

NON-CANONICAL AMINO ACID MUTAGENESIS OF POSITION B28 IN
INSULIN WITH PROLINE ANALOGS

Thesis by
Seth Lieblich

In Partial Fulfillment of the Requirements for the degree
of
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California
2017
(Defended March 15th, 2017)

This thesis is dedicated to my family, friends, and colleagues.

ACKNOWLEDGEMENTS

I am indebted to all of my friends, family and colleagues without which this work could not have been accomplished. Each of you, whether named specifically here or not, played a crucial role in helping me to grow, learn and complete my PhD.

I would first like to thank my two advisors: Prof. Steve Mayo and Prof. David Tirrell. I am grateful to have had the chance to work with both of you over my career here at Caltech. The two of you made for wonderful advisors with brilliant insights and mentoring that helped me to sharpen my skills, my work and myself. I hope to continue my development, in the manner you have taught me, as I continue onward in my career.

I would also like to thank my committee for their diligent efforts in shepherding me through my time here at Caltech. Prof. Frances Arnold and Prof. Doug Rees have both helped me to explore new avenues in my work as well as brought new insights into my work.

I would also like to thank my many lab and office mates over the years. Though many of you have had to put up with my loud typing, I heard few complaints about it. No one has been more instrumental than my long-suffering co-worker Katharine Fang, with whom which I have collaborated with on all aspects of the work described in this thesis, and who has worked tirelessly with me. Without your help Kat, I am sure this project would have been much harder and much less fruitful. Our collaborators at City of Hope were instrumental in helping us to complete the mouse experiments, I'm glad I did not have to inject any diabetic mice myself. I am also thankful for our collaborators at NovoNordisk, not only were you generous in funding this work, you were gracious hosts and key sounding boards for working out the experimental protocols. I would also like to thank the undergraduates and volunteers whom I have had the pleasure of working with over the years. Mentoring you has taught me much about myself and helped me to learn about leadership. There are also innumerable others in the labs, facilities, offices and clubs I have had the pleasure of working with over the years here at Caltech. I enjoyed all of our interactions greatly and learned a great deal from them.

ABSTRACT

Insulin is a protein hormone that is crucial for maintaining the concentration of blood glucose *in vivo* and is used clinically as a drug for the treatment of diabetes.

Chapter I provides for an overview and background on the state of the art in insulin treatment of diabetes and the many attempts, over 95 years, to improve the pharmaceutically relevant properties of insulin and improve our understanding of the model globular protein.

Chapter II demonstrates the incorporation of hydroxyproline analogs into insulin and shares the discovery of insulin with enhanced stability and an accelerated kinetic rate of dissociation. We also provide the highest resolution structure deposited in the PDB of insulin in the T₂ state and the 3rd highest of any insulin to date.

Chapter III extends the incorporation of proline analogs in insulin to include fluorinated insulins. We also provide, for the first time, high-resolution structures of a single globular protein systematically mutated with all possible stereoisomers of fluorination at the 4-position on a single proline residue (4S, 4R, di-substituted).

Chapter IV extends the incorporation of proline analogs in insulin to include ring variant analogs. We also provide, for the first time, high-resolution structures of globular proteins containing pipercolic acid, azetidine-2-carboxylic acid and 3,4 dehydroproline in the polypeptide chain.

Chapter V discusses the significance of the findings described herein and discusses future directions to undertake in further engineering insulin for improved characteristics.

This thesis describes a systematic approach, akin to medicinal chemistry, of altering a particular protein side chain by atomistic changes. I hope that the breadth of different amino acids incorporated into a single globular protein combined with the structural, functional, thermodynamic and kinetic information contained within this set of mutants will provide future protein engineers, computational protein designers and proline enthusiasts with a wealth of new information to be used to improve our understanding of proteins and predictive power.

PUBLISHED CONTENT AND CONTRIBUTIONS

Lieblich, S.A. et al. (2017 anticipated). “4S-Hydroxylation of insulin at ProB28 accelerates hexamer dissociation and delays fibrillation ”. In submission : J.A.C.S.

S.A.L participated in the conception of the project, all aspects of protein preparation, collected and analyzed biophysical data, solved the crystal structures, prepared the data, and participated in writing the manuscript.

TABLE OF CONTENTS

Acknowledgements	iv
Abstract.....	v
Published Content and Contributions	vi
Table of Contents	vii
List of Illustrations and Tables.....	ix
Chapter I: 95 years of insulin in the treatment of diabetes.....	1
Diabetes.....	2
Impact of diabetes on the world.....	3
The history of insulin and its usage in diabetes treatment.....	4
The structure and function of insulin	8
Insulin's globular structure.....	8
Insulin oligomerization.....	9
Insulin binding to the insulin receptor (IR).....	11
Fibrillation and dynamics.....	13
The general mechanism of oligomerization and activity in insulin.....	15
Non-Canonical Amino Acids for protein engineering	15
The structure and conformations of proline.....	16
Figures.....	19
References.....	25
Chapter II: Hydroxylation of insulin at ProB28 accelerates hexamer dissociation and delays fibrillation.....	37
Abstract.....	38
Introduction.....	38
Materials and Methods	39
Results and Discussion.....	46
Conclusion	48
Figures.....	50
References.....	63
Chapter III: Fluorinated proline analogs of insulin at ProB28 modulate hexamer dissociation and rate of fibrillation	69
Abstract.....	70
Introduction.....	71
Results and Discussion.....	72
Conclusions.....	76
Figures.....	78
References.....	94
Chapter IV: Ring size variation of proline analogs in insulin at ProB28 modulates hexamer dissociation and rate of fibrillation.....	95
Abstract.....	96
Introduction.....	97
Results and Discussion.....	98
Conclusions.....	103
Figures.....	104
References.....	123

Chapter V: Future avenues for insulin engineering.....	125
Discussion.....	126

LIST OF ILLUSTRATIONS AND TABLES

<i>Figure</i>	<i>Page</i>
1.1 Insulin sequence and structure	19
1.2 Circular dichroism spectra of representative monomeric and dimeric insulins.....	20
1.3 The insulin hexamer and its dissociation.....	20
1.4 The interaction between insulin and the insulin receptor	22
1.5 Position B28 and proline structures	23
1.6 Position B28 and proline structures.....	24
2.1 Hydroxyinsulins retain activity.....	50
2.2 Hydroxylation at ProB28 modulates insulin dimerization, dissociation kinetics, and stability.....	51
2.3 Crystal structures of HzpI and HypI.....	52
T2.1 Biophysical characteristics of unsulin variants.....	53
S2.1 Insulin expression and incorporation of hydroxyprolines.....	54
S2.2 Immunoblot detection of insulin receptor activation.....	55
S2.3 Example fits from sedimentation analysis.....	56
S2.4 Example fits for analysis of dissociation kinetics.....	57
S2.5 Alignment at position B28.....	58
S2.6 Alignment of R6-AspI and T2-HzpI at position B28.....	59
TS2.1 Expression yields and incorporation levels of hydroxyinsulins.....	60
TS2.2 Data tables and refinement values.....	61
TS2.3 RMSD and proline angles	62
S2.7 Alignment at position B28 highlighting key contacts in the dimer	63
TS2.4 Key B28 contacts in hydroxyproline insulins.....	64
3.1 Fluorinated proline analogs.....	78
3.2 Orientation of mono-fluoroproline in conformers.....	79
3.3 Fluoroproline incorporates into insulin and fluorinated insulins lower glucose in a mouse model of diabetes.....	80
3.4 Circular dichroism spectra of fluorinsulins	81

3.5	The dimer in the R6 hexamer of the fluoroinsulins.....	87
3.6	Alignment at position B28	88
3.7	B28 highlighting key contacts in the dimer interface.....	89
T3.1	Key properties of fluoroinsulins.....	90
T3.2	RMSD and proline angles	91
T3.3	Key B28 contacts in fluoroproline insulins	92
T3.4	Data table and refinement values for fluoroinsulins.....	93
4.1	Ring proline analogs.....	10
4.2	Ring variant prolines incorporate into insulin	105
4.3	Circular dichroism spectra of ring variant insulins	106
4.4	Data table and refinement values for ring variant insulins.....	113
4.5	The dimer in the R ₆ hexamer of the ring variant insulins	115
4.6	Alignment at position B28	116
4.7	B28 highlighting key contacts in the dimer interface.....	117
4.8	B28 highlighting key contacts in the dimer interface.....	118
T4.1	Key properties of ring variant insulins.....	119
T4.2	RMSD and proline angles	120
T4.3	Key B28 contacts in the T2 structures of the ring variant insulins	121
T4.4	Key B28 contacts in the R6 structures of the ring variant insulins.....	122

CHAPTER I

95 years of insulin in the treatment of diabetes

Diabetes

Diabetes mellitus is a chronic disease characterized by a dysregulation in the homeostasis in the concentration of blood glucose. In healthy individuals, the concentration of blood glucose is tightly regulated through the action of several hormones to lie between 3.3 mM and 5 mM¹. The most important of these hormones is insulin; insulin secretion lowers the concentration of blood glucose primarily by up-regulating glucose uptake into peripheral tissues and down regulating hepatic secretion of glucose. Insulin also has a number of secondary effects such as modulating weight gain, altering cell cycle regulation, and regulating neuronal and vascular processes²⁻⁶. Insulin is especially important in the post-prandial response to food consumption. The pancreas is responsible for secreting insulin from β -cells. Insulin binds to the insulin receptor, through which most of its primary effects are mediated⁷. In healthy patients, the release of insulin in response to food consumption can be divided into 3 phases⁸⁻¹⁰. The first phase has an initial burst of insulin to rapidly respond to glucose entering the blood stream. The second phase is characterized by a prolonged release of insulin at progressively lower levels concurrent with gastric emptying. The third phase, termed the basal phase, is essentially upkeep between meals and is characterized by a release profile of insulin that is constant. Diabetes mellitus occurs when insulin secretion, or the response of tissues to insulin, are perturbed resulting in persistently high concentrations of blood glucose (hyperglycemia). Diabetes can be divided into 2 forms on the basis of the underlying cause: type 1 and type 2. In type 1 diabetes, the cells responsible for insulin production are destroyed via an inappropriate autoimmune reaction that occurs typically in early adulthood or adolescents¹¹. In type 2 diabetes, both the production of insulin and response to insulin are compromised due to a variety of factors^{12,13}. Pancreatic β -cells fail to produce healthy levels of insulin in response to glucose spikes and peripheral tissues fail to respond with sufficient strength to a given insulin concentration. Over time, beta-cells' increased demand for insulin and peripheral tissue's failure to respond create a feedback loop that worsens β -cell function and, at an advanced disease state, resembles type 1 diabetes^{12,13}.

Impact of diabetes on the world

Diabetes is a major disease throughout the world and causes significant degradation in patient quality of life, reduces overall life expectancy and imposes substantial economic burdens. As diabetes is a chronic disease with no known cure these negative impacts are incurred over a patient's lifetime.. As of 2011, over 14% of the United States population is afflicted with a form of diabetes, with around 5% undiagnosed ¹⁴. This rate represents a 10-fold increase in prevalence since 1958 and reflects the United States', rapid evolution into a modern economy with modern consumption patterns ^{15, 16}. This rate further represents a 1.5-fold increase since 1991; notably, the increase in diabetes prevalence has plateaued in recent years ¹⁵. As a result of this prevalence in the US, diabetes now represents the 7th leading cause of death, killing over 75,000 people each year ¹⁷. Recent analysis indicates that the true number of diabetes deaths may even exceed 250,000 per year when all deaths attributable to diabetes are accounted for ¹⁸. These deaths are also premature. Diabetes on average decreases life expectancy between 3 and 18 years ¹⁹. Women are disproportionately affected by the loss in life expectancy, with a loss 3 years greater than similar men. In a matched cohort study, the diabetes mortality hazard ratio was found to be 1.86, and represented a median loss of 10.5 years in life expectancy ²⁰. For individuals below the age of 20, and who are presumably type 1 diabetics, 20 years in life expectancy was lost with an associated hazard ratio of 3 ²⁰. The lifetime risk for a child born in the US during the 2000s is 33% for men and 39% for women, rising to 53% for Hispanic women ²¹, staggeringly, this means that more than a third of the US population is expected to contract some form of diabetes in the future. Diabetes is also the 9th leading cause of death worldwide ²² killing 1.3 million individuals in 2010. Expectations are that diabetes will only further rise as incidence rates globally are rising to levels potentially the same as the US. In 2010 6.4% of the world population was afflicted with diabetes, particularly concentrated in the UAE (18.7% in 2010) ²³. It is predicted that by 2030, 7.7% of the world will suffer from diabetes with up to 87 million diabetics in India ²³. Diabetes is also associated with a wide variety of comorbid conditions. Comorbidities such as dementia,

heart disease, cognitive decline and others ^{18, 24, 25} are widely prevalent in diabetics with nearly all diabetics having at least 1 comorbid condition ²⁴.

Diabetes's widespread prevalence and comorbidities also account for a large fraction of overall medical spending. Diabetes is responsible for nearly 20% of medical spending and diabetics spend about 2-fold more than their non-diabetic peers on medical expenses ^{17, 26}. Diabetes is also characterized by a high rate of complication and inpatient hospitalization and results in 3-fold higher spending on inpatient hospitalization and 2-fold higher spending on ER visits. Overall the economic cost of diabetes is \$245 billion with \$69 billion resulting from losses in productivity and sick days away from work ¹⁷. Compared to peers, over \$120 thousand more in medical spending is spent on diabetics over their lifetime ²⁷. Diabetics spend nearly 3 times more on in-patient expenses and 2 times more in emergency room expenses than their non-diabetic peers ²⁶. Over \$16 billion is spent on potentially avoidable complications in diabetics, predominantly these complications arising from poor control of blood glucose concentration or insulin concentrations ²⁸. Approximately 65% to 80% of patients have 1 or more avoidable complications each per year ²⁸. Complications from diabetes are also responsible for 31% of all hospitalizations ²⁹. The vast majority of diabetes patients do not meet their diabetes care goals, with only 14% of all patients meeting all desired care goals ³⁰. Primarily these goals are not met due to non-adherence of the patients to medication regimes. Over half of diabetes patients are non-adherent to their prescribed regime, which can be ascribed to combination of normal patient non-adherence as well as the complicated nature of diabetes treatment requiring daily or sometimes hourly monitoring of glucose and dosing ³¹.

The history of insulin and its usage in diabetes treatment

Insulin has a 95-year history in the treatment of diabetes. To date, over 180 forms of insulin have been discovered and produced for the amelioration of diabetes ³². Diabetes itself has been known since ancient times; however, no highly effective treatment was available prior to the early 1900's ³³. Without insulin, diabetics lived short

lives characterized by horrific wasting and comorbidities before death³³. In 1910 Edward Sharpey-Schater was the first to name insulin, naming it for the specific cells responsible for production in humans; namely the pancreatic islet cells of Langerhans³³. In the midst of much interest in insulin and diabetes, by 1921, Frederick Banting and Charles Best isolated insulin from the pancreas of dogs^{34,35}. They were subsequently the first to treat diabetes in human patients using insulin. Patients, who previously were given but months to live, were able to live years. Further refinement over the many decades since 1921 have enabled patients to lead lives into old age. Eli Lilly began the first large scale productions and purifications of insulin and was able to improve the variability of insulin batches by a substantial amount³⁶. August Krogh later founded Novo laboratories as a non-profit to produce insulin in Denmark³⁴. Later Hagedorn discovered protamine insulin³⁷. Protamine insulin was the first improvement in the formulation of insulin. Protamine, and subsequently zinc, were utilized in order to generate crystalline suspensions of insulin that provided prolonged action after injection through the lack of solubility and depotting at the injection site³⁴. Between the 1940s and 1990s nearly all insulins relied on the crystalline state for formulation and were prepared from either bovine or porcine sources³⁸. While a vast improvement over no insulin, these crystalline formulations of insulin would not be considered today to recapitulate the prandial or basal response. Moreover, these formulations are of limited stability³⁸.

By 1949 Frederick Sanger determined the amino acid sequence of insulin, allowing the first primary sequence analysis of insulin and its homologs^{39,40}. By 1967, Donald Steiner began elucidating pancreatic production of insulin and understanding the precursor protein, termed proinsulin. Proinsulin gives rise to insulin after proteolytic processing⁴¹. In 1969 Dorothy Hodgkin first revealed the 3-D atomic structure of insulin through the usage of x-ray crystallography⁴². Originally deposited as 1INS in the pdb, and updated to 4INS later, this first crystal structure illuminated the T-state of the insulin hexamer that is the natural pancreatic storage configuration. By 1978 Dave Goeddel and others at Genentech produced the first recombinant form of insulin through expression in *Escherichia coli* followed by subsequent purification and refolding *in vitro*^{34 43}. This recombinant insulin was shown to work identically to insulin sourced directly from

mammals⁴⁴ through an ingenious assay wherein insulin is labeled with I¹²⁵ and the displacement of I¹²⁵ insulin from the insulin receptor by a different insulin is measured. Eli Lilly subsequently commercialized the Genentech discovery and was the first to offer a recombinantly sourced insulin on the market. Modern production of insulin occurs primarily in *E. coli* or yeast strains⁴⁵⁻⁴⁸. A separate track for the chemical synthesis of insulin has been pursued for many decades as an alternative source to recombinant production. However, the percentage yields of chemically synthesized insulin remain in the single digits and provide only milligram quantities⁴⁹⁻⁵². Chemical strategies are therefore insufficient to meet the world's need for insulin and are orders of magnitude less efficient than microbial systems, which are able to produce multigram quantities of insulin per hour⁵³.

Later work on insulin, in the 1990's and early 2000's, has focused on discovering insulins that are better able to recapitulate the post-prandial and basal responses in humans. The rapid acting insulins (RAIs) LisPro (in 1996), Aspart (in 2000) and Glulisine (in 2004) were approved for use in humans to better approximate the prandial response³⁴. These RAIs were a substantial improvement over prior insulin as they were better able to recapitulate the first and second phases of normal insulin release. RAIs have a 2-fold increase in max concentration and require half of the time to reach peak activity when compared to wild-type human insulin^{54, 55}. This translates into a 100-minute lag time until peak glucose response for RAIs vs 180 minutes for wild-type insulin^{54, 55}. However glycemic control is still imperfect with RAIs as they do not completely recapitulate post-prandial response and better mealtime control is likely to further help patients⁵⁶. RAIs help to lower the risk of hypoglycemia by as much as 12% due to their shorter lag time, and less insulin remains within the patient and reduces the likelihood of accidental overdosing⁵⁶⁻⁶⁰. The approved RAIs are created by promoting rapid disassociation of the insulin hexamer after subcutaneous injection through mutations that destabilize the hexameric complex found in pharmaceutical formulations⁶¹⁻⁶⁴. The insulin monomer then diffuses into the bloodstream and it is the rate of the dissociation of hexamers that is the rate limiting step for the onset of glucose lowering action from insulin^{65, 66}.

The long acting insulins Glargine (in 2000), Detemir (in 2004)³⁴ and Degludec (in 2016)⁶⁷ were approved to better approximate the basal response in humans. The long duration of these insulins serves to lower the variance in insulin activity during the day and reduces the risk of hypoglycemia overnight when compared to crystalline insulins⁶⁸.⁶⁹ In order to create the long acting insulins, diffusion of the insulin monomer into the bloodstream was slowed by depotting the insulin after injection into either subcutaneous tissues or blood proteins⁷⁰.

Modern insulin treatments can faithfully recapitulate only the basal phase of insulin in a patient without disease⁸. Currently available RAIs span both the first and second phases of normal insulin release while being unable to properly recapitulate either phase.

Other discoveries would alter the treatment of diabetes through the use of insulin focused on monitoring methods and automated pumps. In particular the discovery of glycosylated hemoglobin (A1c) as a long term measure of diabetic treatment was critical in improving the benefits of novel treatments⁷¹. The usage of glucose self-monitoring was also shown to have substantial benefits in helping patients choose the correct insulin-dosing regime. Self-monitoring improved the A1c measure by almost 7% in adults⁷² and continuous monitoring is now recommended for all type 1 diabetic patients⁷³. The introduction of automated insulin pumps has also proven beneficial to patients⁷⁴. Further refinements using closed loop systems are increasingly a focus of both R&D as well as treatment⁷⁵ and great hope has been placed on insulin pumps^{76,77} as the future of the amelioration of diabetes⁷⁸. However in the world of insulin many challenges remain⁷⁹, primarily those focused on improving the rapidity of insulin activity to better approximate prandial insulin release, improving the stability of insulin to better guard against aggregation and injection site intolerance and enable usage in long-term implantable pumps, and developing an insulin variant that is responsive to glucose concentration in real-time without patient involvement⁷. These issues arise from the narrow therapeutic index of insulin, the 30 minutes required for the RAIs to begin taking

effect, the 100 minutes required for peak activity of the RAIs and the fibrillation of insulin as patients carry a supply on their person rather than in a controlled environment with refrigeration. Further the peripheral and non-glucose effects of insulin are exaggerated relative to the glucose lowering effects due to the subcutaneous injection as the route of administration, leading to side effects such as weight gain ².

The structure and function of insulin

Insulin's globular structure

Insulin itself is a 51-amino acid protein composed of 2 covalently linked peptide chains. The 2 chains are termed the A-chain and the B-chain, and have separate N- and C-termini. Throughout this work, residues shall be designated by the letter of the chain and residue number (i.e. the first residue of the A-chain is A1). An insulin monomer contains 2 disulfide bonds between these chains. Residues CysA7-CysB7 and CysA20-CysB19 are the covalent links between the A-chain and the B-chain. An additional disulfide bond exists between CysA6-CysA11. These three disulfide bonds are conserved in all insulins of mammalian origin ⁸⁰. Insulin is thought to have evolved sometime around 500 million years ago ⁸⁰⁻⁸² (Figure 1.1a). The secondary structure of insulin can be characterized as predominantly helical. Insulin has 3 helices and a single beta strand in its monomeric state ⁴² (Figure 1.1b). The helices consist of residues A2-A8, A13-A19 and B9-B19, while the beta strand consists of B9-B19. The 2 helices in the A-chain lie antiparallel to each other and are connected via a turn in residues A9-A12. The presence of the A6-A11 disulfide serves to secure a short turn between the two helices ^{42, 83, 84}. The single beta strand lies antiparallel to the helix in the B-chain (B9-B19) and abuts both termini on the A-chain ⁸⁵. This orientation is held in place by a variety of non-covalent interactions as well as the two interchain disulfide bonds. The hydrophobic core of insulin is held together by a series of leucine residues (A13, A16, B6, B11, B15) and a single valine (B18) that come together in close contact with each other ^{42, 83, 84, 86} (Figure 1.1b). Residues B1-B8 can be found in three or more different conformations, two of

which are termed the R and T states ⁸⁷, a reference to Monod's naming convention for conformations with allosteric control ⁸⁸. In the T state, B1-B8 form a turn hanging off of the B-chain helix. In the R state, B1-B8 form an extension of the B-chain helix, effectively lengthening the helix to comprise B1-B19 (Figure 1.1c). The switch between these states is modulated by both higher order oligomer assembly as well as the mutational identity of residues B1-B5 that comprise the switch region ⁸⁹⁻⁹¹. The movement of B1-B8 pivots around GlyB8. GlyB8 undergoes a change in psi angle from 59° to -67° between the T and R conformations ⁸⁷.

Insulin oligomerization

Insulin monomers can occupy a classical equilibrium of oligomer states. Insulin monomers dimerize ⁹² at a K_d of between 5-20 μM (depending on the method used to characterize K_d, typically spectroscopic measures yield lower values and centrifugation measurements yield higher values). Residues B8, B9, B13, B16 as well as B23-B28 play key roles in building the dimer interface ^{83, 89}. Residues B20-23 form a turn leading into the dimer interface and serve to position the beta strand for self-association. B9, B13 and B23-B26 form multiple hydrogen bonds across the dimer interface. These hydrogen bonds displace solvent interactions present in the monomer ⁶¹. Throughout this work I will refer to dimer related residues with the notation of ' (i.e. SerB9 forms a hydrogen bond, across the dimer interface, with GluB13'). Residues B23-B26 (a span of three highly hydrophobic residues Phe, Phe, Tyr) and B28 also form substantial hydrophobic packing between the dimers. Disruptions in, or deletions from the C-terminus of the B-chain lead to substantial losses in the K_d of the dimer. For example, DesB23-30 insulin exhibits no tendency towards dimerization ⁹³. The dimerization of insulin can be conveniently monitored by inspection of the circular dichroism (CD) spectra. The insulin monomer is associated with a deep minimum at 208nm and a shallow minimum at 222nm. Upon dimerization, the minimum at 222nm is deepened, likely due to the association of the single beta strands from each monomer into a 2-stranded beta sheet. Concurrently the minimum at 208nm is attenuated. This change in CD spectra can be seen in variants with the loss of dimerization as well as changes in the concentration of

the insulin solution near the K_d ⁹³. Changes in CD spectra also correlate closely with other biophysical methods for determining the K_d of dimerization (for example: analytical ultracentrifugation)⁶¹ (Figure 1.2a)

In the presence of zinc, the dimers can further oligomerize into hexamers with 2 zincs bound per hexamer⁹³. One histidine residue from each monomer contributes to coordinating the 2 zinc atoms (3 histidines per zinc). Typically throughout these oligomer states insulin maintains its T-conformation. In this T-conformation the zinc is bound in octahedral geometry with 3 histidines and 3 water molecules⁸³. Key residues that form interactions on the interface of hexamer are A13, A14, A17, B1, B2, B4, B13, B14, B17-B20^{81, 83, 94}. It is only in the presence of phenolic ligands, in addition to zinc, that the insulin hexamer can occupy the R-state. The presence of phenolic ligands can be thought of as driving a conformational switch from T to R states in an allosteric manner. Binding of the phenolic ligand also serves to occlude solvent access to the zinc binding sites⁹⁵⁻⁹⁷. In the R-state, the geometry around the zinc changes, zinc becomes bound in tetrahedral geometry with 3 histidines and an anion (typically chloride)^{98, 99}. In addition to zinc other divalent metal cations, such as cobalt, can bind to the metal binding sites in the insulin hexamer. Cobalt, in particular, serves as a useful measure of T to R transitions as the d-d adsorption line changes between hexameric conformations^{98, 100}. Chelation of cobalt, from R₆ insulin hexamers, showed that in the absence of phenolic ligands the hexamers disassociate in just 1.25 seconds. However in the presence of 30mM phenol this disassociation time jumps to 100 seconds, and jumps to well over 100 seconds with 80mM phenol⁹⁸. Another useful measure of hexamer disassociation is the chelating ligand 2,2',2''-terpyridine (terpy). Upon binding to divalent metal cations, terpy undergoes a spectroscopic change which can be tracked to monitor the disassociation of the insulin hexamer and release of zinc^{98, 101}.

The hexamer can occupy both full T₆ and full R₆ states as well as the intermediate T₃R₃ state. The transition from T to R occurs in the order: T₆ to T₃R₃ to R₆^{100, 102}. This order is reversed during dissociation of the complex driven by chelation of zinc ions. In the presence of EDTA, the T₆ hexamer first loses a single zinc and a T₂ dimer, forming a

short-lived intermediate 1-zinc and 4-insulin complex. Further chelation of the last zinc releases two more T₂ dimers¹⁰³. Increasing concentrations of either the phenolic ligand or the coordinating anion serve to stabilize the R₆ hexamer against disassociation in the presence of chelating agents⁹⁸ (figure 1.3 a,b). Although the two zinc sites in the insulin are nominally identical, two different K_d's can be measured. The binding of the first zinc has a K_d of 0.5 uM and the second has a K_d of 14uM^{93, 104, 105}. Isothermal titration calorimetry (ITC) measures indicated that the two zinc binding sites are associated with a deltaG of -14.0 kcal/mol and -12 kcal/mol respectively¹⁰³. Two complexes of a binding site for zinc with 3-hisitindes would normally be associated with a deltaG of -32 kcal/mol¹⁰⁶. The 6 kcal/mol difference between the expected value and measured value was attributed by Carpenter and Wilcox to be lost in the T to R transition¹⁰³. The R₆ hexamer is itself inactive in binding and activation of the insulin receptor (IR)⁸⁷.

As might be expected, loss of the coordinating histidines (HisB10 to Asp) prevents formation of the hexamer and binding of zinc¹⁰⁷. Strangely this mutation also seems to confer carcinogenic properties to insulin through a mechanism that is not yet fully understood. AspB10 has an enhanced affinity for binding insulin like growth factor receptor 1, and it is thought that through this interaction the carcinogenic character is imparted^{108, 109}.

Insulin binding to the insulin receptor (IR)

The IR is a homodimer, ab₂ transmembrane tyrosine receptor kinase through which insulin's primary effects are mediated^{110, 111}. Binding of insulin to the IR occurs though a 2-site binding mechanism where 2 insulin monomers bind to the IR sequentially. Recent work from the Weiss group has enhanced our understanding of this binding through crystallographic visualization of the insulin-IR binding interface¹¹². Many of the observations from the crystal structure were predictable on the basis of earlier mutational studies (figure 1.4a).

Almost half of the insulin molecule is sensitive to alanine scanning mutations (22/51 amino acids) resulting in either substantial decreases in binding affinity to the IR (11/51) or over 10-fold loss in IR binding affinity (11/51)¹¹³. 16 of these 22 amino acids are also highly conserved across multiple mammalian species⁸⁶. GlyB8, LeuB11, and GluB13 are some of the residues critical for binding to IR. Mutating any of these residues to alanine decreases IR binding affinity by 10-fold or greater, but not all mutations result in such dramatic losses to IR binding affinity¹¹³. For some residues (examples GlyA1, IleA2, ValA3, TyrA19), any mutation, including alanine, serves to drastically reduce IR binding affinity^{86,114}. These four positions are generally regarded as hotspot interactions^{87,115,116}. ValB12 and TyrB16 are also exquisitely sensitive to mutation, most amino acids mutations at these positions degrade the binding of inulin to the IR¹¹⁷. ValB12 is critical not only for binding of insulin to the IR but also for stability¹¹³. Other mutations, such as most to LeuA16, do not impair binding,¹¹³. It is thought that B27-B30 must be displaced upon binding to the IR, and mutations in or deletions of B27-B30 do not impact binding affinity *in vitro* or *in vivo*^{112,118,119}. FDA approved products exist for a desB30 insulin¹²⁰ as well as an insulin with extensive conjugation to the sidechain of B29¹²¹, further highlighting the lack of importance for B27-B30 for insulin binding to the IR¹¹⁷. Evident from both mutational studies and co-crystal structures PheB24 and PheB25 form a hydrophobic interaction that is interactions for receptor binding¹²². Mutations to B24 can increase or decrease IR binding affinity; mutations to B25 are only tolerated by substitution with large hydrophobic residues with rings^{64,109,112,118,123,124}. In a co-crystal structure, residues B26-B30 of insulin are displaced by an alpha helix (The C-terminal segment of the human insulin receptor a-chain or aCT) from the IR. The aCT helix packs against the alpha helix of the B-chain and the termini of the A-chain and the aCT is anchored by interactions with B24-B25^{112,115,118}. From the co-crystal structure residues A1-A4, A19, B3, B8, B9, B11, B12 and B15 are seen to engage in direct contact with the IR¹¹⁵ (figure 1.4a).

ProB28 is a common site of mutation for fast-acting analogs and LysB29 is a common site of modification for long-acting analogs, highlighting their unimportance with regards to IR binding but strong importance for control of oligomerization⁸⁶. The

common rapid-acting insulin mutations break up the beta strand to beta strand interaction of the dimer through the introduction of charged residues that disturb backbone trajectory at residues B28-B30 but do not alter IR binding¹²⁵ (figure 1.5a). It seems that both a combination of charge and backbone perturbations are responsible for breakup of the dimer. ProB28 to Asp results in a 3 order of magnitude loss to the Kd of insulin dimerization while mutation to Ala, or mutation of LysB29 to Pro lead only to a 2-order magnitude loss to the Kd of dimerization⁶¹. The ProB28 to Asp mutation also increased the disorder of the B-chain C-terminus¹⁰¹. Insulin binding to the IR is only related insofar as insulin is able to achieve a monomeric state prior to binding. An insulin analogue that is covalently locked into the dimer state (B25-B25' disulfide bond) is completely inactive and does not bind the IR, but maintains the ability to form insulin hexamers¹²⁶. It is not clear whether it is truly the dimeric nature of this insulin or inability for the C-terminus of the B-chain to swing away from the core IR binding residues that results in lack of detectable IR binding. Unfortunately the impact of the covalent link between dimers upon the rate of hexamer disassembly was not studied.

Fibrillation and dynamics

Molecular dynamics (MD) simulations have revealed that the insulin dimer is a relatively stable protein that does not undergo any substantial movements. Conversely, MD studies have shown that the insulin monomer undergoes substantial movements at both the N- and C-termini of the B-chain¹²⁷. For example, in one simulation, overall RMSD of the dimer structure was 1.4 angstroms whereas the monomer had an RMSD of 3 angstroms during the simulation time¹²⁷. MD simulations indicated that, in particular, the B-chain C-terminus swings out away from the core of insulin pivoting around PheB24¹²⁸. This observation is in line with conformational changes observed in the crystal structure of the T to R transition (N-terminus) and co-crystal structures showing movement of the C-terminus upon binding to the IR^{83, 97, 129}.

Insulin can undergo aggregation, termed fibrillation, into amyloidogenic fibers under circumstances of high heat, time, or agitation. It is thought that the partial unfolding and aggregation of insulin monomers leads to formation of fibrils¹³⁰. Insulin fibrillation is undesirable from a therapeutic standpoint, but few mutations to insulin show enhanced lag time to fibrillate. Currently the primary methods for stabilizing insulin against fibrillation focus almost exclusively on the contents of the formulation through the inclusion of zinc, phenolic ligands, salts, sugars, or other excipients⁷. For mutants with increased lag time, there does not seem to be a correlation with the thermostability of the insulin monomer or dimer and there seems to be a reverse correlation with kinetic disassociation rates of the hexamer^{122, 131, 132}. For example, iodination of TyrB26 results in a 4-fold increase in fibrillation resistance but slows the rate of disassociation of the hexamer by 8-fold⁶¹. Alternative mutations at TyrB26 (such as Glu) can enhance the rate of hexamer disassembly, but impair the stability of the analog¹³³.

Fibrillation seems to highly correlate to the degree with which the B-chain C-terminus separates from the protein core. Mutants in and near the hinge region (B20-B26) of the B-chain, which promote fraying of residues B26-B30, also promote faster fibrillation^{122, 129, 134}. Conversely mutations that reduce fraying promote slower fibrillation^{126, 129}. Other attempts to enhance the resistance to fibrillation have focused on helix capping¹³⁵, single chain variants¹³⁶ and additional disulfide bonds¹³⁷.

The ideal combination of traits, from a therapeutic perspective, would be an insulin analog with increased resistance to fibrillation and faster rates of disassociation from the hexamer. However, it has proven difficult to fully optimize insulin because so many residues on the protein play a substantial role in IR binding and further mediate insulin oligomerization or protection against fibrillation.

The general mechanism of oligomerization and activity in insulin

On the basis of multiple lines of evidence the general mechanism for insulin biophysical and biochemical function upon subcutaneous injection is as follows. R₆ insulin hexamers containing phenol and zinc are injected into the subcutaneous tissue. Phenol diffuses into surrounding tissue and the R₆ hexamer undergoes a conformational shift to the T₆ hexamer. Zinc carrier proteins (i.e. albumin) chelate zinc from the hexamer and T₂ dimers are released into the subcutaneous space^{138, 139}. T₂ dimers disassociate and diffuse to the blood as monomer, the accumulation of monomers in the subcutaneous space is the rate-limiting step for diffusion into the blood^{65, 66}. In the monomer, the C-terminus of the B-chain undergoes a conformational change swinging residues B27-B30 away from the core of the protein. Insulin subsequently binds to the IR and induces the reduction in blood glucose.

Most importantly to the work herein, B28 has proven to be a position of minimal importance for IR binding while of utmost importance for oligomerization control^{7, 54, 61, 62, 70}. Previous work has explored canonical mutations at B28; however, none of these mutations could replicate the unique properties, detailed below, which proline imparts on a polypeptide chain because these mutants lacked a pyrrolidine ring linked to the polypeptide backbone. Below I will detail attempts to improve the pharmaceutically relevant properties and enhance our understanding of insulin.

Non-canonical amino acids for protein engineering

In order to replace proline with a similar amino acid it is necessary to mutate B28 to alternative amino acids that maintain a ring structure connected to the backbone chain. Fortunately for the work described herein, the method of non-canonical mutagenesis (ncAAM) has been discovered and improved over the past decade. ncAAM allows for the substitution of non-canonical amino acids into proteins. This method has been discovered and extensively reviewed by our lab and others¹⁴⁰⁻¹⁴⁴. Generally two main methods for ncAAM have been discovered; residue replacement and nonsense suppression. The

nonsense suppression methods, discovered by Schultz and others ¹⁴⁴, focus on placement of a stop codon into a protein. This stop codon is then recoded, through the introduction of novel tRNA/synthetase pairs, to code for a new amino acid. This method, while powerful, suffers from several drawbacks; termination at the activity stop codon cannot be completely suppressed. Incomplete suppression results in truncated protein products and read through with substitution of, sometimes multiple different, unintended amino acids into the stop codon location. The residue replacement methods, discovered by Tirrell and others ¹⁴⁰, focus on replacing a residue in the proteome with a complexly different amino acid, effectively overwriting the entire incorporation of that amino acid throughout the proteome. This method, while powerful, also suffers from several drawbacks; chiefly it only replaces one amino acid with another and does not add a 21st amino acid to the proteome. The Tirrell method is also frequently contaminated with background incorporation of the original amino acid that was intended to be over written. Luckily, insulin itself contains only a single proline residue at B28. So for our purposes, residue specific replacement is site specific and only affects B28. For proteins with only a single residue to be replaced, the Tirrell method is superior as the potential cross-contaminating residues are limited to only the wild-type amino acid and in optimal cases incorporation levels can be quantitative (approaching 100%) ¹⁴⁵. Truncated proteins as well as the range of amino acid read-through present in the Schultz method can be avoided. Previous work by Conticello and others demonstrated the incorporation of proline analogs with a range of side chain functionalities in collagen using the Tirrell method ¹⁴⁶ and this same range of prolines is utilized in this work to probe insulin (Figure 1.6a).

The structure and conformations of proline

Proline itself is a rich amino acid for replacement with a non-canonical version. Proline is the only canonical amino acid where the amino group is in a pyrrolidine ring with the side chain linked back upon the backbone atoms. This connection means that the phi angles of proline are limited to 65 +/- 25 degrees ^{147, 148}. As a result of this angular limitation proline finds itself in proteins primarily as a helix breaker, in poly proline

helices, in loops, or anchoring turns. The covalent hindrance of proline's backbone conformation means it has a unique set of allowed Ramachandran angles¹⁴⁹. The backbone coupling to the sidechain also means that the amide bond can adopt a cis conformation much more easily in proline. The position of the backbone carbonyl by the sidechain of proline also predisposes proline to participate in an n to π^* interaction¹⁴⁸. The n to π^* interaction occurs when the lone pair of the carbonyl oxygen interacts with the π^* orbital of the next amino acid in the sequence¹⁵⁰. The proline ring can also participate in an aromatic interaction. A C-H/ π interaction has been observed both thermodynamically and by frequent co-occurrence of proline with aromatic rings in the PDB^{151, 152}. The interaction between proline and an aromatic ring is on the order of typical aromatic stacking interactions¹⁴⁷, approaching energies of 1.5 kcal/mol¹⁵³. The C-H/ π interaction is also in addition to proline's more typical hydrophobic contributions to the energetics of protein folding¹⁵²⁻¹⁵⁴.

