Introduction to Revised Application

This is a revised version of 1 R01 EB025892-01, entitled “Glycoengineering of Therapeutic Peptides for Improved Treatment of Human Diseases”. I sincerely thank the reviewers for their thoughtful comments and suggestions. We have made changes to the proposal according to their feedback. In order to keep the text more readable and avoid potential confusion, the changes are marked by vertical lines on the left side of the text. Below, we summarize the key changes and address the specific comments made by the reviewers.

Resume and summary of discussion: In addition to the specific concerns of the reviewers, the overall impression by some reviewers that this proposal should be viewed as a drug development project caused some decrease in enthusiasm for the work. We have altered the goals outlined in the Specific Aims section to make explicit that drug development to the clinic is not our purpose here. Instead, we are attempting to refine a set of guidelines for glycoengineering of peptides and will be using insulin and GLP-1 as model molecules.

Critique 1: This reviewer brought up the possibility of adverse immune responses as a consequence of introducing unnatural glycan structures to the insulin peptide. We will certainly be cognizant of this possibility and we will look for any obvious signs of immune response during experiments with live rats. Detailed studies on the immunogenicity of glycosylated insulin and/or GLP-1 are planned for the future. In general, we do not anticipate this to be a large issue as previous work with glycosylated proteins has shown that glycans tend to reduce a molecule’s immunogenicity overall (see PMID: 1968169 and PMID: 7994026). This reviewer also suggested we justify the number of animals required in the research strategy section. We carried out a power analysis based on predicted effect sizes observed in previous similar work to arrive at the number of rats used in each study. This information is now included in the Research Strategy in addition to the Vertebrate Animal section.

Critique 2: A main concern of this reviewer relates to the *in vitro* measure of proteolytic stability, and specifically the reviewer has raised concerns that the half-life increase observed here is modest, other proteases might act far from the glycosylation sites, and that there are no clear endpoints defined for an acceptable half-life. As noted in the proposal, a doubling of *in vitro* half-life can translate to a much longer extension of half-life *in vivo* (see PMID: 15959882). This reflects the inherent limitations in estimating half-life in live animals based in *in vitro* assays. The preliminary data is meant to serve only as a proof of concept that proteolytic stabilization is possible through our glycoengineering approach, and the live animal studies proposed here would be much more relevant. For insulin stability, α-chymotrypsin was chosen since previous evidence has shown that it specifically is a critical protease responsible for insulin degradation in the small intestine *in vivo* (PMID: 2860300). GLP-1 is rapidly degraded by the protease DPP IV and analogs that are more resistant to this protease have been shown to be more effective in live animals (PMID: 21525469). Again, these studies are only meant as preliminary and the *in vivo* half-lives we are aiming to collect in the proposed work would be much more relevant. The second main concern of this reviewer is the difficulty of the synthesis as proposed. We have already synthesized many of the most difficult glyco-amino acids discussed in the proposal, including the notoriously difficult sialylated species. The remaining structures have been synthesized by other groups and there is good documentation for the procedures. With our synthetic expertise in the area, we should be able to quickly prepare those glycans with minimal optimization or modification of earlier routes. We have also prepared many peptides with densely-spaced O-glycans, including both mucin-type linker domains and chemokine signaling molecules, and this previous experience should allow us to synthesize multiply-glycosylated analogs in this study. As to the use of unnatural linkers to ease the synthesis, previous work with insulin has shown that unnatural glycosylation linkages can lead to immunogenic responses (PMID: 10837703) and work in other systems has shown natural linkage chemistry to decrease immunogenicity (PMID: 1968169 and PMID: 7994026) and increase stability. We thus feel that natural glycosylation chemistry is the best starting point for the project. Other more minor points made by the reviewers are that no rationale was given for the carbohydrate structures included in Figure 5 and that the sex of the rats was not discussed. The glycans chosen for this proposal are all naturally occurring, mammalian O-glycans. Since a goal of the proposed research is to improve our understanding of protein O-glycosylation, we decided to focus on these structures to characterize the ability of naturally occurring protein glycosylation to affect properties of peptides. The sex of the rats will not be used as a criterion for inclusion in the experiment, therefore the rats used will be both male and female.

Critique 3: This reviewer correctly asserted that there is some probability that GLP-1 glyco-variants proposed in the study might not retain full biologic activity. If this occurs, we will (as suggested by the reviewers) try other glycosylation sites. It is likely that modeling of the glycosylated GLP-1 interacting with its receptor would be helpful in this situation.

SPECIFIC AIMS.

Insulin has been widely used to treat diabetes for decades, and recently glucagon-like peptide 1 (GLP-1) has become an important part of treatment for many diabetics as well. Both of these molecules are peptides, a class of drugs that has become increasingly important and influential for not only the treatment of metabolic disorders like diabetes, but also cancer, chronic inflammatory conditions and infectious disease. Compared to traditional small-molecule drugs, peptide therapeutics generally have a couple of advantages: they are more specific and have fewer side-effects due to the precise interactions between biomolecules. On the other hand, the complex and fragile nature of peptides poses significant challenges for their administration. Of particular concern is that peptides tend to oligomerize and aggregate. Additionally, since they are often unstable, peptides can be rapidly degraded by various proteases after dosing. Most peptide drugs, including insulin and GLP-1, are administrated via frequent subcutaneous injections for the treatment of chronic diseases, which can result in low patient compliance due to pain, stress, inconvenience and other associated side effects. To address these inherent problems of peptides, many different technologies including delivery optimization, protease inhibition, and peptide structure modification, have been attempted. These technologies have led to the successful development of many peptide drugs.

Here, we propose to manipulate peptides through an underused method: *O*-glycoengineering. Glycoengineering is the term used to refer to manipulation of sugar molecules. Glycans attached to peptides have the potential to simultaneously and positively impact the proteolytic stability and resistance to both oligomerization and polymerization. This ability to affect multiple properties at once makes glycoengineering an attractive tool for further expanding the therapeutic application of peptides, but detailed guidelines for such work are not currently well-established. While much current work with glycoengineering focuses on *N*-linked glycosylation of proteins, we have chosen to focus on the more chemically diverse and difficult to control enzymatically *O*-linked glycosylation. *Based on our preliminary findings, we hypothesize that glyco-variants of peptides with improved therapeutic properties can be developed by systematically varying amino acid sequences and O-glycosylation patterns (glycosylation sites, glycan sizes and structures) in conformationally flexible, functionally important and/or fragile regions.* In order to test this hypothesis, we plan to bring together the techniques of chemical synthesis, biophysics, biochemistry, and cell biology to systematically study the impact of *O*-glycosylation on two model molecules: human insulin and GLP-1.

Our ***short-term goal*** is to provide practical and detailed guidelines for the glycoengineering of therapeutic peptides, and to identify insulin and GLP-1 glyco-variants with more desirable properties. This goal will be accomplished through the following specific aims:

Aim 1: Glycoengineering as a tool to optimize the properties of insulin for oral delivery. *We have successfully developed a convenient synthetic route for the preparation of O-glycosylated insulin variants and have demonstrated that, without affecting its biological activity, glycosylation can site-specifically and glycan-specifically improve the proteolytic stability while decreasing the oligomerization propensity of human insulin.* In this aim, we will extend our preliminary studies to prepare and characterize additional insulin glycoforms with systematic differences in glycosylation patterns and local amino acid sequences and we will determine the NMR structures of representative insulin glyco-variants. The large set of structure-property data will allow us to develop and optimize computational approaches to assist in the design of better insulin glyco-variants. Successful completion of this aim is expected to identify more potent, orally available insulin variants for animal studies and provide clues as to how *O*-glycosylation can affect therapeutic peptides that are extensively crosslinked by disulfide bonds.

Aim 2: Glycoengineering as a tool to optimize the properties of GLP-1 for oral delivery. GLP-1 and related molecules have attracted substantial attention for their utility in treating type 2 diabetes mellitus. *We have demonstrated that O-linked glycosylation at human GLP-1's N-terminus can improve its proteolytic stability*, yet guidelines for more significantly improving the properties of GLP-1 have not been developed. Here, we propose to develop such guidelines by preparing and comparing the properties of GLP-1 glyco-variants with systematically varied glycosylation patterns and amino acid sequences. We expect these guidelines to pave the way towards improved therapeutics based on GLP-1 and other GLP-1 receptor agonists, including Exenatide, Lixisenatide, and Liraglutide.

The ***anticipated result*** of this work is the identification of a collection of insulin and GLP-1 glyco-variants that have better therapeutic properties, especially oral bioavailability. More importantly, by testing such a wide variety of glycans and glycosylation patterns, this work will help to illuminate the scope and limitations of glycoengineering as a strategy to improve the properties of peptide therapeutics, which is critical to our ***long-term goal*** of improving the therapeutic performance of a wide variety of peptides, such as enfuvirtide, calcitonin, and teduglutide.

A) SIGNIFICANCE.

Importance: peptides are valuable candidates for therapeutic agents. Peptides, which are typically defined as small proteins of up to 50 amino acids, are ubiquitous in nature and more than 7,000 such molecules have been identified so far from a wide variety of natural sources.[1] They have evolved to play important and diverse roles in human physiology; they can provide a first line of defense against invading microorganisms,[2] act as ligands of membrane receptors[3] and ion channels,[4] and serve as a means of intercellular communication in the form of hormones,[5] neurotransmitters,[6] and growth factors.[7] The versatile biological activities of peptides, together with the fact that they generally have high specificity, potency, and safety, yet low toxicity, make them attractive candidates for development as therapeutics.[8] More than 60 peptides are currently approved or in the process of final approval for the treatment of human disease, most commonly indicated for metabolic disease and oncology.[9,10] With more than 600 peptide molecules in clinical and preclinical development, the number of peptide therapeutics is expected to quickly expand in the near future.[11]

Current challenges. Peptide therapeutics have a distinct set of limitations compared to small-molecules, the most prominent of which is high susceptibility to acid/base hydrolysis and proteolytic degradation.[10,12] Because of this low stability, most peptides are quickly destroyed by the digestive system, and thus must be administered via subcutaneous, intravenous, or intramuscular injection.[13] For chronic conditions, a long-term schedule of frequent injections is required for sustained effect. Due to inconvenience, pain, fear and side effects of injections, patient compliance with such administration regimens is generally poor and this increases the prevalence of disease complications.[14] Less invasive oral administration could do a lot to address this issue,[15] but peptides' poor gastrointestinal stability and absorbability in the small intestine act as major challenges preventing such a route.[16] Therefore, optimizing these two properties through engineering for oral delivery purposes represents a major and important step for improving the therapeutic potential of peptides.[8,17] Many different strategies, including PEGylation,[18] fusion to other polymers[19,20] or antibody Fc fragments,[21] and replacing L-amino acids with D-amino acids[22] have been pursued to improve peptide stability. These approaches have varying degrees of success in extending the half-lives of peptides.[8,22,23] However, *polymer conjugation significantly increases the molecular size of a peptide, which can lead to much lower oral bioavailability,[24] and prolonged used of PEGylated and unnatural amino acid-modified peptides raises safety concerns related to the accumulation of non-biodegradable PEG and unnatural amino acids in the liver and other organs*.[8,22] Such issues have limited the broad application of these approaches in peptide engineering.

One possible solution: peptide glycoengineering. Over the past three decades, research from many disciplines has established the importance of glycoengineering in overcoming the limitations of proteins. Glycoengineering covers both optimizing naturally occurring glycan patterns and introducing carbohydrates onto naturally glycosylation-naive peptides.[25] Mounting evidence points to the likelihood that *glycosylation of therapeutic peptides can have profound positive impacts on proteolytic stability,[26,27] resistance to oligomerization or polymerization,[28,29] and immunogenicity.*[30-32] These observations make glycoengineering an attractive tool for adjusting the properties of peptides to the point where oral delivery is an option.[33,34] Regrettably, we currently lack a quantitative understanding of how specific glycosylation patterns (glycosylation sites, glycan sizes and structures) influence different properties. Glycoengineering is therefore still an approach largely based on trial-and-error, which makes research in this area time-consuming and costly.[35] This knowledge deficit is a direct result of the low number of systematic structure-function studies that might link specific glycosylation patterns to physical and biological properties of therapeutic peptides. Traditionally such studies have been extremely difficult to carry out due to the inaccessibility of peptides bearing uniform, structurally well-defined glycans.

