa Cum Laude* **June 2021**

#### **Ivy Tech Community College of Indiana, Indianapolis, IN**

*Associate Degree, General Studies with Magna Cum Laude* **August 2019**

*Technical Certificate, with Magna Cum Laude* **December 2018**

### **Honors and Awards**

American Society for Clinical Laboratory Science (ASCLS): 9/2020 – Present

Phi Theta Kappa Honor Society: 2019 – 2021

Dean's List, Indiana University Purdue University of Indianapolis: Spring and Fall 2020

Dean's List, Ivy Tech Community College of Indiana: 2017-2018

All Indiana Academic Team: 2019