The proline ring itself can adopt three different conformations: endo, exo or planar^{155, 156} (figure 1.5 b-d). The endo conformation is characterized by the ring puckering up and towards the carbonyl oxygen of proline. Endo prolines are more common in protein structural elements of the extended types. The exo conformation is characterized by the opposite ring pucker of endo; namely the ring puckers down and away from the carbonyl oxygen. Exo prolines are more common in helical and polyproline elements in proteins. The exo and endo conformations are roughly equally represented in PDB structures, but from computational measurements proline is thought to prefer the endo conformation by a 2:1 ratio with an energy difference of only 0.5 kcal/mol^{157, 158}. The planar conformation is observed in a minority of protein structures in the PDB (<5%)¹⁵⁹.

The endo and exo conformations also have consequences on the strength of proline's typical interactions. Endo prolines weaken the n to π^* interaction while exo prolines enhance it¹⁶⁰. By biasing prolines towards or away from a particular conformation the C-H/ π interaction can also be tuned by improving packing so as to better stack the proline ring against an aromatic^{153, 154}.

The 9 prolines discussed throughout this work represent a set of proline with altered endo/exo preferences as well as other subtle changes to the interactions that a proline-like residue can form (figure 1.6a). By ncAA mutagenesis at B28 we have shown substantial ability to modulate important properties of insulin without compromising IR binding and activity. The set of prolines, and their insulins, examined here provides for a basis of ongoing insulin engineering as well as enhances our understanding of ring constrained amino acid residues in protein engineering.

I will discuss the 4-substituted variants (hydroxylated and fluorinated) in chapters 2 and 3 followed by the pseudoproline and ring size variants in chapter 4.

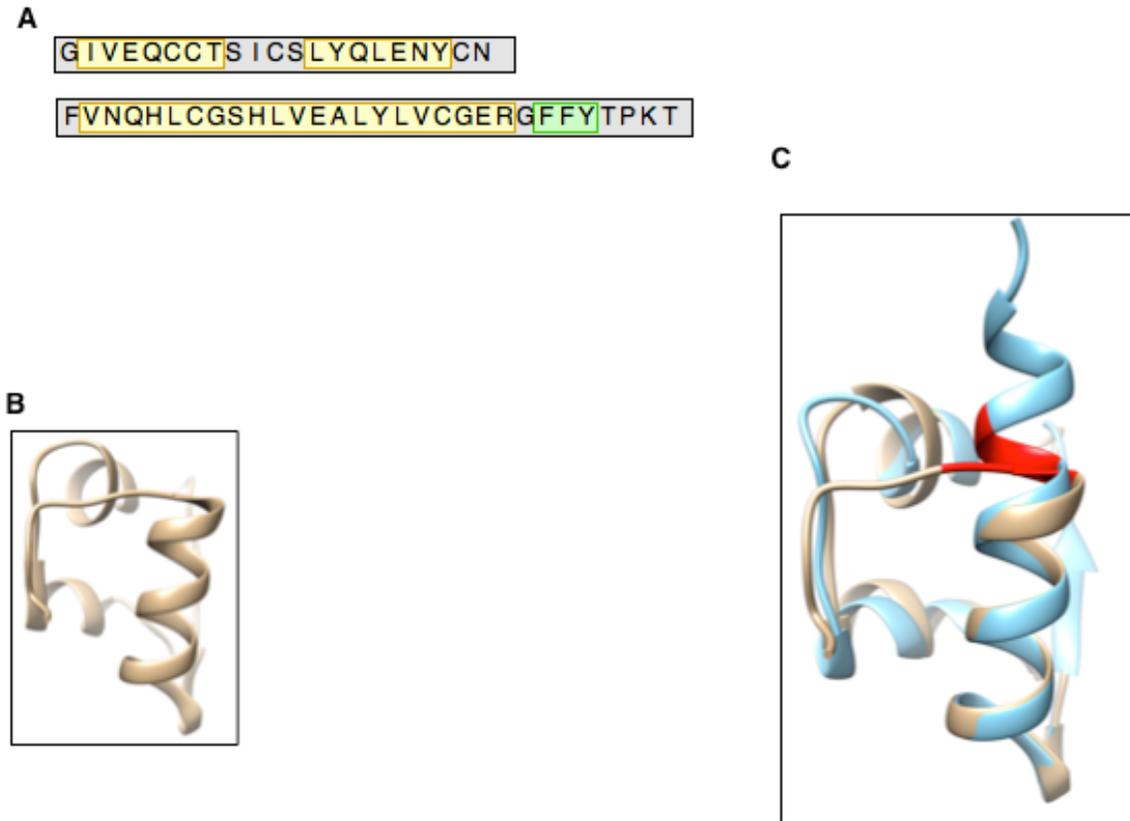


Figure 1.1 | Insulin sequence and structure.

- A) Insulin primary sequence, helices are highlighted in yellow and the triple hydrophobic patch on the B-chain is highlighted in green. The A-chain of insulin comprises 21 amino acids and is listed on top of the B-chain. The B-chain of insulin comprises 30 amino acids and is listed second.
- B) Insulin (PDB: 3T2A). Closest to the reader on the right is the alpha helix of the B-chain. Behind the alpha helix and still to the right is the beta-strand of the B-chain. Closest to the reader, and on the left, is the N-terminal turn segment of the B-chain. Behind the B-chain and on the left are the two helices of the A-chain.
- C) Alignment of T_2 (Tan) and R_6 insulins (Blue) (PDB IDs: 3T2A and 1EV6) highlighting the R vs T state. Residues GlyA8 and CysA7 are colored red. Closest to the reader on the right is the alpha helix of the B-chain. Behind the alpha helix and still to the right is the beta-strand of the B-chain. Closest to the reader, and on the left, is the N-terminal turn segment of the B-chain. This turn is not present in the R_6 monomer, instead the residues lengthen the B-chain helix on the right. Behind the B-chain and on the left are the two helices of the A-chain.

A

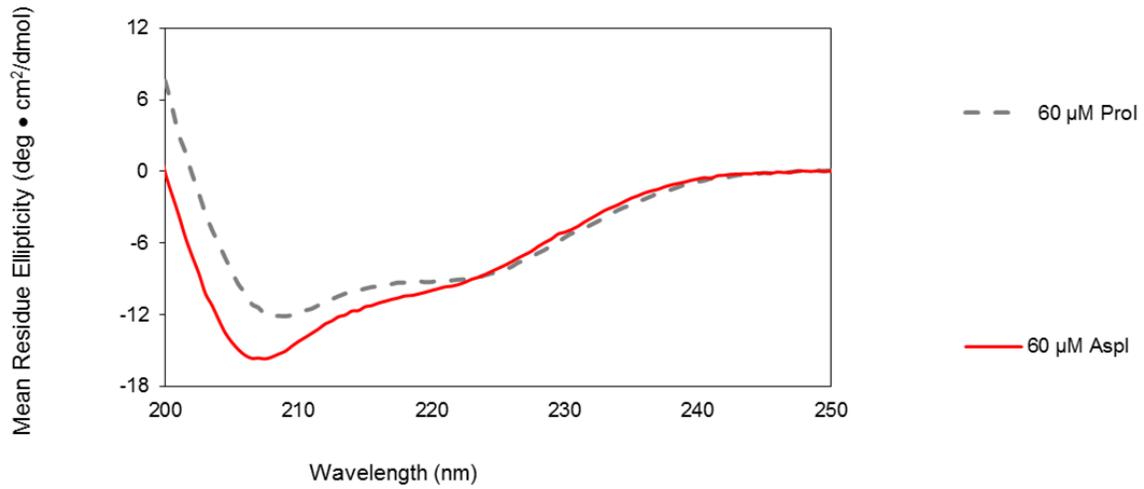
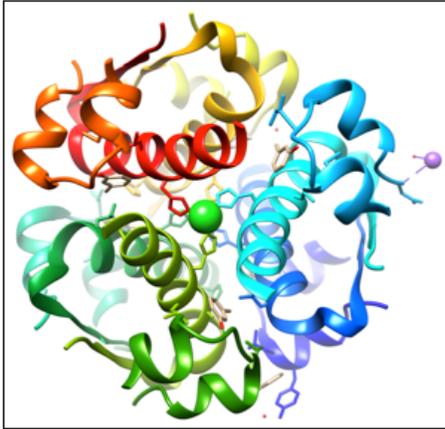


Figure 1.2 | Circular dichroism spectra of representative monomeric and dimeric insulins

A) CD spectra of 60uM insulin in 10mM Phosphate pH 8.0. In dotted gray, Wt-insulin (ProI) and in red, a representative monomeric insulin (Aspart or AspI).

A



B

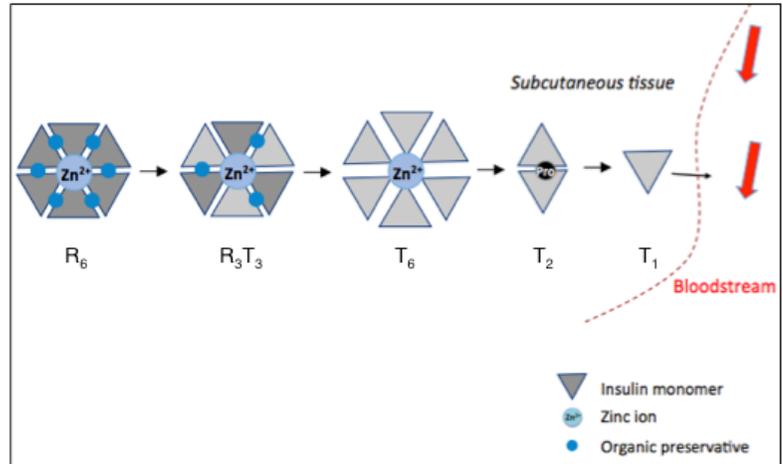


Figure 1.3 | The insulin hexamer and its disassociation

- A) The R₆ insulin hexamer (PDB: 1EV6) viewed down the threefold axis around the bound zinc and chloride ions. Each chain is colored by a different color.
- B) The general scheme of insulin hexamer disassociation from the R₆ state to the free monomer and later fibrillation from the monomer. R-state insulin monomers are dark gray, T-state monomers are in light gray. Figure adapted from: Freeman, J. S. (2009). *J Am Osteopath Assoc* 109(1): 26-36. and Birnbaum, D. T., et al. (1997). *Pharm Res* 14(1): 25-36.

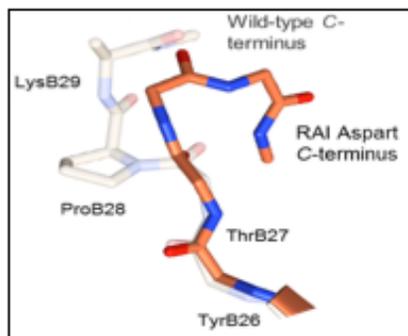
A



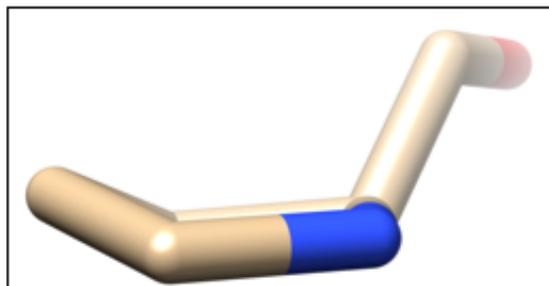
Figure 1.4 | The interaction between insulin and the insulin receptor

- A) The complex between insulin and the insulin receptor (PDB: 4OGA). The two insulin chains are depicted in light blue and dark blue. Residues B26-B30 of the B-chain of insulin (not visualized in the structure) are displaced by a helix originating from the insulin receptor (Red). Other portions of the receptor are colored green, cyan and yellow.

A



B



C

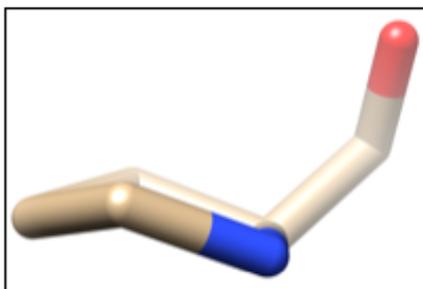


Figure 1.5 | Position B28 and proline structures

- A) The contrast between ProI and AspI as seen at the C-terminus of the B-chain (PDB IDs: 3T2A and 1ZEH)
- B) Proline in an endo pucker
- C) Proline in an exo pucker

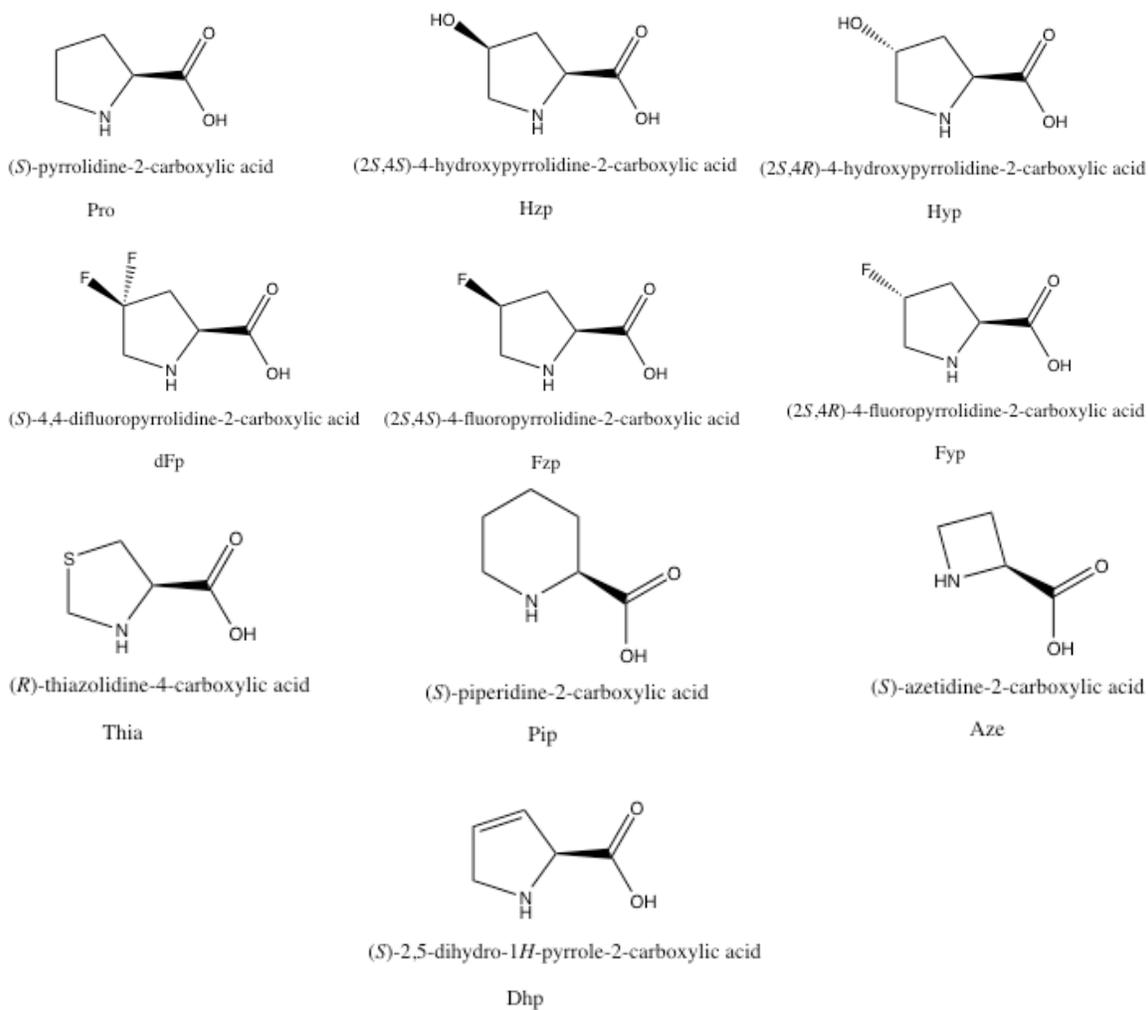


Figure 1.6 | Position B28 and proline structures

A) Proline and the nine analogs discussed in this work and incorporated into insulin. Short amino acid abbreviations are listed below the full IUPAC name for each analog.

References

- 1 Stephen L. Aronoff, Kathy Berkowitz, Barb Shreiner, and Laura Want, 'Glucose Metabolism and Regulation: Beyond Insulin and Glucagon', *Diabetes Spectrum*, 17 (2004), 183-90.
- 2 R. J. Brown, R. C. Wijewickrama, D. M. Harlan, and K. I. Rother, 'Uncoupling Intensive Insulin Therapy from Weight Gain and Hypoglycemia in Type 1 Diabetes', *Diabetes Technol Ther*, 13 (2011), 457-60.
- 3 Z. G. Hua, L. J. Xiong, C. Yan, D. H. Wei, Z. YingPai, Z. Y. Qing, Q. Z. Lin, F. R. Fei, W. Y. Ling, and M. Z. Ren, 'Glucose and Insulin Stimulate Lipogenesis in Porcine Adipocytes: Dissimilar and Identical Regulation Pathway for Key Transcription Factors', *Mol Cells*, 39 (2016), 797-806.
- 4 R. Herring, R. H. Jones, and D. L. Russell-Jones, 'Hepatoselectivity and the Evolution of Insulin', *Diabetes Obes Metab*, 16 (2014), 1-8.
- 5 S. M. Gray, R. I. Meijer, and E. J. Barrett, 'Insulin Regulates Brain Function, but How Does It Get There?', *Diabetes*, 63 (2014), 3992-7.
- 6 J. M. Gregory, G. Kraft, M. F. Scott, D. W. Neal, B. Farmer, M. S. Smith, J. R. Hastings, E. J. Allen, E. P. Donahue, N. Rivera, J. J. Winnick, D. S. Edgerton, E. Nishimura, C. Fledelius, C. L. Brand, and A. D. Cherrington, 'Insulin Delivery into the Peripheral Circulation: A Key Contributor to Hypoglycemia in Type 1 Diabetes', *Diabetes*, 64 (2015), 3439-51.
- 7 A. N. Zaykov, J. P. Mayer, and R. D. DiMarchi, 'Pursuit of a Perfect Insulin', *Nat Rev Drug Discov*, 15 (2016), 425-39.
- 8 I. B. Hirsch, 'Insulin Analogues', *N Engl J Med*, 352 (2005), 174-83.
- 9 W. K. Ward, D. C. Bolgiano, B. McKnight, J. B. Halter, and D. Porte, Jr., 'Diminished B Cell Secretory Capacity in Patients with Noninsulin-Dependent Diabetes Mellitus', *J Clin Invest*, 74 (1984), 1318-28.
- 10 K. Rave, P. N. Sidharta, J. Dingemans, L. Heinemann, and K. Roggen, 'First-Phase Insulin Secretion Has Limited Impact on Postprandial Glycemia in Subjects with Type 2 Diabetes: Correlations between Hyperglycemic Glucose Clamp and Meal Test', *Diabetes Technol Ther*, 12 (2010), 117-23.
- 11 M. A. Atkinson, and N. K. Maclaren, 'The Pathogenesis of Insulin-Dependent Diabetes Mellitus', *N Engl J Med*, 331 (1994), 1428-36.
- 12 S. E. Kahn, M. E. Cooper, and S. Del Prato, 'Pathophysiology and Treatment of Type 2 Diabetes: Perspectives on the Past, Present, and Future', *Lancet*, 383 (2014), 1068-83.
- 13 R. N. Bergman, D. T. Finegood, and S. E. Kahn, 'The Evolution of Beta-Cell Dysfunction and Insulin Resistance in Type 2 Diabetes', *Eur J Clin Invest*, 32 Suppl 3 (2002), 35-45.
- 14 A. Menke, S. Casagrande, L. Geiss, and C. C. Cowie, 'Prevalence of and Trends in Diabetes among Adults in the United States, 1988-2012', *Jama*, 314 (2015), 1021-9.
- 15 Center for Disease Control and Prevention, 'Long-Term Trends in Diabetes', ed. by Health and Human Services Center for Disease Control and Prevention, (2016).

- 16 Center for Disease Control And Prevention, 'Crude and Age-Adjusted Rates of Diagnosed Diabetes Per 100 Civilian, Non-Institutionalized Adult Population, United States, 1980–2014', ed. by Health and Human Services Centers for Disease Control and Prevention, (2016).
- 17 'Economic Costs of Diabetes in the U.S. In 2012', *Diabetes Care*, 36 (2013), 1033-46.
- 18 A. Stokes, and S. H. Preston, 'Deaths Attributable to Diabetes in the United States: Comparison of Data Sources and Estimation Approaches', *PLoS One*, 12 (2017), e0170219.
- 19 M. Y. Leung, L. M. Pollack, G. A. Colditz, and S. H. Chang, 'Life Years Lost and Lifetime Health Care Expenditures Associated with Diabetes in the U.S., National Health Interview Survey, 1997-2000', *Diabetes Care*, 38 (2015), 460-8.
- 20 Z. Wang, and M. Liu, 'Life Years Lost Associated with Diabetes: An Individually Matched Cohort Study Using the U.S. National Health Interview Survey Data', *Diabetes Res Clin Pract*, 118 (2016), 69-76.
- 21 K. M. Narayan, J. P. Boyle, T. J. Thompson, S. W. Sorensen, and D. F. Williamson, 'Lifetime Risk for Diabetes Mellitus in the United States', *Jama*, 290 (2003), 1884-90.
- 22 R. Lozano, M. Naghavi, K. Foreman, S. Lim, K. Shibuya, V. Aboyans, J. Abraham, T. Adair, R. Aggarwal, S. Y. Ahn, M. Alvarado, H. R. Anderson, L. M. Anderson, K. G. Andrews, C. Atkinson, L. M. Baddour, S. Barker-Collo, D. H. Bartels, M. L. Bell, E. J. Benjamin, D. Bennett, K. Bhalla, B. Bikbov, A. Bin Abdulhak, G. Birbeck, F. Blyth, I. Bolliger, S. Boufous, C. Bucello, M. Burch, P. Burney, J. Carapetis, H. Chen, D. Chou, S. S. Chugh, L. E. Coffeng, S. D. Colan, S. Colquhoun, K. E. Colson, J. Condon, M. D. Connor, L. T. Cooper, M. Corriere, M. Cortinovis, K. C. de Vaccaro, W. Couser, B. C. Cowie, M. H. Criqui, M. Cross, K. C. Dabhadkar, N. Dahodwala, D. De Leo, L. Degenhardt, A. Delossantos, J. Denenberg, D. C. Des Jarlais, S. D. Dharmaratne, E. R. Dorsey, T. Driscoll, H. Duber, B. Ebel, P. J. Erwin, P. Espindola, M. Ezzati, V. Feigin, A. D. Flaxman, M. H. Forouzanfar, F. G. Fowkes, R. Franklin, M. Fransen, M. K. Freeman, S. E. Gabriel, E. Gakidou, F. Gaspari, R. F. Gillum, D. Gonzalez-Medina, Y. A. Halasa, D. Haring, J. E. Harrison, R. Havmoeller, R. J. Hay, B. Hoen, P. J. Hotez, D. Hoy, K. H. Jacobsen, S. L. James, R. Jasrasaria, S. Jayaraman, N. Johns, G. Karthikeyan, N. Kassebaum, A. Keren, J. P. Khoo, L. M. Knowlton, O. Kobusingye, A. Koranteng, R. Krishnamurthi, M. Lipnick, S. E. Lipshultz, S. L. Ohno, J. Mabweijano, M. F. MacIntyre, L. Mallinger, L. March, G. B. Marks, R. Marks, A. Matsumori, R. Matzopoulos, B. M. Mayosi, J. H. McAnulty, M. M. McDermott, J. McGrath, G. A. Mensah, T. R. Merriman, C. Michaud, M. Miller, T. R. Miller, C. Mock, A. O. Mocumbi, A. A. Mokdad, A. Moran, K. Mulholland, M. N. Nair, L. Naldi, K. M. Narayan, K. Nasser, P. Norman, M. O'Donnell, S. B. Omer, K. Ortblad, R. Osborne, D. Ozgediz, B. Pahari, J. D. Pandian, A. P. Rivero, R. P. Padilla, F. Perez-Ruiz, N. Perico, D. Phillips, K. Pierce, C. A. Pope, 3rd, E. Porrini, F. Pourmalek, M. Raju, D. Ranganathan, J. T. Rehm, D. B. Rein, G. Remuzzi, F. P. Rivara, T. Roberts, F. R. De Leon, L. C. Rosenfeld, L. Rushton, R. L. Sacco, J. A. Salomon, U. Sampson, E. Sanman, D. C. Schwebel, M. Segui-Gomez, D. S. Shepard, D. Singh, J.

- Singleton, K. Sliwa, E. Smith, A. Steer, J. A. Taylor, B. Thomas, I. M. Tleyjeh, J. A. Towbin, T. Truelsen, E. A. Undurraga, N. Venketasubramanian, L. Vijayakumar, T. Vos, G. R. Wagner, M. Wang, W. Wang, K. Watt, M. A. Weinstock, R. Weintraub, J. D. Wilkinson, A. D. Woolf, S. Wulf, P. H. Yeh, P. Yip, A. Zabetian, Z. J. Zheng, A. D. Lopez, C. J. Murray, M. A. AlMazroa, and Z. A. Memish, 'Global and Regional Mortality from 235 Causes of Death for 20 Age Groups in 1990 and 2010: A Systematic Analysis for the Global Burden of Disease Study 2010', *Lancet*, 380 (2012), 2095-128.
- 23 J. E. Shaw, R. A. Sicree, and P. Z. Zimmet, 'Global Estimates of the Prevalence of Diabetes for 2010 and 2030', *Diabetes Res Clin Pract*, 87 (2010), 4-14.
- 24 A. M. Rawlings, A. R. Sharrett, A. L. Schneider, J. Coresh, M. Albert, D. Couper, M. Griswold, R. F. Gottesman, L. E. Wagenknecht, B. G. Windham, and E. Selvin, 'Diabetes in Midlife and Cognitive Change over 20 Years: A Cohort Study', *Ann Intern Med*, 161 (2014), 785-93.
- 25 J. D. Piette, and E. A. Kerr, 'The Impact of Comorbid Chronic Conditions on Diabetes Care', *Diabetes Care*, 29 (2006), 725-31.
- 26 M. N. Ozieh, K. G. Bishu, C. E. Dismuke, and L. E. Egede, 'Trends in Health Care Expenditure in U.S. Adults with Diabetes: 2002-2011', *Diabetes Care*, 38 (2015), 1844-51.
- 27 X. Zhuo, P. Zhang, L. Barker, A. Albright, T. J. Thompson, and E. Gregg, 'The Lifetime Cost of Diabetes and Its Implications for Diabetes Prevention', *Diabetes Care*, 37 (2014), 2557-64.
- 28 F. B. Bailey, A. Wilson, 'Identifying Sources of Variation in Diabetes Episodes of Care with Prometheus Analytics', (Health Care Incentives Improvement Institute: Health Care Incentives Improvement Institute, 2016), p. 15.
- 29 Y. Y. Meng, M. C. Pickett, S. H. Babey, A. C. Davis, and H. Goldstein, 'Diabetes Tied to a Third of California Hospital Stays, Driving Health Care Costs Higher', *Policy Brief UCLA Cent Health Policy Res* (2014), 1-7.
- 30 M. K. Ali, K. M. Bullard, and E. W. Gregg, 'Achievement of Goals in U.S. Diabetes Care, 1999-2010', *N Engl J Med*, 369 (2013), 287-8.
- 31 L. E. Egede, M. Gebregziabher, C. E. Dismuke, C. P. Lynch, R. N. Axon, Y. Zhao, and P. D. Mauldin, 'Medication Nonadherence in Diabetes: Longitudinal Effects on Costs and Potential Cost Savings from Improvement', *Diabetes Care*, 35 (2012), 2533-9.
- 32 D. R. Owens, 'New Horizons--Alternative Routes for Insulin Therapy', *Nat Rev Drug Discov*, 1 (2002), 529-40.
- 33 K. S. Polonsky, 'The Past 200 Years in Diabetes', *N Engl J Med*, 367 (2012), 1332-40.
- 34 C. C. Quianzon, and I. Cheikh, 'History of Insulin', *J Community Hosp Intern Med Perspect*, 2 (2012).
- 35 D. T. Karamitsos, 'The Story of Insulin Discovery', *Diabetes Res Clin Pract*, 93 Suppl 1 (2011), S2-8.
- 36 L. Rosenfeld, 'Insulin: Discovery and Controversy', *Clin Chem*, 48 (2002), 2270-88.

- 37 H. C. Hagedorn, B. N. Jensen, N. B. Krarup, and I. Wodstrup, 'Landmark Article Jan 18, 1936: Protamine Insulinate. By H.C. Hagedorn, B.N. Jensen, N.B. Krarup, and I. Wodstrup', *Jama*, 251 (1984), 389-92.
- 38 J. R. White, Jr., 'A Brief History of the Development of Diabetes Medications', *Diabetes Spectr*, 27 (2014), 82-6.
- 39 F. Sanger, 'The Terminal Peptides of Insulin', *Biochem J*, 45 (1949), 563-74.
- 40 F. Sanger, 'Species Differences in Insulins', *Nature*, 164 (1949), 529.
- 41 D. F. Steiner, 'On the Discovery of Precursor Processing', *Methods Mol Biol*, 768 (2011), 3-11.
- 42 T. L. Blundell, J. F. Cutfield, S. M. Cutfield, E. J. Dodson, G. G. Dodson, D. C. Hodgkin, D. A. Mercola, and M. Vijayan, 'Atomic Positions in Rhombohedral 2-Zinc Insulin Crystals', *Nature*, 231 (1971), 506-11.
- 43 D. V. Goeddel, D. G. Kleid, F. Bolivar, H. L. Heyneker, D. G. Yansura, R. Crea, T. Hirose, A. Kraszewski, K. Itakura, and A. D. Riggs, 'Expression in Escherichia Coli of Chemically Synthesized Genes for Human Insulin', *Proc Natl Acad Sci U S A*, 76 (1979), 106-10.
- 44 L. M. Keefer, M. A. Piron, and P. De Meyts, 'Human Insulin Prepared by Recombinant DNA Techniques and Native Human Insulin Interact Identically with Insulin Receptors', *Proc Natl Acad Sci U S A*, 78 (1981), 1391-5.
- 45 I. S. Johnson, 'Human Insulin from Recombinant DNA Technology', *Science*, 219 (1983), 632-7.
- 46 L. Thim, M. T. Hansen, K. Norris, I. Hoegh, E. Boel, J. Forstrom, G. Ammerer, and N. P. Fiil, 'Secretion and Processing of Insulin Precursors in Yeast', *Proc Natl Acad Sci U S A*, 83 (1986), 6766-70.
- 47 T. Kjeldsen, 'Yeast Secretory Expression of Insulin Precursors', *Appl Microbiol Biotechnol*, 54 (2000), 277-86.
- 48 C. K. Min, Y. J. Son, C. K. Kim, S. J. Park, and J. W. Lee, 'Increased Expression, Folding and Enzyme Reaction Rate of Recombinant Human Insulin by Selecting Appropriate Leader Peptide', *J Biotechnol*, 151 (2011), 350-6.
- 49 P. G. Katsoyannis, A. M. Tometsko, J. Z. Ginos, and M. A. Tilak, 'Insulin Peptides. Xi. The Synthesis of the B Chain of Human Insulin and Its Combination with the Natural a Chain of Bovine Insulin to Generate Insulin Activity', *J Am Chem Soc*, 88 (1966), 164-6.
- 50 Y. T. Kung, Y. C. Du, W. T. Huang, C. C. Chen, and L. T. Ke, 'Total Synthesis of Crystalline Bovine Insulin', *Sci Sin*, 14 (1965), 1710-6.
- 51 J. G. Menting, J. Gajewiak, C. A. MacRaid, D. H. Chou, M. M. Disotuar, N. A. Smith, C. Miller, J. Erchegy, J. E. Rivier, B. M. Olivera, B. E. Forbes, B. J. Smith, R. S. Norton, H. Safavi-Hemami, and M. C. Lawrence, 'A Minimized Human Insulin-Receptor-Binding Motif Revealed in a Conus Geographus Venom Insulin', *Nat Struct Mol Biol*, 23 (2016), 916-20.
- 52 F. Liu, A. N. Zaykov, J. J. Levy, R. D. DiMarchi, and J. P. Mayer, 'Chemical Synthesis of Peptides within the Insulin Superfamily', *J Pept Sci*, 22 (2016), 260-70.
- 53 N. A. Baeshen, M. N. Baeshen, A. Sheikh, R. S. Bora, M. M. Ahmed, H. A. Ramadan, K. S. Saini, and E. M. Redwan, 'Cell Factories for Insulin Production', *Microb Cell Fact*, 13 (2014), 141.

- 54 D. C. Howey, R. R. Bowsher, R. L. Brunelle, and J. R. Woodworth, '[Lys(B28), Pro(B29)]-Human Insulin. A Rapidly Absorbed Analogue of Human Insulin', *Diabetes*, 43 (1994), 396-402.
- 55 D. C. Howey, R. R. Bowsher, R. L. Brunelle, H. M. Rowe, P. F. Santa, J. Downing-Shelton, and J. R. Woodworth, '[Lys(B28), Pro(B29)]-Human Insulin: Effect of Injection Time on Postprandial Glycemia', *Clin Pharmacol Ther*, 58 (1995), 459-69.
- 56 L. Quagliaro, L. Piconi, R. Assaloni, L. Martinelli, E. Motz, and A. Ceriello, 'Intermittent High Glucose Enhances Apoptosis Related to Oxidative Stress in Human Umbilical Vein Endothelial Cells: The Role of Protein Kinase C and Nad(P)H-Oxidase Activation', *Diabetes*, 52 (2003), 2795-804.
- 57 J. H. Anderson, Jr., R. L. Brunelle, V. A. Koivisto, A. Pfoetzner, M. E. Trautmann, L. Vignati, and R. DiMarchi, 'Reduction of Postprandial Hyperglycemia and Frequency of Hypoglycemia in Iddm Patients on Insulin-Analog Treatment. Multicenter Insulin Lispro Study Group', *Diabetes*, 46 (1997), 265-70.
- 58 A. Pfoetzner, E. Kustner, T. Forst, B. Schulze-Schleppinghoff, M. E. Trautmann, M. Haslbeck, H. Schatz, and J. Beyer, 'Intensive Insulin Therapy with Insulin Lispro in Patients with Type 1 Diabetes Reduces the Frequency of Hypoglycemic Episodes', *Exp Clin Endocrinol Diabetes*, 104 (1996), 25-30.
- 59 A. Siebenhofer, J. Plank, A. Berghold, K. Jeitler, K. Horvath, M. Narath, R. Gfrerer, and T. R. Pieber, 'Short Acting Insulin Analogues Versus Regular Human Insulin in Patients with Diabetes Mellitus', *Cochrane Database Syst Rev* (2006), Cd003287.
- 60 B. Fullerton, A. Siebenhofer, K. Jeitler, K. Horvath, T. Semlitsch, A. Berghold, J. Plank, T. R. Pieber, and F. M. Gerlach, 'Short-Acting Insulin Analogues Versus Regular Human Insulin for Adults with Type 1 Diabetes Mellitus', *Cochrane Database Syst Rev* (2016), Cd012161.
- 61 D. N. Brems, L. A. Alter, M. J. Beckage, R. E. Chance, R. D. DiMarchi, L. K. Green, H. B. Long, A. H. Pekar, J. E. Shields, and B. H. Frank, 'Altering the Association Properties of Insulin by Amino Acid Replacement', *Protein Eng*, 5 (1992), 527-33.
- 62 S. R. Mudaliar, F. A. Lindberg, M. Joyce, P. Beerdsen, P. Strange, A. Lin, and R. R. Henry, 'Insulin Aspart (B28 Asp-Insulin): A Fast-Acting Analog of Human Insulin: Absorption Kinetics and Action Profile Compared with Regular Human Insulin in Healthy Nondiabetic Subjects', *Diabetes Care*, 22 (1999), 1501-6.
- 63 A. Plum, H. Agerso, and L. Andersen, 'Pharmacokinetics of the Rapid-Acting Insulin Analog, Insulin Aspart, in Rats, Dogs, and Pigs, and Pharmacodynamics of Insulin Aspart in Pigs', *Drug Metab Dispos*, 28 (2000), 155-60.
- 64 S. E. Shoelson, Z. X. Lu, L. Parlautan, C. S. Lynch, and M. A. Weiss, 'Mutations at the Dimer, Hexamer, and Receptor-Binding Surfaces of Insulin Independently Affect Insulin-Insulin and Insulin-Receptor Interactions', *Biochemistry*, 31 (1992), 1757-67.
- 65 T. Lauritzen, S. Pramming, T. Deckert, and C. Binder, 'Pharmacokinetics of Continuous Subcutaneous Insulin Infusion', *Diabetologia*, 24 (1983), 326-9.
- 66 D. L. Bakaysa, J. Radziuk, H. A. Havel, M. L. Brader, S. Li, S. W. Dodd, J. M. Beals, A. H. Pekar, and D. N. Brems, 'Physicochemical Basis for the Rapid Time-

- Action of Lysb28prob29-Insulin: Dissociation of a Protein-Ligand Complex', *Protein Sci*, 5 (1996), 2521-31.
- 67 L. Meadowcraft, G. Mospan, T. Morrisette, K. Smart, and M. Janis, 'Drug Updates and Approvals: 2016 in Review', *Nurse Pract*, 41 (2016), 20-27.
- 68 C. Sorli, M. Warren, D. Oyer, H. Mersebach, T. Johansen, and S. C. L. Gough, 'Elderly Patients with Diabetes Experience a Lower Rate of Nocturnal Hypoglycaemia with Insulin Degludec Than with Insulin Glargine: A Meta-Analysis of Phase Iiiia Trials', *Drugs Aging*, 30 (2013), 1009-18.
- 69 J. Rosenstock, V. Fonseca, S. Schinzel, M. P. Dain, P. Mullins, and M. Riddle, 'Reduced Risk of Hypoglycemia with Once-Daily Glargine Versus Twice-Daily Nph and Number Needed to Harm with Nph to Demonstrate the Risk of One Additional Hypoglycemic Event in Type 2 Diabetes: Evidence from a Long-Term Controlled Trial', *J Diabetes Complications*, 28 (2014), 742-9.
- 70 V. Pandyarajan, and M. A. Weiss, 'Design of Non-Standard Insulin Analogs for the Treatment of Diabetes Mellitus', *Curr Diab Rep*, 12 (2012), 697-704.
- 71 D. M. Nathan, D. E. Singer, K. Hurxthal, and J. D. Goodson, 'The Clinical Information Value of the Glycosylated Hemoglobin Assay', *N Engl J Med*, 310 (1984), 341-6.
- 72 W. V. Tamborlane, R. W. Beck, B. W. Bode, B. Buckingham, H. P. Chase, R. Clemons, R. Fiallo-Scharer, L. A. Fox, L. K. Gilliam, I. B. Hirsch, E. S. Huang, C. Kollman, A. J. Kowalski, L. Laffel, J. M. Lawrence, J. Lee, N. Mauras, M. O'Grady, K. J. Ruedy, M. Tansey, E. Tsalikian, S. Weinzimer, D. M. Wilson, H. Wolpert, T. Wysocki, and D. Xing, 'Continuous Glucose Monitoring and Intensive Treatment of Type 1 Diabetes', *N Engl J Med*, 359 (2008), 1464-76.
- 73 M. Langendam, Y. M. Luijf, L. Hooft, J. H. Devries, A. H. Mudde, and R. J. Scholten, 'Continuous Glucose Monitoring Systems for Type 1 Diabetes Mellitus', *Cochrane Database Syst Rev*, 1 (2012), Cd008101.
- 74 W. V. Tamborlane, R. S. Sherwin, M. Genel, and P. Felig, 'Reduction to Normal of Plasma Glucose in Juvenile Diabetes by Subcutaneous Administration of Insulin with a Portable Infusion Pump', *N Engl J Med*, 300 (1979), 573-8.
- 75 H. Thabit, and R. Hovorka, 'Coming of Age: The Artificial Pancreas for Type 1 Diabetes', *Diabetologia*, 59 (2016), 1795-805.
- 76 J. C. Pickup, 'Diabetes: Insulin Pump Therapy for Type 2 Diabetes Mellitus', *Nat Rev Endocrinol*, 10 (2014), 647-9.
- 77 J. C. Pickup, and E. Renard, 'Long-Acting Insulin Analogs Versus Insulin Pump Therapy for the Treatment of Type 1 and Type 2 Diabetes', *Diabetes Care*, 31 Suppl 2 (2008), S140-5.
- 78 F. M. Alsaleh, F. J. Smith, S. Keady, and K. M. Taylor, 'Insulin Pumps: From Inception to the Present and toward the Future', *J Clin Pharm Ther*, 35 (2010), 127-38.
- 79 P. Home, M. Riddle, W. T. Cefalu, C. J. Bailey, R. G. Bretzel, S. Del Prato, D. Leroith, G. Scherthaner, L. van Gaal, and I. Raz, 'Insulin Therapy in People with Type 2 Diabetes: Opportunities and Challenges?', *Diabetes Care*, 37 (2014), 1499-508.
- 80 D. F. Steiner, S. J. Chan, J. M. Welsh, and S. C. Kwok, 'Structure and Evolution of the Insulin Gene', *Annu Rev Genet*, 19 (1985), 463-84.