To facilitate the glycoengineering of peptide therapeutics, it is necessary to *enrich our understanding of the effects of glycosylation and develop practical guidelines that simplify the engineering process, especially for naturally unglycosylated peptides.* As an important step to achieve this ***goal***, we chose to glycoengineer two model therapeutic peptides: human insulin and glucagon-like peptide-1 (GLP-1), each representing one group of

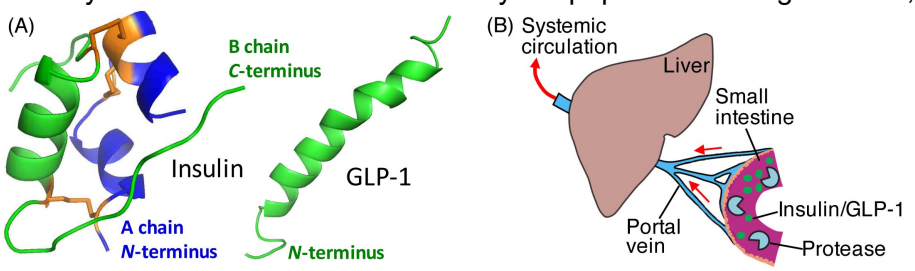


Figure 1. (A) Structure of human insulin, with the disulfide bonds being highlighted in orange, and GLP-1. (B) A schematic illustration of the absorption pathway of insulin or GLP-1 following their oral administration. If not degraded by intestinal proteases, they are absorbed into the portal vein, which transits the liver before passing the insulin into systemic circulation. This more closely mimics the route of pancreas-secreted insulin than subcutaneous injection.

therapeutic peptides.[36,37] Human insulin represents small peptides made up of two separate polypeptide chains – A chain (1-21) and B chain (1-30) – which are intra- and inter-connected through three disulfide bridges.[38] GLP-1 represents short peptides that form stable α-helices in aqueous solution without the aid of disulfide bridges (Fig. 1A).[39] Both of them are widely employed for the management of diabetes. Like most peptides, they also require frequent injections to maintain efficacy, which can cause discomfort or inconvenience to patients and lower compliance.[40] An obvious solution to this issue would be to change the route of administration to oral delivery. In addition to providing greater comfort and convenience to patients, orally ingested insulin has the added benefit of being passed first through the liver before reaching systemic circulation, which closely mimics the path of naturally secreted insulin.[41] The liver plays a major role in glucose homeostasis and is much more sensitive to insulin than muscle tissue. By mimicking the physiological release, oral administration of insulin may lead to much more effective glucose control compared to injected insulin, which would minimize many long-term complications of diabetes (Fig. 1B).[41]

Like most peptides, insulin and GLP-1 taken orally is quickly degraded in the stomach and small intestine before having a chance to pass into the bloodstream and reach their intended targets. Happily, the first of these barriers has been overcome through pH sensitive capsules that reliably protect peptides from the harsh environment of the stomach and selectively deliver them to the intestinal tract.[42] This leaves insulin and GLP-1's low absorption in the small intestine as the major remaining barrier for oral delivery.[43,44] While several factors contribute to this problem, the most important are the relatively large size of oligomerized insulin and the vulnerability of insulin and GLP-1 to the proteases in the small intestine.[17] Therefore, in order to realize an orally available therapy, it is necessary to engineer insulin and GLP-1 analogs that are more resistant to proteolytic degradation and/or self-association.[17,45] *More stable analogs will survive longer in the intestinal track, which increases overall absorption efficiency, and small monomers have higher permeability across the intestinal epithelium.*[46,47]

Based on our preliminary findings,[48,49] *we hypothesize that improved insulin and GLP-1 analogs can be identified by analyzing the properties of many differently glycosylated isoforms (glycoforms) with systematically varied amino acid sequences and glycan structures in conformationally flexible regions*. We recently developed convenient synthetic routes to large collections of insulin and GLP-1 glycoforms that carry systematic variations in both glycan structures/sizes and glycosylation sites.[48] Easier access to libraries of homogeneous glyco-variants is expected to greatly facilitate achieving the goal of the work proposed here.

The ***impact*** of the proposed research will be profound and broad. First, it will provide new insulin and GLP-1 variants with better properties, which will move us closer to orally available insulin and GLP-1 molecules. Furthermore, as an extremely important short peptide, insulin has long been used as a model molecule to develop new strategies for protein sequencing,[50] synthesis,[51] expression,[52] structure determination,[53] and engineering.[54] Similarly, previous studies of GLP-1 have established new approaches to engineer therapeutic peptides.[55,56] By systematically analyzing synthetic, homogeneous glyco-variants of human insulin and GLP-1, we will obtain a better understanding of the scope and capacity of glycoengineering. Such information is expected to greatly facilitate ***our long-term goal*** of using glycoengineering to address the limitations of many different therapeutic peptides, such as enfuvirtide,[57] calcitonin,[58] and teduglutide.[59]

B) INNOVATION.

Glycosylation holds substantial promise to simultaneously change multiple physical properties of small peptides in ways that are advantageous for therapeutic administration.[27] As such, the glycoengineering approach outlined in this research proposal is an important alternative to more traditional means of peptide engineering like secondary structure staples,[60] PEGylation,[61] or amino acid modifications.[62] Much evidence has shown that the carbohydrate residues closest to the peptide backbone have the strongest effects on physical properties of glycopeptide conjugates.[63,64] While *N*-glycans have only a single monosaccharide that is directly linked to the protein, *O*-glycans are initiated by a huge diversity of carbohydrate structures, at least seven of which are common in mammalian systems.[65,66] Such chemical diversity, while challenging to study, suggests that *O*-glycosylation has a potentially greater ability for fine-tuning physical properties in a way that would be difficult with *N*-glycans. *We believe that our innovative use of different types of O-glycosylation has the potential to shift the paradigm of peptide engineering beyond traditional methods.*

Integrated and pioneering approaches are needed to tackle the complexities associated with glycosylation. In the proposed research, chemical synthesis will be used to prepare site-specifically glycosylated peptides.[67,68] Although other methods, such as biological expression,[69] enzymatic synthesis,[70] or "click-like" conjugation methods,[71] are more practical means for large-scale production of glycopeptides, *chemical synthesis offers greater flexibility for introducing variations into glycopeptides and for controlling every aspect of glycan structure and amino acid sequence.*[22,72] This is a direct consequence of the fact that chemical

glycosylation is not dictated by the underlying amino acid sequences, local peptide conformation, or the chemical properties of side chains.[71,73] It thus allows for more diversity in glyco-variant structures, which will enable us to explore the fullest range of possible glycosylation patterns.[48,74] Such a high level of control is particularly relevant in the context of *O*-glycosylation, which is notoriously difficult to enzymatically introduce in a controlled and predictable manner compared to *N*-glycosylation.[75]

A further innovative contribution is that the large set of experimentally derived data generated from the proposed project will allow us to develop computational molecular modeling tools that predict the properties of insulin and GLP-1 variants in silico before synthesis. It is possible that the best analogs could result from a combination of amino acid sequence mutations, glycosylation sites, glycan sizes and structures that have been individually identified as beneficial. However, a comprehensive investigation of this subject would involve the preparation and analysis of a large number of variants with all possible combinations of variations in glycosylation and mutation. Because such a process would be very time-consuming and impractical, it would instead be advantageous to predict, before synthesis, particular glycoforms most likely to result in large improvements. Based on the large set of property and structure data that will be obtained from the proposed study, we will be able to develop such predictive modeling tools that promise to simplify the engineering processes considerably.[76,77] Such advances would be particularly beneficial for engineering peptides that are difficult to synthesize.

C) APPROACH.

The primary goal of this project is to harness glycosylation as a means to improve the properties of therapeutic peptides for oral delivery, and to provide practical tools to facilitate this process. We seek to achieve this goal through the combined use of chemical glycobiology[67] and computational biochemistry.[78] The proposed study is built upon our previous research[48,64,79-81] and encouraging preliminary results.[49] We will address two aims. Aim 1 will focus on the glycoengineering of human insulin.[48,64] It involves chemical synthesis of insulin bearing systematically varied amino acid mutations and glycosylation patterns, studies comparing the influence of these variations on the properties and structure of insulin, and use of the resultant experimental data to develop computational approaches that will ease future engineering of insulin. At the same time, in Aim 2, an identical research strategy will be applied to the engineering of a structurally distinct peptide: GLP-1. *We expect that improved versions of insulin and GLP-1 will be unveiled as a result of these comprehensive systematic studies, and, more importantly, the success of the proposed efforts will significantly advance our understanding of the principles and application of peptide glycoengineering.*

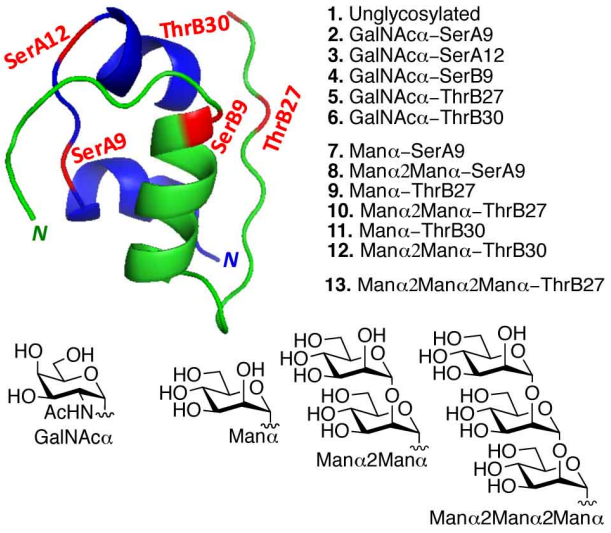


Figure 2. Design of human insulin glyco-variants. The *O*-glycosylated Ser and Thr residues are highlighted in red. The structural feature of each glycoform is indicated by its name, *i.e.* GalNAcα-SerA9 contains a single GalNAc α-linked to the A chain Ser9. Manα2Manα2Manα-ThrB27 representing the glycoform containing an α1,2-linked tri-mannose at the B chain Thr27 site.

C1) Aim 1: Glycoengineering as a tool to optimize the properties of insulin for oral delivery.

C1a) Introduction. We propose herein to develop insulin glyco-variants with significantly enhanced resistance to protease degradation, very low or even completely abolished oligomerization, and unchanged biological function. Our studies will build upon our preliminary results that show the attachment of a saccharide to the *C*-terminal region of insulin's B-chain can improve its properties for oral delivery[49] and previous findings that multiple modifications can act synergistically to further optimize the molecule.[64,82,83] With an appreciation for the sometimes unpredictable ways certain mutations and glycosylation patterns may interact, *we hypothesize that screening a more diverse collection of glycosylated insulin analogs will provide a higher probability to discover specific variants with the most desirable properties.*

C1b) Justification and Feasibility. Many different types of modifications have already been exploited to engineer peptides.[84] Compared to other modifications, glycosylation has a unique potential to simultaneously improve many properties of peptides, including accelerating folding rates, increasing solubility, proteolytic and thermal stability, decreasing propensity for oligomerization or aggregation, and altering biological activity and specificity.[85] This attractive ability of glycans is likely due to a pair of structural features: (1) each glycan

monomer contains many hydrophilic groups and a hydrophobic face; (2) depending on the glycosylation type, glycan oligomers can have a huge variety of molecular compositions and sizes.[86] These features make it possible for glycans to participate in many interactions with surrounding moieties, both adjacent and distal, within the same molecule.[27] For example, glycans can help increase the thermal stability through the formation of CH−π interactions with neighboring aromatic residues[87] or diminish the aggregation tendency by shielding exposed hydrophobic residues that are far from the glycosylation site.[88] Because of the great potential for glycans to modulate peptide properties, we chose glycoengineering as a way to address the natural limitations of insulin.

Although previous studies have uncovered numerous instances of glycosylation altering another molecule's characteristics,[27] there are still many gaps surrounding the detailed links between specific glycosylation patterns and physical properties. Results from several laboratories,[82,89] including our own,[48,81] suggest that *attaching glycans to unstructured regions has the greatest chance of imparting beneficial effects while also minimizing any interference with a protein's folding.* Previous observations also indicate that, while glycans can be quite large, it is predominantly the core structures (the carbohydrate residues closest to the peptide backbone) that mediate most of the physical changes.[48,63,64,82,90] Our first concern was to test if these conclusions, which were made with naturally glycosylated peptides and proteins, could be directly applied to artificially glycosylated systems, namely insulin.

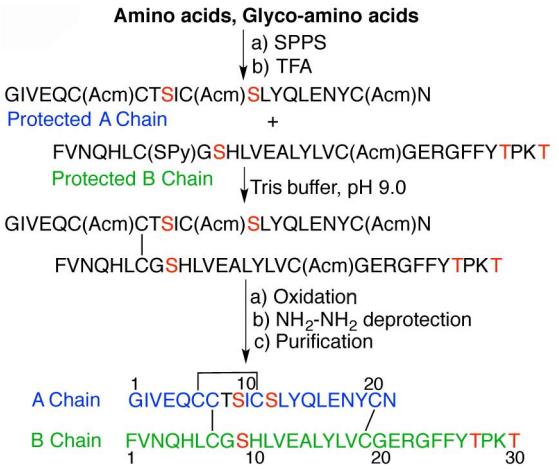


Figure 3. The optimized synthetic route to glycosylated insulin variants. The *O*-glycosylated Ser and Thr residues are highlighted in red. SPPS, solid-phase peptide synthesis.

We started by designing and synthesizing 12 different insulin glycoforms, **2**-**13** (Fig. 2), each containing an *O*-linked *N*-acetylgalactosamine (GalNAcα), mono-mannose (Manα), di-mannose (Manα2Manα), or tri-mannose (Manα2Manα2Manα), at either SerA9, SerA12, SerB9, ThrB27 or ThrB30. ThrA8 was not used in this initial study because it is adjacent to SerA9 and the effects of its glycosylation can be roughly represented by those of SerA9 glycosylation. Due to the presence of difficult-to-synthesize sequences and the possibility of forming non-native disulfide bonds, chemical synthesis has yet to prove itself as a convenient technique for the preparation of insulin glyco-variants.[91,92] By systematically optimizing each of the synthetic steps, we developed a robust, practical and efficient one-pot process to synthesize, fold, deprotect, and purify uniform insulin glycoforms (Fig. 3). Using this method, we were able to quickly generate highly pure unglycosylated insulin **1** and insulin glycoforms **2**-**13** in sufficient amounts for biophysical and biological characterization. Correct folding, identity, and homogeneity of the synthetic products were each confirmed by a combination of circular dichroism (CD), ultra-performance liquid chromatography mass-spectrometry, and enzymatic (Glu-C) digestion.[49]

With the synthetic insulin glycoforms in hand, we first investigated if *O*-linked glycans at any one of the five glycosylation sites affects the stability of human insulin in the presence of α-chymotrypsin, a protease synthesized by the pancreas and secreted into the lumen of the small intestine. α-Chymotrypsin cleaves human insulin with in the *C*-terminus of its B chain, an important region for receptor binding and activation, thus diminishing its biological activity.[93,94] This cleavage also causes an easily detectable change in molecular mass, and therefore each insulin glycoform's half-life towards proteolytic degradation can be calculated by monitoring the first-order exponential decay of the full-length glycoform using quantitative Matrix Assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry (MALDI-TOF MS).[48,64] These factors make α-chymotrypsin an ideal test of glycosylated insulin's resistance to digestion in the small intestine. The role of *O*-linked glycans in human insulin proteolytic stability can be established by comparing the half-lives of synthetic glycoforms with that of the unglycosylated insulin. As shown in Figure 4A, *O*-glycosylation with a GalNAcα (**2**-**6**) or Manα (**7**,**8**,**9**,**11**,**12**) moiety does not positively impact the proteolytic stability. Encouragingly, we found that either di-mannosylation (**10**) and tri-mannosylation (**13**) at ThrB27, which is adjacent to one of the cleavage sites, leads to noticeable improvement in proteolytic stability. The half-life of tri-mannosylated **13** is twice as long as that of unglycosylated insulin **1**.

In addition to improving the proteolytic stability of insulin, we found that *O*-glycosylation could also decrease the oligomerization propensity of insulin. Oligomerization is a critical regulatory factor in insulin absorption and it is well known that aggregation of insulin can decrease its absorption.[95] By analyzing the sedimentation velocity though analytical ultracentrifugation (AUC, Biophysics Core Facilities, University of Colorado Anschutz Medical Campus), we derived a distribution of insulin molecular species that have different

degrees of self-association (Fig. 4B).[96] The area under each peak gives the relative concentration of that species. As illustrated by the data, tri-mannosylation at ThrB27 increased the amount of monomer to more than 70 percent of the sample, whereas amounts of monomer and dimer in samples of unglycosylated **1** were almost equal. Insulin dimers are doubly undesirable in the context of membrane permeability because, not only are they twice as big as monomeric insulin, but they are also the first step in formation of hexamers in the presence of Zn^{2+} ions, which will severely limit transit across the intestinal membrane.[97]

The promise of glycoengineering design is to create insulin products of increased beneficial delivery properties while not sacrificing biological activity. In order to understand the relationship between the glycosylation and the biological activity of human insulin, we used a quantitative fluorescence assay to compare the trafficking of hemagglutinin (HA)-tagged glucose transporter type 4 (GLUT4) in 3T3-L1 adipocytes.[98,99] GLUT4 is an insulin-regulated glucose transporter that is responsible for insulin-regulated glucose uptake into liver, muscle, and fat cells. Insulin can increase the cell-surface level of GLUT4 by stimulating the translocation of GLUT4 to the plasma membrane. Therefore, the biological activity of each insulin analog can be evaluated by analyzing the level of cell-surface GLUT4. As shown in Figure 4C, the attachment of tri-mannose to Thr27 slightly increases the natural biological activity of insulin.

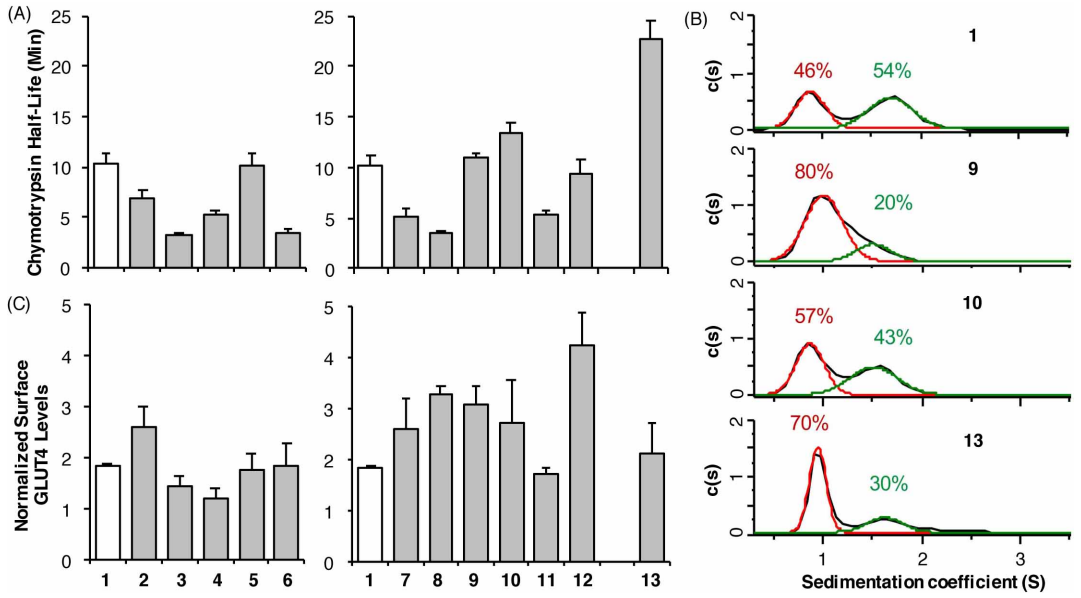


Figure 4. Characterization of synthetic insulin glyco-variants. (A) The effects of *O*-glycosylation on the proteolytic stability (half-life to α-chymotrypsin degradation). (B) The effects on oligomerization propensity (sedimentation coefficient distribution). The monomer is highlighted in red and the dimer in green. (C) The effects on insulin-stimulated translocation of HA-GLUT4. Error bars show the standard deviations from three trials.

Together with previous studies, *our preliminary work has confirmed the feasibility and effectiveness of the glycoengineering approach in increasing the beneficial properties of human insulin.* In a site-specific manner, *O*-mannosylation at B-chain Thr27 was shown to make insulin more resistant to protease attack and self-association. More importantly, the artificial *O*-glycans at this site were found not to block the natural receptor interactions that mediate insulin's biological functions. These observations make us confident that our proposed research represents a unique opportunity to improve the performance of human insulin in a therapeutic context.

C1c) Research design. Based on what has been learned from studying natural glycoproteins and our findings that insulin can be enhanced by *O*-glycosylation of ThrB27,[49,82] *we hypothesize that insulin analogs with further improved properties for oral delivery can be identified by systematically varying the amino acid sequences, glycosylation sites, glycan sizes and structures at the C-terminus of its B chain*.[94,100] To test this hypothesis, we will first prepare and characterize a larger library of glycosylated insulin analogs. By comparing the properties and structures of chemically diverse insulin variants, we will be able to determine the molecular basis of the observed effects of insulin glycosylation.[90] Next, we will use our experimentally derived data sets to develop computational tools that can predict the properties of insulin glyco-variants *in silico* and apply those tools to design, test and identify additional insulin glyco-variants with improved properties. The insulin glyco-variants identified at each research stage will be further studied and verified in animal models.

Prepare and analyze glycosylated insulin analogs with varied amino acid sequences and glycosylation patterns. Given previous evidence of the importance of interactions between glycans and local amino acids in modulating peptide properties, it is likely that amino acid changes in glycosylated human insulin will lead to better outcomes.[80,82,90] To test this possibility, we propose to synthesize and characterize a collection of insulin variants that are tri-mannosylated at ThrB27 to determine the roles of different amino acid residues in changing the properties of insulin.

We will generate the first collection of insulin glyco-variants by introducing tri-mannose to the ThrB27 sites of fast-acting insulin drugs like Aspart, Glulisine, and Lispro, which were previously developed through mutations of the amino acid sequence at the *C*-terminus of insulin B chain (Fig. 5A).[54] The proteolytic stability, oligomerization propensity, and biological activity of these synthetic glyco-variants will be initially screened *in vitro*. We will determine the stability of each glyco-variant in the presence of major intestinal proteases (pepsin, trypsin, α-chymotrypsin and elastase) using the MALDI-TOF MS-based quantitative assay developed in our laboratory.[48,64] Since even a doubling of *in vitro* half-life can translate into significantly longer increases in serum half-life, we expect that observing 2 to 3 fold increases in resistance to these enzymes will be sufficient for improvements in the animal studies detailed below.[26] Because previous research with orally delivered insulin has shown α-chymotrypsin to be highly responsible for degradation in the small intestine,[101] resistance to α-chymotrypsin will be particularly important. The effects of different amino acid sequences on insulin's biological activity will be monitored using the GLUT4 translocation assay described above.[98] The influence of mutations on insulin's tendency towards oligomerization will be probed by AUC.[102] These simple *in vitro* assays and our familiarity with them will allow for rapid assessment of the effects of amino acid mutations to the wild-type sequence. Based on previous work and our preliminary studies,[80,82,90] we expect that the glycosylation of insulin Aspart and Lispro at Thr27 will lead to the most beneficial outcomes. The *C*-terminal B chain glycosylation sites in these two insulin analogs are flanked by aromatic residues (Phe, Tyr), which are capable of crosstalk with adjacent glycans, and non-Pro residues, which are generally not harmful to the stabilizing effect of protein glycosylation.[48,82] If no glyco-variants derived from Aspart, Glulisine, or Lispro are found to have significantly improved physical properties that suggest greater oral availability, other glycosylated insulin analogs will be generated by systematically varying the *C*-terminal sequence of the B chain. Specifically, the amino acids Asn, Asp, Glu, Gln, Arg, and His, which have planar polar side-chain groups and are known to interact favorably with glycans, will be used to replace ProB28 and LysB29 (Fig. 5A).[103] Again, these glyco-variants will be carefully characterized using the above-described *in vitro* assays to identify the most stable, least aggregation-prone glycoforms.

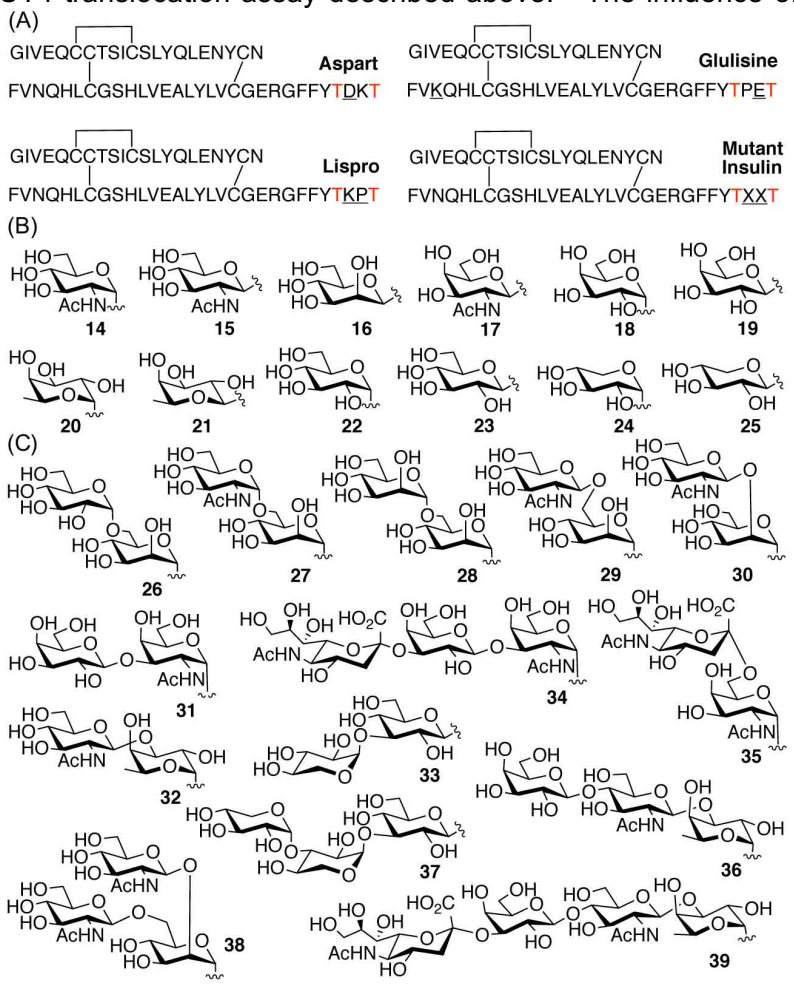


Figure 5. Insulin glyco-variants possessing varied amino acid sequences and glycan structures. (A) Amino acid sequences of insulin variants. The mutated amino acids are underlined and the *O*-glycosylated residues are highlighted in red. (B) (C) Additional *O*-linked glycans to attach to insulin to determine their effects.

We will also investigate the effects of more chemically diverse *O*-linked glycans than those used in the preliminary studies. There are many different types of naturally occurring *O*-linked glycans.[65] Although α-linked mannose clearly introduces many benefits, other glycan structures may also prove capable of enhancing the properties of insulin. To test this possibility, we propose to systematically study the effects of additional glycosylation patterns, including a wide range of chemically diverse glycan structures, sizes, and glycosylation sites. To achieve this research goal, we will first synthesize and characterize 12 variants by mono-glycosylating insulin at ThrB27 site with α- and β-linked *N*-acetylglucosamine (GlcNAc), β-Man, β-GalNAc, α- and β-galactose (Gal), α- and β-fucose (Fuc), α- and β-glucose (Glc), and α- and β-xylose (Xyl) (Fig. 5B).[65] It is our hope that retaining natural linker chemistry between the carbohydrate and peptide moieties will lead to the most pronounced effects on physical properties, as evidenced by our earlier work.[48] The O-glycan structures are found naturally occurring in mammals, which should decrease the likelihood of adverse immune responses *in vivo*.[104-106] *We have already*

synthesized most of the glycoamino acid building blocks required for the synthesis of these glyco-variants[48,64] *and the rest can be readily synthesized based on previous work.*[107-117] We also wish to be able to directly compare the benefits of glycosylation to those of PEGylation, which has been explored by others previously.[118-120] We will thus synthesize PEGylated insulin analogs as well. Care will be taken to ensure that the PEG moiety is small enough that it will not significantly hinder the absorption of insulin. Concurrent with the synthesis of these insulin variants, we will use the above-mentioned screening assays to quickly assess their properties.