- 81 T. N. Vinther, T. B. Kjeldsen, K. J. Jensen, and F. Hubalek, 'The Road to the First, Fully Active and More Stable Human Insulin Variant with an Additional Disulfide Bond', *J Pept Sci*, 21 (2015), 797-806.
- 82 J. D. Peterson, and D. F. Steiner, 'The Amino Acid Sequence of the Insulin from a Primitive Vertebrate, the Atlantic Hagfish (*Myxine Glutinososa*)', *J Biol Chem*, 250 (1975), 5183-91.
- 83 E. N. Baker, T. L. Blundell, J. F. Cutfield, S. M. Cutfield, E. J. Dodson, G. G. Dodson, D. M. Hodgkin, R. E. Hubbard, N. W. Isaacs, C. D. Reynolds, and et al., 'The Structure of 2zn Pig Insulin Crystals at 1.5 a Resolution', *Philos Trans R Soc Lond B Biol Sci*, 319 (1988), 369-456.
- 84 J. L. Whittingham, D. J. Edwards, A. A. Antson, J. M. Clarkson, and G. G. Dodson, 'Interactions of Phenol and M-Cresol in the Insulin Hexamer, and Their Effect on the Association Properties of B28 Pro --> Asp Insulin Analogues', *Biochemistry*, 37 (1998), 11516-23.
- 85 H. Marshall, M. Venkat, N. S. Seng, J. Cahn, and D. H. Juers, 'The Use of Trimethylamine N-Oxide as a Primary Precipitating Agent and Related Methylamine Osmolytes as Cryoprotective Agents for Macromolecular Crystallography', *Acta Crystallogr D Biol Crystallogr*, 68 (2012), 69-81.
- 86 J. P. Mayer, F. Zhang, and R. D. DiMarchi, 'Insulin Structure and Function', *Biopolymers*, 88 (2007), 687-713.
- 87 L. Kosinova, V. Veverka, P. Novotna, M. Collinsova, M. Urbanova, N. R. Moody, J. P. Turkenburg, J. Jiracek, A. M. Brzozowski, and L. Zakova, 'Insight into the Structural and Biological Relevance of the T/R Transition of the N-Terminus of the B-Chain in Human Insulin', *Biochemistry*, 53 (2014), 3392-402.
- 88 J. Monod, J. Wyman, and J. P. Changeux, 'On the Nature of Allosteric Transitions: A Plausible Model', *J Mol Biol*, 12 (1965), 88-118.
- 89 M. A. Weiss, 'The Structure and Function of Insulin: Decoding the Tr Transition', *Vitam Horm*, 80 (2009), 33-49.
- 90 G. D. Smith, W. A. Pangborn, and R. H. Blessing, 'The Structure of T6 Human Insulin at 1.0 a Resolution', *Acta Crystallogr D Biol Crystallogr*, 59 (2003), 474-82.
- 91 G. D. Smith, W. A. Pangborn, R. H. Blessing. 'The Structure of T6 Bovine Insulin', *Acta Crystallogr D Biol Crystallogr*, 61 (2005), 1476-82.
- 92 P. T. Grant, T. L. Coombs, and B. H. Frank, 'Differences in the Nature of the Interaction of Insulin and Proinsulin with Zinc', *Biochem J*, 126 (1972), 433-40.
- 93 J. Goldman, and F. H. Carpenter, 'Zinc Binding, Circular Dichroism, and Equilibrium Sedimentation Studies on Insulin (Bovine) and Several of Its Derivatives', *Biochemistry*, 13 (1974), 4566-74.
- 94 G. D. Smith, E. Ciszak, L. A. Magrum, W. A. Pangborn, and R. H. Blessing, 'R6 Hexameric Insulin Complexed with M-Cresol or Resorcinol', *Acta Crystallogr D Biol Crystallogr*, 56 (2000), 1541-8.
- 95 D. T. Birnbaum, S. W. Dodd, B. E. Saxberg, A. D. Varshavsky, and J. M. Beals, 'Hierarchical Modeling of Phenolic Ligand Binding to 2zn--Insulin Hexamers', *Biochemistry*, 35 (1996), 5366-78.

- 96 K. Huus, S. Havelund, H. B. Olsen, B. W. Sigurskjold, M. van de Weert, and S. Frokjaer, 'Ligand Binding and Thermostability of Different Allosteric States of the Insulin Zinc-Hexamer', *Biochemistry*, 45 (2006), 4014-24.
- 97 C. G. Frankaer, M. V. Knudsen, K. Noren, E. Nazarenko, K. Stahl, and P. Harris, 'The Structures of T6, T3r3 and R6 Bovine Insulin: Combining X-Ray Diffraction and Absorption Spectroscopy', *Acta Crystallogr D Biol Crystallogr*, 68 (2012), 1259-71.
- 98 S. Rahuel-Clermont, C. A. French, N. C. Kaarsholm, M. F. Dunn, and C. I. Chou, 'Mechanisms of Stabilization of the Insulin Hexamer through Allosteric Ligand Interactions', *Biochemistry*, 36 (1997), 5837-45.
- 99 P. S. Brzovic, W. E. Choi, D. Borchardt, N. C. Kaarsholm, and M. F. Dunn, 'Structural Asymmetry and Half-Site Reactivity in the T to R Allosteric Transition of the Insulin Hexamer', *Biochemistry*, 33 (1994), 13057-69.
- 100 M. Roy, M. L. Brader, R. W. Lee, N. C. Kaarsholm, J. F. Hansen, and M. F. Dunn, 'Spectroscopic Signatures of the T to R Conformational Transition in the Insulin Hexamer', *J Biol Chem*, 264 (1989), 19081-5.
- 101 N. C. Kaarsholm, H. C. Ko, and M. F. Dunn, 'Comparison of Solution Structural Flexibility and Zinc Binding Domains for Insulin, Proinsulin, and Miniproinsulin', *Biochemistry*, 28 (1989), 4427-35.
- 102 M. F. Dunn, 'Zinc-Ligand Interactions Modulate Assembly and Stability of the Insulin Hexamer -- a Review', *Biometals*, 18 (2005), 295-303.
- 103 M. C. Carpenter, and D. E. Wilcox, 'Thermodynamics of Formation of the Insulin Hexamer: Metal-Stabilized Proton-Coupled Assembly of Quaternary Structure', *Biochemistry*, 53 (2014), 1296-301.
- 104 A. H. Pekar, and B. H. Frank, 'Conformation of Proinsulin. A Comparison of Insulin and Proinsulin Self-Association at Neutral Ph', *Biochemistry*, 11 (1972), 4013-6.
- 105 B. H. Frank, A. H. Pekar, and A. J. Veros, 'Insulin and Proinsulin Conformation in Solution', *Diabetes*, 21 (1972), 486-91.
- 106 S. M. Liao, Q. S. Du, J. Z. Meng, Z. W. Pang, and R. B. Huang, 'The Multiple Roles of Histidine in Protein Interactions', *Chem Cent J*, 7 (2013), 44.
- 107 J. Brange, U. Ribbel, J. F. Hansen, G. Dodson, M. T. Hansen, S. Havelund, S. G. Melberg, F. Norris, K. Norris, L. Snel, and et al., 'Monomeric Insulins Obtained by Protein Engineering and Their Medical Implications', *Nature*, 333 (1988), 679-82.
- 108 G. Milazzo, L. Sciacca, V. Papa, I. D. Goldfine, and R. Vigneri, 'Aspb10 Insulin Induction of Increased Mitogenic Responses and Phenotypic Changes in Human Breast Epithelial Cells: Evidence for Enhanced Interactions with the Insulin-Like Growth Factor-I Receptor', *Mol Carcinog*, 18 (1997), 19-25.
- 109 K. E. Bornfeldt, R. A. Gidlof, A. Wasteson, M. Lake, A. Skottner, and H. J. Arnqvist, 'Binding and Biological Effects of Insulin, Insulin Analogues and Insulin-Like Growth Factors in Rat Aortic Smooth Muscle Cells. Comparison of Maximal Growth Promoting Activities', *Diabetologia*, 34 (1991), 307-13.
- 110 C. W. Ward, and M. C. Lawrence, 'Ligand-Induced Activation of the Insulin Receptor: A Multi-Step Process Involving Structural Changes in Both the Ligand and the Receptor', *Bioessays*, 31 (2009), 422-34.

- 111 A. Papaioannou, S. Kuyucak, and Z. Kuncic, 'Elucidating the Activation Mechanism of the Insulin-Family Proteins with Molecular Dynamics Simulations', *PLoS One*, 11 (2016), e0161459.
- 112 T. I. Croll, B. J. Smith, M. B. Margetts, J. Whittaker, M. A. Weiss, C. W. Ward, and M. C. Lawrence, 'Higher-Resolution Structure of the Human Insulin Receptor Ectodomain: Multi-Modal Inclusion of the Insert Domain', *Structure*, 24 (2016), 469-76.
- 113 C. Kristensen, T. Kjeldsen, F. C. Wiberg, L. Schaffer, M. Hach, S. Havelund, J. Bass, D. F. Steiner, and A. S. Andersen, 'Alanine Scanning Mutagenesis of Insulin', *J Biol Chem*, 272 (1997), 12978-83.
- 114 S. H. Nakagawa, and H. S. Tager, 'Importance of Aliphatic Side-Chain Structure at Positions 2 and 3 of the Insulin a Chain in Insulin-Receptor Interactions', *Biochemistry*, 31 (1992), 3204-14.
- 115 J. G. Menting, J. Whittaker, M. B. Margetts, L. J. Whittaker, G. K. Kong, B. J. Smith, C. J. Watson, L. Zakova, E. Kletvikova, J. Jiracek, S. J. Chan, D. F. Steiner, G. G. Dodson, A. M. Brzozowski, M. A. Weiss, C. W. Ward, and M. C. Lawrence, 'How Insulin Engages Its Primary Binding Site on the Insulin Receptor', *Nature*, 493 (2013), 241-5.
- 116 P. De Meyts, and J. Whittaker, 'Structural Biology of Insulin and Igf1 Receptors: Implications for Drug Design', *Nat Rev Drug Discov*, 1 (2002), 769-83.
- 117 T. Glendorf, A. R. Sorensen, E. Nishimura, I. Pettersson, and T. Kjeldsen, 'Importance of the Solvent-Exposed Residues of the Insulin B Chain Alpha-Helix for Receptor Binding', *Biochemistry*, 47 (2008), 4743-51.
- 118 M. A. Weiss, '169 How Insulin Binds: Structure of a Micro-Receptor Complex and Implications for Analog Design', *J Biomol Struct Dyn*, 33 Suppl 1 (2015), 110-1.
- 119 P. G. Katsoyannis, J. Ginos, G. P. Schwartz, and A. Cosmatos, 'Synthesis of a Biologically Active Truncated Insulin. Des(Pentapeptide B23-30) Human (Porcine) Insulin', *J Chem Soc Perkin I*, 11 (1974), 1311-7.
- 120 I. Jonassen, S. Havelund, T. Hoeg-Jensen, D. B. Steensgaard, P. O. Wahlund, and U. Ribel, 'Design of the Novel Protraction Mechanism of Insulin Degludec, an Ultra-Long-Acting Basal Insulin', *Pharm Res*, 29 (2012), 2104-14.
- 121 S. Havelund, A. Plum, U. Ribel, I. Jonassen, A. Volund, J. Markussen, and P. Kurtzhals, 'The Mechanism of Protraction of Insulin Detemir, a Long-Acting, Acylated Analog of Human Insulin', *Pharm Res*, 21 (2004), 1498-504.
- 122 J. G. Menting, Y. Yang, S. J. Chan, N. B. Phillips, B. J. Smith, J. Whittaker, N. P. Wickramasinghe, L. J. Whittaker, V. Pandeyarajan, Z. L. Wan, S. P. Yadav, J. M. Carroll, N. Strokes, C. T. Roberts, Jr., F. Ismail-Beigi, W. Milewski, D. F. Steiner, V. S. Chauhan, C. W. Ward, M. A. Weiss, and M. C. Lawrence, 'Protective Hinge in Insulin Opens to Enable Its Receptor Engagement', *Proc Natl Acad Sci U S A*, 111 (2014), E3395-404.
- 123 S. H. Nakagawa, and H. S. Tager, 'Role of the Phenylalanine B25 Side Chain in Directing Insulin Interaction with Its Receptor. Steric and Conformational Effects', *J Biol Chem*, 261 (1986), 7332-41.
- 124 V. Pandeyarajan, B. J. Smith, N. B. Phillips, L. Whittaker, G. P. Cox, N. Wickramasinghe, J. G. Menting, Z. L. Wan, J. Whittaker, F. Ismail-Beigi, M. C.

- Lawrence, and M. A. Weiss, 'Aromatic Anchor at an Invariant Hormone-Receptor Interface: Function of Insulin Residue B24 with Application to Protein Design', *J Biol Chem*, 289 (2014), 34709-27.
- 125 E. Ciszak, J. M. Beals, B. H. Frank, J. C. Baker, N. D. Carter, and G. D. Smith, 'Role of C-Terminal B-Chain Residues in Insulin Assembly: The Structure of Hexameric Lysb28prob29-Human Insulin', *Structure*, 3 (1995), 615-22.
- 126 T. N. Vinther, M. Norrman, H. M. Strauss, K. Huus, M. Schlein, T. A. Pedersen, T. Kjeldsen, K. J. Jensen, and F. Hubalek, 'Novel Covalently Linked Insulin Dimer Engineered to Investigate the Function of Insulin Dimerization', *PLoS One*, 7 (2012), e30882.
- 127 V. Zoete, M. Meuwly, and M. Karplus, 'A Comparison of the Dynamic Behavior of Monomeric and Dimeric Insulin Shows Structural Rearrangements in the Active Monomer', *J Mol Biol*, 342 (2004), 913-29.
- 128 A. Papaioannou, S. Kuyucak, and Z. Kuncic, 'Molecular Dynamics Simulations of Insulin: Elucidating the Conformational Changes That Enable Its Binding', *PLoS One*, 10 (2015), e0144058.
- 129 Q. X. Hua, B. Xu, K. Huang, S. Q. Hu, S. Nakagawa, W. Jia, S. Wang, J. Whittaker, P. G. Katsoyannis, and M. A. Weiss, 'Enhancing the Activity of a Protein by Stereospecific Unfolding: Conformational Life Cycle of Insulin and Its Evolutionary Origins', *J Biol Chem*, 284 (2009), 14586-96.
- 130 J. Brange, L. Andersen, E. D. Laursen, G. Meyn, and E. Rasmussen, 'Toward Understanding Insulin Fibrillation', *J Pharm Sci*, 86 (1997), 517-25.
- 131 Y. Yang, A. Petkova, K. Huang, B. Xu, Q. X. Hua, I. J. Ye, Y. C. Chu, S. Q. Hu, N. B. Phillips, J. Whittaker, F. Ismail-Beigi, R. B. Mackin, P. G. Katsoyannis, R. Tycko, and M. A. Weiss, 'An Achilles' Heel in an Amyloidogenic Protein and Its Repair: Insulin Fibrillation and Therapeutic Design', *J Biol Chem*, 285 (2010), 10806-21.
- 132 L. Nielsen, S. Frokjaer, J. Brange, V. N. Uversky, and A. L. Fink, 'Probing the Mechanism of Insulin Fibril Formation with Insulin Mutants', *Biochemistry*, 40 (2001), 8397-409.
- 133 V. Pandeyarajan, N. B. Phillips, N. Rege, M. C. Lawrence, J. Whittaker, and M. A. Weiss, 'Contribution of Tyrb26 to the Function and Stability of Insulin: Structure-Activity Relationships at a Conserved Hormone-Receptor Interface', *J Biol Chem*, 291 (2016), 12978-90.
- 134 K. Krizkova, V. Veverka, L. Maletinska, R. Hexnerova, A. M. Brzozowski, J. Jiracek, and L. Zakova, 'Structural and Functional Study of the Glnb22-Insulin Mutant Responsible for Maturity-Onset Diabetes of the Young', *PLoS One*, 9 (2014), e112883.
- 135 N. C. Kaarsholm, K. Norris, R. J. Jorgensen, J. Mikkelsen, S. Ludvigsen, O. H. Olsen, A. R. Sorensen, and S. Havelund, 'Engineering Stability of the Insulin Monomer Fold with Application to Structure-Activity Relationships', *Biochemistry*, 32 (1993), 10773-8.
- 136 N. B. Phillips, J. Whittaker, F. Ismail-Beigi, and M. A. Weiss, 'Insulin Fibrillation and Protein Design: Topological Resistance of Single-Chain Analogs to Thermal Degradation with Application to a Pump Reservoir', *J Diabetes Sci Technol*, 6 (2012), 277-88.

- 137 T. N. Vinther, M. Norrman, U. Ribel, K. Huus, M. Schlein, D. B. Steensgaard, T. A. Pedersen, I. Pettersson, S. Ludvigsen, T. Kjeldsen, K. J. Jensen, and F. Hubalek, 'Insulin Analog with Additional Disulfide Bond Has Increased Stability and Preserved Activity', *Protein Sci*, 22 (2013), 296-305.
- 138 M. R. DeFelippis, R. E. Chance, and B. H. Frank, 'Insulin Self-Association and the Relationship to Pharmacokinetics and Pharmacodynamics', *Crit Rev Ther Drug Carrier Syst*, 18 (2001), 201-64.
- 139 D. F. Berenson, A. R. Weiss, Z. L. Wan, and M. A. Weiss, 'Insulin Analogs for the Treatment of Diabetes Mellitus: Therapeutic Applications of Protein Engineering', *Ann N Y Acad Sci*, 1243 (2011), E40-e54.
- 140 J. A. Johnson, Y. Y. Lu, J. A. Van Deventer, and D. A. Tirrell, 'Residue-Specific Incorporation of Non-Canonical Amino Acids into Proteins: Recent Developments and Applications', *Curr Opin Chem Biol*, 14 (2010), 774-80.
- 141 L. Leisle, F. Valiyaveetil, R. A. Mehl, and C. A. Ahern, 'Incorporation of Non-Canonical Amino Acids', *Adv Exp Med Biol*, 869 (2015), 119-51.
- 142 V. E. Cox, and E. A. Gaucher, 'Molecular Evolution Directs Protein Translation Using Unnatural Amino Acids', *Curr Protoc Chem Biol*, 7 (2015), 223-8.
- 143 J. W. Chin, 'Expanding and Reprogramming the Genetic Code of Cells and Animals', *Annu Rev Biochem*, 83 (2014), 379-408.
- 144 C. C. Liu, and P. G. Schultz, 'Adding New Chemistries to the Genetic Code', *Annu Rev Biochem*, 79 (2010), 413-44.
- 145 F. Truong, T. H. Yoo, T. J. Lampo, and D. A. Tirrell, 'Two-Strain, Cell-Selective Protein Labeling in Mixed Bacterial Cultures', *J Am Chem Soc*, 134 (2012), 8551-6.
- 146 W. Kim, A. George, M. Evans, and V. P. Conticello, 'Cotranslational Incorporation of a Structurally Diverse Series of Proline Analogues in an Escherichia Coli Expression System', *Chembiochem*, 5 (2004), 928-36.
- 147 K. Madhusudan Makwana, and R. Mahalakshmi, 'Implications of Aromatic-Aromatic Interactions: From Protein Structures to Peptide Models', *Protein Sci*, 24 (2015), 1920-33.
- 148 Robert W. Newberry, and Ronald T. Raines, '4-Fluoroproline: Conformational Analysis and Effects on the Stability and Folding of Peptides and Proteins', (Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 1-25.
- 149 C. M. Venkatachalam, and G. N. Ramachandran, 'Conformation of Polypeptide Chains', *Annu Rev Biochem*, 38 (1969), 45-82.
- 150 R. W. Newberry, G. J. Bartlett, B. VanVeller, D. N. Woolfson, and R. T. Raines, 'Signatures of N \rightarrow PI* Interactions in Proteins', *Protein Sci*, 23 (2014), 284-8.
- 151 R. Bhattacharyya, and P. Chakrabarti, 'Stereospecific Interactions of Proline Residues in Protein Structures and Complexes', *J Mol Biol*, 331 (2003), 925-40.
- 152 N. J. Zondlo, 'Aromatic-Proline Interactions: Electronically Tunable Ch/Pi Interactions', *Acc Chem Res*, 46 (2013), 1039-49.
- 153 A. K. Pandey, K. M. Thomas, C. R. Forbes, and N. J. Zondlo, 'Tunable Control of Polyproline Helix (Ppii) Structure Via Aromatic Electronic Effects: An Electronic Switch of Polyproline Helix', *Biochemistry*, 53 (2014), 5307-14.
- 154 A. K. Pandey, D. Naduthambi, K. M. Thomas, and N. J. Zondlo, 'Proline Editing: A General and Practical Approach to the Synthesis of Functionally and

- Structurally Diverse Peptides. Analysis of Steric Versus Stereoelectronic Effects of 4-Substituted Prolines on Conformation within Peptides', *J Am Chem Soc*, 135 (2013), 4333-63.
- 155 S. Donnini, G. Groenhof, R. K. Wierenga, and A. H. Juffer, 'The Planar Conformation of a Strained Proline Ring: A Qm/Mm Study', *Proteins*, 64 (2006), 700-10.
- 156 E. J. Milner-White, L. H. Bell, and P. H. Maccallum, 'Pyrrolidine Ring Puckering in Cis and Trans-Proline Residues in Proteins and Polypeptides. Different Puckers Are Favoured in Certain Situations', *J Mol Biol*, 228 (1992), 725-34.
- 157 R. Improta, C. Benzi, and V. Barone, 'Understanding the Role of Stereoelectronic Effects in Determining Collagen Stability. 1. A Quantum Mechanical Study of Proline, Hydroxyproline, and Fluoroproline Dipeptide Analogues in Aqueous Solution', *J Am Chem Soc*, 123 (2001), 12568-77.
- 158 M. L. DeRider, S. J. Wilkens, M. J. Waddell, L. E. Bretscher, F. Weinhold, R. T. Raines, and J. L. Markley, 'Collagen Stability: Insights from Nmr Spectroscopic and Hybrid Density Functional Computational Investigations of the Effect of Electronegative Substituents on Prolyl Ring Conformations', *J Am Chem Soc*, 124 (2002), 2497-505.
- 159 B. K. Ho, E. A. Coutsiias, C. Seok, and K. A. Dill, 'The Flexibility in the Proline Ring Couples to the Protein Backbone', *Protein Sci*, 14 (2005), 1011-8.
- 160 A. Choudhary, D. Gandla, G. R. Krow, and R. T. Raines, 'Nature of Amide Carbonyl--Carbonyl Interactions in Proteins', *J Am Chem Soc*, 131 (2009), 7244-6.

CHAPTER II

Hydroxylation of insulin at ProB28 accelerates
hexamer dissociation and delays fibrillation

This chapter was adapted from a manuscript in submission at the Journal of the American Chemical Society.

Abstract

Daily injections of insulin provide lifesaving benefits to millions of diabetics. But currently available prandial insulins are suboptimal: The onset of action is delayed by slow dissociation of the insulin hexamer in the subcutaneous space, and insulin forms amyloid fibrils upon storage in solution. Here we show, through the use of non-canonical amino acid mutagenesis, that replacement of the proline residue at position 28 of the insulin B-chain (ProB28) by (4S)-hydroxyproline (Hzp) yields an active form of insulin that dissociates more rapidly, and fibrillates more slowly, than the wild-type protein. Crystal structures of dimeric and hexameric insulin preparations suggest that a hydrogen bond between the hydroxyl group of Hzp and a backbone amide carbonyl positioned across the dimer interface may be responsible for the altered behavior. The effects of hydroxylation are stereospecific; replacement of ProB28 by (4R)-hydroxyproline (Hyp) causes little change in the rates of fibrillation and hexamer disassociation. These results demonstrate a new approach that fuses the concepts of medicinal chemistry and protein design, and paves the way to further engineering of insulin and other therapeutic proteins.

Introduction

Blood glucose levels are tightly controlled in mammals through a sensitive regulatory system mediated by insulin, a 51-amino acid endocrine hormone composed of two disulfide-linked polypeptide chains (designated A and B). Upon binding to its receptor, insulin initiates a signaling cascade that accelerates glucose uptake and glycogen production. In diabetic patients, this system malfunctions, and glucose levels must be controlled through subcutaneous injections of insulin¹. The C-terminus of the B-chain is important in mediating dimerization of the hormone^{2,3}, and the flexibility of the

B-chain C-terminus is believed to contribute to aggregation through formation of amyloid fibrils⁴⁻⁶. Pharmaceutical formulations of insulin are stabilized with respect to fibrillation by addition of zinc and phenolic preservatives, which drive assembly of the R₆ hexamer (Figure 2.1a)⁷⁻⁹. The R₆ form of insulin is inactive and dissociates slowly to its active monomeric form after subcutaneous injection; the lag time for dissociation delays the onset of action¹⁰. Mutation of ProB28 yields rapid-acting insulins (RAIs) by disrupting contacts that are critical for dimer formation⁸, but replacement of Pro through conventional mutagenesis also increases the flexibility and perturbs the trajectory of the protein backbone (Figure 2.1b). We sought a means to disrupt the dimer interface without releasing the conformational constraints characteristic of proline by using non-canonical amino acid (ncAA) mutagenesis¹¹⁻¹³. Specifically, we introduced hydroxyl groups at the 4-position of ProB28 (Figure 2.1 b,c) by replacing Pro with Hzp or Hyp. In addition to introducing a polar functional group and the capacity for hydrogen-bonding, hydroxylation at the 4-position is known to alter the *endo/exo* preference of the pyrrolidine ring and the *cis/trans* equilibrium of the amide backbone¹⁴.

Materials and methods

Materials. All canonical amino acids and (4*R*)-hydroxy-L-proline (Hyp) were purchased from Sigma. (4*S*)-hydroxy-L-proline (Hzp) was purchased from Bachem Americas. All solutions and buffers were made using double-distilled water (ddH₂O).

Strains and plasmids. The proinsulin (PI) gene with an *N*-terminal hexahistidine tag (6xHIS), and flanked by *Eco*R1 and *Bam*H1 cut sites was ordered as a gBlock (Integrated DNA Technologies). Both the gBlock and vector pQE80L for IPTG-inducible expression were digested with *Eco*RI and *Bam*HI. Linearized vector pQE80L was dephosphorylated by alkaline phosphatase (NEB). Ligation of the digested PI gene and linearized vector yielded plasmid pQE80PI (to produce ProI). To make plasmid pQE80PI-proS (to produce HzpI and HypI): Genomic DNA was extracted from *E. coli* strain DH10β using DNeasy Blood and Tissue Kit (Qiagen). Primers (Integrated DNA Technologies) were designed to amplify the *E. coli proS* gene, encoding prolyl-tRNA

synthetase, under constitutive control of its endogenous promoter, from purified genomic DNA, and to append *NheI* and *NcoI* sites. The digested *proS* gene was then inserted into pQE80PI between transcription termination sites by ligation at *NheI* and *NcoI* restriction sites. Proline-auxotrophic *E. coli* strain CAG18515 was obtained from the Coli Genetic Stock Center at Yale University. Prototrophic *E. coli* strain BL-21 was used for rich media expression of canonical insulins (ProI, AspI). Site-directed mutagenesis of pQE80PI at B28 was performed to make plasmid pQE80PI-asp, which differs from pQE80PI by three nucleotides that specify a single amino acid mutation to aspartic acid. All genes and plasmids were confirmed by DNA sequencing.

Protein expression. Plasmids pQE80PI and pQE80PI-asp were transformed into BL21 cells and grown on ampicillin-selective agar plates. A single colony was used to inoculate 5 mL of Luria-Bertani (LB) medium and grown overnight; the resulting saturated culture was used to inoculate another 1 L of LB medium. All expression experiments were conducted at 37°C, 200 RPM in shake flasks (Fernbach 2.8 L flasks, Pyrex®). Each culture was induced with 1 mM IPTG at mid-exponential phase ($OD_{600} \sim 0.8$). For incorporation of Hyp and Hzp, pQE80PI-*proS* was transformed into CAG18515 cells, which were grown on ampicillin-selective agar plates. To facilitate growth, a single colony was used to inoculate 25 mL of LB medium and the culture was grown overnight prior to dilution into 1 L of 1X M9, 20 amino acids (8.5 mM NaCl, 18.7 mM NH_4Cl , 22 mM KH_2PO_4 , 47.8 mM Na_2HPO_4 , 0.1 mM $CaCl_2$, 1 mM $MgSO_4$, 3 mg/L $FeSO_4$, 1 $\mu g/L$ of trace metals (Cu^{2+} , Mn^{2+} , Zn^{2+} , MoO_4^{2-}), 35 mg/L thiamine hydrochloride, 10 mg/L biotin, 20 mM D-glucose, 200 mg/L ampicillin with 50 mg/L of L-amino acids, each). At an appropriate cell density ($OD_{600} \sim 0.8$), the culture was subjected to a medium shift; briefly, cells were centrifuged and washed with saline prior to resuspension into 0.8 L of 1.25X M9, 19 aa (M9, 20 aa medium without L-proline). After cells were further incubated for 30 min to deplete intracellular proline, 200 mL of 5X additives (1.5 M NaCl, 2.5 mM Hyp or Hzp) was added to the culture. After another 15 min of incubation at 37°C to allow amino acid uptake prior to induction, IPTG was added to a final concentration of 1 mM. At the end of 2 h, cells were harvested by centrifugation and stored at -80°C until further use.

Cell lysis and refolding from inclusion bodies. Cells were thawed on the benchtop for 15 min prior to resuspension in lysis buffer (B-PER®, 0.5 mg/mL lysozyme, 50 U/mL benzonase nuclease). Cells were gently agitated at RT for 1 h prior to centrifugation (10 000 g, 10 min, RT); supernatant was discarded and the pellet was washed thrice: once with wash buffer (2 M urea, 20 mM Tris, 1% Triton X-100, pH 8.0) and twice with sterile ddH₂O; centrifugation followed each wash and the supernatant was discarded. The final washed pellet containing inclusion bodies (IBs, ~50% PI) was re-suspended in Ni-NTA binding buffer (8 M urea, 300 mM NaCl, 50 mM NaH₂PO₄, pH 8.0) overnight at 4°C or at RT for 2 h, both with gentle agitation. The suspension was centrifuged to remove insoluble debris; the remaining pellet was discarded and the supernatant was mixed with pre-equilibrated Ni-NTA resin (Qiagen) at RT for 1 h in order to purify PI from the IB fraction. Unbound proteins in the IB fraction were collected in the flow-through (FT), and the resin was washed with Ni-NTA wash buffer (8 M urea, 20 mM Tris base, 5 mM imidazole, pH 8.0) and Ni-NTA rinse buffer (8 M urea, 20 mM Tris base, pH 8.0) prior to stripping PI from the resin with Ni-NTA elution buffer (8 M urea, 20 mM Tris base, pH 3.0). Fractions (IBs, FT, W, elution) were collected and run under reducing conditions on SDS-PAGE (Bis/Tris gels, Novex®); elution fractions containing PI were pooled and solution pH was adjusted to 9.6 with 6 N NaOH in preparation for oxidative sulfitolysis. Oxidative sulfitolysis was performed at RT for 4 h, with the addition of sodium sulfite and sodium tetrathionate (0.2 M Na₂SO₃, 0.02 M Na₂S₄O₆); the reaction was quenched by 10-fold dilution with ddH₂O. To isolate PI from the quenched solution, the pH was adjusted to between 3.5 and 4.5 by adding 6 N HCl dropwise; the solution became cloudy. The solution was centrifuged (10 000 g, 10 min, RT) and supernatant discarded. The PI pellet was then re-suspended in refolding buffer (0.3 M urea, 50 mM glycine, pH 10.6) and protein concentration was estimated by the bicinchoninic acid assay (BCA assay, Pierce®). The concentration of PI was adjusted to 0.5 mg/mL. Refolding was initiated by addition of β-mercaptoethanol to a final concentration of 0.5 mM and allowed to proceed at 12°C overnight with gentle agitation (New Brunswick® shaker, 100 RPM). Post-refolding, soluble PI was harvested by adjusting the pH of the solution to 4-5 by dropwise addition of 6 N HCl and by high speed centrifugation to remove insoluble proteins. The supernatant was adjusted to pH 8-

8.5 by dropwise addition of 6 N NaOH and dialyzed against fresh PI dialysis buffer (7.5 mM sodium phosphate buffer, pH 8.0) at 4°C with five buffer changes to remove urea. The retentate (PI in dialysis buffer) was then lyophilized and subsequently stored at -80°C until further processing. Typical yields were 25-50 mg PI per L of culture (25-30 mg/L for non-canonical PI, 40-50 mg/L for canonical PI expression in rich media)

Proteolysis and chromatographic (HPLC) purification. The dry PI powder was re-dissolved in water to a final concentration of 5 mg/mL PI (final concentration of sodium phosphate buffer is 100 mM, pH 8.0). Trypsin (Sigma-Aldrich) and carboxypeptidase-B (Worthington Biochemical) were added to final concentrations of 20 U/mL and 10 U/mL, respectively to initiate proteolytic cleavage. The PI/protease solution was incubated at 37°C for 2.5 h; proteolysis was quenched by addition of 0.1% trifluoroacetic acid (TFA) and dilute HCl to adjust the pH to 4. Matured insulin was purified by reversed phase high-performance liquid chromatography (HPLC) on a C₁₈ column using a gradient mobile phase of 0.1% TFA in water (solvent, A) and 0.1% TFA in acetonitrile (ACN; solvent, B). Elution was carried from 0% B to 39% B with a gradient of 0.25% B per minute during peak elution. Fractions were collected and lyophilized, and the dry powder was re-suspended into 10 mM sodium phosphate, pH 8.0. Insulin-containing fractions were verified by matrix-assisted laser desorption/ionization-mass spectrometry (MALDI-MS; Voyager MALDI-TOF, Applied Biosystems) and SDS-PAGE to ensure identify and purity. Typical yields were 5-10 mg insulin per 100 mg PI. Fractions were stored at -80°C in 10 mM phosphate buffer, pH 8.0 until further use.

Verification of Hyp and Hzp incorporation levels and maturation. A 30 µL aliquot of PI solution (8 M urea, 20 mM Tris, pH 8) was subjected to cysteine reduction and alkylation (5 mM DTT, 55°C, 20 min; 15 mM iodoacetamide, RT, 15 min, dark) prior to 10-fold dilution into 100 mM NH₄HCO₃, pH 8.0 (100 µL final volume). Peptide digestion was initiated with 0.6 µL of gluC stock solution (reconstituted at 0.5 µg/µL with ddH₂O, Promega) at 37°C for 2.5 h. The reaction was quenched by adding 10 µL of 5% TFA and immediately subjected to C₁₈ ZipTip (Millipore) peptide purification and desalting according to the manufacturer's protocol. Peptides were eluted in 50% ACN, 0.1% TFA; the eluent was then diluted three-fold into matrix solution (saturated α-

cyanohydroxycinnamic acid in 50% ACN, 0.1% TFA) and analyzed by mass spectrometry (Voyager MALDI-TOF, Applied Biosystems). Hyp and Hzp incorporation levels were analyzed prior to and after refolding; incorporation percentage was calculated by comparing total AUC (area under the curve, arbitrary units) of the non-canonical peak (1573 Da for the proinsulin peptide containing B28Hzp or B28Hyp, ~5824 Da for intact HzpI and HypI) with total AUC of its wild-type counterpart (1557 Da and 5808 Da, respectively). Maturation of HypI and HzpI was analyzed after HPLC purification. TFA (1.6 μ L, 5%) was added to 15 μ L mature insulin solution (10 mM phosphate buffer pH 8.0) and subjected to C₁₈ ZipTip (Millipore) peptide purification and desalting according to the manufacturer's protocol. MALDI-MS conditions described above were used to confirm insulin maturation.

Insulin receptor (IR) phosphorylation immunoblot. *In vitro* analysis of insulin receptor (IR) phosphorylation was performed using HEK293 cells according to a previous report¹⁵. Briefly, HEK293 cells were maintained in a 37°C, 5% CO₂ humidified incubator chamber using Dulbecco's modified Eagle's medium with 4.5 g/L glucose, 2 mM L-glutamine and phenol red (DMEM, Life Technologies) supplemented with 10% fetal bovine serum (FBS, Life Technologies), 5% penicillin/streptomycin (P/S, Life Technologies). Every 3 days, at approximately 80% confluency, cells were subcultured and seeded in a 6-well plate at a cell density of 8×10^3 cells / cm² (or 8×10^4 cells per well) for 24 h prior to insulin addition. Insulins or vehicle were added directly to the medium at 200 nM (10 μ L of a 50 μ M solution in vehicle PBS) and incubated for 10 min prior to PBS washes to remove excess medium. HEK293 cells were lysed on-plate using IP Lysis Buffer (ThermoFisher, Pierce) with 50 U/mL benzonase nuclease (Sigma-Aldrich) for 20 min at 4°C; lysates were precipitated using ice cold acetone and re-suspended in 8 M urea, 20 mM Tris, pH 10.0. The protein concentration in the lysate was quantified by the BCA assay (ThermoFisher, Pierce) according to the manufacturer's protocol and normalized for even protein loading across lanes. Lysates were separated by SDS-PAGE (4-12% Novex Bis/Tris SDS-PAGE gels, Life Technologies) in duplicate and transferred to a nitrocellulose membrane (Hybond ECL, GE Healthcare) using a wet transfer system. The membrane was blocked at RT in 5% nonfat milk in Tris-buffer saline with 0.1% Tween 20 (TBS/Tween) and washed with TBS/Tween prior to blotting with antibodies.