After determining the effects of various types of *O*-linked mono-saccharides, we will investigate the effects of more complex *O*-linked glycans with a wide range of structures (Fig. 5C). Although we expect that addition of the core substructures shown in Figure 5B will have considerable impacts,[63,64] our pilot study suggests that the outer saccharide residues (*e.g.* tri-mannose) can increase the beneficial effects beyond the effects seen with small glycans (*e.g.* mono-mannose).[49] To test this hypothesis, we propose to synthesize and systematically study the effects of extended oligosaccharide structures on the properties of insulin. To accomplish this task, we will synthesize and characterize another collection of mono-glycosylated insulin variants bearing the complex O-linked glycans **26-39** at ThrB27 (Fig. 5C). Sialic acid terminated O-glycans were included due to their well-documented ability to impact important properties including reducing immunogenicity of proteins[121] and increasing stability.[122] *We have extensive experience in the synthesis of complex glycoamino acid building blocks*[123-125] *and have already synthesized the sialylated O-linked glyco-amino acids **34** and **35** in a gram scale.*[48] The knowledge obtained from these prior experiences makes us uniquely prepared to overcome the synthetic challenges that may come up during the research process.[111,126,127]

Previous studies of multi-glycosylated proteins suggest that glycosylation sites can act synergistically to affect the properties of glycoproteins.[27,64,128] Therefore, in addition to systematically study the effects of single-site occupancy glycosylation, we will also elucidate the properties of insulin variants that are doubly glycosylated at both ThrB27 and ThrB30 sites. We will design the glyco-variants with the results obtained from the above studies in mind. Amino acid sequences and glycan structures that led to better effects when attached to ThrB27 in isolation will also be attached to the ThrB27 site of the di-glycosylated variants. Mono-, di-, and tri-mannoses, which are relatively easy to prepare and behaved well in our pilot study, will be added to ThrB30 in the di-glycosylated variants. We have previously prepared and studied densely O-glycosylated peptides, and the synthetic experience gained for those efforts will allow us to quickly prepare the proposed multiply-glycosylated insulin analogs. As with earlier stages of the proposed study, the properties of the di-glycosylated ones will be quickly evaluated through *in vitro* assays.

Determine the molecular basis of the effects of amino acid mutations and glycosylation. Engineering therapeutic peptides with enhanced properties requires detailed insight into the structural factors that lead to any changes observed after amino acid mutation or post-translational modification. This is tied to the commonly held belief that a protein's properties are determined to a large extent by its three dimensional structure. To better understand the mechanisms by which different side chains and carbohydrates can affect the properties of insulin, it is important to analytically compare the structures of variants with systematically varied compositions and properties. Understanding the structural changes that occur as a result of glycosylation and being able to correlate those with the physical properties of insulin is a valuable step towards being able to predict how specific engineering choices will affect the peptide's behavior as an oral agent.[80,82,90] This type of knowledge is expected to significantly improve our understanding of the strengths and limitations of a glycoengineering approach to therapeutic peptide optimization.

We will choose several glycosylated insulin analogs for in-depth structural characterization by NMR. These analogs will be chosen based on the physical properties we observe them to have during the course of the experiments detailed above, and will cover the full range of results we get.[80] For example, we will include analogs that are found to be vulnerable and resistant to protease digestion and compare them to better understand how the structure may lead to the different outcomes. Similar pairs of analogs will be selected for the other two properties important for oral delivery: propensity for self-association and biological activity. The unglycosylated, wild-type insulin peptide will also be included as a control, and the differences between each of the glycosylated analogs and the wild-type insulin will tell us how different glycosylation patterns influence the structure of insulin.

We plan on using an approach similar to what we have recently published to look at the structural details of glycosylated insulin analogs.[80] Since this approach is relatively simple and not technically demanding, it would allow us to quickly characterize the differences in glycosylated CBM variants. Briefly, each insulin sample will be dissolved in an acetate buffer (pH 5.0) in both 90% H_2O/10% D_2O and 100% D_2O, which allows for easy

comparison of solvent-exposed, exchangeable protons in the structure to those involved in the more stable bonds of the peptide secondary structure. 0.01% (w/v) sodium 2,2-dimethylsilapentane-5-sulfonate (DSS) will be used as an internal standard for chemical shift reference.[129,130] Spectral processing will be performed using the program NMRPipe[131] and analysis will be done with NMRViewJ.[132] Peak assignments for protons will be identified with two-dimensional (2D) ^{1}H-^{1}H DQF-COSY, ^{1}H-^{1}H TOCSY, and ^{1}H-^{1}H NOESY spectra, while assignments for carbon and nitrogen resonances will be identified through 2D ^{1}H-^{13}C HSQC, ^{1}H-^{13}C HSQC-TOCSY, and ^{1}H-^{15}N HSQC spectra. Sugar ring atom assignments can be made with 2D ^{1}H-^{13}C HSQC, H2BC, and ^{1}H-^{1}H COSY spectra.[133] Peak relaxation times will be measured using a Carr-Purcell-Meiboom-Gill pulse sequence with presaturation water suppression and these data will be analyzed through the T1/T2 module of Topspin.[134]

The program CNS will allow us to take distance constraints obtained during the 2D ^{1}H-^{1}H NOESY experiments[135] and backbone torsion angles obtained by TALOS-N analysis[136] to calculate initial three-dimensional structures for each glycosylated insulin. The program SANE will be used to semi-automatically assign the NOE peaks.[137] Refinement of these initial structures with explicit water calculations will then be done using RECOODScript.[138] The 20 lowest energy structures after refinement will be validated and analyzed using PROCHECK-NMR,[139] and WHATCHECK.[140]

After these extensive NMR experiments, we will have high-resolution solution-phase structures of each glycosylated insulin. Such structures will allow us to see, in great detail, exactly where and how the glycan units contact the insulin peptide. Comparing individual contacts, including potential hydrogen bonds, CH−π, and hydrophobic interactions between the glycans and amino acid side chains or the peptide backbone, with those of the unglycosylated insulin (Fig. 6) will give us a much better understanding of the molecular-scale causes of any changes to the molecule's physical properties. We will also be able to understand the dynamics of the glycosylated insulin variants at the level of individual amino residues. By comparing the unglycosylated, wild-type insulin with each of the glycosylated variants, we can see if any glycosylation patterns significantly alter the dynamic flexibility of the molecule in either a global or local sense. This information would help better understand the data measured through above-mentioned *in vitro* experiments on the physical properties of each glycosylated insulin.[80]

Rational design and *in vitro* evaluation of insulin analogs with combined variations. The very best insulin analogs might consist of several amino acid mutations and glycosylation patterns that individually gave good results.[48,82] It is also entirely possible that unforeseen synergies between amino acid mutations and certain glycosylation patterns could be uncovered. To exhaustively test all possible combinations of glycosylation patterns and mutations, an enormous number of insulin variants would need to be synthesized and tested. This would require an unreasonable expenditure of time and effort and it would be advantageous to predict, before synthesis, particular combinations that are likely to improve the molecule. Professor Heather Mayes at the University of Michigan (see Letter of Support) will collaborate with us and use the trove of empirical data gathered during the course of the study to develop computational tools capable of predicting a theoretical glyco-variant's properties *in silico*.

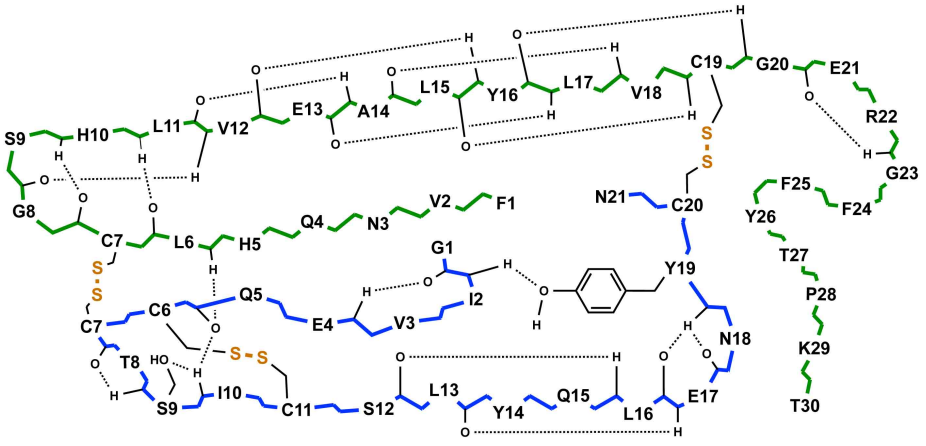


Figure 6. Schematic representation of the hydrogen bonding networks for unglycosylated human insulin. Hydrogen bonds identified in the NMR structures are based on the distance and angle constraints proposed by Mills and Dean.[129] A chain and B chain are shown in blue and green respectively. Hydrogen bonds are shown with dashed lines.

The most important properties for developing oral insulin are the proteolytic stability, oligomerization propensity, and biological activity. An analog that is very resistant to protease digestion, does not form oligomers, and possesses the biological activity of wild-type insulin is ideal candidate we are aiming for. Previous studies have demonstrated that it is feasible to use computational approaches to study each of these properties.[119,141-144] Therefore, we will use previously developed approaches to perform the proposed studies. The proteolytic stability will be studied using molecular dynamics (MD) simulations. The aggregation propensity will be studied using predictive algorithms based on empirical data and also MD simulations. The biological activity will be studied using MD

and Monte Carlo (MC) simulations. These computational approaches will be honed using the NMR structures that we solved in the previous section. The large set of structural and property data we will gather during the many experiments in this proposal will be used to validate and optimize our computational approach.

The effects of the engineering designs on the proteolytic stability will be investigated through MD simulations of the glycosylated and mutated insulin molecules in water. Similar methods have been used before with PEGylated insulin[119] and other proteins,[145] which will inform the design of our computational methods here. By using calculated measures of the molecule's dynamics, we will be able to predict how certain modifications will stabilize or destabilize the insulin molecule. Specifically, we will look at the root-mean squared deviation (RMSD), root-mean square fluctuation (RMSF), and radius of gyration values for each simulated molecule.[146]

Protein aggregation will be examined through a combination of predictive algorithms and MD simulations. Much like proteolytic stability, the aggregation of proteins is intrinsically linked to flexibility, and several groups have taken advantage of this to predict aggregation propensity based on RMSF measures calculated during MD simulations.[147,148] Solvent-accessible-surface (SAS) area has also been used to effectively predict aggregation propensity from simulated molecules. Predictive algorithms, based on experimentally-obtained, real-world aggregation propensities for many proteins, are also quite useful for our purpose here. In particular, the AGGRESCAN3D approach will be applied. This approach uses the 3D structure of the molecule as well as intrinsic aggregation values for each amino acid, which are based on empirical data, to predict aggregation-prone proteins.[149]

Numerous computational methods have been combined with detailed structural data to study the interaction of insulin with its receptor.[150] In particular, the application of both MD and MC simulations has greatly advanced our understanding of how the conformational dynamics of both insulin and its receptor lead to binding and activation.[151] We will use these previous explorations to guide our simulations of glycosylated insulin variants and help us to predict how the glycans contribute, and in some cases hinder, the biological activity of the molecule. Previous work has shown that the unfolding of *C*-terminus of the B-Chain is absolutely critical to receptor engagement.[152] We believe this is a good starting point, and will begin by focusing on the dynamics of this region for glycosylated insulin.

These *in silico* tests will allow us to test a much greater number of insulin glyco-variants than chemical synthesis and *in vitro* characterization would permit. Based on the results of the *in silico* screens, we will be able to choose a small collection that shows the most promise for oral delivery. We will then use chemical synthesis to prepare these selected variants and carry out *in vitro* assays to experimentally characterize their properties. Insulin variants that perform well in the *in vitro* assays will be carried on to the later phases of the study for further testing in rats.

Test insulin glyco-variants in animal studies. After the *in vitro* phase of our study, we will move on to testing the identified insulin glyco-variant in rats. *In vivo* potency and the ability of insulin glyco-variants to regulate glucose disposal will be tested in streptozotocin (STZ) diabetic rats as previously reported.[153] Briefly, pH-sensitive, gelatin-based capsules containing individual insulin glyco-variants will be orally fed to diabetic rats.[154] These capsules are coated with hydroxypropyl methylcellulose phthalate (HP55) that allows for selective release of insulin in the small intestine at pH ~7. The blood glucose levels will be measured at different time intervals to evaluate the effects of glycosylation on the action of orally administrated insulin glyco-variants, which should be directly related to their proteolytic stability, absorbability, and biological activity. During this study we will also monitor the concentration of each glycosylated insulin variant in the blood over time, which will give us valuable pharmacokinetic data including how fast each variant is absorbed into and cleared from systemic circulation.[154]

We are also interested in studying the distribution of the molecules in rats after oral administration. We will use the methods previously employed by other groups to perform the research task.[155] Briefly, ^{125}I-labeled insulin glyco-variants will be administered to rats in HP55-coated capsules. Live-animal images will be collected at certain time points after administration using computed tomography (CT) scanning equipment available in the Small Animal Imaging facility at the University of Colorado Anschutz Medical Campus. These high-resolution images will tell us where in the rat body the insulin is located at each time point after administration, and will allow us to verify its transit through the small intestinal epithelium into systemic circulation, and its eventual clearance. Encouragingly, power analysis based on observed effect sizes from previous work suggested that only a small number of rats (50) would be required to obtain significant, statistically sound data for both of these studies.[156,157] Rat sex will not be used as a criterion for inclusion in this study, so both male and female rats will be used.

C1d) Expected outcomes, potential problems, and alternative approaches. The proposed study will provide more insight into the power of glycoengineering and its ability to overcome the challenges faced by insulin and peptides with similar structures, and may lead to the identification of insulin glyco-variants that have the potential to be further developed into orally administered products.

Solid-phase peptide synthesis should be able to generate mono-glycosylated insulin chains with minimal difficulty. It is possible, however, that the efficient SPPS of glyco-variants bearing two large glycans at ThrB27 and ThrB30 could be complicated by steric hindrance. In these cases, we will either use a chemoenzymatic approach[158] or use manual, solution-phase peptide coupling[125] for the densely glycosylated regions.

C2) Aim 2: Glycoengineering as a tool to optimize the properties of GLP-1 for oral delivery.

C2a) Introduction. We propose herein to use glycoengineering to improve the properties of GLP-1, another peptide therapeutic that can be used to treat type 2 diabetes mellitus. GLP-1 has been shown to significantly lower the occurrence of hypoglycemia,[37] the most commonly observed adverse effect during therapeutic intervention with traditional treatment options. GLP-1 has also been shown to help decrease the weight gain commonly associated with insulin-based therapy options.[37] These added clinical benefits come as a result of GLP-1's action on not only insulin release, but also its ability to regulate glucagon activity, satiety, and gastric emptying rate.[159] Since it has the potential to address many of the currently unmet needs of diabetes patients on traditional therapy, GLP-1 and related molecules have attracted substantial attention as potentially lucrative new drug molecules. As with insulin and most other therapeutic peptides, the proteolytic stability of orally administered GLP-1 in the gastrointestinal tract is low. Because of their great therapeutic value, the development of orally available GLP-1 and GLP-1 receptor agonists is being widely investigated.[160,161] *In our preliminary efforts, we have established N-terminal O-glycosylation as a feasible strategy for improving its proteolytic stability. Based on such supportive data, we hypothesize that GLP-1 analogs with further improved properties for oral delivery can be identified by systematically varying the glycosylation patterns at its N-terminal region.*

C2b) Justification and feasibility. GLP-1 is naturally circulated as a pair of equally active isoforms: GLP-1 (7-37) and GLP-1 (7-36) amide (Fig. 7A).[162] Unfortunately, as a short peptide, GLP-1 is rapidly inactivated by dipeptidyl peptidase IV (DPP IV)[37] and other digestive enzymes.[163] In order to develop orally active GLP-1, it is necessary to improve the stability of GLP-1 against these enzymes. To achieve this objective, we first conducted a pilot study to determine the feasibility of using glycoengineering to decrease the degradation rate of GLP-1 (7-37) in the presence of DPP IV.

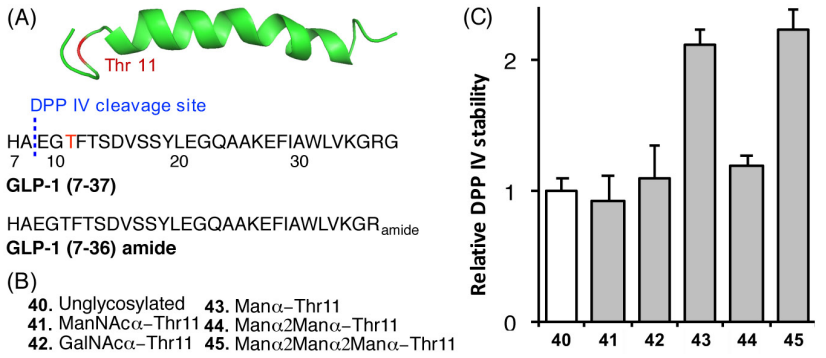


Figure 7. Design and characterization of synthetic GLP-1 glyco-variants. (A) The NMR structure, amino acid sequence, and DPP IV cleavage site of human GLP-1. The glycosylated Thr residue is highlighted in red. (B) Synthesized GLP-1 variants. (C) The effects of *O*-glycosylation on the proteolytic stability (relative half-life to DPP IV degradation). The error bars report the standard deviations from three separate trials.

Our study began with the design and synthesis of differently glycosylated GLP-1 analogs. Similar to the glycoengineering of insulin, α-linked *N*-acetylmannosamine (ManNAc), GalNAc and Man was first use to modify the GLP-1 molecule. Previous work has clearly demonstrated that DPP IV deactivates the GLP-1 by cleaving the "HA" dipeptide from the flexible *N*-terminal region of this peptide (Fig. 7A).[164,165] *From our experience, we expected that the attachment of O-linked saccharides to amino acids that are close to this cleavage site may be capable of slowing the process of GLP-1 degradation.* Based on this hypothesis, we designed 5 glycosylated GLP-1 analogs with varied glycan structures at Thr11 (Fig. 7B). These glycoforms have relatively simple structures, which allowed for very reliable and efficient synthesis. After the synthesis, the stability of these GLP-1 glyco-variants in the presence of DPP IV was tested in a similar way to that of the insulin variants detailed in Aim 1. As show in Figure 7C, our preliminary results clearly demonstrate that *O*-mannosylation of Thr11 confers protease protection.

C2c) Research design. Encouraged by the preliminary results, we will continue our studies of GLP-1 *O*-glycosylation. Similar to the proposed studies involving insulin, we will begin our research by designing, preparing and then testing new GLP-1 glycoforms using cell-free and cell-based approaches, followed by in-depth structural characterization of representative glyco-variants. The accumulated experimental data will then be used as input to develop computational approaches to predict the properties of as-yet unsynthesized GLP-1

analogs. Synthesizing and testing analogs predicted to have interesting properties is expected to lead to novel candidates for orally available GLP-1 variants.

Design, synthesize and characterize a large collection of GLP-1 analogs with systematically varied *N*-terminal glycosylation patterns. We will begin by focusing on the preparation of GLP-1 variants with glycans shown in Figure 5 being attached to Thr11. With the ease of GLP-1's synthesis and the ready availability of glycoamino acid building blocks (as a result of the work proposed in Aim 1), we anticipate that the preparation of these analogs can be accomplished in a relatively short period of time.

Very similar to that outlined in Aim 1 for the study of insulin, after the synthesis of the GLP-1 library, the vulnerability of each library member to digestion by the many proteases present in the small intestine, including DPP IV, pepsin, trypsin, chymotrypsin and elastase, will be determined by comparing its half-life with that of the unglycosylated, wild-type GLP-1. Again, the half-lives will be measured by monitoring the first-order exponential decay of the full-length molecules.[48,64]

The most protease-resistant peptide in the world is not clinically relevant if it cannot bind its native receptor and initiate signaling. It is thus critical to verify that the glycoengineering-based modifications found to increase stability do not interfere unacceptably with the biological activity of GLP-1. GLP-1 exerts its function mainly by binding to and activating its native cell surface receptor: GLP-1 receptor (GLP-1R).[166] Based on previous studies, we will use HEK293 cells that stably express GLP-1R and carry a cAMP response element (CRE)-driven luciferase reporter gene construct to characterize the initial activation of GLP-1R by GLP-1 glyco-variants.[167] Briefly, transformed HEK293 cells will be seeded and cultured in multi-well plates. The cells will then be treated with GLP-1 glyco-variants at several different doses for 24 hours, and the GLP-1R activation will be quantified through luminescence intensity using One-Glo™ (Promega) luciferase reagent according to the manufacturer's instructions. Plotting response against concentration will yield a curve from which the EC50 of each analog can be determined using Origin software.[168]

Activation of GLP-1R leads to activation of several downstream pathways, culminating in the coordinated expression and secretion of insulin. Glycosylation may change the ability of GLP-1 to induce the secretion of insulin. This will be tested using a mouse pancreatic β-cell line: MIN6. Based on previous work, the MIN6 cells will be cultured and treated with GLP-1 glyco-variants along with 10 mM glucose for different lengths of time. Supernatant from the cell culture will then be collected and analyzed for insulin concentration using a commercially available mouse insulin ELISA kit.[163]

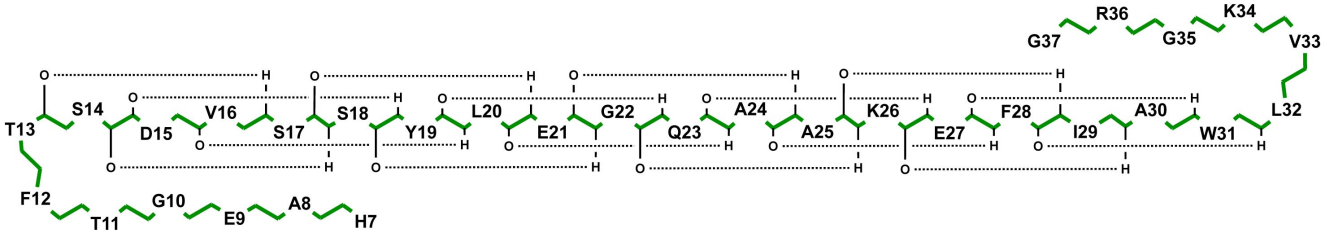


Figure 8. Schematic representation of the hydrogen bonding networks for unglycosylated human GLP-1. The peptide chain is shown in green. Hydrogen bonds are shown with dashed lines.

In addition to stimulating the secretion of insulin, GLP-1 receptor signaling can also lead to therapeutically beneficial outcomes like the proliferation pancreatic β-cells.[169] Based on earlier work, the ability of each of our GLP-1 glyco-variants to induce proliferation will also be determined using the MIN6 cells.[163] Briefly, cells will be cultured and plated in flat-bottomed multi-well plates. They will then be treated with various concentrations of each GLP-1 analog and incubated for 2 days. A commercially available cell-counting kit will be used to dye the cells according to manufacturer's instructions and the cells will be quantified by measuring the optical density at 450nm (OD450) with a plate reader. Comparing OD450 values for each concentration to an untreated control and the wild-type GLP-1 will allow us to correlate the glycosylation pattern with ability to induce β-cell proliferation.

Completing the research proposed above will produce a large amount of data relating different glycan sizes and structures to functional performance of GLP-1. By systematically analyzing these data, we will identify GLP-1 glyco-variants with the highest stability and biologically activities for further study.

In order to better understand how glycosylation can either enhance or diminish properties of GLP-1 important for oral delivery, we will examine the structural differences of representative GLP-1 variants. This is similar to the structural research proposed for insulin. We anticipate that comparing NMR structures and conformational dynamics of many GLP-1 glyco-variants with that of the unglycosylated peptide (Fig. 8) will help us to elucidate the molecular basis for effects we observe upon *O*-glycosylation. Again, similar to the studies

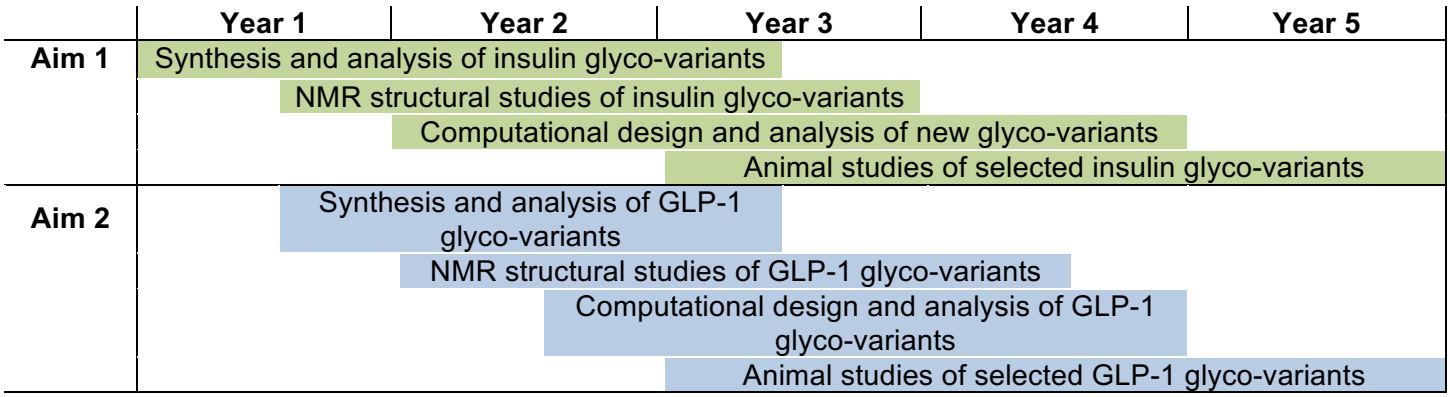
described for insulin, the obtained wealth of structural information will be combined with *in vitro* assay results to serve as the basis for developing computational tools that can predict physical properties and biological function of GLP-1. This is especially relevant for analogs that are difficult to access by synthetic chemistry like those with large *O*-linked glycans at both Thr11 and Ser14 sites and/or mutations at the *N*-terminus of GLP-1. Once computational analysis of the analogs is completed, we will use chemical synthesis to prepare and characterize the small number of candidates that exhibit the most desirable properties.

Evaluate better GLP-1 analogs in animal models. GLP-1 analogs that are found to be both more resistant to proteases and have comparable or better *in vitro* activity to that of wild type GLP-1 will be further characterized in rats. The *in vivo* biological activity and oral availability of selected GLP-1 glyco-variants will be tested in rats in much the same way as insulin variants. Briefly, GLP-1 glyco-variants in pH-sensitive, gelatin-based capsules will be orally fed to SD rats, and blood samples will be collected at different time intervals. A rat insulin radioimmunoassay (RIA) kit will be used to measure insulin levels in the blood samples.[170,171] Concentration of insulin in the blood serum over time will allow us to quantify each analog's *in vivo* activity. At the same time, concentration of GLP-1 glyco-variants in the blood samples over time will allow us to quantify each analog's rate of absorption and rate of clearance. As in Aim1, we will also examine explicitly the biodistribution of GLP-1 glyco-variants in rats after oral administration. This will be done with ^{125}I-GLP-1 glyco-variants and CT scanning technology. High-resolution images obtained from these experiments will allow us to visually monitor the absorption/distribution/clearance process for each glyco-variant of GLP-1 and understand how glycosylation patterns of the peptide might influence each step.

C2d) Expected outcomes, potential problems, and alternative approaches. By synthesizing and systematically comparing the properties of structurally well-defined GLP-1 glycoforms, the proposed research is expected to lead to the identification GLP-1 glyco-variants with enhanced oral bioavailability. Moreover, the studies will provide new information about the ability of glycans to adjust the properties of GLP-1, which could be applied to glycoengineering of other GLP-1R agonists like Exenatide, Lixisenatide, and Liraglutide.[166]

All GLP-1 glycoforms proposed in this Aim will be synthesized by attaching different glycans to Thr11 and Ser14, and by changing the amino acid sequences of the *N*-terminus. It is possible that glyco-engineering of this single region will be able to improve the stability of GLP-1 against DPP IV, but might be insufficient to lead to analogs that are stable towards all tested proteolytic enzymes. Additionally, it might result in analogs that do not retain sufficient levels of biological activity. If such problems arise, the amino acid sequence in the middle or at the *C*-terminus of GLP-1 will be systematically varied and/or glycosylated to determine if amino acid mutations and glycosylation in these regions have additive effects on its stability and biological function.[163,166]

TIMELINE.