Primary antibodies for insulin receptor, phosphorylated insulin receptor (from Cell Signaling Technologies) and β -actin (as loading control, from Invitrogen) were added at 1:1000 dilution in TBS/Tween with gentle agitation either at RT for 4 h or overnight at 4°C. Blots were washed and secondary antibodies (Invitrogen) were added at 1:2000 dilution in TBS/Tween. Blots were washed again prior to fluorescence imaging on a Typhoon Trio (GE Healthcare).

Reduction of blood glucose in diabetic animals. NODscid (NOD.CB17-*Prkdc^{scid}/J*) mice were obtained from Jax Mice (Bar Harbor, Maine). Mice were maintained under specific pathogen-free conditions, and experiments were conducted according to procedures approved by the Institutional Animal Care and Use Committee at the City of Hope. Adult (8-12 week old) male NODscid mice were injected intraperitoneally (50 mg/kg/day for 3 consecutive days) with freshly prepared streptozotocin (STZ) in 0.05 M citrate buffer, pH 4.5 to induce diabetes. Diabetes was confirmed 3 weeks after the last dose of STZ by detection of high glucose levels (defined as >200 mg/dL), measured by using a glucomonitor (FreeStyle; Abbott Diabetes Care, Alameda, CA) in blood (10 μ L) sampled from the lateral tail vein. Insulin analogs concentrations were determined from A_{280} measurements using a molar extinction coefficient of 6080 $M^{-1} cm^{-1}$ and diluted to 100 μ g/mL into a formulation buffer according to a previous report ¹⁶. Insulin analogs in solution were injected subcutaneously at the scruff and blood glucose was measured at the indicated time points.

Hexamer dissociation assay. Insulins were quantified by both UV absorbance (NanoDrop Lite, ThermoFisher) and BCA assay, and normalized to 125 μ M insulin prior to dialysis against 50 mM Tris/perchlorate, 25 μ M zinc sulfate, pH 8.0 overnight at 4°C using a D-tube dialyzer (Millipore Corp.) with MWCO of 3.5 kDa. Aliquots of dialyzed insulin solution were mixed with phenol to yield samples of the following composition: 100 μ M insulin, 20 μ M zinc sulfate, 100 mM phenol. Dissociation was initiated by addition of terpyridine (Sigma-Aldrich) to a final concentration of 0.3 mM from a 0.75 mM stock solution. A Varioskan multimode plate reader (Thermo Scientific) was used to monitor absorbance at 334 nm. Kinetic runs were done at least in triplicate, and the data

were fit to a mono-exponential function using Origin software. Post assay insulin samples were pooled and sample quality was determined by SDS-PAGE.

Fibrillation assay. Insulin samples (60 μ M in 10 mM phosphate, pH 8.0) were centrifuged at 22 000 g for 1 h immediately after addition of thioflavin T (ThT) (EMD Millipore) to a final concentration of 1 μ M. Samples were continuously shaken at 960 rpm on a Varioskan multimode plate reader at 37°C, and fluorescence readings were recorded every 15 min for 48 h (excitation 444 nm, emission 485 nm). Assays were run in quadruplicate, in volumes of 200 μ L in sealed (Perkin-Elmer), black, clear-bottom 96 well plates (Grenier BioOne).

Circular dichroism. Spectra were collected in a 1 cm quartz cuvette at an insulin concentration of 60 μ M in 50 mM sodium phosphate buffer pH 8.0. Data were collected from 185 nm to 250 nm, with step size of 0.25 nm and averaging time of 1 s on a Model 410 Aviv Circular Dichroism Spectrophotometer; spectra were averaged over 3 repeat scans. A reference buffer spectrum was subtracted from the sample spectra for conversion to mean residue ellipticity.

Analytical ultracentrifugation. Sedimentation velocity (SV) and sedimentation equilibrium (SE) experiments were carried out on an XL-1 AUC (BeckmanCoulter). SV experiments were conducted with insulin samples dialyzed against 50 mM Tris, 0.1 mM EDTA, pH 8.0, which also served as the reference buffer. Two sector cells with sapphire windows were filled with sample and reference buffer. These cells were centrifuged at 50,000 rpm with absorbance data collected at 280 nm, or for concentrations above 1 mg/mL, 281 nm or 287 nm. SV data were analyzed in SEDFIT with the c(s) algorithm for a continuous distribution¹⁷. Buffer density and viscosity were calculated from SEDNTERP; the partial specific insulin volume used was 0.735¹⁸. SE experiments were conducted with insulin samples dialyzed against 50 mM Tris, 0.1 mM EDTA, pH 8.0, which also served as the reference buffer. Two sector cells with sapphire windows were filled with sample and reference buffer and centrifuged at 15,000, 24,000, 36,000 and 50,000 rpm with absorbance data collected at 280 nm. Equilibrium was ascertained by analysis in SEDFIT and non-equilibrated scan speeds were excluded from data analysis. SE and SV data from multiple concentrations were fitted to a monomer-dimer-hexamer reversible self-association model in SEDPHAT with best model chosen by inspection of

residuals as well as critical χ value deviation¹⁹. Radial dependent baselines were computationally determined using TI noise. Figures were generated using GUSSEI²⁰.

Crystallographic studies. Insulin crystals were obtained from sitting drop trays set using a Mosquito robot (TTP Labtech). Drops were set by mixing 0.4 μ L insulin solution with 0.4 μ L well solution. Well solution conditions were as follows: 462.5 mM sodium citrate, 100 mM HEPES, pH 8.25 for 5HQI; 300 mM Tris, 0.5 mM zinc acetate, 8.5% acetone, 0.5 M sodium citrate pH 8.0 for 5HPR; 300 mM Tris, 17 mM zinc acetate, 1% phenol, 7.5% acetone, 2.675 M sodium citrate pH 8.0 for 5HRQ; 300 mM Tris, 17 mM zinc acetate, 1% phenol, 7.5% acetone, 1.95 M sodium citrate pH 8.0 for 5HPU. Cells were cryoprotected in a mother liquor containing 30% glycerol prior to looping and flash freezing in liquid nitrogen. Data were collected at SSRL beamline BL12-2 using a DECTRIS PILATUS 6M pixel detector. Initial indexing and scaling was performed with XDS; for some structures, data were re-scaled in alternative space groups using Aimless²¹. Initial phases were generated by molecular replacement in PHASER with 3T2A (5HQI and 5HPR) or 1EV3 (5HRQ and 5HPU)²². Structure refinement was carried out in Coot and Refmac5^{23,24}. Data were deposited in the PDB with the following codes: 5HQI (T₂-HzpI), 5HPR (T₂-HypI), 5HRQ (R₆-HzpI), 5HPU (R₆-HypI).

Results and discussion

We expressed modified proinsulins (PIs) in the proline-auxotrophic *E. coli* strain CAG18515 in M9 minimal media supplemented with Hyp or Hzp. The extent of replacement of Pro by either Hyp or Hzp was approximately 90%^{25,26} as determined by matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS; figure S2.1). Denatured PIs were purified by Ni-NTA affinity chromatography in yields of 32 mg/L for Hzp-PI and 29 mg/L for Hyp-PI (table S1) from the inclusion body fraction. The PIs were refolded and cleaved with trypsin and carboxypeptidase B²⁷. The resulting mature insulins were purified by reversed phase HPLC^{26,27}, and proper proteolytic processing of each variant was verified by MALDI-MS (table S2.1). Wild-type insulin (ProI) and RAI Aspart (AspI, in which ProB28 is replaced by aspartic acid) were produced similarly. All of the variants caused similar reductions in blood glucose upon subcutaneous injection into diabetic mice (Figure 2.1d)^{2,28-30}. RAIs cannot be distinguished from ProI in rodent models³¹.

In the absence of Zn^{2+} and phenolic preservatives, insulins dimerize with a dissociation constant (K_D) of approximately 10 μM . In contrast, K_D for RAIs is typically $>500 \mu M$, and it is believed that destabilization of the dimer interface causes the accelerated onset of insulin action after subcutaneous injection^{28,32,33}. Monomeric forms of insulin give rise to characteristic circular dichroism (CD) spectra with distinct minima at 208 and 222 nm (e.g., AspI; Figure 2.2a). Dimerization causes a loss of negative ellipticity at 208 nm (e.g., ProI; Figure 2.2a). At a concentration of 60 μM , HypI appears to be monomeric (with a CD spectrum nearly identical to that of AspI; Figure 2.2a) while the spectrum of HzpI suggests a dimeric insulin (Figure 2.2a). Sedimentation velocity (SV) and sedimentation equilibrium (SE) experiments were consistent with the results of the CD analysis (figure S2.3). SE data were fitted to a model of monomer-dimer-hexamer self-association (SEDPHAT)^{34,35}, and yielded monomer-dimer dissociation constants (K_D) of $>200 \mu M$ and 25 μM for HypI and HzpI, respectively.

Previous studies of RAIs have shown that destabilization of the dimer interface correlates with accelerated dissociation of the hexamer and rapid onset of insulin action^{8,13}. Triggered dissociation of Zn^{2+} -hexamers by addition of terpyridine³⁶ revealed nearly identical rates of dissociation for HypI and ProI, ($\tau_{1/2} = 87.0 \pm 10$ s and 90.4 ± 4.2 s, respectively; Figure 2.2b S2.4) while HzpI exhibited kinetics similar to those of AspI ($\tau_{1/2} = 53.6 \pm 3.7$ s and 42.7 ± 3.0 s, respectively; Figure 2.2b and S2.4). We found these results surprising – replacement of Pro by Hyp destabilizes the dimer but has essentially no effect on hexamer dissociation, while introduction of Hzp causes little change in dimer stability but a substantial increase in the rate of hexamer disassembly.

Each of the insulin variants was subjected to fibrillation lag time analysis (Figure 2.2c)³⁷. We found similar times to onset of fibrillation for HypI, ProI and AspI; in contrast, HzpI is markedly more resistant to aggregation, with a mean time to onset more than three-fold longer than that observed for ProI. The behavior of HzpI is especially striking, in that it combines fast hexamer dissociation with enhanced stability toward fibrillation.

Each subunit in the insulin hexamer adopts one of two conformational states (T or R), depending on the concentration of phenolic ligand (Figure 2.1a)¹³. Pharmaceutical formulations are prepared in the more stable R_6 form, whereas the T-state is observed in

the absence of phenolic ligands, most commonly in the form of T₂-dimers³⁸. To elucidate the molecular origins of the dissociation and fibrillation behavior of HypI and HzpI, we examined crystal structures of both states.

Hydroxylation at ProB28 does not cause substantial perturbation of the overall insulin structure (Figure 2.3 and S2.5). In comparison with ProI, the backbone RMSD values of HypI and HzpI are 0.31 Å (T₂- HypI), 0.44 Å (T₂- HypI), 0.35 Å (R₆- HypI) and 0.68 Å (R₆- HypI)³⁹. The ring puckers for HzpI are endo in both the T₂ and R₆ forms (Table S2.3). For HypI, surprisingly, the T₂ form adopts an endo pucker while the R₆ form adopts a close to planar pucker (Table S2.3). Hyp in free solution has preference for an exo pucker but the local structural environment dissuades the Hyp from adopting this conformer in HypI. The most notable feature of the HzpI structures is the proximity of the hydroxyl group of Hzp to the backbone carbonyl oxygen atom of GluB21', which lies across the dimer interface (denoted by prime; figure 2.3 b,e). The inter-oxygen distances (2.8 Å in the T₂ structure, 2.7 Å in R₆), are consistent with the formation of strong hydrogen bonds between the hydroxyl group of HzpB28 and the backbone carbonyl of GluB21' in both structures. An analogous hydrogen bond has been observed in a structure (PDB ID: 1ZEH) of R₆-AspI co-crystallized with *m*-cresol⁴⁰; here the phenolic ligand serves as the hydrogen-bond donor (figure S2.6). Although the significance of this hydrogen bond has not been discussed in the literature, we suggest that it may play an important role in determining the relative stabilities of the insulin species involved in dissociation and fibrillation. In contrast to the (4*S*)-hydroxyl group of Hzp, the (4*R*)-hydroxyl of HypB28 does not contact any crystallographically resolved hydrogen bond acceptor in the T₂-structure (figure 2.3c), and appears to bond to an ordered water molecule in the R₆-hexamer (figure 2.3f). The absence of new hydrogen-bonding interactions is consistent with the unaltered dissociation and fibrillation kinetics of HypI.

Conclusion

Taken together, the results of CD, sedimentation and crystallographic analyses show that replacement of Pro by Hzp at position 28 of the insulin B-chain introduces a new hydrogen bond across the inter-subunit interface, accelerates hexamer dissociation and delays the onset of fibrillation (Table 2.1). We suggest that the hydrogen bond between Hzp and Glu21' may stabilize the dimer relative to the hexamer, or perhaps

reduce the energy of the transition state for the conformational change from the R-state to T-state, and thereby speed disassembly. In addition, we propose that this hydrogen bond may prevent the C-terminus of the B chain from fraying and initiating fibril formation. Whether or not these hypotheses are correct, the results described here demonstrate the power of nCAA mutagenesis to control functionally relevant biophysical properties of therapeutic proteins. We anticipate that this approach will find increasing application in the design of antibody-drug conjugates, bispecific antibodies, and other novel protein therapeutics.

The work described in this chapter was completed by multiple individuals. Katharine Fang completed the fibrillation, kinetics and immunoblot experiments as well as jointly expressed and purified the insulins described here. Mouse injection assays were conducted by Teresa Ku, Jeanne LeBon and Jeffery Rawson.

I would also like to thank J. T. Kaiser, P. Nikolovski, S. Russi, S. Virgil, M. Shahgholi, A. Lakshmanan and the scientific staff of Beamline 12-2 at the Stanford Synchrotron Radiation Laboratory for assistance.

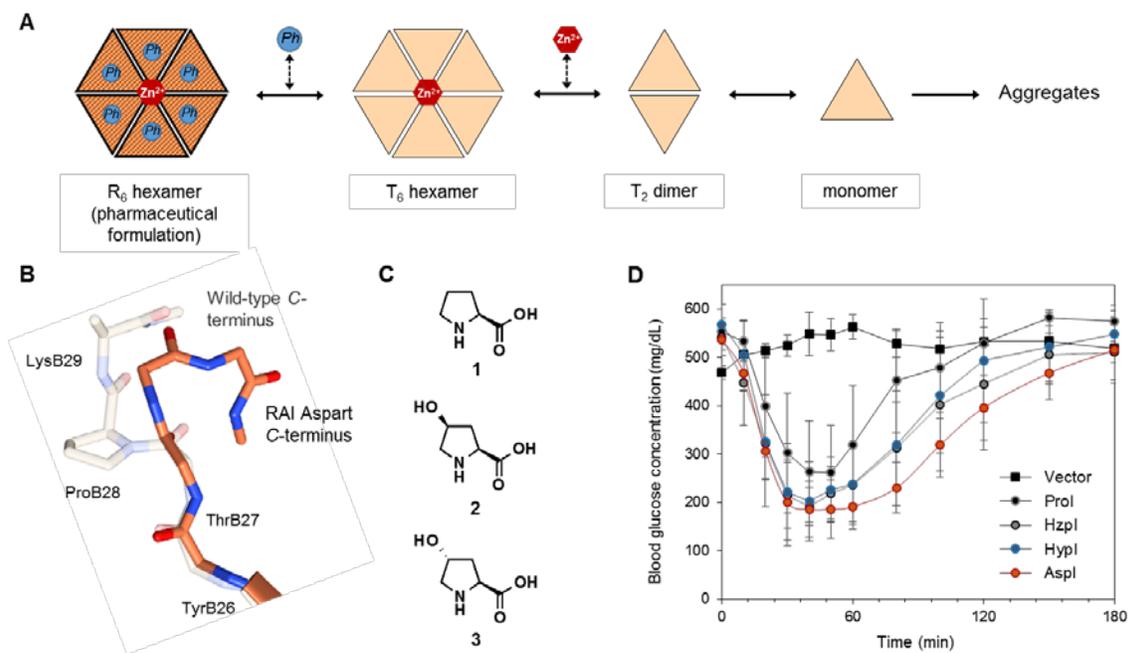


Figure 2.1 | Hydroxyinsulins retain activity

A) Schematic of hexamer disassembly (adapted from mechanism previously described⁸). Phenolic ligand (Ph), zinc ion (Zn^{2+}), insulin monomer (triangle). Darker shading indicates the R-state of the hexamer.

B) Structures of the B-chain C-termini of wild-type insulin (ProI) and RAI Aspart (AspI).

C) Chemical structures of L-proline (**1**), (2*S*,4*S*)-hydroxyproline (Hzp; **2**), (2*S*,4*R*)-hydroxyproline (Hyp; **3**)

D) Reduction of blood glucose following subcutaneous injection of 35 μ g/kg insulins into streptozotocin-induced diabetic mice. Glucose levels were measured post-injection via tail vein sampling. ProI, AspI, HzpI, HypI or vector were formulated as described²⁰. Error bars denote one standard deviation ($n = 3$).

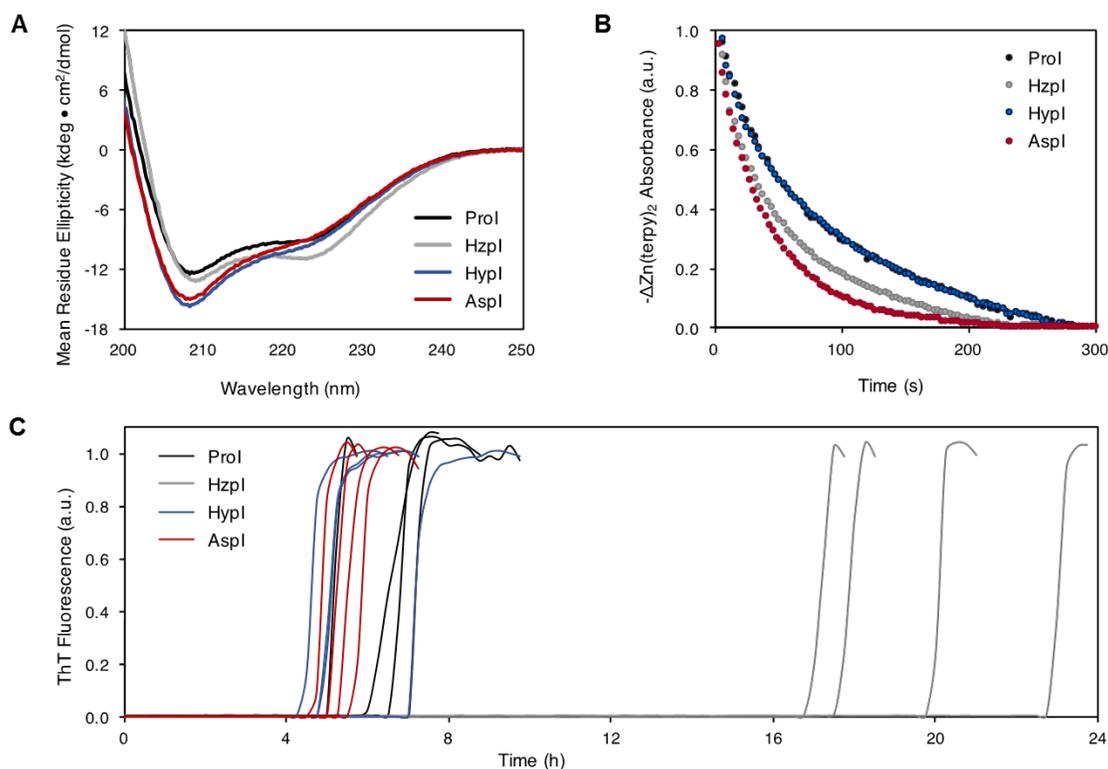


Figure 2.2 | Hydroxylation at ProB28 modulates insulin dimerization, dissociation kinetics, and stability.

A) Far UV CD spectra collected on 60 μM insulins in 10 mM phosphate buffer, pH 8.0 at 25°C.

B) Insulin hexamer dissociation following sequestration of Zn^{2+} by terpyridine. Zn^{2+} -terpy signal was monitored at 334 nm and fitted to a mono-exponential decay. HzpI and HypI contain 10% ProI. The curves for HypI and ProI are indistinguishable in this plot.

C) Representative fibrillation curves for 60 μM insulins (37°C, 960 RPM; n=4). Insulin fibrils were detected by the rise in Thioflavin T (ThT) fluorescence that accompanies binding to fibrillar aggregates.

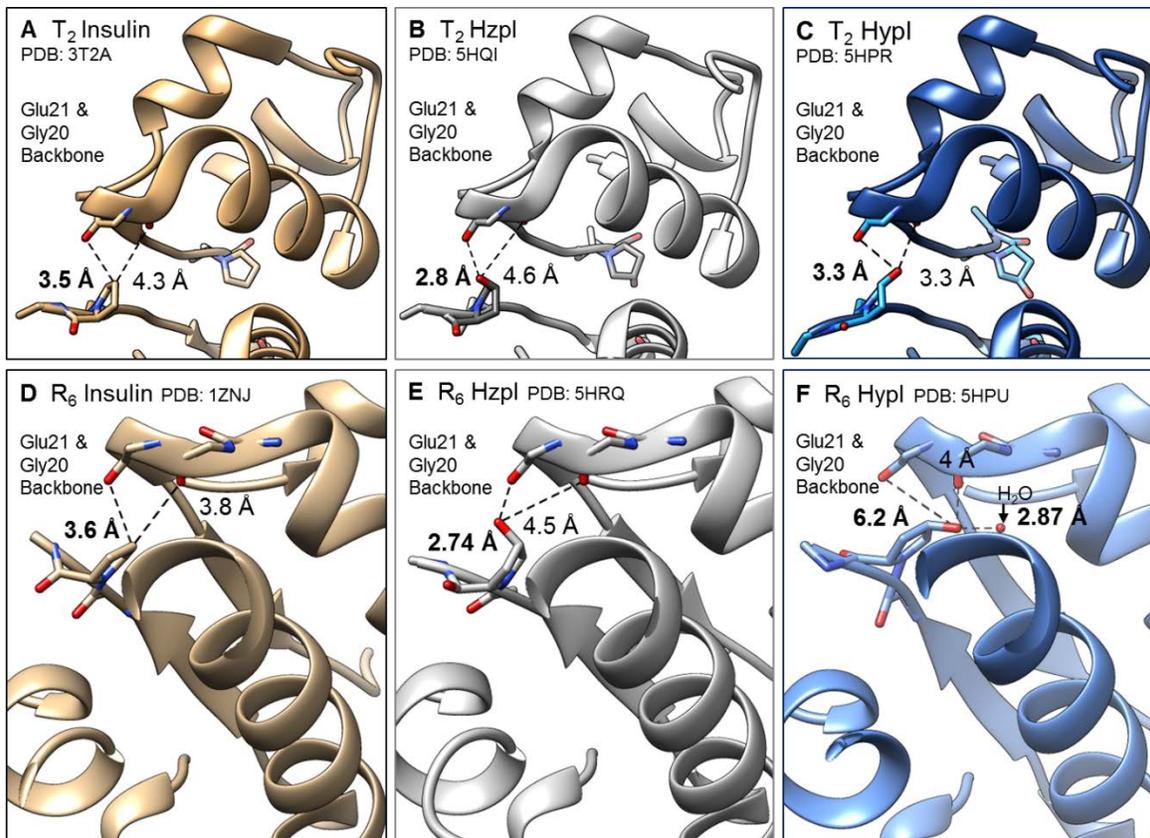


Figure 2.3 | Crystal structures of HZI and HPI.

In tan (left), wt insulins from PDB (3T2A, 1ZNI) highlighting the distance between the carbon atom at the 4th position of ProB28 and its closest neighbors, backbone carbonyl oxygen atoms of GlyB20' and GluB21 in the T₂ dimer (A) and R₆ hexamer (D) forms. In grey (middle), HZI in the T₂ dimer (B) and R₆ hexamer (E) forms. In blue (right), HPI in the T₂ dimer (C) and R₆ hexamer (G) forms.

Table 1 | Biophysical Characteristics of Insulin Variants

Protein	Hexamer $\tau_{1/2}$ (s)	Fibrillation lag time (h)	K_D dimer (μ M)
Prol	90.4 \pm 4.2	5.1 \pm 1.5	9 μ M
Hzpl*	53.6 \pm 3.7	19.6 \pm 2.6	~25 μ M
Hysl*	87.0 \pm 10	5.5 \pm 1.2	>200 μ M
Aspl	42.7 \pm 4.3	5.3 \pm 1.0	>500 μ M

Errors are given as one standard deviation ($n \geq 4$).

*10% Prol present.

Table 2.1 | Biophysical Characteristics of Insulin Variants

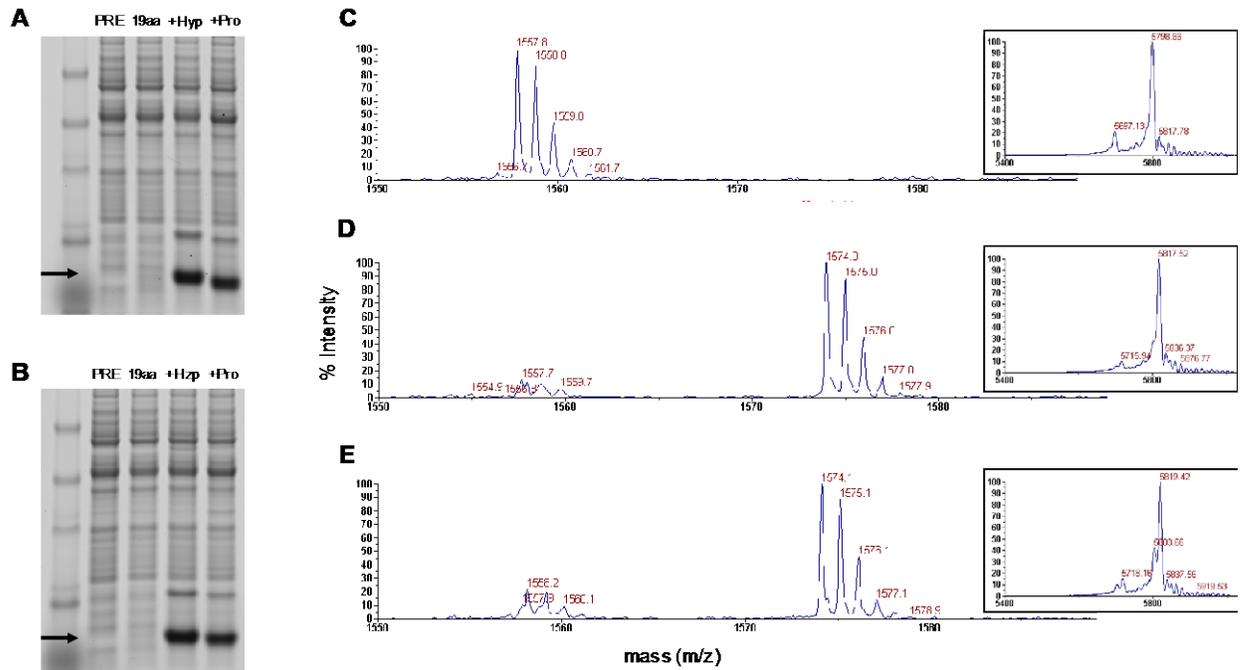


Figure S2.1 | Insulin expression and incorporation of hydroxyprolines.

A, B) SDS-PAGE of cell lysates with lanes labeled for pre-induction (PRE) and post-induction in minimal media supplemented with either nothing (19aa), Hyp (**A**), Hzp (**B**), or Pro at 0.5 mM.

C-E) MALDI-MS traces of isolated proinsulin peptide fragment $^{46}\text{RGFFYTPKTRRE}^{57}$ obtained by gluC digestion. Peptide fragment masses correspond to either wild type mass (1558 Da) (**C**) or shifted mass (1574 Da) if Hyp (**D**) or Hzp (**E**) is incorporated. Inset is whole protein MALDI-MS. All MALDI-MS spectra contain ion counts $>10^3$.

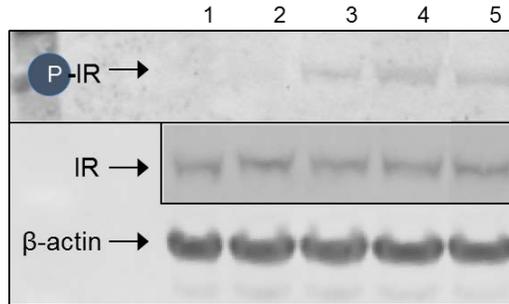


Figure S2.2 | Immunoblot detection of insulin receptor activation.

HEK293 cells treated with insulin (200 nM in PBS, pH 7.4) or vehicle. Whole cell lysates were then run on an SDS-PAGE gel and transferred to nitrocellulose membrane to detect insulin receptor (IR) and IR phosphorylation. β -actin immunoblot shown as loading control. Lane 1: Vehicle (PBS); Lane 2: 10% ProI serving as a second negative control due to presence of 10% wt in HzpI and HypI preparations; Lane 3: HzpI; Lane 4: HypI; Lane 5: ProI

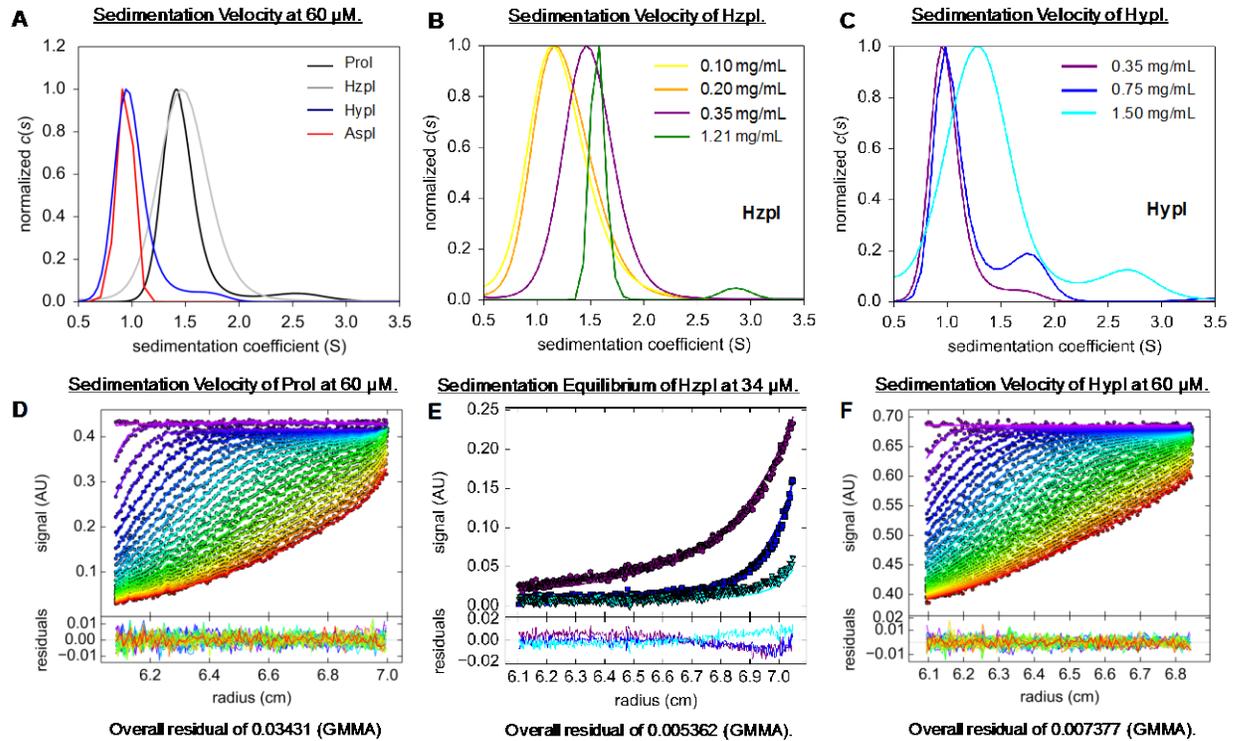


Figure S2.3 | Example fits from sedimentation analysis.

Insulin samples in 50 mM Tris pH 8.0.

A) $c(s)$ curves overlaid for ProI, AspI, HzpI and HypI at 60 μM .

B, C) $c(s)$ curves for HzpI (**B**) and HypI (**C**) at indicated concentrations.

D-F) Example fits for 60 μM ProI (**D**), 34 μM HzpI (**E**), and 60 μM HypI (**F**), overlaid on top of noise corrected velocity (**D, F**) or equilibrium data (**E**). The SEDPHAT monomer-dimer-hexamer model was utilized over a range of velocity and equilibrium experiments. Global multi-method analysis residuals for the dataset are displayed below each plot.

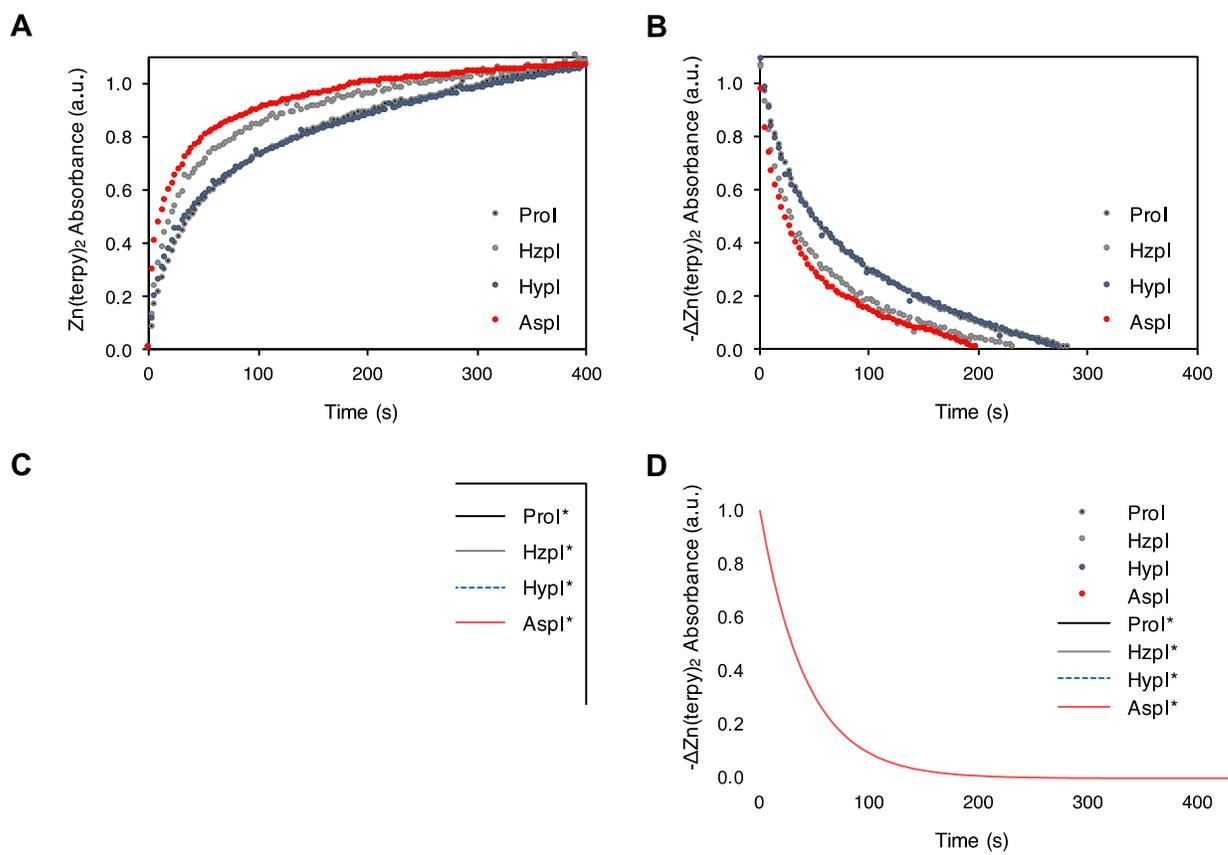


Figure S2.4 | Example fits for analysis of dissociation kinetics.

A) Representative dissociation kinetic traces for Zn²⁺ sequestration. Raw data shown and used to fit to a mono-exponential using Origin Software ($y-y_0 = Ae^{(-t/\tau)}$), where fitted value τ is the characteristic dissociation time constant.

B, C) Fitted mono-exponential decay traces for dissociation kinetics, corresponding to (A) shown in (B). Fitted values for y_0 , A used to convert raw data (A) to mono-exponential decay representation shown in (C).

D) Overlay of (B) and (C) show fitted and raw data to demonstrate quality of fits. *Denotes fitted curves

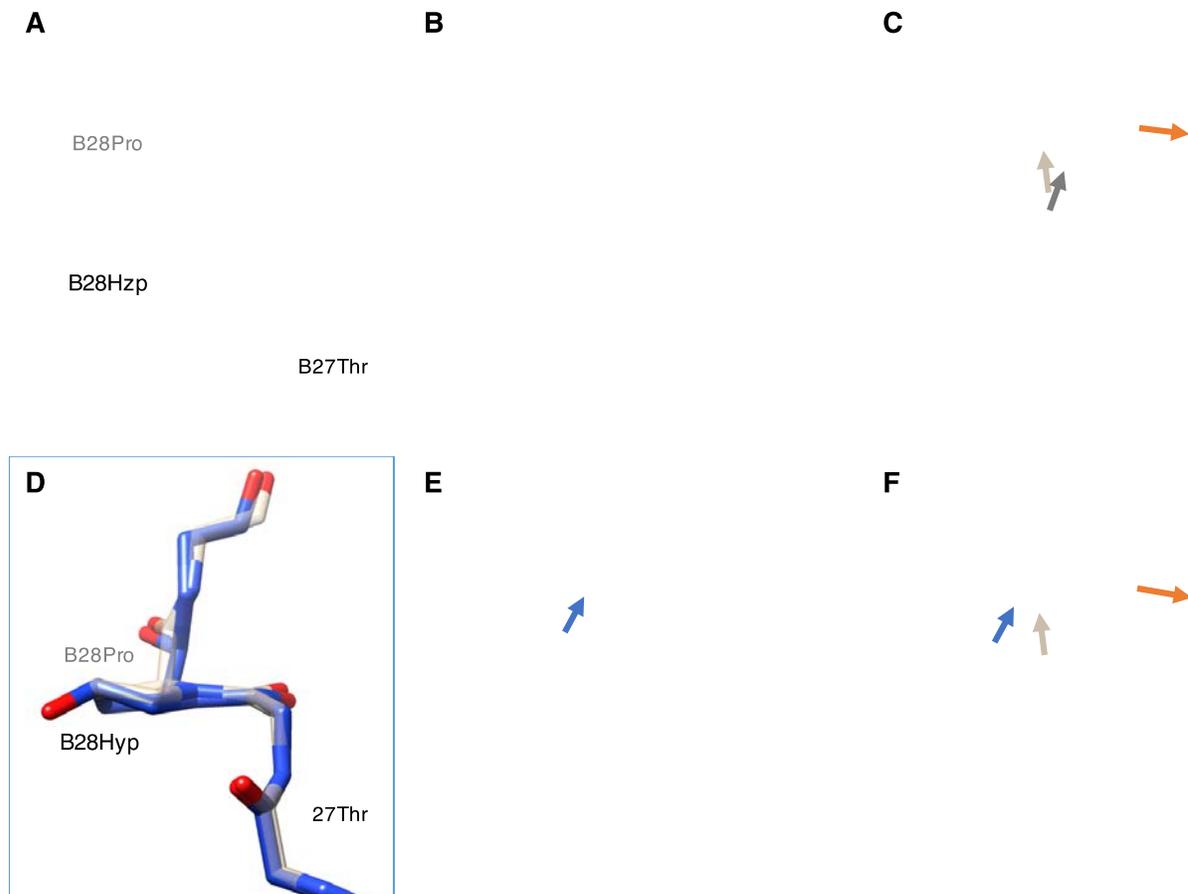


Figure S2.5 | Alignment at position B28.

A, D) Alignment of T₂ ProI (tan, PDB:3T2A), and T₂-HzpI (grey) or T₂-HypI (blue) centered on position B28.

B, E) Alignment of R₆-ProI (tan) and R₆-HzpI (grey) or R₆-HypI (blue) highlighting the overlap of the backbone at the C-terminus. B29 not shown in **(E)** due to lack of electron density.

C, F) Alignment of R₆ insulins (ProI, and HzpI or HypI), and AspI (orange, PDB: 1ZEG) centered on position B28 illustrates the similarity of the polypeptide backbones of ProI, HzpI and HypI, and the distinct backbone trajectory of AspI. B29 **(C, F)** and B30 **(A-F)** amino acids not shown for clarify. Arrows denote the N-to-C terminal direction of the backbone originating from carbonyl carbon.

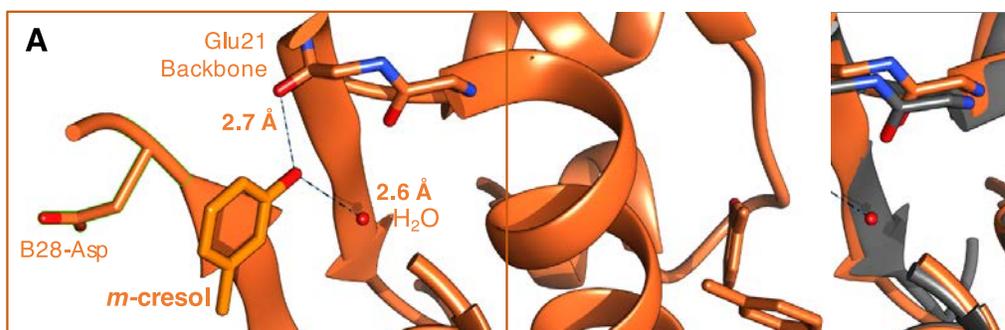


Figure S2.6 | Alignment of R₆-AspI and T₂-HzpI at position B28.

- A)** R₆-AspI (dark orange; PDB: 1ZEH) does not maintain the backbone trajectory of ProI at position B28. The C-terminus of the AspI B-chain is shifted, and a *m*-cresol ligand (light orange) fills the site occupied by B28Pro in ProI. The hydroxyl group of *m*-cresol forms hydrogen bonds with the backbone carbonyl of Glu21' and a nearby water molecule.
- B)** The same representation of R₆-AspI overlaid with R₆-HzpI (dark grey). Interatomic distances were determined using Chimera. Amino acid B30 is not shown.

B28 Amino Acid	Incorporation ¹	Yield ²
L-proline (Pro)	--	50 mg/L
(2S,4S)-4-hydroxy-L-proline (Hzp)	91.3 ± 1.8 %	32 mg/L
(2S,4R)-4-hydroxy-L-proline (Hyp)	88.2 ± 1.2 %	29 mg/L

Table S2.1 | Expression Yields and Incorporation Levels of Hydroxyinsulins.

¹Quantified by MALDI-MS from shake flask expressions (n≥4) using *proinsulin peptide* (⁴⁶*RGFFYTPKTRRE*⁵⁷) obtained by *gluC* digestion. ²Yield quantified by BCA assay post-refolding at the proinsulin level

	HypI-T ₂ 5HPR	HypI-R ₆ 5HPU	Hzp-T ₂ 5HQI	Hzip-R ₆ 5HRQ
Data collection				
Space group	<i>I</i> 2 ₁ 3	<i>H</i> 3	<i>I</i> 2 ₁ 3	<i>P</i> 1 2 ₁ 1
Cell dimensions				
a, b, c (Å)	78.25, 78.25, 78.25	77.76, 77.76, 40.40	78.23, 78.23, 78.23	47.16, 60.68, 60.66
α, β, γ (°)	90.00, 90.00, 90.00	90.00, 90.00, 120.00	90.00, 90.00, 90.00	90.00, 110.82, 90.00
Resolution (Å)	55.33-1.33 (1.36-1.33)	38.88-2.20 (2.26-2.20)	55.32-0.97 (0.99-0.97)	56.80-1.28 (1.30-1.28)
R _{p.i.m.} (%)	3.3 (20.4)	4.1 (100.1)	3.4 (209.0)	3.2 (40.8)
Mn(I /sd)	17.4 (4.0)	7.4 (0.7)	13.2 (0.4)	13.6 (2.1)
Completeness (%)	99.4 (88.8)	97.0 (87.8)	99.4 (88.2)	96.0 (62.1)
Redundancy	5.9 (3.3)	3.0 (2.5)	4.7 (2.7)	3.0 (2.2)
Refinement				
No. reflections	17298 (1295)	4307 (325)	44555 (3056)	75363 (4122)
R _{work} / R _{free} (%)	12.5/16.4 (16.1/17.2)	16.1/21.7 (27.4/29.5)	13.9/15.8 (35.6/37.1)	13.5/16.8 (20.6/23.5)
No. atoms				
Protein	419	752	414	2451
Ligand/ion	7	18	0	46
Water	60	17	62	160
RMSD				
Bond lengths (Å)	0.034	0.006	0.026	0.014
Bond angles (°)	1.386	1.981	1.584	1.259
Ramachandran map analysis				
Favored	48	84	49	288
Allowed	2	3	1	2
Outliers	0	2	0	2

Table S2.2. | Data Tables and Refinement Values.

	RMSD vs ProI (Å) ^a	Ring conformation ^b	Chi 1 angle ^c	Phi ^c
ProI T ₂		Endo	22.2	-75
ProI R ₆		Endo (5/6)	21.2	-60
HypI T ₂	0.44	Endo	22.9	-62
HzipI T ₂	0.31	Endo	31.7	-68
HypI R ₆	0.68	Planar (2/2)	-1.75	-82
HzipI R ₆	0.35	Endo (6/6)	26.48	-57

Table S2.3 | RMSD and proline angles

A) HypI and HzipI T₂ insulins were aligned against T₂ ProI (PDB:3T2A) and R₆ insulins were aligned against R₆-ProI (PDB: 1EV6) RMSD was calculated over backbone atoms.

B) Ring conformation was determined on the basis of the Chi1 angles. Where more than one B28 residue is present in the asymmetric unit the predominant conformation is listed. The fraction in parentheses denotes the number of B28 residues in one asymmetric unit with the predominant conformation.

C) Chi 1 and Phi angles are reported as an average over all B28 residues present in the asymmetric unit for each crystal structure. For T₂ insulins only a single B28 is averaged as there is only a single insulin monomer in the asymmetric unit, for R₆ insulins the average value is calculated over multiple prolines as there are 2 (HypI) or 6 (HzipI) prolines present in the asymmetric unit.

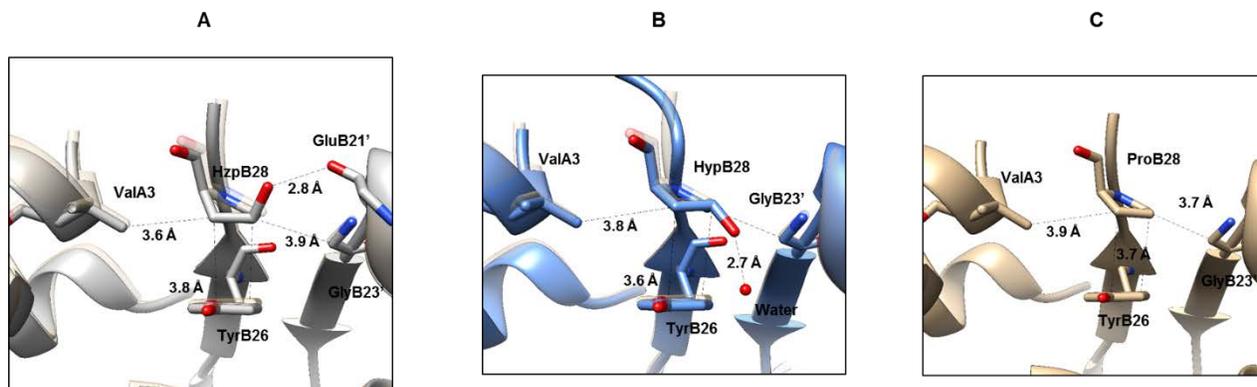


Figure S2.7 | Alignment at position B28 highlighting key contacts in the dimer.

A) Alignment of T_2 ProI (tan, PDB:3T2A) and T_2 -HzpI (grey) centered on position B28 highlighting the important interactions made between B28 and surrounding residues in the dimer. Interaction distances are listed in angstroms.

B) Alignment of T_2 ProI (tan, PDB:3T2A) and T_2 -HypI (blue) centered on position B28 highlighting the important interactions made between B28 and surrounding residues in the dimer. Interaction distances are listed in angstroms.

C) T_2 ProI (tan, PDB:3T2A) centered on position B28 highlighting the important interactions made between B28 and surrounding residues in the dimer. Interaction distances are listed in angstroms

	Proline atom	Other residue	VDW Overlap (Å) ^a	Distance (Å) ^b
T ₂ Prol	CD	Gly23' CA	0.13	3.6
	CG	Tyr26 CE	-0.14	3.8
	CA	Val3 CG	-0.14	3.9
	Proline atom	Other residue	VDW Overlap (Å) ^a	Distance (Å) ^b
T ₂ HypI	OD1	HOH	0.27	2.7
	CD	Gly23' CA	0.10	3.7
	CD	Tyr26 C	-0.01	3.6
	CB	Val3 CG	-0.03	3.8
	Proline atom	Other residue	VDW Overlap (Å) ^a	Distance (Å) ^b
T ₂ Hzpl	OD1	Glu21' O	0.13	2.8
	CD	Gly23' CA	-0.13	3.9
	CG	Tyr26 CE	-0.14	3.8
	CB	Val3 CG	0.21	3.6

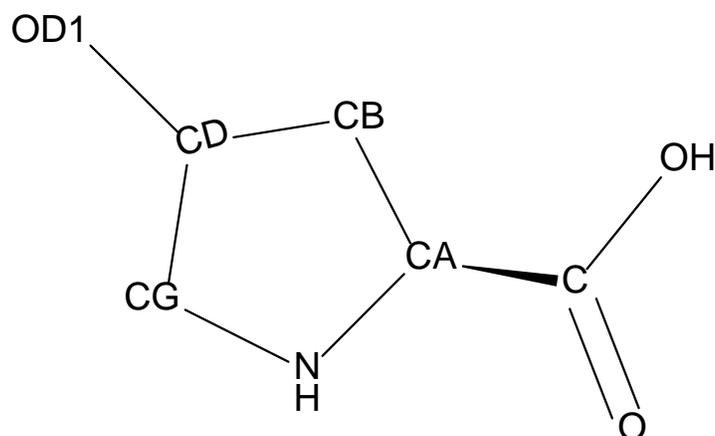


Table S2.4 | Key B28 contacts in hydroxyproline insulins

- VDW overlap was calculated in Chimera as the distance between the van der waals radii of each participating atom. Positive values indicate VDW radii overlapping. Negative values indicate the distance between VDW radii. The strongest interactions are bolded.
- Distances are measured as the linear distance between the center point of the two atoms listed in each row. Nomenclature of the atoms in the residues follows standard protein atomic naming as utilized in the PDB. Below the table is a guide to the proline ring atoms nomenclature.

References

1. Zaykov, A.N., Mayer, J.P. & DiMarchi, R.D. Pursuit of a perfect insulin. *Nat. Rev. Drug Discov* **15**, 425-439 (2016).
2. Ciszak, E. et al. Role of C-terminal B-chain residues in insulin assembly: the structure of hexameric LysB28ProB29-human insulin. *Structure* **3**, 615-622 (1995).
3. Menting, J.G. et al. How insulin engages its primary binding site on the insulin receptor. *Nature* **493**, 241-5 (2013).
4. Huang, K., Maiti, N.C., Phillips, N.B., Carey, P.R. & Weiss, M.A. Structure-specific effects of protein topology on cross- β assembly: studies of insulin fibrillation. *Biochemistry* **45**, 10278-10293 (2006).
5. Menting, J.G. et al. Protective hinge in insulin opens to enable its receptor engagement. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E3395-404 (2014).
6. Ivanova, M.I., Sievers, S.A., Sawaya, M.R., Wall, J.S. & Eisenberg, D. Molecular basis for insulin fibril assembly. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 18990-18995 (2009).
7. Freeman, J.S. Insulin analog therapy: improving the match with physiologic insulin secretion. *J. Am. Osteopath. Assoc.* **109**, 26-36 (2009).
8. Birnbaum, D.T., Kilcomons, M.A., DeFelippis, M.R. & Beals, J.M. Assembly and dissociation of human insulin and LysB28ProB29-insulin hexamers: a comparison study. *Pharm. Res.* **14**, 25-36 (1997).
9. Carpenter, M.C. & Wilcox, D.E. Thermodynamics of formation of the insulin hexamer: metal-stabilized proton-coupled assembly of quaternary structure. *Biochemistry* **53**, 1296-301 (2014).
10. Bakaysa, D.L. et al. Physicochemical basis for the rapid time-action of LysB28ProB29-insulin: dissociation of a protein-ligand complex. *Protein Sci.* **5**, 2521-2531 (1996).
11. Pandey, A.K., Naduthambi, D., Thomas, K.M. & Zondlo, N.J. Proline editing: a general and practical approach to the synthesis of functionally and structurally diverse peptides. Analysis of steric versus stereoelectronic effects of 4-substituted prolines on conformation within peptides. *JACS* **135**, 4333-4363 (2013).
12. Liu, C.C. & Schultz, P.G. Adding new chemistries to the genetic code. *Annu. Rev. Biochem* **79**, 413-44 (2010).
13. Pandeyarajan, V. & Weiss, M.A. Design of ⁶⁵non-standard insulin analogs for the

- treatment of diabetes mellitus. *Curr. Diab. Rep.* **12**, 697-704 (2012).
14. Shoulders, M.D., Kotch, F.W., Choudhary, A., Guzei, I.A. & Raines, R.T. The aberrance of the 4S diastereomer of 4-hydroxyproline. *JACS* **132**, 10857-10865 (2010).
 15. Wharton, J., Meshulam, T., Vallega, G. & Pilch, P. Dissociation of insulin receptor expression and signaling from Caveolin-1 expression. *J. Biol. Chem.* **280**, 13483-13486 (2005).
 16. Pandeyarajan, V. et al. Aromatic anchor at an invariant hormone-receptor interface: Function of insulin residue B24 with application to protein design. *J. Biol. Chem.* (2014).
 17. Schuck, P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys. J.* **78**, 1606-1619 (2000).
 18. Laue, T.M., Shah, B.D., Ridgeway, T.M. & Pelletier, S.L. *Analytical ultracentrifugation in biochemistry and polymer science*, (Royal Society of Chemistry, Cambridge [England], 1992).
 19. Vistica, J. et al. Sedimentation equilibrium analysis of protein interactions with global implicit mass conservation constraints and systematic noise decomposition. *Anal. Biochem* **326**, 234-256 (2004).
 20. Brautigam, C.A. Chapter five - calculations and publication-quality illustrations for analytical ultracentrifugation data. in *Methods Enzymol.*, Vol. Volume 562 (ed. James, L.C.) 109-133 (Academic Press, 2015).
 21. Winn, M.D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr., Sect D: Biol. Crystallogr.* **67**, 235-242 (2011).
 22. McCoy, A.J. et al. Phaser crystallographic software. *J Appl Crystallogr.* **40**, 658-674 (2007).
 23. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr., Sect D: Biol. Crystallogr.* **66**, 486-501 (2010).
 24. Murshudov, G.N., Vagin, A.A. & Dodson, E.J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr., Sect D: Biol. Crystallogr.* **53**, 240-255 (1997).
 25. Kim, W., George, A., Evans, M. & Conticello, V.P. Cotranslational incorporation of a structurally diverse series of proline analogues in an *Escherichia coli* expression system. *ChemBioChem* **5**, 928-936 (2004).

26. Kemmler, W., Peterson, J.D. & Steiner, D.F. Studies on the conversion of proinsulin to insulin. *J. Biol. Chem.* **246**, 6786-6791 (1971).
27. Min, C.-K., Son, Y.-J., Kim, C.-K., Park, S.-J. & Lee, J.-W. Increased expression, folding and enzyme reaction rate of recombinant human insulin by selecting appropriate leader peptide. *J. Biotechnol.* **151**, 350-356 (2011).
28. Brems, D.N. et al. Altering the association properties of insulin by amino acid replacement. *Protein Eng.* **5**, 527-533 (1992).
29. Sakata, N., Yoshimatsu, G., Tsuchiya, H., Egawa, S. & Unno, M. Animal models of diabetes mellitus for islet transplantation. *Exp. Diabetes Res.* **2012**(2012).
30. Pandyarajan, V. et al. Biophysical optimization of a therapeutic protein by non-standard mutagenesis: studies of an iodo-insulin derivative. *J. Biol. Chem.* **289**, 23367-23381 (2014).
31. Plum, A., Agersø, H. & Andersen, L. Pharmacokinetics of the Rapid-Acting Insulin Analog, Insulin Aspart, in Rats, Dogs, and Pigs, and Pharmacodynamics of Insulin Aspart in Pigs. *Drug Metab Dispos.* **28**, 155-160 (2000).
32. Attri, A.K., Fernández, C. & Minton, A.P. pH-dependent self-association of zinc-free Insulin characterized by concentration-gradient static light scattering. *Biophys. Chem.* **148**, 28-33 (2010).
33. Antolikova, E. et al. Non-equivalent role of inter- and intramolecular hydrogen bonds in the insulin dimer interface. *J. Biol. Chem.* **286**, 36968-77 (2011).
34. Zhao, H. & Schuck, P. Combining biophysical methods for the analysis of protein complex stoichiometry and affinity in SEDPHAT. *Acta Crystallogr., Sect D: Biol. Crystallogr.* **71**, 3-14 (2015).
35. Brown, P.H. & Schuck, P. Macromolecular size-and-shape distributions by sedimentation velocity analytical ultracentrifugation. *Biophys. J.* **90**, 4651-61 (2006).
36. Rahuel-Clermont, S., French, C.A., Kaarsholm, N.C. & Dunn, M.F. Mechanisms of stabilization of the insulin hexamer through allosteric ligand interactions. *Biochemistry* **36**, 5837-5845 (1997).
37. Vinther, T.N. et al. Novel covalently linked insulin dimer engineered to investigate the function of insulin dimerization. *PLoS ONE* **7**, e30882 (2012).
38. Palmieri, L.C., Favero-Retto, M.P., 67Lourenco, D. & Lima, L.M. A T3R3

- hexamer of the human insulin variant B28Asp. *Biophys. Chem.* **173-174**, 1-7 (2013).
39. Marshall, H., Venkat, M., Seng, N.S., Cahn, J. & Juers, D.H. The use of trimethylamine N-oxide as a primary precipitating agent and related methylamine osmolytes as cryoprotective agents for macromolecular crystallography. *Acta Crystallogr., Sect D: Biol. Crystallogr.* **68**, 69-81 (2012).
40. Smith, G.D., Ciszak, E., Magrum, L.A., Pangborn, W.A. & Blessing, R.H. R6 hexameric insulin complexed with *m*-cresol or resorcinol. *Acta Crystallogr., Sect D: Biol. Crystallogr.* **56**, 1541-1548 (2000).

CHAPTER III

Replacement of proline at ProB28 in insulin with
fluorinated proline analogs modulates hexamer dissociation
and rate of fibrillation

Abstract

Fluorination is a common modification to enhance pharmacokinetics and stability in drug-like molecules. Here we show that non-canonical amino acid mutagenesis can be used to introduce fluorine into insulin at position B28. The monofluoro 4R analog (FypI) yielded a severely destabilized insulin with fast dissociation kinetics. Crystallographic analysis suggests that the C-F bond of FypI clashes with an adjacent tyrosine and destabilizes the oligomeric complexes of insulin. This destabilization is stereospecific, the 4S (FzpI) and difluoro 4,4 (DfpI) analogs were not similarly destabilized. Our results highlight the potential importance of a TyrB26-ProB28 stacking interaction and further extend ncAA mutagenesis of insulin with fluorinated analogs complementing our earlier hydroxylation study.

Introduction

Fluorination has a long history in medicinal chemistry and, sometimes, confers surprising properties to molecules¹. Fluoroproline, for example, have been used to elucidate the nature of the triple helix in collagen and to explore protein structure more broadly²⁻⁵. (2S, 4S)-Fluoroproline (Fzp), (2S, 4R)-Fluoroproline (Fyp), and (2S)-difluoroproline (Dfp) have been used most widely and here we utilize these same amino acids in insulin to explore the structure and activity of insulin more fully (figure 3.1).

Replacement of the hydrogen with alternative substituents on the 4-position of proline provides for a potential steric overlap when in a gauche orientation. In the gauche orientation the steric overlap positions a σ C-H to overlap with the σ^* orbital of the 4-substituent (figure 3.2)³⁻⁵. This effect is dramatic in fluoroproline. Fzp favors the endo conformation by 20:1, an order of magnitude increase compared to Pro⁶. In contrast, Fyp favors the endo conformation by 1:6, reversing the preference of Pro⁶. The exo preference of Fyp also enhances both the trans preference of the residue as well as the n to π^* interaction by enforcing a trans orientation which allows for better overlap of the n to π^* involved orbitals^{4,5,7}. Dfp does not alter the preference for either ring pucker, however it does enhance the interconversion rate between the endo and exo states. Modifying the proline ring with fluorines both enhances the hydrophobicity of the residue and introduces a dipole^{8,9}.

The C-F bond can participate in a wealth of interactions in protein structures, including halogen bonding (akin to hydrogen bonding), aliphatic interactions and aryl interactions. Aromatic interactions with fluorine have a strong positional and angular dependence¹⁰. C-F bonds oriented in the plane of the aromatic must be shifted away from the aromatic centroid while C-F bonds oriented perpendicular to the aromatic plane approach at a closest distance of 2.6 Å.

Fluororolines have been incorporated into a small range of proteins with varied effect. Introduction of Fyp and Fzp in collagen can stabilize or destabilize the triple helix

through stereoelectronic effects that depend on the position of introduction for the fluoroproline⁹. In the villin headpiece protein, mutation of Pro to Fzp improved packing efficiency and stabilized the protein. Conversely, Fyp destabilized a Pro-Tyr interaction in the villin headpiece by enforcing an exo pucker¹¹. Introduction of Fyp into elastins stabilized an exo conformation and thereby stabilized the overall polymer¹². Mutating ubiquitin with Fyp similarly stabilized the protein by enforcing an exo conformation¹³. Fyp has also been used to stabilize mRFP and taq via its exo preference^{14, 15}. Only Fyp stabilized the protein when both Fyp and Fzp were incorporated into scfvs,¹⁶. Introducing Fzp into gfp and thioredoxin resulted in the only known and correctly folded globular proteins with Fzp^{17, 18}. Crystallographic analysis of thioredoxin revealed that both Fyp and Fzp adopted the endo conformation due to local structural constraints. Fyp, Fzp and Dfp were all introduced into beta-2 microglobin. Dfp enhanced fibrillation by lowering the barrier for the cis/trans isomerization through enhanced fraying of the beta-2 microglobin¹⁹. To date, no structures deposited in the PDB include Dfp in the polypeptide chain of a globular protein.

Fyp, Fzp and Dfp all allow for a robust test of several theories that arise out of earlier work on insulin. It seems clear that the enhancement in fibrillation resistance and kinetic disassociation rates seen in HzpI arises, at least in part, from the hydrogen bond across the dimer interface. However, HypI, while monomerized, does not exhibit any of the more typical RAI characteristics. Luckily the fluoroprolines allow for such a test. Fyp strongly favors exo puckers, Fzp favors endo and Dfp favors neither, while most hydrogen bonding interactions should remain unavailable to the C-F bond in the insulin dimer interface.

Results and Discussion

Towards this end, modified proinsulins were expressed with Fyp (FypI), Fzp (FzpI), or Dfp (DfpI) in proline auxotrophic *E. coli*. PI was refolded, matured and purified as described in chapter 2. Maldi-MS analysis revealed that replacement levels were similarly high for all 3 fluoroinsulins (figure 3.3a, Table 3.1). All 3 fluoro insulins

were also active in rodent models and the onset of action as well as glucose lowering potentials were within error of ProI (figure 3.3b).

Insulin dimers and monomers exhibit characteristic and different CD spectra with minima at both 208 nm and 222 nm. The insulin monomer has a deepening of the 208 nm minimum relative to the dimer²⁰. At a concentration of 60 μ M FypI and DfpI exhibit spectra consistent with a monomerized insulin while FzpI exhibits a spectrum consistent with a dimerized insulin (figure 3.4a).

The change in minima between insulin dimer and monomer forms can also be expressed as the ratio of ellipticities at 208 nm and 222 nm. Monomeric insulins approach a ratio of 1.6 while dimeric insulins approach 1.3²¹. A series of CD spectra were collected at concentrations both above and below the typical K_d for ProI with the objective of collecting multiple concentrations for each insulin variant in both monomer and dimer form (figure 3.4b-j). The 208:222 ratios arising out of these experiments are shown in a dot plot (figure 3.4k). Both FypI and DfpI show clear monomer forms until concentrations of around 150 μ M, implying the K_d for these variants is similar to that of HypI (figure 3.4b-k). Unfortunately, FzpI never reached a ratio of 1.6 even at concentrations as low as 3 μ M, implying that both FzpI and ProI have similar K_d 's around 5 μ M.

The rate of onset of action for insulin after subcutaneous injection has been shown to be related to the disassociation rate of the R_6 hexamer²²⁻²⁷. In order to measure this disassociation rate, the fluoroinsulins were subjected to Terpy-driven disassociation as described in chapter 2. FypI had a disassociation rate ($\tau^{1/2} = 40s \pm 4$) indistinguishable from AspI (a model RAI with $\tau^{1/2} = 43s \pm 4$), while FzpI had a disassociation rate ($\tau^{1/2} = 98s \pm 5$) indistinguishable from ProI ($\tau^{1/2} = 90s \pm 4$). DfpI had a disassociation rate ($\tau^{1/2} = 67s \pm 4$) intermediate between that of ProI and AspI (table 3.1). These results are consistent with the approximate K_d 's derived from the CD dilution experiments. The monomerized insulins (DfpI, FypI, AspI) all dissociated at a rate faster than that of ProI. FzpI retains a ProI-like rate of hexamer dissociation and ProI-like dimer K_d .

Fibrillation lag time of the fluoroinsulins was also assessed via the THT assay as described in chapter 2. FzpI and DfpI exhibited lag times that were 1.5-2X longer ($8.4 \text{ h} \pm 1.2$ and $9.0 \text{ h} \pm 0.5$ respectively) compared with the lag times for ProI and AspI ($5.1 \text{ h} \pm 1.5$ and $5.3 \text{ h} \pm 1.0$ respectively). FypI exhibited a dramatically faster fibrillation rate ($2.6 \text{ h} \pm 0.3$) with a lag time one-half that of ProI or AspI (table 3.1).

In sum, these 3 fluoroinsulins present an intriguing array of insulin-like behavior. FypI represents a monomerized insulin with RAI-like dissociation rates of the hexamer and rapid aggregation, strongly implying a structural destabilization underlies these RAI-like characteristics. In contrast DfpI is a monomerized insulin with an intermediate rate of hexamer disassociation and slight enhancement of resistance against fibrillation. In order to further our understanding of the structural and stereoelectronic rationales underpinning these behaviors, the x-ray structure of each variant was determined.

FypI crystals were obtained of the T_2 state to a resolution of 1.17 \AA while crystals of DfpI and FzpI were obtained in the R_6 state to a resolution of 2.25 \AA and 1.31 \AA respectively (table 3.4). All three crystal structures were compared to representative ProI structures (PDB:3T2A for the T_2 state and PDB:1EV6 for the R_6 state) and found to have low overall RMSDs values (table 3.2) and strong visual overlap with ProI dimers (figure 3.4a). Examining the proline rings directly, it is observed that FypI occurs in an exo pucker, FzpI is in an endo pucker and DfpI can be seen in both endo and exo puckers with endo predominating 2:1 (figure 3.6, table 3.2). These endo/exo preferences track with the known pucker preferences of fluoroproline and are consistent with the observation that fluorine, as a strongly electronegative substituent, typically overrides local structural preferences in favor of its intrinsic stereoelectronic endo/exo preferences^{5, 9, 28}. DfpI and FzpI also appear to be displaced slightly from the ProI position whereas FypI overlays near perfectly to B28 in ProI (figure 3.6).

By examining key residues in close proximity to B28 we see a striking picture of subtle structural changes underlying the observed oligomeric and kinetic behaviors.

In ProI TyrB26 sits in close contact with ProB28 and is likely participating in an aromatic-proline C-H/ π interaction²⁹. The exo pucker of FypI forces the 4R fluorine into proximity (3 Å) with the aromatic ring (figure 3.7, table 3.3). However, the angle of the C-F bond for close contact of the FypI is poor for a close contacting C-F to an aromatic. Relative to the plane of the tyrosine ring, the C-F of FypI is 31 degrees, not the ideal perpendicular angles seen in beneficial C-F/aromatic interactions in close proximity to an aromatic ring. ProB28 in ProI can also be seen to be in close packing with Gly23' with just 3.6 Å separating the residues. FypI also seems to pack subtly worse as the distance between FypB28 and Gly23' is increased to 4.0 Å (figure 3.7, table 3.3).

The endo pucker of FzpI closely resembles ProI structures and maintains close contact with TyrB26 similar to ProI (figure 3.7, table 3.3). FzpI also subtly overlaps GluB21' with 0.26 Å of overlap. In FzpI the C-F seems to pack efficiently into the dimer interface without any strong negative interactions.

In contrast the enhanced hydrophobic bulk of DfpI does not seem capable of packing without negative interactions. The C-F on DfpI is in close contact with several atoms on Glu21' and Gly20', interactions that are not canonically considered beneficial. A very large separation of DfpB28 and ValB3 as well as separation from TyrB26 is also apparent in the structure, depriving B28 of close packing internal to the monomer.

In conjunction with the hydroxyinsulins, the fluoroinsulins present intriguing trends relating packing and conformation of B28 with kinetic and oligomeric activities.

All insulins with lower barriers to entering the exo pucker conformation (HypI, FypI, DfpI) are observed to have increased K_d 's of dimerization. Similarly all insulins (HzpI, FzpI, ProI) with endo preferences (or higher barriers to entering exo pucker), are observed to maintain WT-like dimerization constants. FypI is the only insulin with a faster rate of fibrillation, implying destabilization or fraying of the B-chain in solution. The structure of FypI provides an explanation for this behavior, the exo pucker of FypI

brings the C-F into a rare orientation that is expected to be destabilizing. Presumably this destabilization also results in faster hexamer disassociation by destabilizing all forms of insulin with the B-chain packed against TyrB26. The B-chain in the monomer may be able to swing out in a more extended conformation that relaxes this clash. Frayed B-chain in monomers may even be favored to swing out prematurely as the exo preference leads to n to π^* stabilization. Finally the fast disassociation of DfpI can be explained on the basis of minor clashes between the fluorines against atoms on the opposing monomer (Glu21' and Gly20'). Overlaps of only 0.2Å - 0.4Å are at most minor clashes³⁰, and may drive monomerization and kinetic disassociation away from higher order oligomers but would not impact the organization of the B-chain in the monomer state. This overlap for DfpI is seen in the R₆ structure, one can speculate that the T₂ structure may not contain a similar overlap and that such a difference would lead to all the observed behaviors of DfpI. Unfortunately all attempts to obtain a T₂ crystal of DfpI have, so far, failed to yield a diffracting crystal.

Conclusions

In sum, the fluoroinsulins demonstrate the importance of local structural contacts influencing overall protein behavior. Insulin itself is a model system for structural and biochemical analysis of 4-substituted prolines. High-resolution crystal structures have been obtained for all variants herein regardless of the stabilization or destabilization that may be present in these insulin analogs. Future interest in 4-substituted prolines, for both improving insulin as well as basic biochemical exploration, might be complemented by further systematic introduction into insulin. Indeed the Tirrell group is continuing these efforts by looking into a range of further substituent groups (amino, azido etc) on the 4-position. The results from this and future proline studies will inform ncAA mutagenesis for a wide range of potential applications where local structural control around prolines would be of maximum utility (for example; proline in the loop of an antibody paratope). Medicinal chemistry operates on the general principle that atomic interactions can be tuned to improve binding, pharmacokinetics or other important pharmaceutical

properties. We anticipate that such tuning of proteins down to an atomic level, akin to medicinal chemistry, will become commonplace.

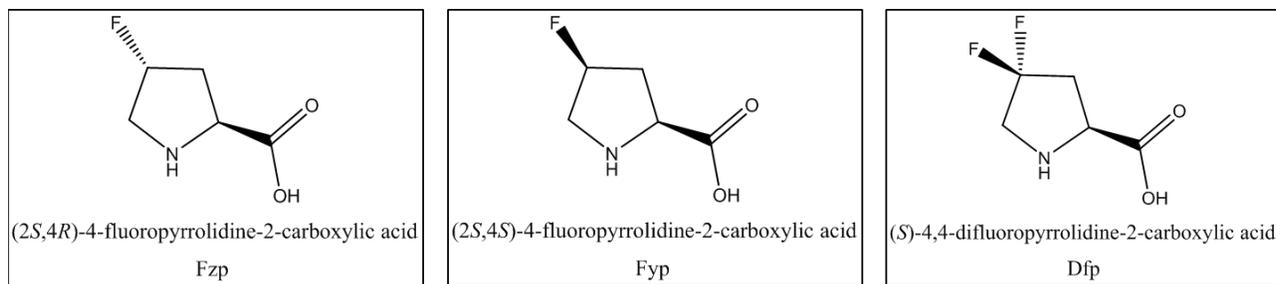


Figure 3.1 | Fluorinated proline analogs
The structures of the fluoroproline analogs discussed in this chapter.

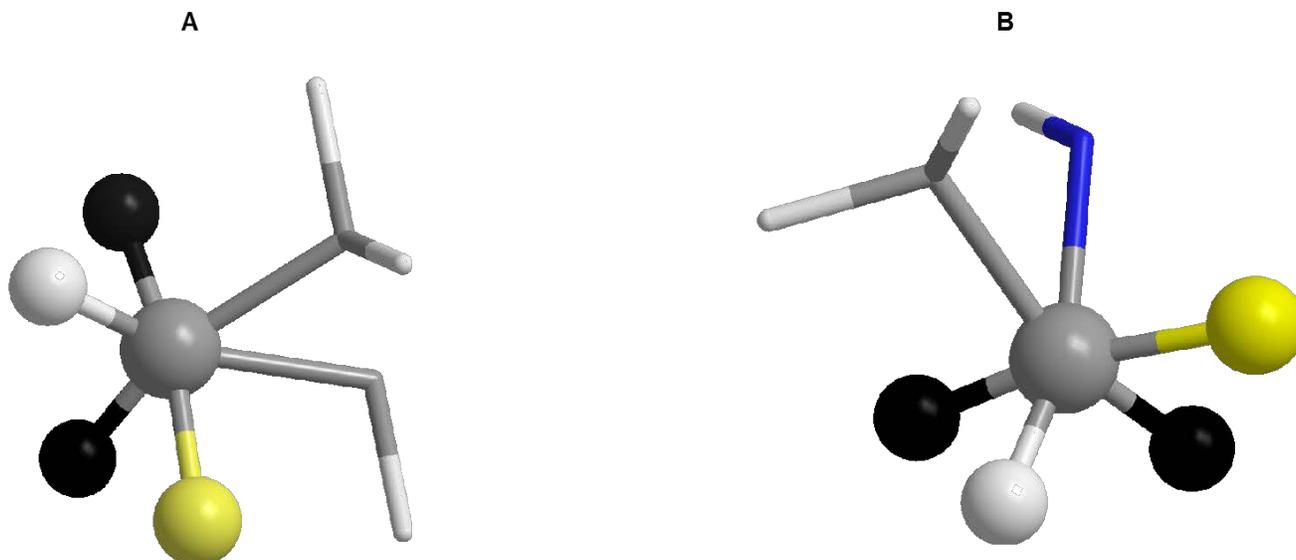


Figure 3.2 | Orientation of mono-fluoroprolines in conformers

Visualization of the bond orientations in the preferred conformations of fluoroprolines. Carbon (grey), fluorine (yellow), and hydrogen (white, black) are shown with the 4-position carbon closest to the reader. Hydrogen atoms which are colored black are bonded to the 3-position (**A**) or 5-position (**B**) and are the potential orbital partners for the C-F antibonding orbital.

A) Fyp with an exo pucker.

B) Fzp with an endo pucker

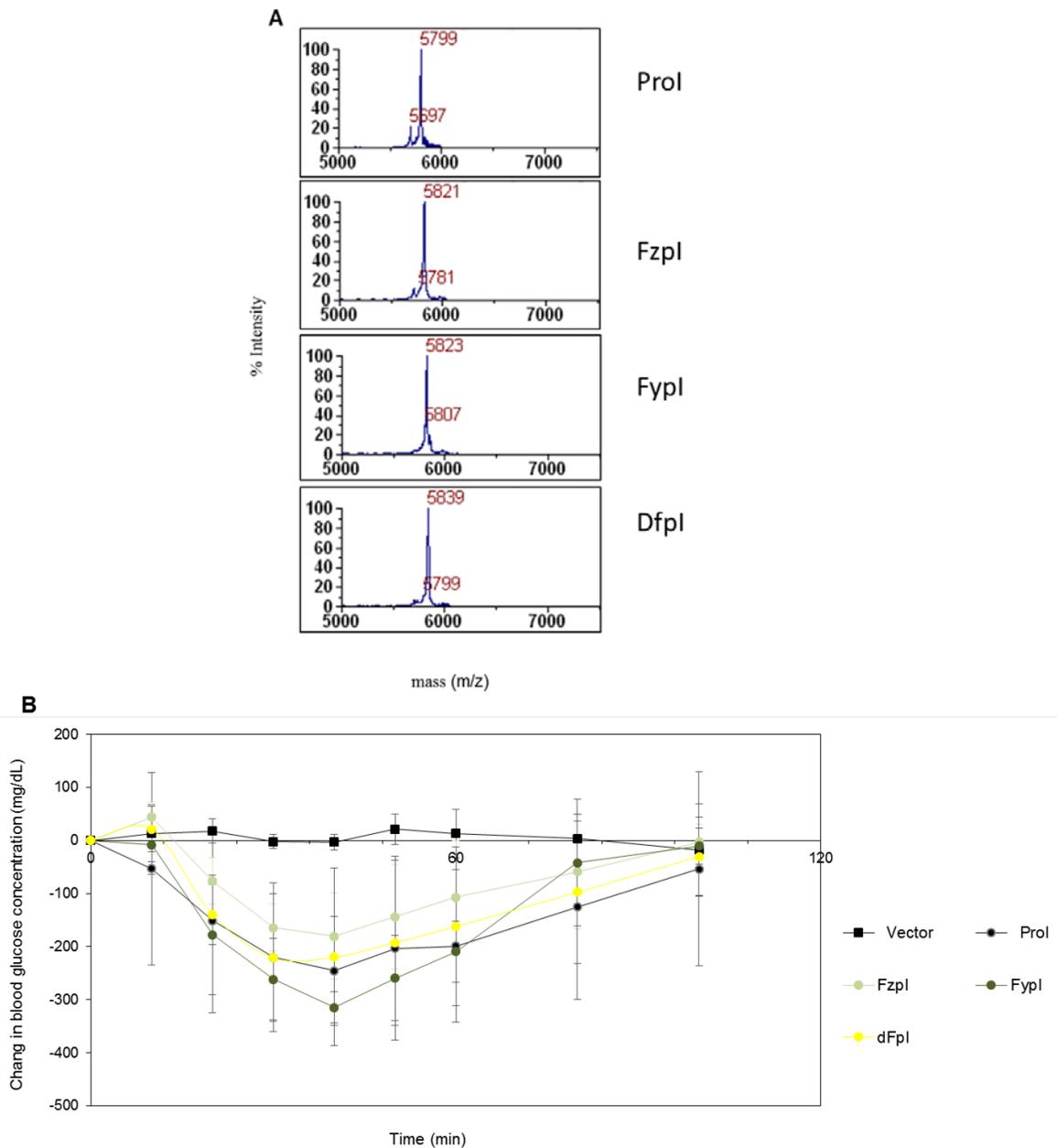


Figure 3.3| Fluoroprolines incorporate into insulin and do not perturb the glucose lowering activity of insulin in a mouse model of diabetes

A) MALDI mass spectra showing successful incorporation and maturation of the fluoroinsulins. The WT mass of 5799 is visible in ProI and the monofluoro or difluoro masses are visible in Fzpl (+18 expected), FypI (+18 expected) or Dfpl (+36 expected).