FUTURE DIRECTION.

In the future, we will further investigate the effects of glycosylation on the in vivo behavior, receptor binding specificity[172] *and immunogenicity*[173] *of insulin and GLP-1 variants*. Anti-drug antibodies may be a problem after sustained use of any biologic therapy. Future work to measure the anti-drug antibody titer in animals for particularly promising insulin and GLP-1 analogs would therefore be of great interest. Given the rising popularity of peptide-based therapeutics in recent years, there are numerous well-documented protocols for these types of studies.[174] *We will also develop better technologies for large-scale and cost effective production of promising candidates for clinical assessment.*[175] One such scale-up technology we are particularly interested in exploring is chemo-enzymatic synthesis of specific glycoforms. Moreover, in *the long term*, we will apply the lessons derived from the proposed studies to the glycoengineering of other interesting therapeutic peptides, with the goal of addressing their possible therapeutic challenges.[9]

Vertebrate Animals

Approved Protocol Summary (IACUC Office, University of Colorado Boulder):
Title: Study of *in vivo* bioavailability and stability of glycosylated insulin and GLP-1 peptides
Protocol Number: 2538
Protocol Approved Date: 5/2/2017
Protocol Expiration Date: 5/2/2020

Description of Procedures:
Fifty pathogen-free, adult Sprague–Dawley (SD) rats (200-250g) will be used to test the stability and bioavailability of promising insulin glyco-variants in the small intestine. SD rats will be housed at a controlled temperature (23 ± 3 °C) and in a 12 h day–night cycle. Diabetes will be induced by injection of a single dose of 65 mg/kg streptozotocin (STZ) in 0.01 M citrate buffer (pH = 4.5). Rats displaying symptoms of diabetes mellitus (increased thirst, frequent urination, weight loss, and increased blood glucose level) will be selected for the experiments. They will be randomly grouped into several experimental groups, with five rats per group. Rats in each group will be used to characterize the properties of one identified insulin variant. Prior to experiments, the STZ diabetic rats will be fasted for 12 h but will have free access to water. Insulin glyco-variants (1.8 mg/kg body weight) in HP55-coated gelatin capsules will be orally administrated to diabetic rats. Blood samples will be collected from the tail veins of the rats before drug administration and at eleven time intervals (0, 0.5, 1, 2, 3, 4, 6, 8, 12, 18 and 24 h) after insulin feeding. The blood glucose level of each sample will be determined using a glucometer.

Fifty pathogen-free, adult SD rats (200-250g) will be used to examine the stability and bioavailability of GLP-1 glyco-variants in the small intestine. The SD rats will be randomly grouped into ten groups (five in each group) to test up to 8 GLP-1 variants. GLP-1 molecules (1.0 mg/kg body weight) in HP55-coated gelatin capsules will be orally administrated to diabetic rats. The blood samples will be collected before drug administration and at different time intervals (0, 0.5, 1, 2, 3, 4, 6, 8, 12, 18 and 24 h) after GLP-1 feeding. The insulin level of each blood sample will be measured using a rat insulin radioimmunoassay (RIA) kit.

For each study a power analysis was carried out based on expected effect sizes, from previous similar studies, to ensure statistically meaningful results could be achieved with the proposed number of animals.

Justifications:
One objective of the proposed research is to identify insulin and GLP-1 glyco-variants with improved properties for oral administration. Although there are many *in vitro* assays that can be used to study the stability, aggregation propensity, and the capability of stimulating GLUT4 to be expressed on the plasma membrane, currently, there is no technique that is sophisticated enough to accurately mimic each step in the process of oral drug absorption, including proteolytic degradation, small intestine absorption, portal circulation, hepatic glucose production, systemic circulation, *et al*. Therefore, in order to identify insulin and GLP-1 variants with better properties for future development of oral drugs, it is necessary to use animal models to test the promising candidates. In diabetes research, one of the most frequently used animal models is rats because they are relatively easy to handle, have relatively large volume of blood that can allow the measurement of a range of biochemical markers, and to some extent, respond similarly to insulin and GLP-1 as humans. Based on all the above-mentioned reasons, we choose to use SD rats to test the promising insulin and GLP-1 glyco-variants.

Minimization of Pain and Distress:
For each of the injections and blood sample collections, rats will be restrained and the operation will take place over the course of 20s-30s. There is minimal discomfort and anesthesia/analgesia is not required. All procedures will be carried out following the guidelines of the Institutional Animal Care and Use Committee (IACAC) of the University of Colorado Boulder to ensure handling of animals is in accordance with the NIH guidelines for humane use.

Euthanasia:
After the experiments, all rats will be euthanized in full accordance with AVMA guidelines. Rats will be heavily anesthetized under barbiturates and decapitated. Corpses will be disposed of according to the guidelines of the IACAC of the University of Colorado Boulder.

References Cited:

1. Sable R, Parajuli P, Jois S (2017) Peptides, peptidomimetics, and polypeptides from marine sources: A wealth of natural sources for pharmaceutical applications. *Mar Drugs* 15(4):124.
2. Hancock RE, Haney EF, Gill EE (2016) The immunology of host defence peptides: beyond antimicrobial activity. *Nat Rev Immunol* 16(5):321-334.
3. Krumm BE, Grisshammer R (2015) Peptide ligand recognition by G protein-coupled receptors. *Front Pharmacol* 6:48.
4. Beraud E, Chandy KG (2011) Therapeutic potential of peptide toxins that target ion channels. *Inflamm Allergy Drug Targets* 10(5):322-342.
5. Baldo BA (2016) Peptide hormones. *Safety of Biologics Therapy: Monoclonal Antibodies, Cytokines, Fusion Proteins, Hormones, Enzymes, Coagulation Proteins, Vaccines, Botulinum Toxins*, (Springer International Publishing, Cham), pp 309-400.
6. van den Pol AN (2012) Neuropeptide transmission in brain circuits. *Neuron* 76(1):98-115.
7. Lim DW, Wales PW, Turner JM, Bigam DL, Brubaker PL (2016) On the horizon: trophic peptide growth factors as therapy for neonatal short bowel syndrome. *Expert Opin Ther Targets* 20(7):819-830.
8. Buchanan A, Revell JD (2015) Novel therapeutic proteins and peptides. *Novel Approaches and Strategies for Biologics, Vaccines and Cancer Therapies*, (Academic Press, San Diego), pp 171-197.
9. Vlieghe P, Lisowski V, Martinez J, Khrestchatisky M (2010) Synthetic therapeutic peptides: science and market. *Drug Discov Today* 15(1-2):40-56.
10. Fosgerau K, Hoffmann T (2015) Peptide therapeutics: current status and future directions. *Drug Discov Today* 20(1):122-128.
11. Kaspar AA, Reichert JM (2013) Future directions for peptide therapeutics development. *Drug Discov Today* 18(17-18):807-817.
12. Craik DJ, Fairlie DP, Liras S, Price D (2013) The future of peptide-based drugs. *Chem Biol Drug Des* 81(1):136-147.
13. Mathur D, Prakash S, Anand P, Kaur H, Agrawal P, Mehta A, Kumar R, Singh S, Raghava GP (2016) PEPlife: A Repository of the Half-life of Peptides. *Sci Rep* 6:36617.
14. Cox D, Stone J (2006) Managing self-injection difficulties in patients with relapsing-remitting multiple sclerosis. *J Neurosci Nurs* 38(3):167-171.
15. Lax R (2013) The future of peptide development in the pharmaceutical industry. *PharManufacturing: Int Peptide Rev* 10–15.
16. Bruno BJ, Miller GD, Lim CS (2013) Basics and recent advances in peptide and protein drug delivery. *Ther Deliv* 4(11):1443-1467.
17. Muheem A, Shakeel F, Jahangir MA, Anwar M, Mallick N, Jain GK, Warsi MH, Ahmad FJ (2016) A review on the strategies for oral delivery of proteins and peptides and their clinical perspectives. *Saudi Pharm J* 24(4):413-428.
18. Turecek PL, Bossard MJ, Schoetens F, Ivens IA (2016) PEGylation of Biopharmaceuticals: A Review of Chemistry and Nonclinical Safety Information of Approved Drugs. *J Pharm Sci* 105(2):460-475.
19. Schellenberger V, Wang CW, Geething NC, Spink BJ, Campbell A, To W, Scholle MD, Yin Y, Yao Y, Bogin O, Cleland JL, Silverman J, Stemmer WP (2009) A recombinant polypeptide extends the in vivo half-life of peptides and proteins in a tunable manner. *Nat Biotechnol* 27(12):1186-1190.
20. Liu Y, Lee J, Mansfield KM, Ko JH, Sallam S, Wesdemiotis C, Maynard HD (2017) Trehalose glycopolymer enhances both solution sability and pharmacokinetics of a therapeutic protein. *Bioconjug Chem* 28(3):836-845.
21. Wu B, Sun YN (2014) Pharmacokinetics of Peptide-Fc fusion proteins. *J Pharm Sci* 103(1):53-64.
22. Di L (2015) Strategic approaches to optimizing peptide ADME properties. *AAPS J* 17(1):134-143.
23. Karsdal MA, Riis BJ, Mehta N, Stern W, Arbit E, Christiansen C, Henriksen K (2015) Lessons learned from the clinical development of oral peptides. *Br J Clin Pharmacol* 79(5):720-732.
24. Jitendra, Sharma PK, Bansal S, Banik A (2011) Noninvasive routes of proteins and peptides drug delivery. *Indian J Pharm Sci* 73(4):367-375.
25. Elliott S*, et al.* (2003) Enhancement of therapeutic protein in vivo activities through glycoengineering. *Nat Biotechnol* 21(4):414-421.

26. Sinclair AM, Elliott S (2005) Glycoengineering: the effect of glycosylation on the properties of therapeutic proteins. *J Pharm Sci* 94(8):1626-1635.
27. Sola RJ, Griebenow K (2009) Effects of glycosylation on the stability of protein pharmaceuticals. *J Pharm Sci* 98(4):1223-1245.
28. Yuzwa SA, Shan X, Macauley MS, Clark T, Skorobogatko Y, Vosseller K, Vocadlo DJ (2012) Increasing O-GlcNAc slows neurodegeneration and stabilizes tau against aggregation. *Nat Chem Biol* 8(4):393-399.
29. Frenkel-Pinter M, Richman M, Belostozky A, Abu-Mokh A, Gazit E, Rahimipour S, Segal D (2016) Selective inhibition of aggregation and toxicity of a Tau-derived peptide using its glycosylated analogues. *Chem Eur J* 22(17):5945-5952.
30. Fournillier A, Wychowski C, Boucreux D, Baumert TF, Meunier JC, Jacobs D, Muguet S, Depla E, Inchauspe G (2001) Induction of hepatitis C virus E1 envelope protein-specific immune response can be enhanced by mutation of N-glycosylation sites. *J Virol* 75(24):12088-12097.
31. Wei X*, et al.* (2003) Antibody neutralization and escape by HIV-1. *Nature* 422(6929):307-312.
32. Helle F, Vieyres G, Elkrief L, Popescu CI, Wychowski C, Descamps V, Castelain S, Roingeard P, Duverlie G, Dubuisson J (2010) Role of N-linked glycans in the functions of hepatitis C virus envelope proteins incorporated into infectious virions. *J Virol* 84(22):11905-11915.
33. Peyrot M, Rubin RR, Kruger DF, Travis LB (2010) Correlates of insulin injection omission. *Diabetes Care* 33(2):240-245.
34. Pawar VK, Meher JG, Singh Y, Chaurasia M, Surendar Reddy B, Chourasia MK (2014) Targeting of gastrointestinal tract for amended delivery of protein/peptide therapeutics: strategies and industrial perspectives. *J Control Release* 196:168-183.
35. Dicker M, Strasser R (2015) Using glyco-engineering to produce therapeutic proteins. *Expert Opin Biol Ther* 15(10):1501-1516.
36. Owens DR (2002) New horizons--alternative routes for insulin therapy. *Nat Rev Drug Discov* 1(7):529-540.
37. Garber AJ (2011) Long-acting glucagon-like peptide 1 receptor agonists: a review of their efficacy and tolerability. *Diabetes Care* 34 Suppl 2:S279-284.
38. Hua Q (2010) Insulin: a small protein with a long journey. *Protein & cell* 1(6):537-551.
39. Underwood CR, Garibay P, Knudsen LB, Hastrup S, Peters GH, Rudolph R, Reedtz-Runge S (2010) Crystal Structure of Glucagon-like Peptide-1 in Complex with the Extracellular Domain of the Glucagon-like Peptide-1 Receptor. *J Biol Chem* 285(1):723-730.
40. Garcia-Perez LE, Alvarez M, Dilla T, Gil-Guillen V, Orozco-Beltran D (2013) Adherence to therapies in patients with type 2 diabetes. *Diabetes Ther* 4(2):175-194.
41. Arbit E, Kidron M (2009) Oral insulin: the rationale for this approach and current developments. *J Diabetes Sci Technol* 3(3):562-567.
42. Balamuralidhara V, Pramodkumar TM, Srujana N, Venkatesh MP, Gupta NV, K.L. K, Gangadharappa HV (2011) pH Sensitive drug delivery systems: A review. *Am J Drug Discovery Dev* 1(24-28).
43. Kalra S, Kalra B, Agrawal N (2010) Oral insulin. *Diabetol Metab Syndr* 2:66.
44. Elsayed AM (2012) *Oral delivery of insulin: Novel approaches* (INTECH Open Access Publisher, InTech) pp 282-314.
45. Yin N, Brimble MA, Harris PW, Wen J (2014) Enhancing the oral bioavailability of peptide drugs by using chemical modification and other approaches. *Med Chem* 4(12):763-769.
46. Renukuntla J, Vadlapudi AD, Patel A, Boddu SH, Mitra AK (2013) Approaches for enhancing oral bioavailability of peptides and proteins. *Int J Pharm* 447(1-2):75-93.
47. Choonara BF, Choonara YE, Kumar P, Bijukumar D, du Toit LC, Pillay V (2014) A review of advanced oral drug delivery technologies facilitating the protection and absorption of protein and peptide molecules. *Biotechnol Adv* 32(7):1269-1282.
48. Guan X, Chaffey PK, Zeng C, Greene ER, Chen L, Drake MR, Chen C, Groobman A, Resch MG, Himmel ME, Beckham GT, Tan Z (2015) Molecular-scale features that govern the effects of O-glycosylation on a carbohydrate-binding module. *Chem Sci* 6:7185 - 7189.
49. Guan X, Chaffey PK, Wei X, Gulbranson DR, Ruan Y, Wang X, Ouyang Y, Chen L, Zeng C, Koelsch TN, Tran AH, Liang W, Shen J, Tan Z (2017) Chemically precise glycoengineering improves human insulin. *ACS Chem Biol*:in press.