B) Reduction of blood glucose following subcutaneous injection of 35 $\mu\text{g}/\text{kg}$ insulins into streptozotocin-induced diabetic mice. Glucose levels were measured post-injection via tail vein sampling. ProI, dFpI, Fzpl, FypI or vector were formulated as described in chapter 2. Error bars denote one standard deviation ($n = 3$).

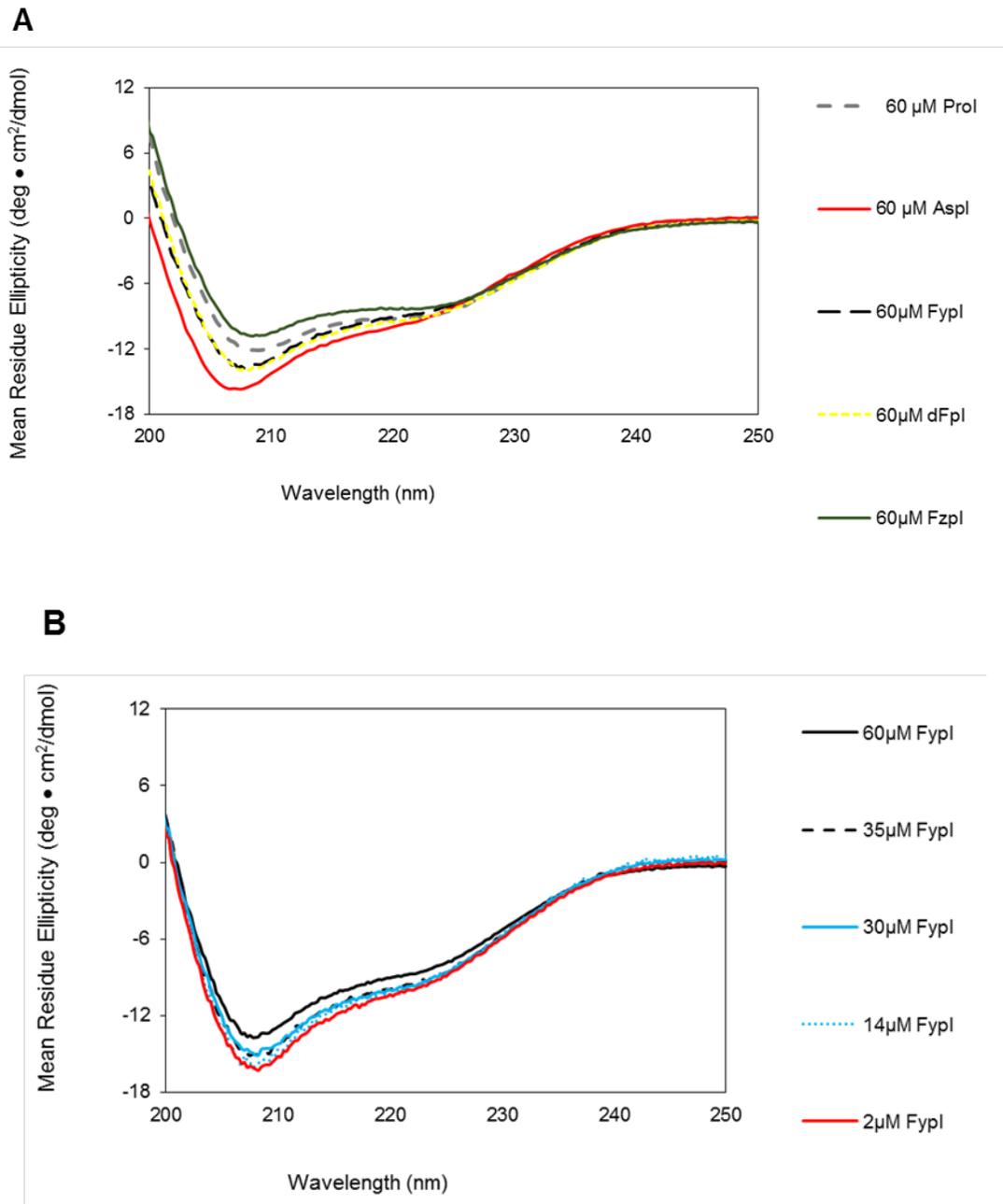


Figure 3.4 Circular Dichroism Spectra of Fluoroinsulins

- (A) Far UV CD spectra collected on 60 μM insulins in 10 mM phosphate buffer, pH 8.0 at 25°C.
- (B) Far UV CD spectra collected on multiple concentrations of FypI as indicated in the legend. CD spectra collected in 10 mM phosphate buffer, pH 8.0 at 25°C.

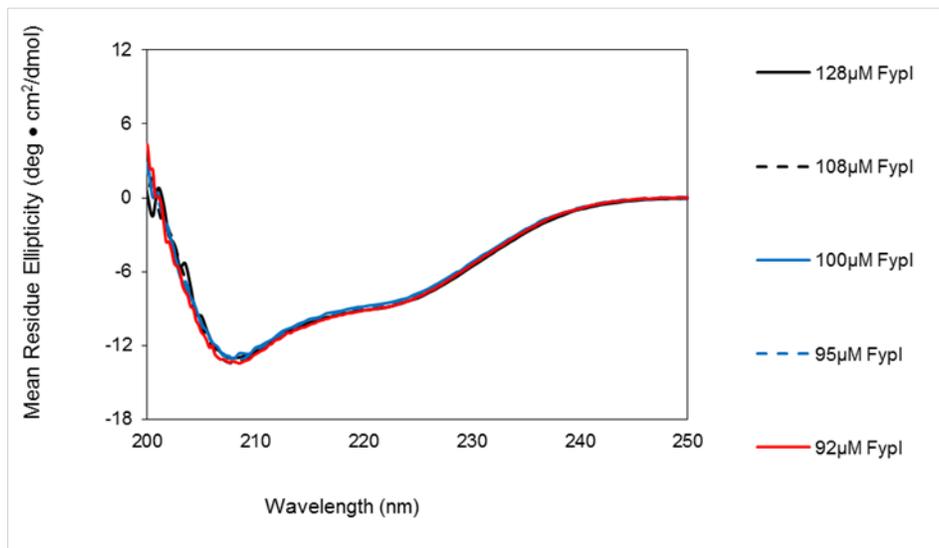
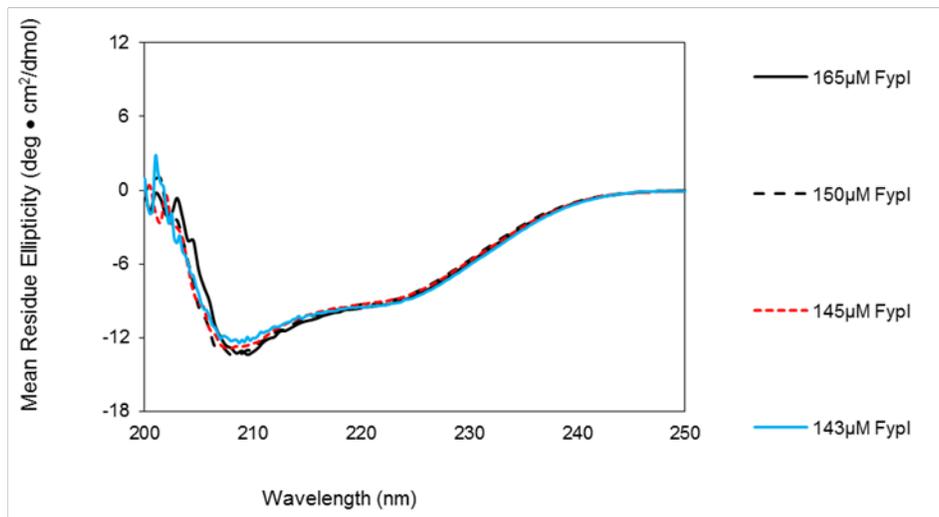
C**D**

Figure 3.4 Circular Dichroism Spectra of Fluoroinsulins

- (C) Far UV CD spectra collected on multiple concentrations of FypI as indicated in the legend. CD spectra collected in 10 mM phosphate buffer, pH 8.0 at 25°C.
- (D) Far UV CD spectra collected on multiple concentrations of FypI as indicated in the legend. CD spectra collected in 10 mM phosphate buffer, pH 8.0 at 25°C.

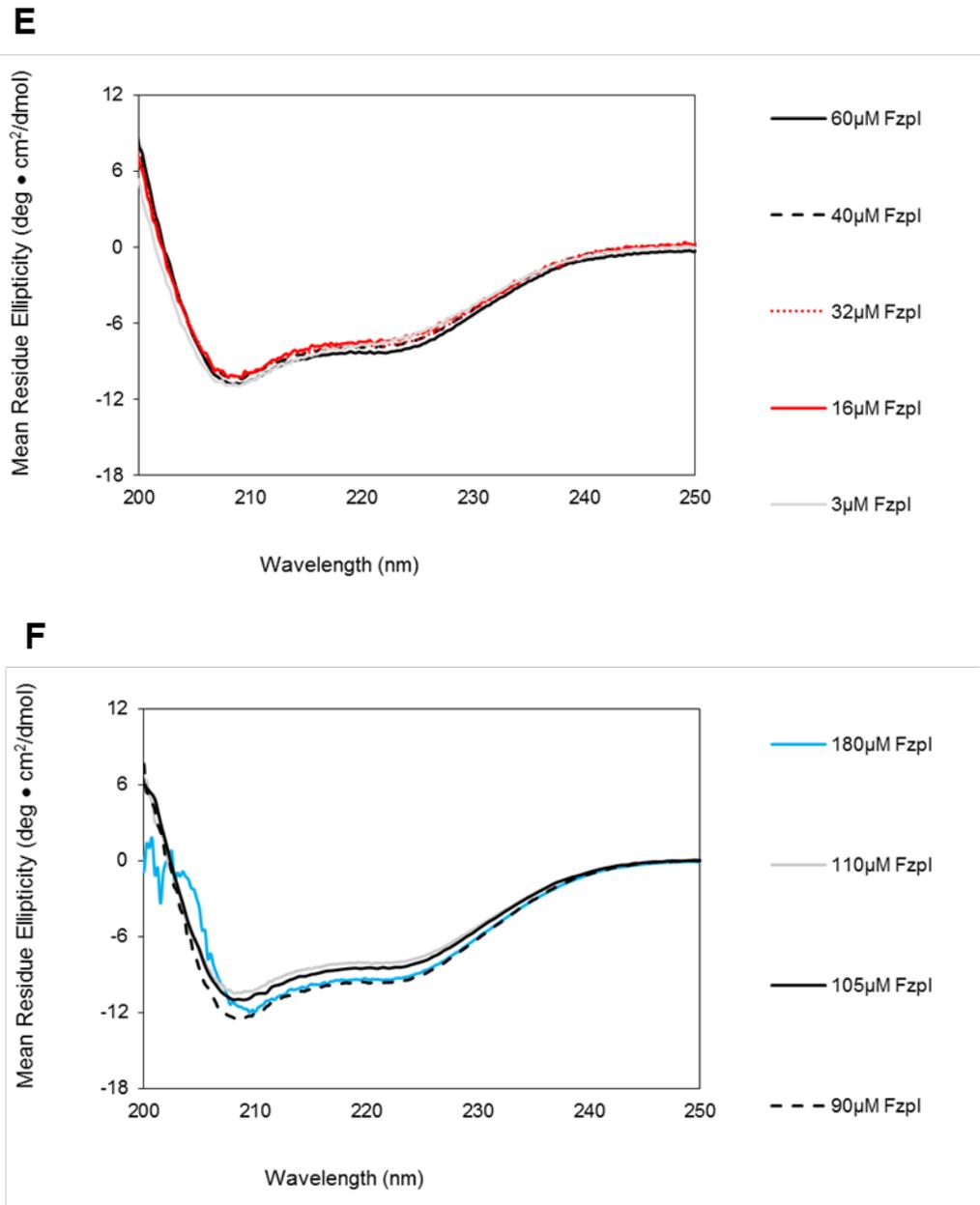


Figure 3.4 Circular Dichroism Spectra of Fluoroinsulins

- (E) Far UV CD spectra collected on multiple concentrations of Fzpl as indicated in the legend. CD spectra collected in 10 mM phosphate buffer, pH 8.0 at 25°C.
- (F) Far UV CD spectra collected on multiple concentrations of Fzpl as indicated in the legend. CD spectra collected in 10 mM phosphate buffer, pH 8.0 at 25°C.

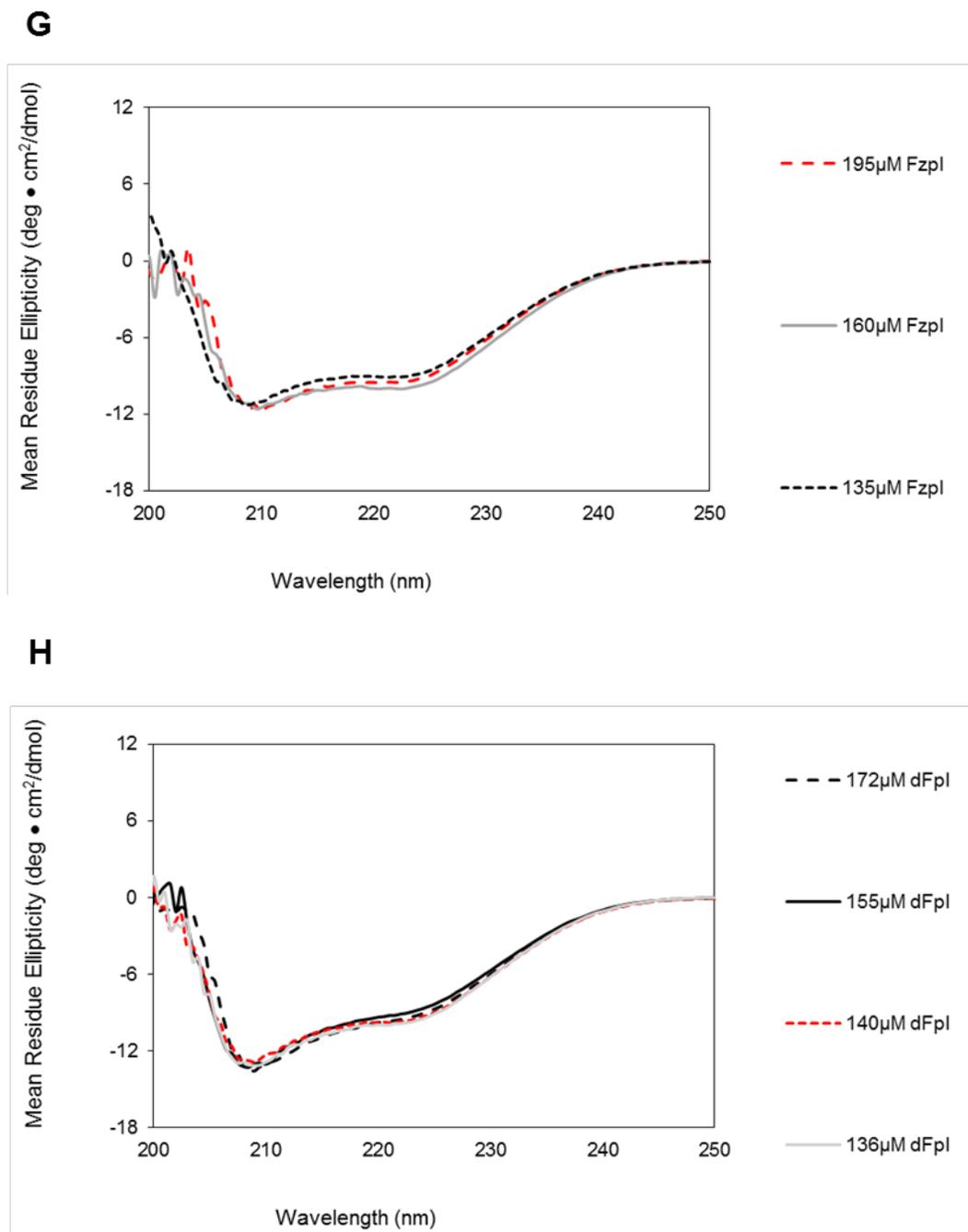


Figure 3.4 Circular Dichroism Spectra of Fluoroinsulins

- (G) Far UV CD spectra collected on multiple concentrations of Fzpl as indicated in the legend. CD spectra collected in 10 mM phosphate buffer, pH 8.0 at 25°C.
- (H) Far UV CD spectra collected on multiple concentrations of dFpl as indicated in the legend. CD spectra collected in 10 mM phosphate buffer, pH 8.0 at 25°C.

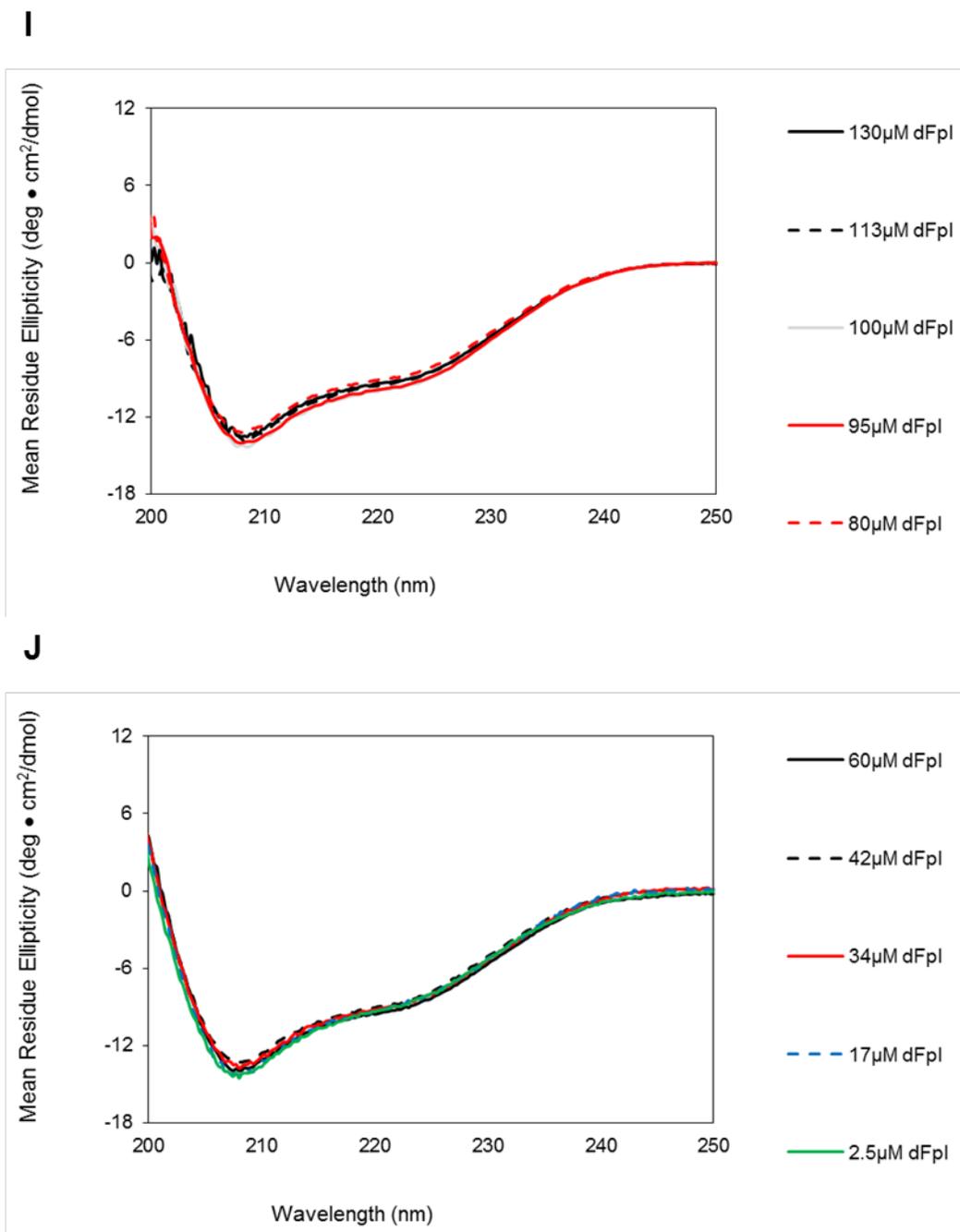


Figure 3.4 Circular Dichroism Spectra of Fluoroinsulins

- (I) Far UV CD spectra collected on multiple concentrations of dFpI as indicated in the legend. CD spectra collected in 10 mM phosphate buffer, pH 8.0 at 25°C.
- (J) Far UV CD spectra collected on multiple concentrations of dFpI as indicated in the legend. CD spectra collected in 10 mM phosphate buffer, pH 8.0 at 25°C.

K

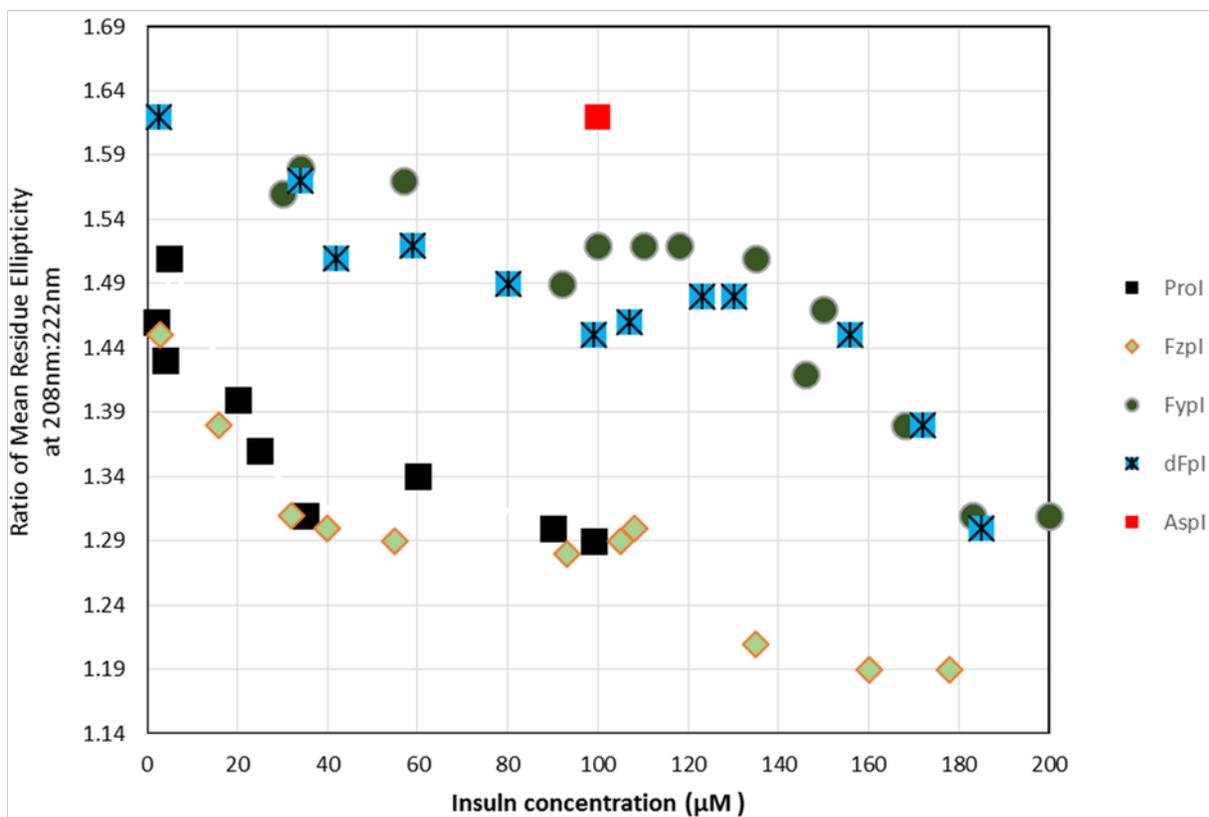


Figure 3.4 Circular Dichroism Spectra of Fluoroinsulins

(K) Ratio of the Far UV CD spectra 208nm minima to 222nm minima for fluoroinsulin variants over a range of concentrations. Ratios were calculated from the 208nm and 222nm data points on the preceding Far UV spectra in this figure.

A

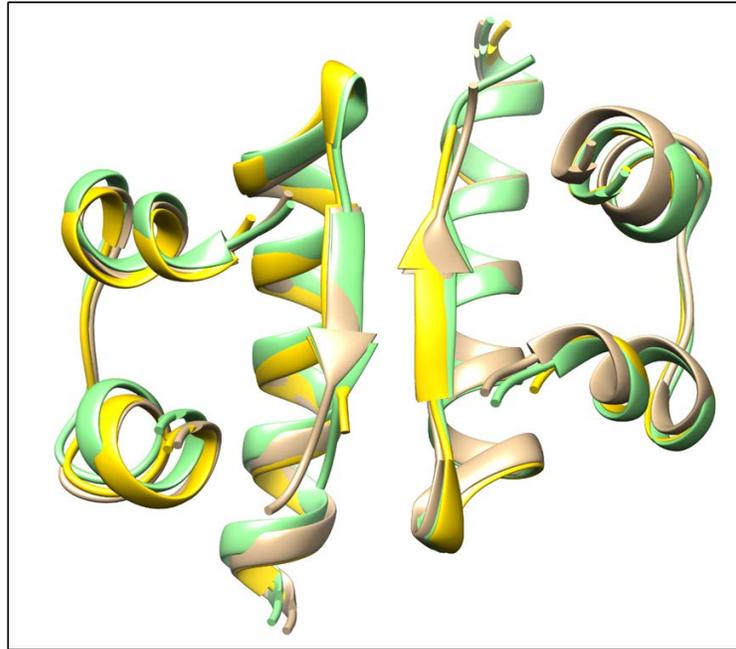


Figure 3.5 | The dimer in the R₆ hexamer of the fluoroinsulins.

A) Alignment of R₆ ProI (tan, PDB:1EV6), R₆-FzpI (light green), and R₆-dFpI (Gold) showing a single dimer in the hexamer.

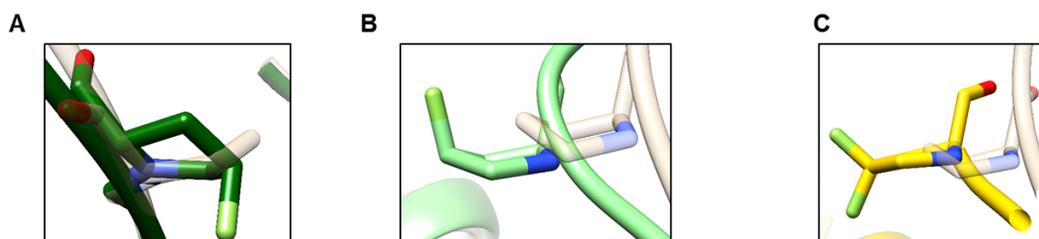


Figure 3.6 | Alignment at position B28.

A) Alignment of T₂-ProI (tan, PDB ID: 3T2A) and T₂-FypI (dark Green) highlighting the pucker of the proline ring.

B) Alignment of R₆-ProI (tan, PDB ID: 1EV6) and R₆-Fzpl (light green) highlighting the pucker of the proline ring.

C) Alignment of R₆-ProI (tan, PDB ID: 1EV6) and R₆-dFpl (gold) highlighting the pucker of the proline ring.

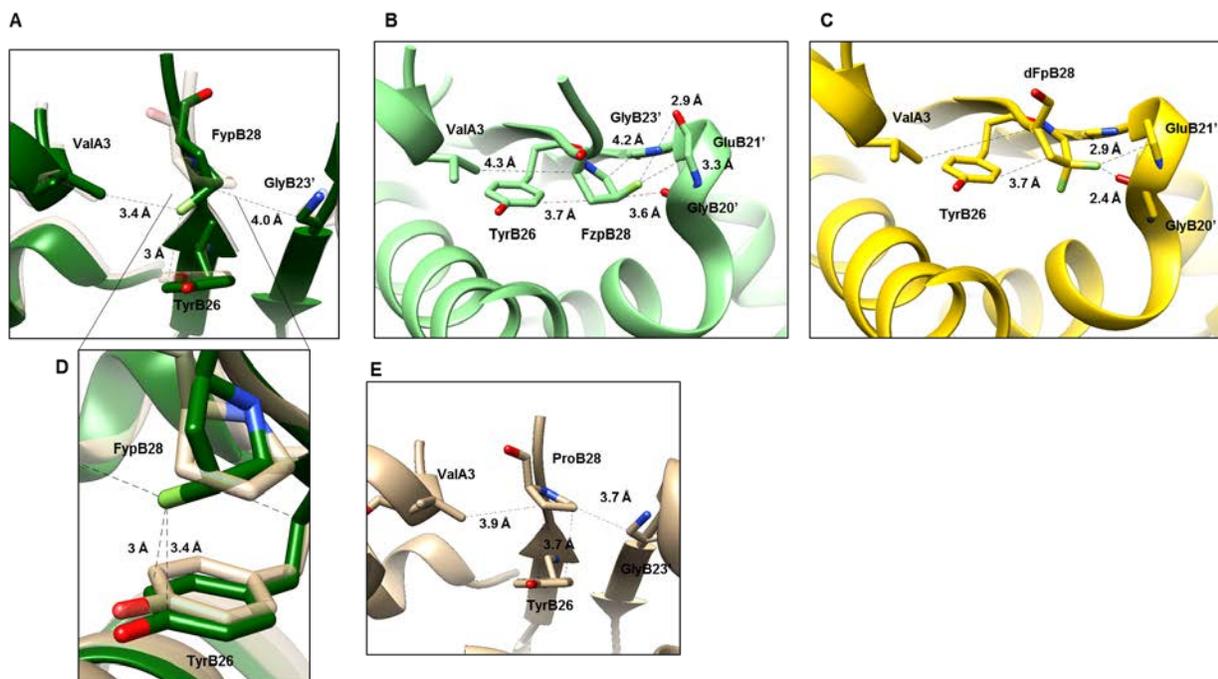


Figure 3.7 | B28 highlighting key contacts in the dimer interface.

A) Alignment of T_2 ProI (tan, PDB:3T2A), and T_2 -FypI (dark green) centered on position B28 highlighting the important interactions made between B28 and surrounding residues in the dimer. Interaction distances are listed in angstroms and are averages where more than one monomer is present in the asymmetric unit.

B,C) R_6 -FzpI (light green) and R_6 -dFpI (gold) centered on position B28 highlighting the important interactions made between B28 and surrounding residues in the dimer. Interaction distances are listed in angstroms.

D) Alignment of T_2 ProI (tan, PDB:3T2A), and T_2 -FypI (dark green) centered on the interaction between the fluorine from FypB28 and the phenyl ring of TyrB26. The centroid of the phenyl ring on B26 is offset 1.4 Å from the fluorine. The angle of the C-F in B28 relative to the phenyl ring in B26 is 31 degrees.

E) T_2 ProI (tan, PDB:3T2A) centered on position B28 highlighting the important interactions made between B28 and surrounding residues in the dimer. Interaction distances are listed in angstroms

Protein	%ncPro‡	Hexamer $\tau_{1/2}$ (s)	Fibrillation lag time (h)
Prol	--	90.4 ± 4.2	5.1 ± 1.5
Aspl	--	42.7 ± 4.3	5.3 ± 1.0
Hzpl*	91%	55.1 ± 2.4	19 ± 4.0
HypI*	88%	87.4 ± 6.3	5.5 ± 1.2
Fzpl*	97%	98.3 ± 5.5	8.4 ± 1.2
FypI*	96%	40.0 ± 4.1	2.6 ± 0.3
dFpl*	86%	67.1 ± 4.3	9.0 ± 0.5

Table 3.1| Key properties of Fluoroinsulins

Errors are given as one s.d. ($n \geq 4$) unless otherwise noted

‡Incorporation error <2% and quantified using MALDI-MS on gluC digested peptide containing B28

*Active in vivo ($n \geq 3$)

	RMSD vs ProI (Å) ^a	Ring conformation ^b	Chi 1 angle ^c	Phi ^c
ProI T ₂		Endo	22	-75
ProI R ₆		Endo (5/6)	21	-60
Fypl T ₂	0.38	Exo	-26.	-66
Fzpl R ₆	0.79	Exo (5/6)	14	-74
dFpl R ₆	0.66	Endo (8/12), Exo (4/12)	3.6	-77

Table 3.2| RMSD and proline angles

^a Fluoroproline T₂ insulins were aligned against T₂ ProI (PDB:3T2A) and R₆ insulins were aligned against R₆-ProI (PDB: 1EV6). RMSD was calculated over backbone atoms.

^b Ring conformation was determined on the basis of the Chi1 angles. Where more than one B28 residue is present in the asymmetric unit the predominant conformation is listed. The fraction in parentheses denotes the number of B28 residues in one asymmetric unit with the predominant conformation.

^c Chi1 and Phi angles are reported as an average over all B28 residues present in the asymmetric unit for each crystal structure. For T₂ insulins only a single B28 is averaged, for R₆ insulins there are 6 or 12 prolines present in the asymmetric unit and the value reported is the average value over all prolines in the asymmetric unit.

T ₂ Prol	Proline atom	Other residue	VDW Overlap (Å) ^a	Distance (Å)
	CD	Gly23' CA	0.13	3.6
	CG	Tyr26 CE	-0.14	3.8
	CA	Val3 CG	-0.14	3.9

T ₂ Fypl	Proline atom	Other residue	VDW Overlap (Å) ^a	Distance (Å)
	CD	Gly23' CA	-0.22	4.0
	CG	Tyr26 CZ	-0.18	3.8
	CB	Val3 C	-0.33	4.1
	FD	Tyr26 CZ	-0.11	3
	FD	Val3 CG	-0.24	3.4
	FD	Tyr26 CE	-0.33	3.4

R ₆ Prol	Atomic interaction ^b	Other residue	VDW Overlap (Å) ^a
	C-C	Gly23'	-0.61
	C-C	Tyr26	0.03
	C-C	Val3	-0.07
	C-C	Glu21'	-0.07
	C-O	Gly20'	-0.63

R ₆ Fzpl	Atomic interaction ^b	Other residue	VDW Overlap (Å) ^a
	C-C	Gly23'	-0.41
	C-C	Tyr26	-0.05
	C-C	Val3	-0.62
	F-C	Glu21'	0.26
	F-C	Gly20'	-0.29
	F-O	Glu21'	-0.08

R ₆ dFpl	Atomic interaction ^b	Other residue	VDW Overlap (Å) ^a
	C-C	Gly23'	-0.17
	C-C	Tyr26	-0.34
	C-C	Val3	None modeled close
	F-C	Glu21'	0.29
	F-C	Gly20'	0.20
	F-O	Glu21'	0.23
	F-O	Gly20'	0.39

Table 3.3 | Key B28 contacts in fluoroproline insulins

^a VDW overlap was calculated in Chimera as the distance between the van der waals radii of each participating atom. Where more than one monomer is present in the asymmetric unit of the crystal structure, the overlap value listed in the table is an average over all B28-contact overlaps. Positive values indicate VDW radii overlapping. Negative values indicate the distance between VDW radii. Distances are measured as the linear distance between the center point of the two atoms listed in each row. Nomenclature of the atoms in the residues follows standard protein atomic naming as utilized in the PDB.

^b Atomic interaction indicates the closest pairing of atoms involved in the B28 to other residue contact. The first atom listed is the atomic identity from B28. The second atom listed is located on the residue listed in the other residue column.

^c The strongest interactions are bolded.

	Fypl-T₂ 5UOZ	Fzpl-R₆ 5UQA	Dfpl-R₆
Data collection			
Space group	<i>I</i> 2 ₁ 3	<i>P</i> 1 2 ₁ 1	<i>P</i> 1
Cell dimensions a, b, c (Å)	78.094 78.094 78.094	46.936, 61.258, 60.390	40.599, 47.425, 79.705
α, β, γ (°)	90.00, 90.00, 90.00	90.00, 111.69, 90.00	89.96, 89.93, 73.41
Resolution (Å)	31.88 - 1.175 (1.217 - 1.175)	35.54-1.31 (1.326-1.31)	39.35-2.25 (2.35- 2.25)
R_{p.i.m.} (%)	3.2 (29.7)	4.0	2.6 (29.8)
Mn(I /sd)	18.03 (3.88)	8.3 (1.7)	13.2 (2.6)
Completeness (%)	98.81 (89.84)	97.43 (97.6)	87.9 (88.8)
Redundancy	8.9 (5.7)	8.3 (3.4)	6.3 (6.3)
Refinement			
No. reflections	26219(2352)	74368 (2587)	23460 (2132)
R_{work} / R_{free} (%)	14.09/15.49 (17.96/19.65)	15.24/18.21 (25.94/29.4)	23.1/29.4 (35.3/37.9)
No. atoms			
Protein	409	2147	4047
Ligand/ion	0	301	83
Water	72	240	2
RMSD			
Bond lengths (Å)	0.048	0.006	0.009
Bond angles (°)	1.47	1.103	1.193
Ramachandran map analysis			
Favored	50	294	500
Allowed	0	1	0
Outliers	0	2	0

Table 3.4 | Data Table and Refinement Values for Fluoroinsulins

References

- 1 E. P. Gillis, K. J. Eastman, M. D. Hill, D. J. Donnelly, and N. A. Meanwell, 'Applications of Fluorine in Medicinal Chemistry', *J Med Chem*, 58 (2015), 8315-59.
- 2 W. Kim, A. George, M. Evans, and V. P. Conticello, 'Cotranslational Incorporation of a Structurally Diverse Series of Proline Analogues in an Escherichia Coli Expression System', *Chembiochem*, 5 (2004), 928-36.
- 3 A. K. Pandey, K. M. Thomas, C. R. Forbes, and N. J. Zondlo, 'Tunable Control of Polyproline Helix (Ppii) Structure Via Aromatic Electronic Effects: An Electronic Switch of Polyproline Helix', *Biochemistry*, 53 (2014), 5307-14.
- 4 A. K. Pandey, D. Naduthambi, K. M. Thomas, and N. J. Zondlo, 'Proline Editing: A General and Practical Approach to the Synthesis of Functionally and Structurally Diverse Peptides. Analysis of Steric Versus Stereoelectronic Effects of 4-Substituted Prolines on Conformation within Peptides', *J Am Chem Soc*, 135 (2013), 4333-63.
- 5 Robert W. Newberry, and Ronald T. Raines, '4-Fluoroproline: Conformational Analysis and Effects on the Stability and Folding of Peptides and Proteins', (Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 1-25.
- 6 R. Improta, C. Benzi, and V. Barone, 'Understanding the Role of Stereoelectronic Effects in Determining Collagen Stability. 1. A Quantum Mechanical Study of Proline, Hydroxyproline, and Fluoroproline Dipeptide Analogues in Aqueous Solution', *J Am Chem Soc*, 123 (2001), 12568-77.
- 7 R. W. Newberry, G. J. Bartlett, B. VanVeller, D. N. Woolfson, and R. T. Raines, 'Signatures of N \rightarrow Π^* Interactions in Proteins', *Protein Sci*, 23 (2014), 284-8.
- 8 L. Moroder, and N. Budisa, 'Synthetic Biology of Protein Folding', *Chemphyschem*, 11 (2010), 1181-7.
- 9 M. D. Shoulders, K. J. Kamer, and R. T. Raines, 'Origin of the Stability Conferred Upon Collagen by Fluorination', *Bioorg Med Chem Lett*, 19 (2009), 3859-62.
- 10 C. Bissantz, B. Kuhn, and M. Stahl, 'A Medicinal Chemist's Guide to Molecular Interactions', *J Med Chem*, 53 (2010), 5061-84.
- 11 T. Y. Zheng, Y. J. Lin, and J. C. Horng, 'Thermodynamic Consequences of Incorporating 4-Substituted Proline Derivatives into a Small Helical Protein', *Biochemistry*, 49 (2010), 4255-63.
- 12 W. Kim, R. A. McMillan, J. P. Snyder, and V. P. Conticello, 'A Stereoelectronic Effect on Turn Formation Due to Proline Substitution in Elastin-Mimetic Polypeptides', *J Am Chem Soc*, 127 (2005), 18121-32.
- 13 M. D. Crespo, and M. Rubini, 'Rational Design of Protein Stability: Effect of (2s,4r)-4-Fluoroproline on the Stability and Folding Pathway of Ubiquitin', *PLoS One*, 6 (2011), e19425.
- 14 B. Holzberger, and A. Marx, 'Replacing 32 Proline Residues by a Noncanonical Amino Acid Results in a Highly Active DNA Polymerase', *J Am Chem Soc*, 132 (2010), 15708-13.
- 15 K. Deepankumar, S. P. Nadarajan, N. Ayyadurai, and H. Yun, 'Enhancing the Biophysical Properties of Mrfp1 through Incorporation of Fluoroproline', *Biochem Biophys Res Commun*, 440 (2013), 509-14.