50. Sanger F (1988) Sequences, sequences, and sequences. *Annu Rev Biochem* 57:1-28.
51. Du YC, Zhang YS, Lu ZX, Tsou CL (1961) Resynthesis of insulin from its glycyl and phenylalanyl chains. *Scientia Sinica* 10:84-104.
52. Goeddel DV, Kleid DG, Bolivar F, Heyneker HL, Yansura DG, Crea R, Hirose T, Kraszewski A, Itakura K, Riggs AD (1979) Expression in Escherichia coli of chemically synthesized genes for human insulin. *Proc Natl Acad Sci U S A* 76(1):106-110.
53. Blundell TL, Cutfield JF, Dodson EJ, Dodson GG, Hodgkin DC, Mercola DA (1972) The crystal structure of rhombohedral 2 zinc insulin. *Cold Spring Harb Symp Quant Biol* 36:233-241.
54. Berenson DF, Weiss AR, Wan ZL, Weiss MA (2011) Insulin analogs for the treatment of diabetes mellitus: therapeutic applications of protein engineering. *Ann N Y Acad Sci* 1243:E40-E54.
55. Yabe D, Seino Y (2014) Defining the role of GLP-1 receptor agonists for individualized treatment of Type 2 diabetes. *Expert Rev Endocrinol Metab* 9(6):659-670.
56. Ryan D, Acosta A (2015) GLP-1 receptor agonists: Nonglycemic clinical effects in weight loss and beyond. *Obesity* 23(6):1119-1129.
57. Mirza RA, Turiansky GW (2012) Enfuvirtide and cutaneous injection-site reactions. *J Drugs Dermatol* 11(10):e35-38.
58. Poursina N, Vatanara A, Rouini MR, Gilani K, Najafabadi AR (2016) The effect of excipients on the stability and aerosol performance of salmon calcitonin dry powder inhalers prepared via the spray freeze drying process. *Acta Pharm* 66(2):207-218.
59. Alters SE, McLaughlin B, Spink B, Lachinyan T, Wang CW, Podust V, Schellenberger V, Stemmer WP (2013) GLP2-2G-XTEN: a pharmaceutical protein with improved serum half-life and efficacy in a rat Crohn's. *PLoS One* 8(6):e50630.
60. Lau YH, de Andrade P, Wu Y, Spring DR (2015) Peptide stapling techniques based on different macrocyclisation chemistries. *Chem Soc Rev* 44(1):91-102.
61. Hamley IW (2014) PEG−peptide conjugates. *Biomacromolecules* 15(5):1543-1559.
62. Galán A, Comor L, Horvatić A, Kuleš J, Guillemin N, Mrljak V, Bhide M (2016) Library-based display technologies: where do we stand? *Mol Biosyst* 12(8):2342-2358.
63. Hanson SR, Culyba EK, Hsu TL, Wong CH, Kelly JW, Powers ET (2009) The core trisaccharide of an N-linked glycoprotein intrinsically accelerates folding and enhances stability. *Proc Natl Acad Sci U S A* 106(9):3131-3136.
64. Chen L, Drake MR, Resch MG, Greene ER, Himmel ME, Chaffey PK, Beckham GT, Tan Z (2014) Specificity of *O*-glycosylation in enhancing the stability and cellulose binding affinity of Family 1 carbohydrate-binding modules. *Proc Natl Acad Sci U S A* 111(21):7612-7617.
65. Spiro RG (2002) Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology* 12(4):43R-56R.
66. Chaffey PK, Chi L, Tan Z (2017) Chemical biology of protein O-glycosylation. *Chemical Biology of Glycoproteins*, (The Royal Society of Chemistry), pp 48-93.
67. Bertozzi CR, Kiessling LL (2001) Chemical glycobiology. *Science* 291(5512):2357-2364.
68. Guo Z, Shao N (2005) Glycopeptide and glycoprotein synthesis involving unprotected carbohydrate building blocks. *Med Res Rev* 25(6):655-678.
69. Durocher Y, Butler M (2009) Expression systems for therapeutic glycoprotein production. *Curr Opin Biotechnol* 20(6):700-707.
70. Wang LX, Amin MN (2014) Chemical and chemoenzymatic synthesis of glycoproteins for deciphering functions. *Chem Biol Drug Des* 21(1):51-66.
71. Gamblin DP, Scanlan EM, Davis BG (2009) Glycoprotein synthesis: an update. *Chem Rev* 109(1):131-163.
72. Chen W, Enck S, Price JL, Powers DL, Powers ET, Wong CH, Dyson HJ, Kelly JW (2013) Structural and energetic basis of carbohydrate-aromatic packing interactions in proteins. *J Am Chem Soc* 135(26):9877-9884.
73. Kiessling LL, Splain RA (2010) Chemical approaches to glycobiology. *Annu Rev Biochem* 79:619-653.
74. Yoshida K, Yang B, Yang W, Zhang Z, Zhang J, Huang X (2014) Chemical synthesis of syndecan-3 glycopeptides bearing two heparan sulfate glycan chains. *Angew Chem Int Ed Engl* 53(34):9051-9058.
75. Tan Z, Wang LX (2017) *Chemical Biology of Glycoproteins* (Royal Society of Chemistry).

76. Shental-Bechor D, Levy Y (2009) Folding of glycoproteins: toward understanding the biophysics of the glycosylation code. *Curr Opin Struct Biol* 19(5):524-533.
77. Childers MC, Daggett V (2017) Insights from molecular dynamics simulations for computational protein design. *Mol Syst Des Eng* 2(1):9-33.
78. Tsai CS (2003) *An Introduction to Computational Biochemistry* (John Wiley & Sons, Inc.).
79. Happs RM, Guan X, Resch MG, Davis MF, Beckham GT, Tan Z, Crowley MF (2015) O-glycosylation effects on family 1 carbohydrate-binding module solution structures. *FEBS J* 282(22):4341-4356.
80. Chaffey PK, Guan X, Chen C, Ruan Y, Wang X, Tran AH, Koelsch TN, Cui Q, Feng Y, Tan Z (2017) Structural insight into the stabilizing effect of O-glycosylation. *Biochemistry* 56(23):2897-2906.
81. Guan X, Chaffey PK, Ruan Y, Wang X, Greene ER, Shoemaker RK, Tran AH, Koelsch TN, Hurd CK, Himmel ME, Beckham GT, Long H, Tan Z (2017) Carbohydrate binding module O-mannosylation alters binding selectivity to cellulose and lignin. *submitted*.
82. Price JL, Culyba EK, Chen W, Murray AN, Hanson SR, Wong CH, Powers ET, Kelly JW (2012) N-glycosylation of enhanced aromatic sequons to increase glycoprotein stability. *Biopolymers* 98(3):195-211.
83. Tong Q*, et al.* (2015) An acetyl-methyl switch drives a conformational change in p53. *Structure* 23(2):322-331.
84. Boutureira O, Bernardes GJ (2015) Advances in chemical protein modification. *Chem Rev* 115(5):2174-2195.
85. Goettig P (2016) Effects of glycosylation on the enzymatic activity and mechanisms of proteases. *Int J Mol Sci* 17(12):1969.
86. Imberty A, Gautier C, Lescar J, Perez S, Wyns L, Loris R (2000) An unusual carbohydrate binding site revealed by the structures of two Maackia amurensis lectins complexed with sialic acid-containing oligosaccharides. *J Biol Chem* 275(23):17541-17548.
87. Chen W, Kong L, Connelly S, Dendle JM, Liu Y, Wilson IA, Powers ET, Kelly JW (2016) Stabilizing the CH2 domain of an antibody by engineering in an enhanced aromatic sequon. *ACS Chem Biol* 11(7):1852-1861.
88. Karpusas M, Whitty A, Runkel L, Hochman P (1998) The structure of human interferon-β: implications for activity. *Cell Mol Life Sci* 54(11):1203-1216.
89. Lawrence PB*, et al.* (2014) Criteria for selecting PEGylation sites on proteins for higher thermodynamic and proteolytic stability. *J Am Chem Soc* 136(50):17547-17560.
90. Imperiali B, O'Connor SE (1999) Effect of N-linked glycosylation on glycopeptide and glycoprotein structure. *Curr Opin Chem Biol* 3(6):643-649.
91. Belgi A, Hossain MA, Tregear GW, Wade JD (2011) The chemical synthesis of insulin: from past to present. *Immunol Endocr Metab Agents Med Chem* 11(1):40-47.
92. Liu F, Luo EY, Flora DB, Mezo AR (2014) A synthetic route to human insulin using isoacyl peptides. *Angew Chem Int Ed Engl* 53(15):3983-3987.
93. Schilling RJ, Mitra AK (1991) Degradation of insulin by trypsin and alpha-chymotrypsin. *Pharm Res* 8(6):721-727.
94. Ciszak E, Beals JM, Frank BH, Baker JC, Carter ND, Smith GD (1995) Role of C-terminal B-chain residues in insulin assembly: the structure of hexameric LysB28ProB29-human insulin. *Structure* 3(6):615-622.
95. Wong CY, Martinez J, Dass CR (2016) Oral delivery of insulin for treatment of diabetes: status quo, challenges and opportunities. *J Pharm Pharmacol* 68(9):1093-1108.
96. Berkowitz SA (2006) Role of analytical ultracentrifugation in assessing the aggregation of protein biopharmaceuticals. *AAPS J* 8(3):E590-605.
97. Vinther TN, Norrman M, Strauss HM, Huus K, Schlein M, Pedersen TA, Kjeldsen T, Jensen KJ, Hubalek F (2012) Novel covalently linked insulin dimer engineered to investigate the function of insulin dimerization. *PLoS One* 7(2):e30882.
98. Govers R, Coster AC, James DE (2004) Insulin increases cell surface GLUT4 levels by dose dependently discharging GLUT4 into a cell surface recycling pathway. *Mol Cell Biol* 24(14):6456-6466.
99. van Dam EM, Govers R, James DE (2005) Akt activation is required at a late stage of insulin-induced GLUT4 translocation to the plasma membrane. *Mol Endocrinol* 19(4):1067-1077.

100. Mirmira RG, Nakagawa SH, Tager HS (1991) Importance of the character and configuration of residues B24, B25, and B26 in insulin-receptor interactions. *J Biol Chem* 266(3):1428-1436.
101. Shinomiya M, Shirai K, Saito Y, Yoshida S, Matsuoka N (1985) Effect of new chymotrypsin inhibitor (FK-448) on intestinal absorption of insulin. *Lancet* 1(8437):1092-1093.
102. Liu J, Andya JD, Shire SJ (2006) A critical review of analytical ultracentrifugation and field flow fractionation methods for measuring protein aggregation. *AAPS J* 8(3):E580-589.
103. Qulocho FA (1989) Protein-carbohydrate interactions: Basic molecular features. *Pure Appl Chem* 61:1293–1306.
104. Uchio T, Baudys M, Liu F, Song SC, Kim SW (1999) Site-specific insulin conjugates with enhanced stability and extended action profile. *Adv Drug Deliv Rev* 35(2-3):289-306.
105. Gribben JG, Devereux S, Thomas NS, Keim M, Jones HM, Goldstone AH, Linch DC (1990) Development of antibodies to unprotected glycosylation sites on recombinant human GM-CSF. *Lancet* 335(8687):434-437.
106. Ragnhammar P, Friesen HJ, Frodin JE, Lefvert AK, Hassan M, Osterborg A, Mellstedt H (1994) Induction of anti-recombinant human granulocyte-macrophage colony-stimulating factor (Escherichia coli-derived) antibodies and clinical effects in nonimmunocompromised patients. *Blood* 84(12):4078-4087.
107. Elofsson M, Roy S, Salvador LA, Kihlberg J (1996) Building blocks for glycopeptide synthesis: Preparation of α-O-fucosylated fmoc serine and threonine in one step from L-fucose tetraacetate. *Tetrahedron Lett* 37(42):7645-7648.
108. Uchiro H, Kurusu N, Mukaiyama T (1997) A convenient method for the preparation of disaccharides by transglycosylation of methyl glycosides. *Isr J Chem* 37(1):87-96.
109. Crich D, Wu B (2006) 1-naphthylpropargyl ether group: a readily cleaved and sterically minimal protecting system for stereoselective glycosylation. *Org Lett* 8(21):4879-4882.
110. Guo Z, Wang Q (2009) Recent development in carbohydrate-based cancer vaccines. *Curr Opin Chem Biol* 13(5-6):608-617.
111. Hiruma-Shimizu K, Hosoguchi K, Liu Y, Fujitani N, Ohta T, Hinou H, Matsushita T, Shimizu H, Feizi T, Nishimura S-I (2010) Chemical synthesis, folding, and structural insights into O-fucosylated epidermal growth factor-like repeat 12 of mouse Notch-1 receptor. *J Am Chem Soc* 132(42):14857-14865.
112. Keyari CM, Polt R (2010) Serine and threonine Schiff base ssters react with β-anomeric peracetates in the presence of BF3·Et2O to Produce β-glycosides. *J Carbohydr Chem* 29(4):181-206.
113. Yao N, Fung G, Malekan H, Ye L, Kurth MJ, Lam KS (2010) Facile synthesis of glycosylated Fmoc amino acid building blocks assisted by microwave irradiation. *Carbohydr Res* 345(15):2277-2281.
114. Talat S, Thiruvikraman M, Kumari S, Kaur KJ (2011) Glycosylated analogs of formaecin I and drosocin exhibit differential pattern of antibacterial activity. *Glycoconj J* 28(8-9):537-555.
115. Dulaney SB, Huang X (2012) Strategies in synthesis of heparin/heparan sulfate oligosaccharides: 2000-present. *Adv Carbohydr Chem Biochem* 67:95-136.
116. Bohe L, Crich D (2015) A propos of glycosyl cations and the mechanism of chemical glycosylation: the current state of the art. *Carbohydr Res* 403:48-59.
117. Das R, Mukhopadhyay B (2016) Chemical O-glycosylations: An overview. *ChemistryOpen* 5(5):401-433.
118. Dou H, Zhang M, Zhang Y, Yin C (2007) Synthesis and purification of mono-PEGylated insulin. *Chem Biol Drug Des* 69(2):132-138.
119. Yang C, Lu D, Liu Z (2011) How PEGylation enhances the stability and potency of insulin: a molecular dynamics simulation. *Biochemistry* 50(13):2585-2593.
120. Caparrotta TM, Evans M (2014) PEGylated insulin Lispro, (LY2605541) - a new basal insulin analogue. *Diabetes Obes Metab* 16(5):388-395.
121. Hudak JE, Canham SM, Bertozzi CR (2014) Glycocalyx engineering reveals a Siglec-based mechanism for NK cell immunoevasion. *Nat Chem Biol* 10(1):69-75.
122. Bork K, Horstkorte R, Weidemann W (2009) Increasing the sialylation of therapeutic glycoproteins: the potential of the sialic acid biosynthetic pathway. *J Pharm Sci* 98(10):3499-3508.
123. Chen J, Chen G, Wu B, Wan Q, Tan Z, Hua Z, Danishefsky SJ (2006) Mature homogeneous erythropoietin-level building blocks by chemical synthesis: the EPO 114-166 glycopeptide domain, presenting the O-linked glycophorin. *Tetrahedron Lett* 47(46):8013-8016.