- 16 S. Edwardraja, S. Sriram, R. Govindan, N. Budisa, and S. G. Lee, 'Enhancing the Thermal Stability of a Single-Chain Fv Fragment by in Vivo Global Fluorination of the Proline Residues', *Mol Biosyst*, 7 (2011), 258-65.
- 17 D. Roderer, R. Glockshuber, and M. Rubini, 'Acceleration of the Rate-Limiting Step of Thioredoxin Folding by Replacement of Its Conserved Cis-Proline with (4S)-Fluoroproline', *Chembiochem*, 16 (2015), 2162-6.
- 18 M. Rubini, M. A. Scharer, G. Capitani, and R. Glockshuber, '(4r)- and (4s)-Fluoroproline in the Conserved Cis-Prolyl Peptide Bond of the Thioredoxin Fold: Tertiary Structure Context Dictates Ring Puckering', *Chembiochem*, 14 (2013), 1053-7.
- 19 V. Y. Torbeev, and D. Hilvert, 'Both the Cis-Trans Equilibrium and Isomerization Dynamics of a Single Proline Amide Modulate Beta2-Microglobulin Amyloid Assembly', *Proc Natl Acad Sci U S A*, 110 (2013), 20051-6.
- 20 J. Goldman, and F. H. Carpenter, 'Zinc Binding, Circular Dichroism, and Equilibrium Sedimentation Studies on Insulin (Bovine) and Several of Its Derivatives', *Biochemistry*, 13 (1974), 4566-74.
- 21 S. H. Nakagawa, D. F. Tager Hs Fau - Steiner, and D. F. Steiner, 'Mutational Analysis of Invariant Valine B12 in Insulin: Implications for Receptor Binding'.
- 22 D. N. Brems, L. A. Alter, M. J. Beckage, R. E. Chance, R. D. DiMarchi, L. K. Green, H. B. Long, A. H. Pekar, J. E. Shields, and B. H. Frank, 'Altering the Association Properties of Insulin by Amino Acid Replacement', *Protein Eng*, 5 (1992), 527-33.
- 23 S. R. Mudaliar, F. A. Lindberg, M. Joyce, P. Beerdsen, P. Strange, A. Lin, and R. R. Henry, 'Insulin Aspart (B28 Asp-Insulin): A Fast-Acting Analog of Human Insulin: Absorption Kinetics and Action Profile Compared with Regular Human Insulin in Healthy Nondiabetic Subjects', *Diabetes Care*, 22 (1999), 1501-6.
- 24 A. Plum, H. Ageroso, and L. Andersen, 'Pharmacokinetics of the Rapid-Acting Insulin Analog, Insulin Aspart, in Rats, Dogs, and Pigs, and Pharmacodynamics of Insulin Aspart in Pigs', *Drug Metab Dispos*, 28 (2000), 155-60.
- 25 S. E. Shoelson, Z. X. Lu, L. Parlautan, C. S. Lynch, and M. A. Weiss, 'Mutations at the Dimer, Hexamer, and Receptor-Binding Surfaces of Insulin Independently Affect Insulin-Insulin and Insulin-Receptor Interactions', *Biochemistry*, 31 (1992), 1757-67.
- 26 V. Pandyarajan, and M. A. Weiss, 'Design of Non-Standard Insulin Analogs for the Treatment of Diabetes Mellitus', *Curr Diab Rep*, 12 (2012), 697-704.
- 27 T. Lauritzen, S. Pramming, T. Deckert, and C. Binder, 'Pharmacokinetics of Continuous Subcutaneous Insulin Infusion', *Diabetologia*, 24 (1983), 326-9.
- 28 J. A. Hodges, and R. T. Raines, 'Stereochemical Effects on Collagen Stability: The Dichotomy of 4-Fluoroproline Diastereomers', *J Am Chem Soc*, 125 (2003), 9262-3.
- 29 N. J. Zondlo, 'Aromatic-Proline Interactions: Electronically Tunable Ch/Pi Interactions', *Acc Chem Res*, 46 (2013), 1039-49.
- 30 Srinivas Ramachandran S Fau - Ramachandran, Pradeep Kota P Fau - Kota, Feng Ding F Fau - Ding, and Nikolay V. Dokholyan Nv Fau - Dokholyan, 'Automated Minimization of Steric Clashes in Protein Structures'.

CHAPTER IV

Replacement of ProB28 in insulin by proline analogs with ring size variation modulates hexamer dissociation and rate of fibrillation

Abstract

Here we show that non-canonical amino acid mutagenesis can be used to introduce variation into the ring composition of position ProB28 in insulin. Addition or removal of a methylene from the proline ring was associated with an overall increase or decrease in stability, respectively. Crystallographic analysis suggests that the packing of these rings against an adjacent tyrosine and in a hydrophobic pocket influences the kinetic dissociation and resistance to fibrillation in insulin. Thiazolidine and dehydro-variants subtly altered packing and also influenced kinetic dissociation and resistance to fibrillation. Our results highlight the potential importance of a TyrB26-ProB28 stacking interaction and further extend ncAA mutagenesis of insulin with ring variants demonstrating a complete range of medicinal chemistry-like manipulations on a globular protein.

Introduction

In the work detailed in the previous chapters, we investigated the effects of 4-substituted prolines on insulin structure and function. We found that that these altered proline rings modulate the kinetic disassociation of the insulin hexamer as well as oligomerization and aggregation. We sought to further understand the role of residue ProB28 by replacing ProB28 with prolines containing altered rings. Pipecolic acid (Pip), azetidine-2-carboxylic acid (Aze), 3,4 Dehydroproline (Dhp) and Thiazolidine-4-carboxylic acid (Thia) are all analogs of proline with alterations to the ring (figure 4.1). Pip and Aze increase and decrease the size of the proline ring by a single carbon, respectively, while Thia replaces the 4-position carbon with a sulfur atom. Thia effectively replaces the gamma hydrogens with lone pairs of electrons from the sulfur atom. Dhp introduces a double bond in the proline ring.

Together these 4 proline variants have been lightly studied in peptides and all but unstudied in proteins. Only a single protein structure, 1SAV with Thia, has been deposited for any of these proline analogs¹.

Accordingly, Thia has been the most studied of all the variants. The carbon-sulfur bond in Thia is elongated relative to Pro and the angles of the ring are slightly reduced to accommodate this elongation². Thia's conformational preferences are strikingly similar to the parent proline^{2,3}. Thia prefers the endo pucker as well as trans-peptide conformation, although the degree to which Thia can isomerize to the cis-peptide depends on the system of analysis used. Simulations and polyproline test peptides disagree as to the tendency of Thia to cis/trans isomerize²⁻⁵. Thia seems to decrease the $n \rightarrow \pi^*$ interaction energy by 0.2 kcal/mol relative to Pro^{2,3}. Replacement of Thia in polyproline peptides lead to a disruption in the polyproline conformation, and a single substitution lead to a 16⁰C loss in T_m ⁴. Replacement of 5 Pro residues by Thia in annexin V leads to a loss of T_m by 4.5⁰C, although functionality of the protein was not compromised¹. In the crystal structure, Thia was modeled as a mixture of endo and exo preferences¹. In an elastin-like-polypeptide model, replacement of Pro by Thia leads to a decrease of 60⁰C in the transition temperature⁶.

Pip has been studied primarily in the contexts of peptides ⁷, ion channel activation ⁸ and *in silico* models ⁹. The 6-membered ring of Pip can orient itself into the common chair and boat conformations. From quantum mechanical calculations, the chair conformation is several kcal/mol favored over the boat conformation ⁹. Notably, an ion channel with Pip incorporated in place of an important Pro retained functionality, demonstrating that incorporation of Pip and loss of the 5-membered proline ring is not inherently detrimental to protein folding ⁸.

Aze has been studied in the contexts of protein degradation ¹⁰, growth inhibition ¹¹, *in silico* models ^{9, 12, 13}, ion channel activation ⁸, collagen formation ¹⁴ and elastin function ⁶. Aze can orient itself into both an endo and exo pucker, although planar displacement is constrained by the smaller ring size and the endo pucker is preferred over the exo ^{9, 13}. Aze also has an enhanced propensity to isomerize the peptide bond from trans to cis conformations ^{9, 13}. C-terminal Aze can influence tagging of a protein (i.e., YbeL) for degradation ¹⁰. Proteome wide incorporation of aze typically leads to growth defects ^{11, 15}. An ion channel with Aze incorporated in place of an important Pro retained functionality ⁸. In an elastin-like-polypeptide model, replacement of Pro by Aze leads to a loss of 10⁰C in the transition temperature ⁶.

Dhp has been studied in the contexts of protein degradation ¹⁰, elastins ⁶ and *in silico* models ^{12, 16}. Dhp is a planar version of proline with little to no pucker present on the ring. Chi¹ angles are limited to between 11° and 3° ¹⁶. In an elastin-like-polypeptide model, replacement of pro by Dhp leads to a loss of 20⁰C in the transition temperature ⁶.

Results and discussion

Modified Proinsulins were expressed with Aze, Dhp, Pip or Thia in proline auxotrophic *E. coli*. PI was refolded, matured and purified as described in chapter 2. MALDI-MS analysis revealed that replacement levels were similarly high (>90%) for all 4 ring variant insulins (figure 4.2, table 4.1).

Analysis to assess the extent of dimerization for the insulin variants was carried out as described in chapter 3. At 60 μM , all 4 ring variants (AzeI, PipI, DhpI, ThiaI) exhibit CD spectra more similar to ProI than AspI, indicating dimerization at 60 μM (figure 4.3a). CD dilution experiments were also carried out as described in chapter 3 to obtain 208:222nm ratios. (figure 4.3 b-m). Both ThiaI, AzeI and DhpI show substantially similar ratios as ProI, indicating substantially unperturbed K_d 's of dimerization. PipI shows substantially lower ratios relative to ProI ratios at similar concentrations, indicating that PipI forms a slightly tighter dimer and thus has a lower K_d than ProI. None of the variants appeared to be in fully monomeric forms even at the lowest concentrations tested.

To measure the rate of disassociation from the hexameric form, the ring variants were subjected to Terpy-driven disassociation as described in chapter 2 (table 4.1). ThiaI had a disassociation rate ($\tau_{1/2} = 83.9\text{s} \pm 1.3$) essentially identical to ProI ($\tau_{1/2} = 90.4\text{s} \pm 4.2$). AzeI and DhpI had disassociation rates ($\tau_{1/2} = 64.5\text{s} \pm 4.5$ and $\tau_{1/2} = 72.3\text{s} \pm 1.3$ respectively) between those measured for ProI and AspI ($\tau_{1/2} = 90.4\text{s} \pm 4.2$ and $\tau_{1/2} = 42.7\text{s} \pm 4.3$ respectively). PipI had a dissociation rate ($\tau_{1/2} = 117\text{s} \pm 19$) slower than ProI. The PipI result, in particular, is consistent with the approximate K_d 's derived from CD dilution experiments. A tighter binding dimer might be expected to stabilize formation of the higher order hexamers and retard exit from the hexameric states.

Fibrillation lag time of the ring variant insulins was also measured via the THT assay as described in chapter 2 (table 4.1). ThiaI and PipI exhibited lag times that were more than 2X longer ($12.8\text{ h} \pm 1.3$ and $18.8\text{ h} \pm 3.6$, respectively) compared with the lag times for ProI and AspI ($5.1\text{ h} \pm 1.5$ and $5.3\text{ h} \pm 1.0$, respectively). AzeI exhibited a faster fibrillation rate ($4.5\text{ h} \pm 1.5$) with a lag time slightly faster than that of ProI or AspI. DhpI exhibited a fibrillation rate ($10.7\text{ h} \pm 3.5$) closer to ProI.

PipI represents the only known mutation at B28 on insulin to clearly slow the dissociation rate out of the hexamer from the R_6 state. Other mutations throughout insulin

(e.g., iodination at TyrB26) have been known to decrease hexamer disassociation rate but no other tested B28 mutation. Similar to iodoTyrB26, PipI also enhances resistance to fibrillation. AzeI and DhpI both represent intermediate enhancement to the disassociation rate of the insulin hexamer. Intriguingly, AzeI and DhpI likely generate this enhancement through different mechanisms, as DhpI retains closer to WT-like fibrillation rates while AzeI accelerates fibrillation. ThiaI behaves similarly to ProI in both kinetic and dimeric respects, but also enhances resistance to fibrillation. In order to further our understanding of the structural rationales underpinning these behaviors, the x-ray structure of each variant was determined.

AzeI crystals of the T₂ state were obtained to a resolution of 1.30 Å while crystals of DhpI, PipI and ThiaI were obtained in both the T₂ and R₆ states to resolutions of 1.17 Å (DhpI T₂), 2.4 Å (DhpI R₆), 1.17 Å (PipI T₂), 2.06 Å (PipI R₆), 1.22 Å (ThiaI T₂), and 1.96 Å (ThiaI R₆) (figure 4.4). AzeI crystals were also obtained in conditions (phenol and zinc) expected to yield an R₆ structure. Although crystals diffracted to 2.4 Å, the structure was unable to be phased and thus was not solved.

All T₂ structures were compared to a representative ProI structure (PDB: 3T2a) and all R₆ structures were compared to a representative ProI structure (PDB: 1EV6). In all cases, the ring variant insulin structures closely resemble WT insulin structures (figure 4.5). Backbone RMSDs measured against ProI range from 0.27 to 0.53 indicating strong overall overlap (table 4.2).

The position of B28 is essentially unchanged between any of the mutants and/or oligomer states visually (figure 4.6 a-g). AzeI adopts a planar conformation for its pucker in the T₂ structure (table 2, figure 6a). PipI adopts the expected conformation of chair for its pucker in the T₂ structure and adopts a mixture of chair and boat in the R₆ structure (table 4.2, figure 4.6 b,e). DhpI adopts the expected conformation of planar for its pucker in both T₂ and R₆ structures (table 4.2, figure 4.6 c,f). ThiaI adopts the expected conformation of endo for its pucker in both T₂ and R₆ structures (table 4.2, figure 4.6 d,g). The pucker conformations of AzeI, DhpI and ThiaI all conform to the predicted

conformations from computational modeling. For PipI, given the several kcal/mol difference between chair and boat predicted by computational modeling, we expected an all chair set of pucker conformations. The mixture of states in the structure may indicate that local steric or electronic considerations are lowering the difference in energies between the chair and boat conformations in the R₆ hexamer.

ProB28 makes a series of local contacts in ProI that are primarily hydrophobic in nature. In the T₂ dimer, the closest contacts of B28 are to GlyB23', Tyr26 and ValA3 (table 4.3, figure 4.7a). In the R₆ hexamer the closest contacts are gain GlyB23', TyrB26, ValA3 as well as GluB21' and GlyB20' (table 4.4, figure 4.8e). The TyrB26 contact is oriented and positioned such that a C-H/ π aromatic interaction is likely present. The ring variants have a set of important variation on this constellation of contacts that helps to explain their altered behaviors.

Overall ThiaI maintains nearly all the interactions present in ProI with minimal changes. In both the T₂ and R₆ structures the GlyB23' and TyrB26 are all maintained within 0.2 Å of the ProI distances (tables 4.3, 4.4, figures 4.7d, 4.8d). The ValA3 contact is 0.3 Å closer for ThiaI in the T₂ structure, a small but meaningful difference. The distance for the Glu21' contact in in the R₆ structure is also indistinguishable. The largest difference lies in the GlyB20' contact, in Thia it is approximately 0.4 Å farther away than the same contact distance in ProI. Given that the GlyB20' contact is between the carbonyl oxygen of GlyB20' and the sulfur (or C δ in ProI), this difference is expected to be more favorable. ThiaI's WT-like kinetics and dimerization behavior can be easily explained by the lack of substantial changes to the contacts that the pseudoproline participates in relative to the WT-protein. The two-fold increase in fibrillation resistance might be ascribed to a combination of improvements in hydrophobic packing (ValA3 closer) and reduction in nonproductive interactions (Gly20' distance lengthened). Even though the TyrB26 contact does not change relative distance to B28, there exists the possibility that the lone pairs of the sulfur produce a positive interaction via a lone pair- π aromatic interaction¹⁷ that is greater than the C-H/ π aromatic interaction.

DhpI is another insulin analog where the interactions present around B28 are generally not perturbed by the mutation (tables 4.3, 4.4, figures 4.7c, 4.8c). In the T₂ structure all three close contacts are all maintained within 0.2 Å of the interaction distances seen in ProI. In the R₆ the close contact at TyrB26 is maintained at a similar interaction distance seen in ProI. The GlyB23' distance has contracted by almost 0.6 Å while the ValA3 contact has lengthened by almost 1 Å. DhpI's Wt-like dimerization and close to Wt-like fibrillation can be explained by the T₂ contacts made by B28 remaining essentially unchanged relative to those of WT. Fibrillation is thought to be initiated from an unfolded monomer; maintenance of the normal T₂ contacts would neither perturb the existing fibrillation tendencies nor the K_d of dimerization. The slightly faster kinetic disassociation of DhpI might be explained by the improved GlyB23' contact and worse ValA3 contact. In particular the loss of the ValA3 contact is only seen in the R₆ structure and might serve to destabilize the R₆ state but not the T₂ state.

In contrast to Thial and DhpI, AzeI fibrillates slightly faster than ProI. It is apparent that the key TyrB26 interaction is lost in AzeI (table 4.3, figure 4.7a). Further the distances for the other T₂ contacts, GlyB23' and ValA3, are also lengthened from 3.6 Å and 3.9 Å in ProI to 4.0 Å and over 4 Å in AzeI. It is hardly surprising that the smaller azetidine ring does not fully occupy the packing space previously occupied by the pyrrolidine ring. However, there does not seem to be compensatory moves by any other residue in the region to fill up the void volume left by removal of the methylene group. One might expect that this destabilizes all states of the protein and leads to faster disassociation, faster fibrillation and loss in the K_d for dimerization; which are the observed data for AzeI.

Similar to AzeI, the constellation of residues around PipI remains unperturbed relative to Wt (tables 4.3, 4.4, figures 4.7b, 4.8b). However, PipI has an additional methylene group leading to improved packing and closer interactions. The ValA3 interaction in the T₂ structure is 0.3 Å closer than the same interaction in ProI. In the R₆ structure the TyrB26, GlyB23' and the GluB21' are all slightly closer in distance relative to the same distances measured in ProI. Improving the hydrophobic packing in a globular

protein can be expected to stabilize the protein. Most globular proteins contain some cavities that are not filled, either as a result of steric or evolutionary/functional constraints¹⁸⁻²¹. The improved packing and filling of any void volume in the folded state of insulin manifests by seemingly stabilizing PipI across the range of oligomeric states, which causes greater association of dimers, enhancement against fibrillation and slow disassociation of the hexamer. One can also speculate that the closer packing present around B28 in PipI helps to equalize the position ring conformations. Both the chair and boat conformers are present in the R₆ structure and possible interchange between these states in solution might allow for improved average packing distances while mitigating any potential for clashes.

Conclusions

The work described herein demonstrates the capabilities of ncAA mutagenesis to manipulate structure and function on atomistic levels. Subtle structural manipulations on the order of ± 1 carbon in a ring can be exploited to improve or harm overall packing in insulin. As might be expected, altering packing efficiency, even at the C-terminus can have profound consequences on stability, kinetics and thermodynamics of oligomeric assembly. The work also completes and extends earlier work performed on insulin through ncAA mutagenesis at ProB28, demonstrating a complete range of medicinal chemistry-like manipulations on a globular protein.

The work described in this chapter was completed by multiple individuals. Katharine Fang completed the fibrillation and kinetics experiments as well as jointly expressed and purified the insulins described here.

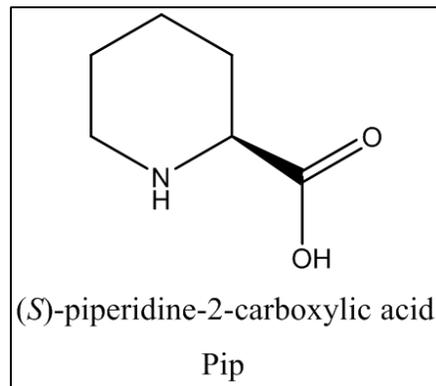
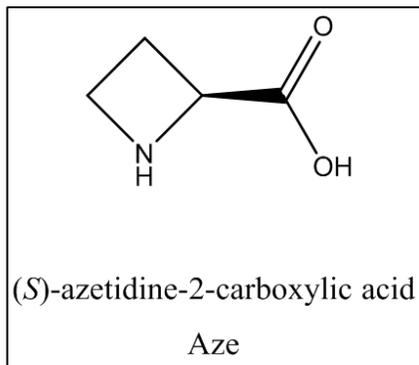
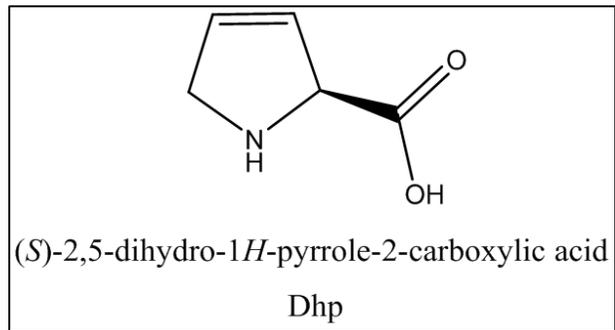
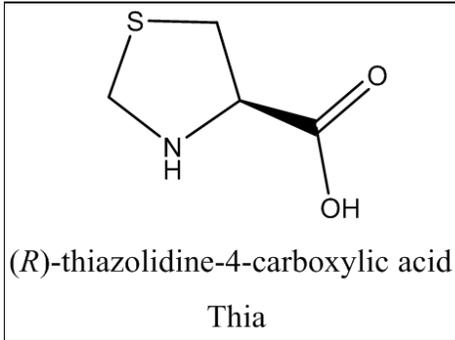


Figure 4.1| Ring proline analogs
The structures of the ring analogs discussed in this chapter.

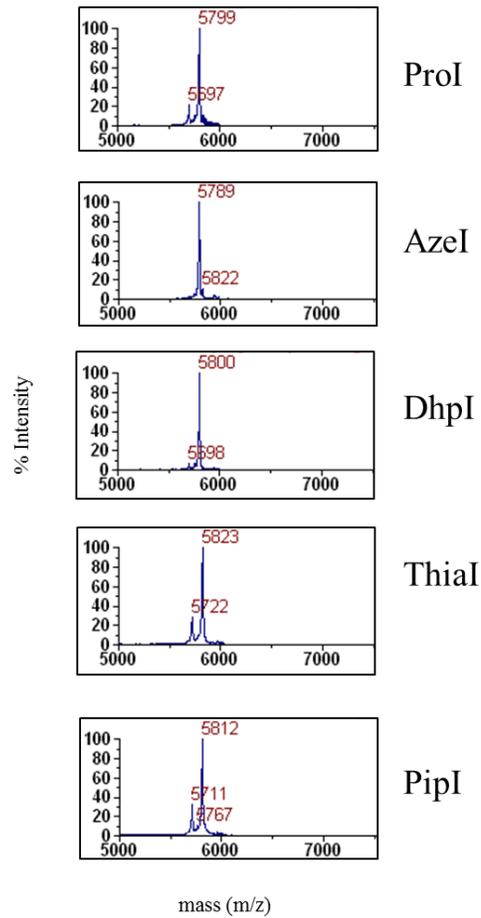


Figure 4.2| Ring variant prolines incorporate into insulin

Maldi mass spectra showing successful incorporation and maturation of the ring variant prolines in mature insulin. The WT mass of 5799 is visible in ProI and the expected mass increases or decreases are visible in PipI (+14 expected), AzeI (-14 expected), DhpI (-2 expected) or ThiaI (+18 expected).

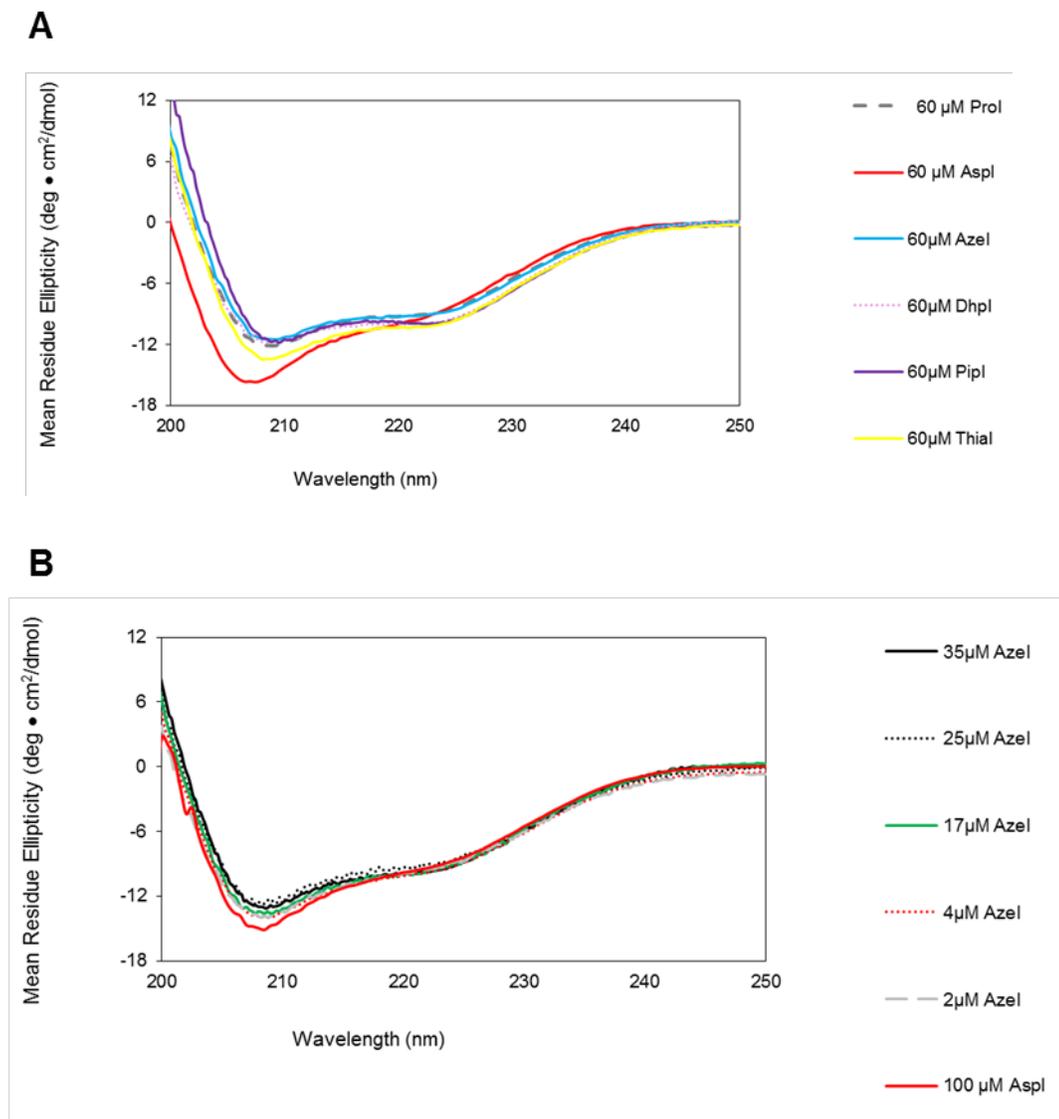


Figure 4.3 Circular Dichroism Spectra of Ring Variant insulins

(A) Far UV CD spectra collected on 60 μM insulins in 10 mM phosphate buffer, pH 8.0 at 25°C.

(B) Far UV CD spectra collected on multiple concentrations of AzeI as indicated in the legend. CD spectra collected in 10 mM phosphate buffer, pH 8.0 at 25°C.

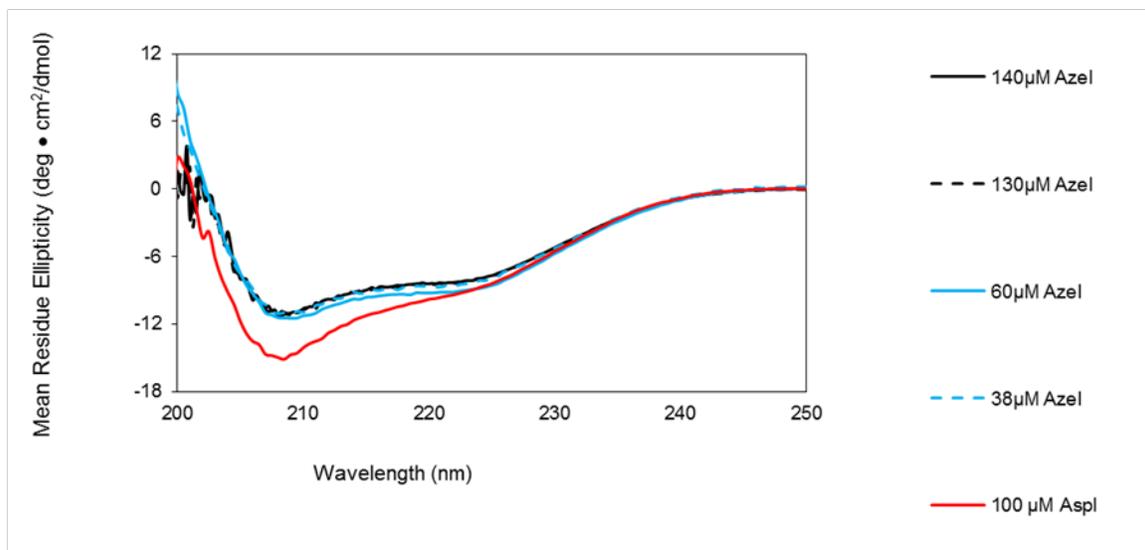
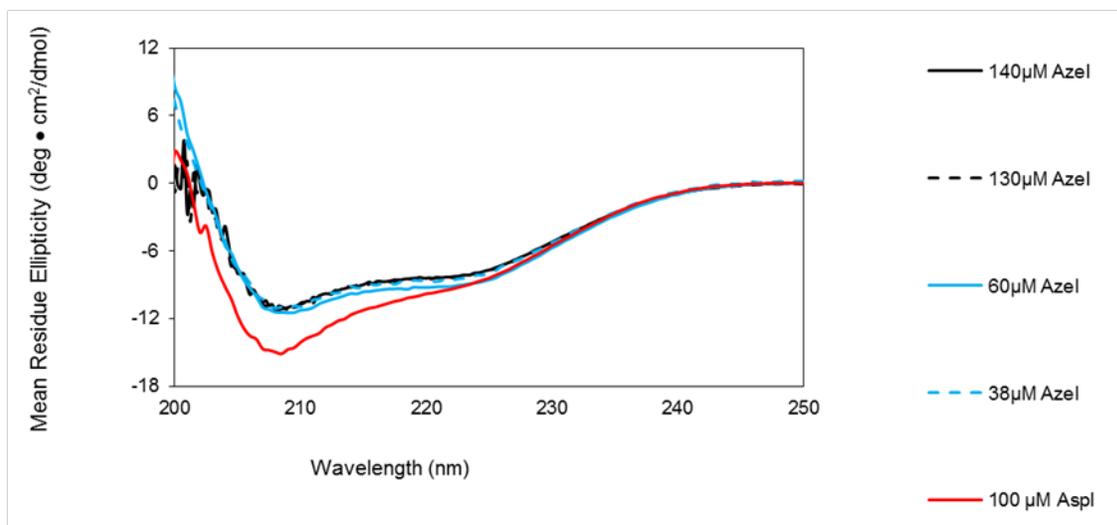
C**D**

Figure 4.3 Circular Dichroism Spectra of Ring Variant insulins

(C) Far UV CD spectra collected on multiple concentrations of AzeI as indicated in the legend. CD spectra collected in 10 mM phosphate buffer, pH 8.0 at 25°C.

(D) Far UV CD spectra collected on multiple concentrations of AzeI as indicated in the legend. CD spectra collected in 10 mM phosphate buffer, pH 8.0 at 25°C.

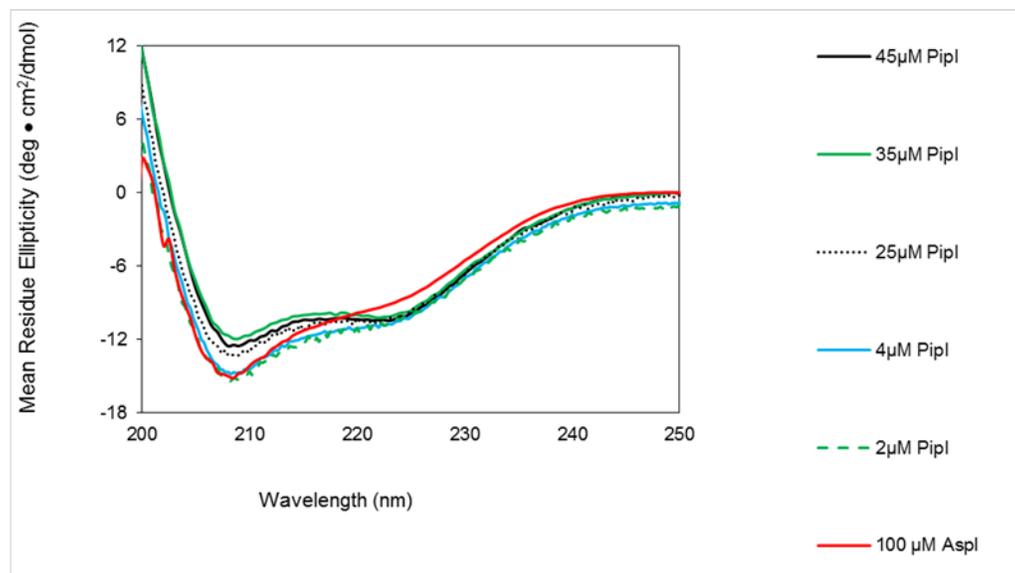
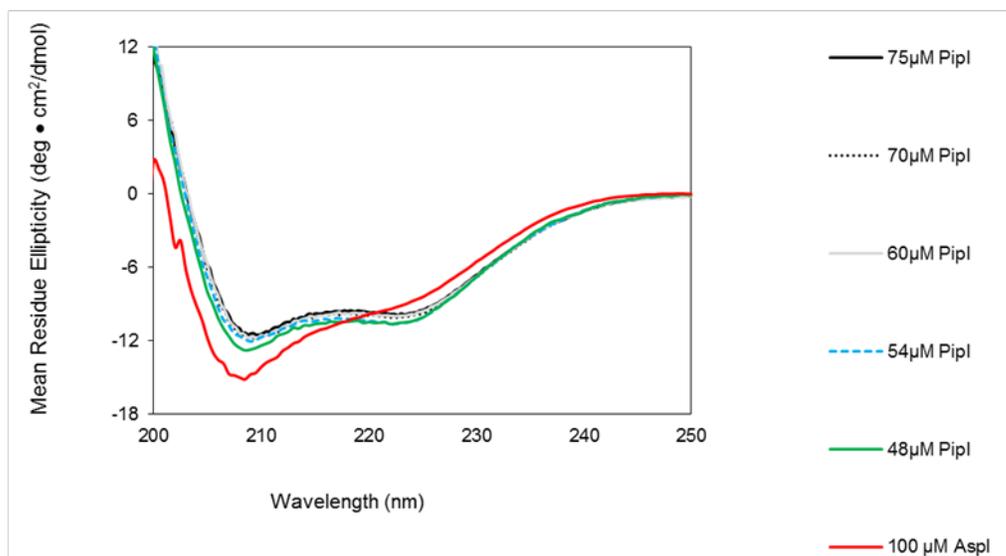
E**F**

Figure 4.3 Circular Dichroism Spectra of Ring Variant insulins

(E) Far UV CD spectra collected on multiple concentrations of PipI as indicated in the legend. CD spectra collected in 10 mM phosphate buffer, pH 8.0 at 25°C.

(F) Far UV CD spectra collected on multiple concentrations of PipI as indicated in the legend. CD spectra collected in 10 mM phosphate buffer, pH 8.0 at 25°C.