124. Wu B, Hua Z, Warren JD, Ranganathan K, Wan Q, Chen G, Tan Z, Chen J, Endo A, Danishefsky SJ (2006) Synthesis of the fucosylated biantennary N-glycan of erythropoietin. *Tetrahedron Lett* 47:5577-5579.
125. Tan Z, Shang S, Halkina T, Yuan Y, Danishefsky SJ (2009) Toward homogeneous erythropoietin: Non-NCL-based chemical synthesis of the Gln(78)-Arg(166) glycopeptide domain. *J Am Chem Soc* 131(15):5424-5431.
126. Seifert J, Ogawa T, Kurono S, Ito Y (2000) Syntheses of alpha-dystroglycan derived glycosyl amino acids carrying a novel mannosyl serine/threonine linkage. *Glycoconj J* 17(6):407-423.
127. Benito-Alifonso D, Jones RA, Tran AT, Woodward H, Smith N, Galan MC (2013) Synthesis of mucin-type O-glycan probes as aminopropyl glycosides. *Beilstein J Org Chem* 9:1867-1872.
128. Tams JW, Vind J, Welinder KG (1999) Adapting protein solubility by glycosylation. N-glycosylation mutants of Coprinus cinereus peroxidase in salt and organic solutions. *Biochim Biophys Acta* 1432(2):214-221.
129. Mills JE, Dean PM (1996) Three-dimensional hydrogen-bond geometry and probability information from a crystal survey. *J Comput Aided Mol Des* 10(6):607-622.
130. Bocian W, Sitkowski J, Bednarek E, Tarnowska A, Kawecki R, Kozerski L (2008) Structure of human insulin monomer in water/acetonitrile solution. *J Biomol NMR* 40(1):55-64.
131. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6(3):277-293.
132. Johnson BA, Blevins RA (1994) NMR View: A computer program for the visualization and analysis of NMR data. *J Biomol NMR* 4(5):603-614.
133. Petersen BO, Vinogradov E, Kay W, Wurtz P, Nyberg NT, Duus JO, Sorensen OW (2006) H2BC: a new technique for NMR analysis of complex carbohydrates. *Carbohydr Res* 341(4):550-556.
134. Bhunia A, Domadia PN, Torres J, Hallock KJ, Ramamoorthy A, Bhattacharjya S (2010) NMR structure of pardaxin, a pore-forming antimicrobial peptide, in lipopolysaccharide micelles: mechanism of outer membrane permeabilization. *J Biol Chem* 285(6):3883-3895.
135. Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* 54(Pt 5):905-921.
136. Shen Y, Bax A (2013) Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J Biomol NMR* 56(3):227-241.
137. Duggan BM, Legge GB, Dyson HJ, Wright PE (2001) SANE (Structure Assisted NOE Evaluation): an automated model-based approach for NOE assignment. *J Biomol NMR* 19(4):321-329.
138. Nederveen AJ, Doreleijers JF, Vranken W, Miller Z, Spronk CA, Nabuurs SB, Guntert P, Livny M, Markley JL, Nilges M, Ulrich EL, Kaptein R, Bonvin AM (2005) RECOORD: a recalculated coordinate database of 500+ proteins from the PDB using restraints from the BioMagResBank. *Proteins* 59(4):662-672.
139. Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J Biomol NMR* 8(4):477-486.
140. Hooft RW, Vriend G, Sander C, Abola EE (1996) Errors in protein structures. *Nature* 381(6580):272.
141. Redler RL, Shirvanyants D, Dagliyan O, Ding F, Kim DN, Kota P, Proctor EA, Ramachandran S, Tandon A, Dokholyan NV (2014) Computational approaches to understanding protein aggregation in neurodegeneration. *J Mol Cell Biol* 6(2):104-115.
142. Morriss-Andrews A, Shea JE (2015) Computational studies of protein aggregation: methods and applications. *Annu Rev Phys Chem* 66:643-666.
143. Karshikoff A, Nilsson L, Ladenstein R (2015) Rigidity versus flexibility: the dilemma of understanding protein thermal stability. *FEBS J* 282(20):3899-3917.
144. Du X, Li Y, Xia YL, Ai SM, Liang J, Sang P, Ji XL, Liu SQ (2016) Insights into protein-ligand interactions: Mechanisms, models, and methods. *Int J Mol Sci* 17(2).
145. Radibratovic M, Minic S, Stanic-Vucinic D, Nikolic M, Milcic M, Cirkovic Velickovic T (2016) Stabilization of human serum albumin by the bBinding of phycocyanobilin, a bioactive chromophore of blue-green Alga Spirulina: Molecular dynamics and experimental study. *PLoS One* 11(12):e0167973.

146. Bonito CA, Nunes J, Leandro J, Louro F, Leandro P, Ventura FV, Guedes RC (2016) Unveiling the pathogenic molecular mechanisms of the most common variant (p.K329E) in medium-chain Acyl-CoA dehydrogenase deficiency by in Vitro and in silico approaches. *Biochemistry* 55(51):7086-7098.
147. Brubaker WD, Freites JA, Golchert KJ, Shapiro RA, Morikis V, Tobias DJ, Martin RW (2011) Separating instability from aggregation propensity in gammaS-crystallin variants. *Biophys J* 100(2):498-506.
148. Baheri M, Dayer MR (2016) Temperature and pH effects on insulin sructure: A molecular dynamic approach. *Jentashapir J Health Res* 7(4):e36931.
149. Zambrano R, Jamroz M, Szczasiuk A, Pujols J, Kmiecik S, Ventura S (2015) AGGRESCAN3D (A3D): server for prediction of aggregation properties of protein structures. *Nucleic Acids Res* 43(W1):W306-313.
150. Tatulian SA (2015) Structural Dynamics of Insulin Receptor and Transmembrane Signaling. *Biochemistry* 54(36):5523-5532.
151. Vashisth H (2015) Theoretical and computational studies of peptides and receptors of the insulin family. *Membranes (Basel)* 5(1):48-83.
152. Papaioannou A, Kuyucak S, Kuncic Z (2015) Molecular Dynamics Simulations of Insulin: Elucidating the Conformational Changes that Enable Its Binding. *PLoS One* 10(12):e0144058.
153. Wu ZM, Zhou L, Guo XD, Jiang W, Ling L, Qian Y, Luo KQ, Zhang LJ (2012) HP55-coated capsule containing PLGA/RS nanoparticles for oral delivery of insulin. *Int J Pharm* 425(1-2):1-8.
154. Wu ZM, Ling L, Zhou LY, Guo XD, Jiang W, Qian Y, Luo KQ, Zhang LJ (2012) Novel preparation of PLGA/HP55 nanoparticles for oral insulin delivery. *Nanoscale Res Lett* 7(1):299.
155. Sonaje K, Lin KJ, Wey SP, Lin CK, Yeh TH, Nguyen HN, Hsu CW, Yen TC, Juang JH, Sung HW (2010) Biodistribution, pharmacodynamics and pharmacokinetics of insulin analogues in a rat model: Oral delivery using pH-responsive nanoparticles vs. subcutaneous injection. *Biomaterials* 31(26):6849-6858.
156. Furman BL (2015) Streptozotocin-induced diabetic models in mice and rats. *Curr Protoc Pharmacol* 70:5 47 41-20.
157. Cao Y, Gao W, Jusko WJ (2012) Pharmacokinetic/pharmacodynamic modeling of GLP-1 in healthy rats. *Pharm Res* 29(4):1078-1086.
158. Yu H, Lau K, Li Y, Sugiarto G, Chen X (2012) One-pot multienzyme synthesis of Lewis x and sialyl Lewis x antigens. *Curr Protoc Chem Biol* 4:233-247.
159. Marathe CS, Rayner CK, Jones KL, Horowitz M (2011) Effects of GLP-1 and incretin-based therapies on gastrointestinal motor function. *Exp Diabetes Res* 2011:279530.
160. Araujo F, Fonte P, Santos HA, Sarmento B (2012) Oral delivery of glucagon-like peptide-1 and analogs: alternatives for diabetes control? *J Diabetes Sci Technol* 6(6):1486-1497.
161. Kumar A, Bharti SK, Kumar A (2017) Therapeutic molecules against type 2 diabetes: What we have and what are we expecting? *Pharmacol Rep,* 69(5):959-970.
162. Holst JJ (2007) The physiology of glucagon-like peptide 1. *Physiological reviews* 87(4):1409-1439.
163. Xu F, Wang KY, Wang N, Li G, Liu D (2017) Bioactivity of a modified human glucagon-like peptide-1. *PLoS One* 12(2):e0171601.
164. Chang X, Keller D, Bjørn S, Led JJ (2001) Structure and folding of glucagon-like peptide-1-(7–36)-amide in aqueous trifluoroethanol studied by NMR spectroscopy. *Magn Reson Chem* 39(8):477-483.
165. McIntosh CH, Demuth HU, Pospisilik JA, Pederson R (2005) Dipeptidyl peptidase IV inhibitors: how do they work as new antidiabetic agents? *Regul Pept* 128(2):159-165.
166. Lorenz M, Evers A, Wagner M (2013) Recent progress and future options in the development of GLP-1 receptor agonists for the treatment of diabesity. *Bioorg Med Chem Lett* 23(14):4011-4018.
167. Yang PY*, et al.* (2016) Engineering a long-acting, potent GLP-1 analog for microstructure-based transdermal delivery. *Proc Natl Acad Sci U S A* 113(15):4140-4145.
168. Dods Rachel L, Donnelly D (2016) The peptide agonist-binding site of the glucagon-like peptide-1 (GLP-1) receptor based on site-directed mutagenesis and knowledge-based modelling. *Biosci Rep* 36(1):e00285.
169. Buteau J (2008) GLP-1 receptor signaling: effects on pancreatic beta-cell proliferation and survival. *Diabetes Metab* 34 Suppl 2:S73-77.

170. Li Y, Xu W, Tang L, Gong M, Zhang J (2011) A novel GLP-1 analog exhibits potent utility in the treatment of type 2 diabetes with an extended half-life and efficient glucose clearance in vivo. *Peptides* 32(7):1408-1414.
171. Yang X, Li Y, Wang Y, Zheng X, Kong W, Meng F, Zhou Z, Liu C, Li Y, Gong M (2014) Long-acting GLP-1 analogue in V-shaped conformation by terminal polylysine modifications. *Mol Pharm* 11(11):4092-4099.
172. Zhao M, Wan ZL, Whittaker L, Xu B, Phillips NB, Katsoyannis PG, Ismail-Beigi F, Whittaker J, Weiss MA (2009) Design of an insulin analog with enhanced receptor binding selectivity: rationale, structure, and therapeutic implications. *J Biol Chem* 284(46):32178-32187.
173. De Groot AS, Scott DW (2007) Immunogenicity of protein therapeutics. *Trends Immunol* 28(11):482-490.
174. Shankar G, Arkin S, Cocea L, Devanarayan V, Kirshner S, Kromminga A, Quarmby V, Richards S, Schneider CK, Subramanyam M, Swanson S, Verthelyi D, Yim S, American Association of Pharmaceutical S (2014) Assessment and reporting of the clinical immunogenicity of therapeutic proteins and peptides-harmonized terminology and tactical recommendations. *AAPS J* 16(4):658-673.
175. Bray BL (2003) Large-scale manufacture of peptide therapeutics by chemical synthesis. *Nat Rev Drug Discov* 2(7):587-593.