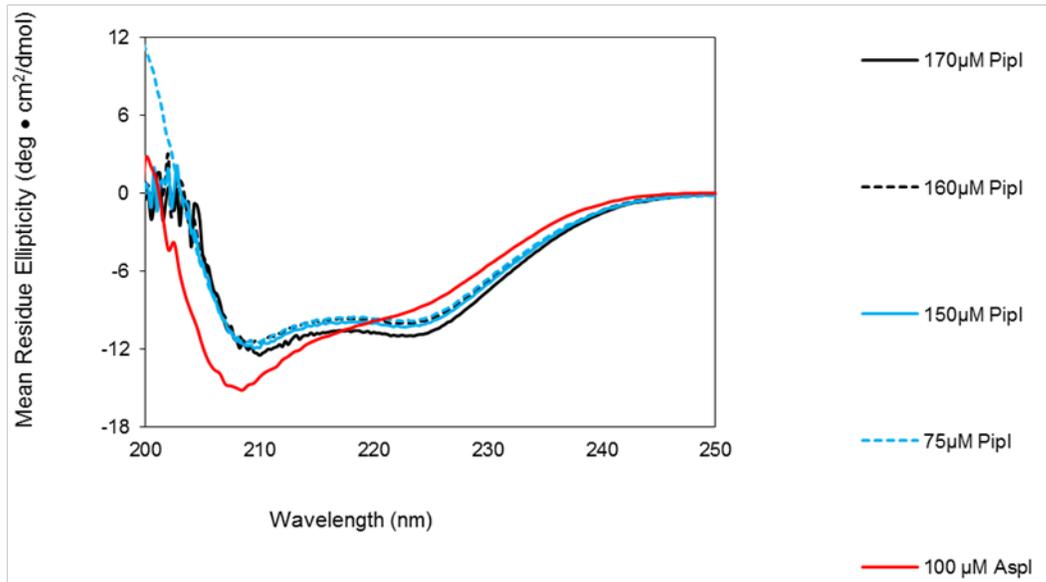
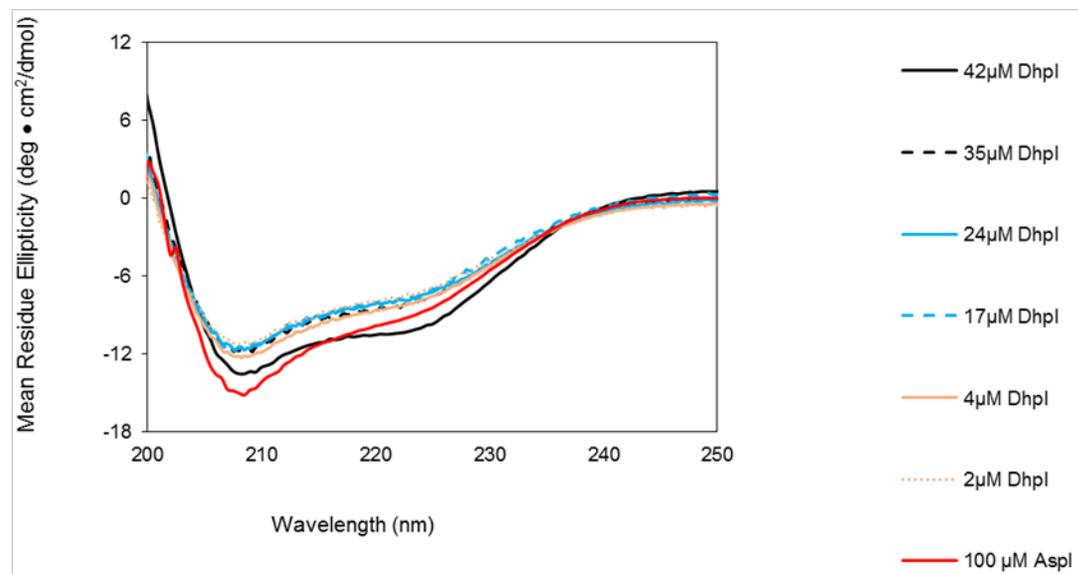
G**H**

Figure 4.3 Circular Dichroism Spectra of Ring Variant insulins

(G) Far UV CD spectra collected on multiple concentrations of PipI as indicated in the legend. CD spectra collected in 10 mM phosphate buffer, pH 8.0 at 25°C.

(H) Far UV CD spectra collected on multiple concentrations of Dhpl as indicated in the legend. CD spectra collected in 10 mM phosphate buffer, pH 8.0 at 25°C.

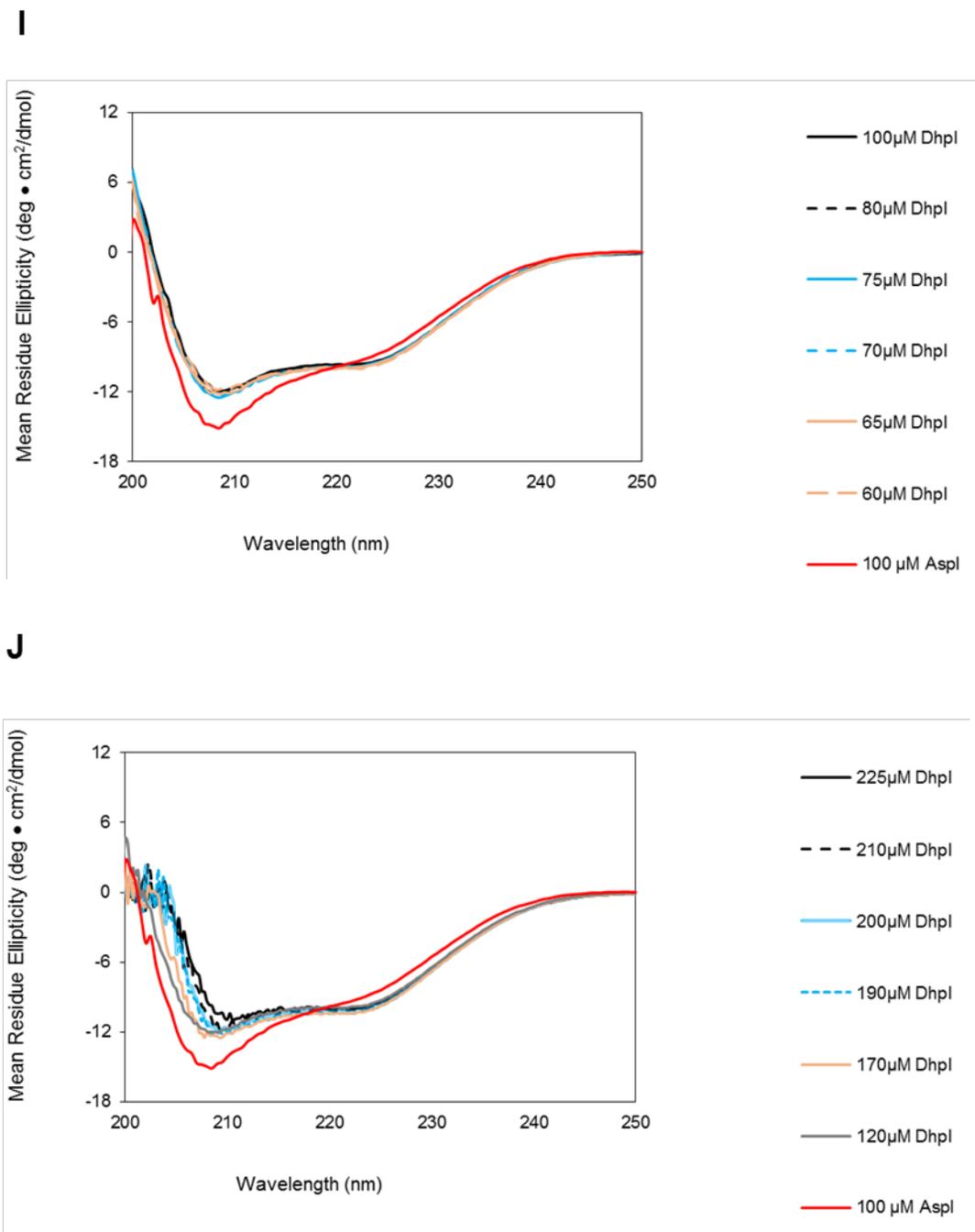


Figure 4.3 Circular Dichroism Spectra of Ring Variant insulins

(I) Far UV CD spectra collected on multiple concentrations of DhpI as indicated in the legend. CD spectra collected in 10 mM phosphate buffer, pH 8.0 at 25°C.

(J) Far UV CD spectra collected on multiple concentrations of DhpI as indicated in the legend. CD spectra collected in 10 mM phosphate buffer, pH 8.0 at 25°C.

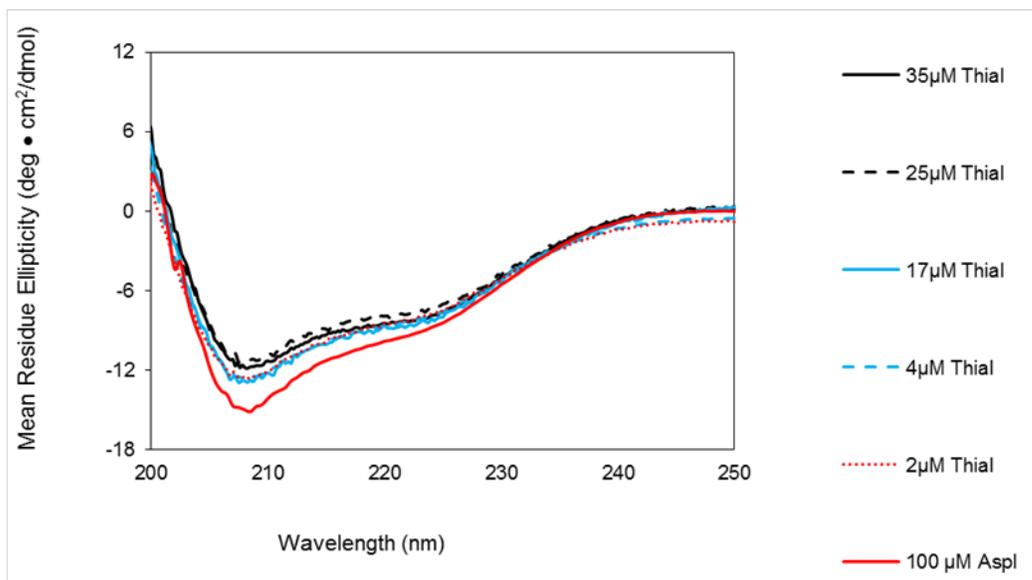
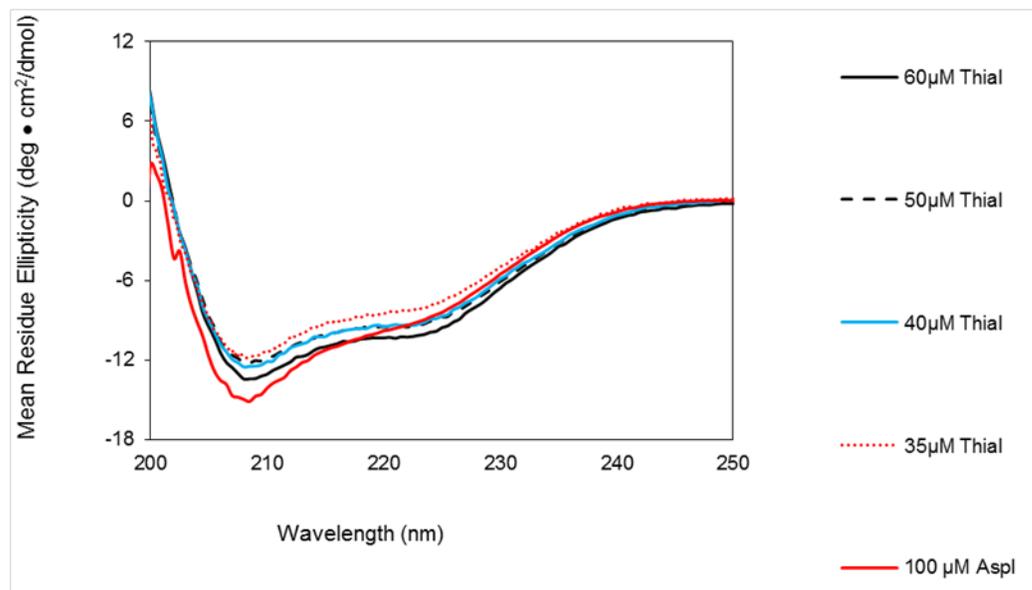
K**L**

Figure 4.3 Circular Dichroism Spectra of Ring Variant insulins

(**K**) Far UV CD spectra collected on multiple concentrations of Thial as indicated in the legend. CD spectra collected in 10 mM phosphate buffer, pH 8.0 at 25°C.

(**L**) Far UV CD spectra collected on multiple concentrations of Thial as indicated in the legend. CD spectra collected in 10 mM phosphate buffer, pH 8.0 at 25°C.

M

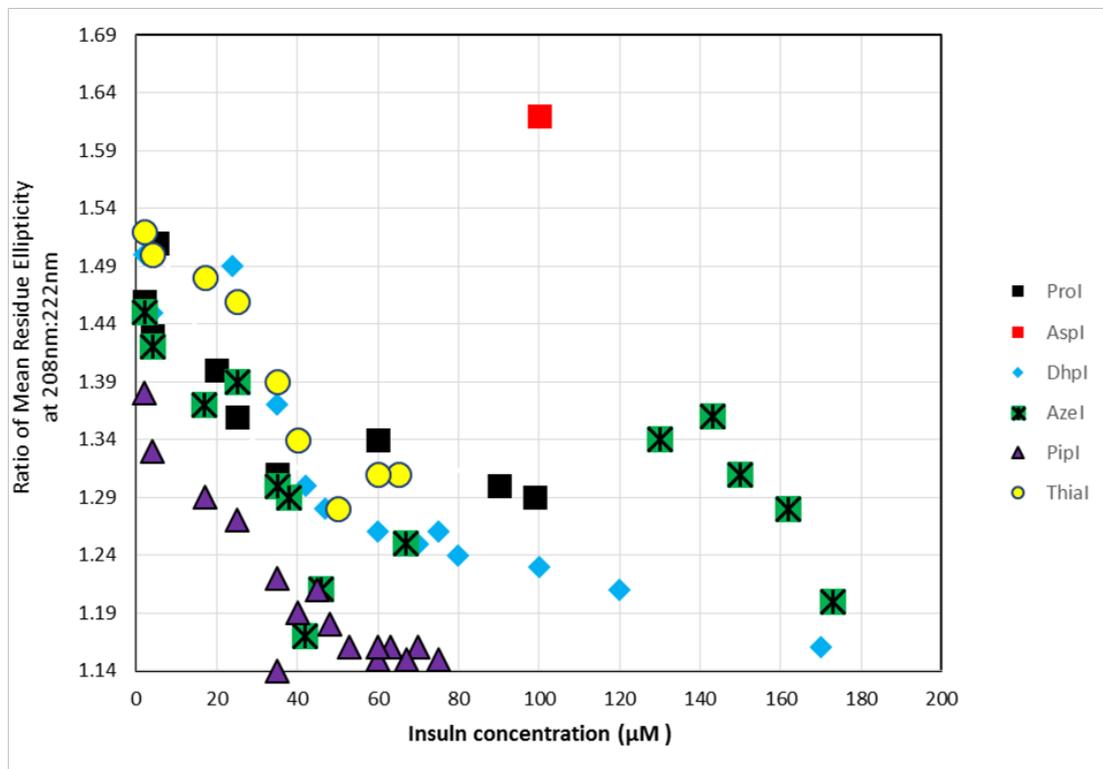


Figure 4.3 Circular Dichroism Spectra of Ring Variant insulins

(M) Ratio of the Far UV CD spectra 208nm minima to 222nm minima for ring size variants over a range of concentrations. Ratios were calculated from the 208nm and 222nm data points on the preceding Far UV spectra in this figure.

	Dhpl-T ₂ SURT	Dhpl-R ₆ SURU	Azel-T ₂ 5USV
Data collection			
Space group	<i>I</i> 2 ₁ 3	<i>H</i> 3	<i>I</i> 2 ₁ 3
Cell dimensions a, b, c (Å)	77.95, 77.95, 77.95	79.259, 79.259, 79.874	77.979, 77.979, 77.979
α, β, γ (°)	90, 90, 90	90, 90, 120	90, 90, 90
Resolution (Å)	55.12 - 1.17 (1.19-1.17)	34.52-2.41 (2.47-2.41)	31.83-1.3 (1.33-1.3)
R _{p.i.m.} (%)	2.1 (53.6)	4.2 (30.8)	3.7 (12.1)
Mn(I /sd)	18.1 (1.3)	5.4 (2.1)	13.2 (1.2)
Completeness (%)	99.7 (94.6)	99.76 (99.6)	99.4 (100)
Redundancy	18.6	4.6	18.4 (17.2)
Refinement			
No. reflections	26449(1328)	6816 (469)	26650 (1288)
R _{work} / R _{free} (%)	11.9 / 14.1 (23.9/27.2)	19.0 / 23.7 (22.2/28.7)	14.2 / 15.9 (23.2 / 23.3)
No. atoms			
Protein	397	1439	392
Ligand/ion	1	32	1
Water	67	7	58
RMSD			
Bond lengths (Å)	0.028	0.022	0.009
Bond angles (°)	2.273	2.04	1.978
Ramachandran map analysis			
Favored	50	194	50
Allowed	0	0	0
Outliers	0	0	0

	Pipl-T ₂ 5USP	Pipl-R ₆ 5USS	Thial-T ₂ 5UU2	Thial-R ₆ 5UU4
Data collection				
Space group	<i>I</i> 2 ₁ 3	<i>H</i> 3	<i>I</i> 2 ₁ 3	<i>H</i> 3
Cell dimensions a, b, c (Å)	78.047, 78.047, 78.047	79.024, 79.024, 39.509	78.067, 78.067, 78.067	78.09, 78.09, 39.93
α, β, γ (°)	90, 90, 90	90, 90, 120	90, 90, 90	90, 90, 120
Resolution (Å)	39.02-1.17 (1.19-1.17)	25.87-2.06 (2.11-2.06)	31.87-1.22 (1.25-1.22)	39.05-1.96 (2.01-1.96)
R _{p.i.m.} (%)	2.0 (34.2)	2.6 (34.2)	1.6 (25.1)	2.2 (14.5)
Mn(I /sd)	17.8 (2.0)	12.9 (3.2)	23.5 (3.4)	19.4 (0.5)
Completeness (%)	99.75 (93.3)	99.8 (99.8)	99.6 (91.5)	98.1 (76.3)
Redundancy	10 (5.4)	5.3 (5.0)	8.1 (4.5)	5.5 (3.6)
Refinement				
No. reflections	26485 (1198)	9265 (422)	23424 (1064)	6396 (350)
R _{work} / R _{free} (%)	13.9/15.26 (20.9/22.3)	16.9/23.8 (22.9/30.7)	13.7/15.5 (17.4/22.9)	23.2/26.4 (33.0/36.3)
No. atoms				
Protein	467	708	397	661
Ligand/ion	1	18	0	22
Water	71	1	68	0
RMSD				
Bond lengths (Å)	0.008	0.008	0.017	0.018
Bond angles (°)	1.07	1.047	1.53	1.800
Ramachandran map analysis				
Favored	50	97	50	93
Allowed	0	0	0	1
Outliers	0	0	0	0

Figure 4.4 | Data Table and Refinement Values for Ring Variant Insulins

A

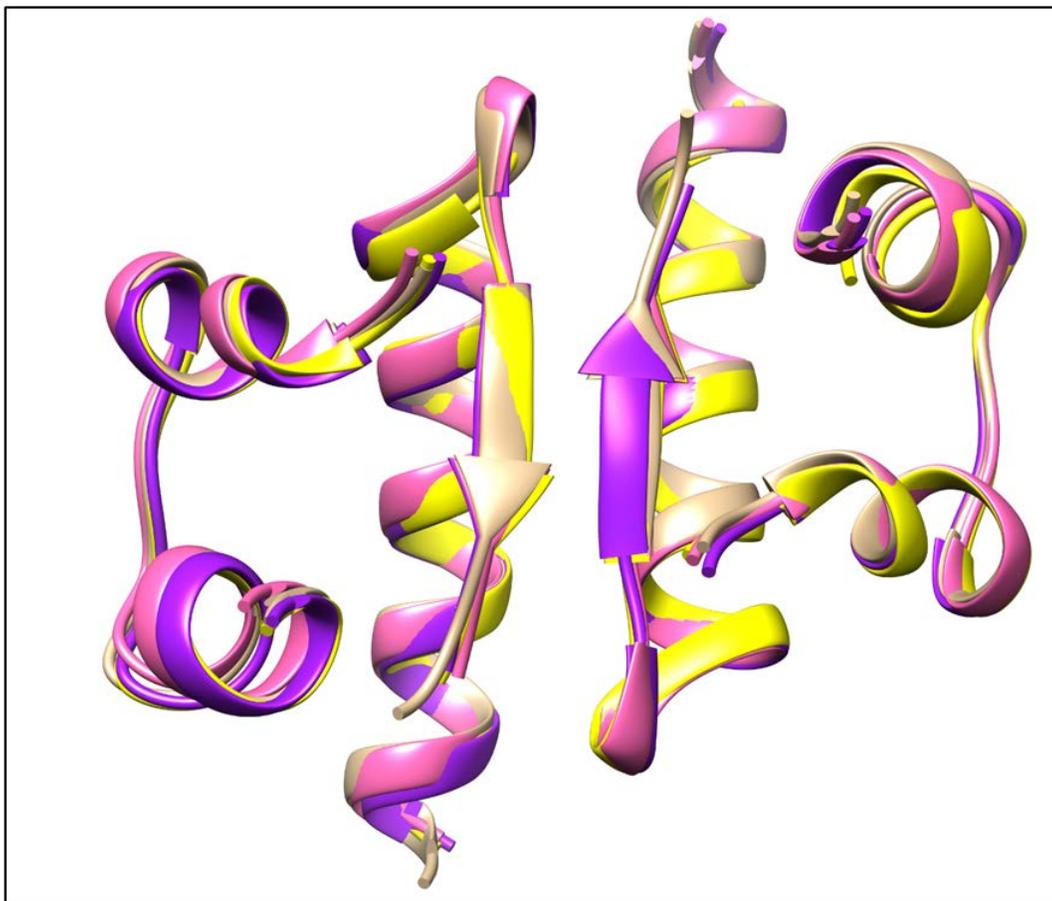


Figure 4.5 | The dimer in the R₆ hexamer of the ring variant insulins.

A) Alignment of R₆ ProI (tan, PDB:1EV6), R₆-PipI (Purple), R₆-DhpI (Pink) and R₆-ThiaI (Yellow) showing a single dimer in the hexamer.

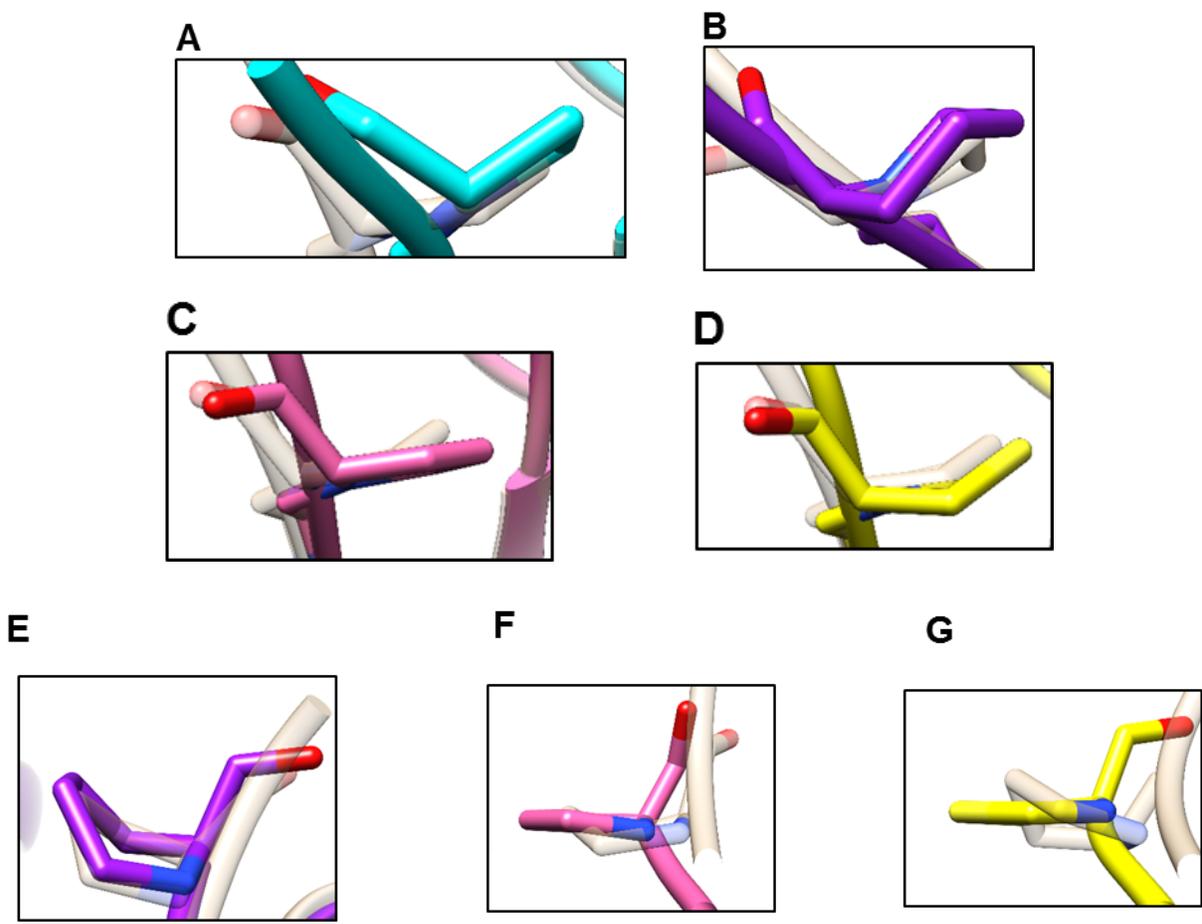


Figure 4.6 | Alignment at position B28.

A,B,C,D) Alignment of T_2 -ProI (tan, PDB ID: 3T2A) and T_2 -AzeI (Cyan) highlighting the pucker of the proline ring.

B) Alignment of T_2 -ProI (tan, PDB ID: 3T2A) and T_2 -PipI (Purple) highlighting the pucker of the proline ring.

C) Alignment of T_2 -ProI (tan, PDB ID: 3T2A) and T_2 -DhpI (Pink) highlighting the pucker of the proline ring.

D) Alignment of T_2 -ProI (tan, PDB ID: 3T2A) and T_2 -Thial (Yellow) highlighting the pucker of the proline ring.

E) Alignment of R_6 -ProI (tan, PDB ID: 1EV6) and R_6 -PipI (Purple) highlighting the pucker of the proline ring.

F) Alignment of R_6 -ProI (tan, PDB ID: 1EV6) and R_6 -DhpI (Pink) highlighting the pucker of the proline ring.

G) Alignment of R_6 -ProI (tan, PDB ID: 1EV6) and R_6 -Thial (yellow) highlighting the pucker of the proline ring.

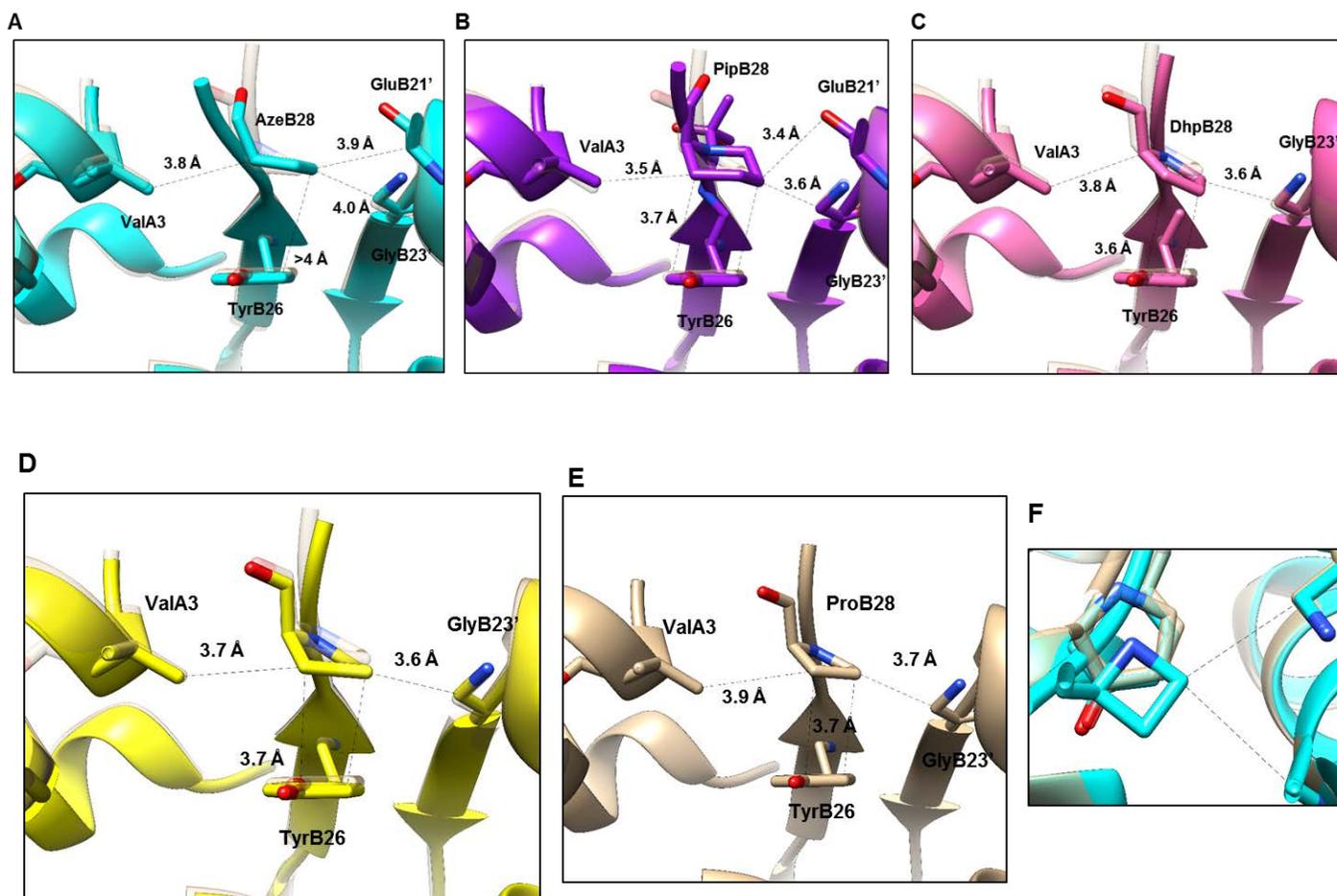


Figure 4.7 | B28 highlighting key contacts in the dimer interface.

A, B, C, D) Alignment of T₂ ProI (tan, PDB:3T2A), and T₂-AzeI (Cyan) or T₂-PipI (Purple) or T₂-DhpI (Pink) or T₂-ThiaI (Yellow) centered on position B28 highlighting the important interactions made between B28 and surrounding residues in the dimer. Interaction distances are listed in angstroms.

E) T₂ ProI (tan, PDB:3T2A) centered on position B28 highlighting the important interactions made between B28 and surrounding residues in the dimer. Interaction distances are listed in angstroms

F) Alignment of T₂ ProI (tan, PDB:3T2A), and T₂-AzeI (Cyan) highlighting the lateral shift of 1.4 Å in the position of B28 in AzeI vs ProI. The preceding TyrB26 has no shift in position between AzeI and ProI.

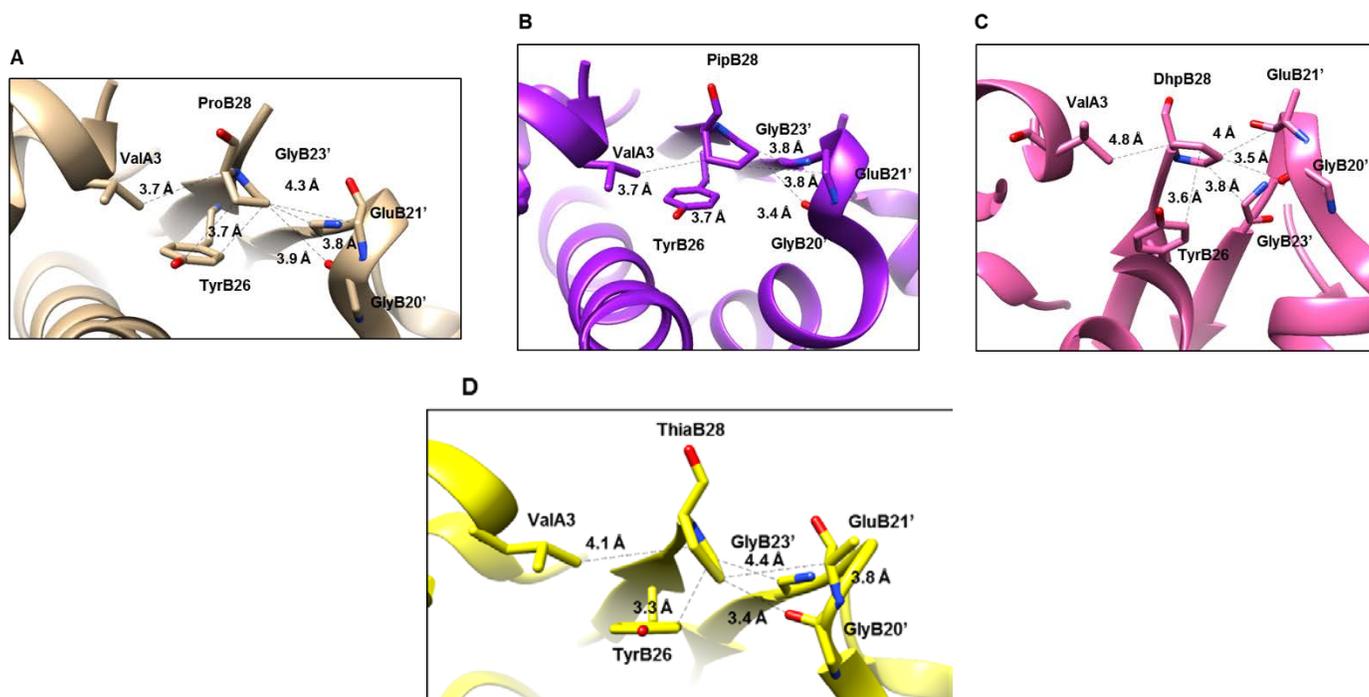


Figure 4.8 | B28 highlighting key contacts in the dimer interface.

A) R_6 ProI (tan, PDB:1EV6) centered on position B28 highlighting the important interactions made between B28 and surrounding residues in the dimer. Interaction distances are listed in angstroms

B, C, D) R_6 -PipI (Purple) or R_6 -DhpI (Pink) or R_6 -ThiaI (Yellow) centered on position B28 highlighting the important interactions made between B28 and surrounding residues in the dimer. Interaction distances are listed in angstroms.

Protein	%ncPro‡	Hexamer $\tau_{1/2}$ (s)	Fibrillation lag time (h)
Prol	--	90.4 ± 4.2	5.1 ± 1.5
Aspl	--	42.7 ± 4.3	5.3 ± 1.0
Hzpl*	91%	55.1 ± 2.4	19 ± 4.0
HypI*	88%	87.4 ± 6.3	5.5 ± 1.2
Dhpl	≤100%	72.3 ± 1.3	10.7 ± 1.5
Azel	≤100%	64.5 ± 4.5	4.5 ± 1.5
Pipl	93%	117 ± 19	18.8 ± 3.6
Thzl	85%	83.9 ± 1.3	12.8 ± 1.9

Table 4.1| Key properties of Ring Variant insulins

Errors are given as one s.d. ($n \geq 4$) unless otherwise noted

‡Incorporation error <2% and quantified using MALDI-MS on gluC digested peptide containing B28

*Active in vivo ($n \geq 3$)

	RMSD vs ProI (Å) ^a	Ring conformation ^b	Chi 1 angle ^c	Phi ^c
ProI T ₂		Endo	22	-75
Dhpl T ₂	0.34	Planar	5.5	-71
Thial T ₂	0.31	Endo	32	-75
Azel T ₂	0.41	Planar	-1.7	-71
Pipl T ₂	0.28	Chair	54	-101
ProI R ₆		Endo (5/6)	21	-60
Dhpl R ₆	0.38	Planar (2/4)	10	-67
Thial R ₆	0.53	Endo	11	-107
Pipl R ₆	0.27	Chair (1/2), Boat (1/2)	64	-101

Table 4.2 | RMSD and proline angles

^a Ring variant T₂ insulins were aligned against T₂ ProI (PDB:3T2A) and R₆ insulins were aligned against R₆-ProI (PDB: 1EV6) RMSD was calculated over backbone atoms.

^b Ring conformation was determined on the basis of the Chi1 angles. Where more than one B28 residue is present in the asymmetric unit the predominant conformation is listed. The fraction in parentheses denotes the number of B28 residues in one asymmetric unit with the predominant conformation.

^c Chi1 and Phi angles are reported as an average over all B28 residues present in the asymmetric unit for each crystal structure. For T₂ insulins only a single B28 is averaged as there is only a single insulin monomer in the asymmetric unit, for R₆ insulins the average value is calculated over multiple prolines as there are 2 (HypI) or 6 (HzpI) prolines present in the asymmetric unit.

T ₂ Prol	Proline atom	Other residue	VDW Overlap (Å) ^a	Distance (Å)
	CD	Gly23' CA	0.13	3.6
	CG	Tyr26 CE	-0.14	3.8
	CA	Val3 CG	-0.14	3.9

T ₂ Thial	Proline atom	Other residue	VDW Overlap (Å) ^a	Distance (Å)
	CD	Gly23' CA	0.02	3.7
	CD	Tyr26 CD	-0.01	3.7
	CB	Val3 CG	0.12	3.6

T ₂ PipI	Proline atom	Other residue	VDW Overlap (Å) ^a	Distance (Å)
	CD	Gly23' CA	0.16	3.6
	CD	Tyr26 CE	-0.07	3.7
	CB	Val3 CG	0.24	3.5
	CD	Glu21' O	-0.05	3.4

T ₂ Azel	Proline atom	Other residue	VDW Overlap (Å) ^a	Distance (Å)
	CD	Gly23' CA	-0.26	4.0
	Any	Tyr26 Ring	None	>4
	CA	Val3 CG	-0.00	3.8
	CD	Glu21' CA	-0.13	3.9

T ₂ Dhpl	Proline atom	Other residue	VDW Overlap (Å) ^a	Distance (Å)
	CD	Gly23' CA	0.15	3.6
	CG	Tyr26 CE	0.15	3.6
	CA	Val3 CG	-0.03	3.8

Table 4.3 | Key B28 contacts in the T₂ structures of the ring-variant insulins

^a VDW overlap was calculated in Chimera as the distance between the Van der Waals radii of each participating atom. Positive values indicate VDW radii overlapping. Negative values indicate the distance between VDW radii. Distances are measured as the linear distance between the center point of the two atoms listed in each row. Nomenclature of the atoms in the residues follows standard protein atomic naming as utilized in the PDB. Below the table is a guide to the proline ring atoms nomenclature.

R ₆ Prol	Atomic interaction ^b	Other residue	VDW Overlap (Å) ^a
	C-C	Gly23'	-0.61
	C-C	Tyr26	0.03
	C-C	Val3	-0.07
	C-C	Glu21'	-0.07
	C-O	Gly20'	-0.63

R ₆ Thial	Atomic interaction ^b	Other residue	VDW Overlap (Å) ^a
	C-C	Gly23'	-0.64
	C-C	Tyr26	0.30
	C-C	Val3	-0.37
	S-C	Glu21'	-0.17
	S-O	Gly20'	-0.22

R ₆ PipI	Atomic interaction ^b	Other residue	VDW Overlap (Å) ^a
	C-C	Gly23'	-0.06
	C-C	Tyr26	-0.10
	C-C	Val3	0.02
	C-C	Glu21'	-0.40
	C-O	Gly20'	-1.17

R ₆ Dhpl	Atomic interaction ^b	Other residue	VDW Overlap (Å) ^a
	C-C	Gly23'	-0.03
	C-C	Tyr26	0.04
	C-C	Val3	-0.97
	C-C	Glu21'	-0.38
	C-O	Gly20'	-0.30

Table 4.4 | Key B28 contacts in the R₆ structures of the ring-variant insulins

^a VDW overlap was calculated in Chimera as the distance between the van der waals radii of each participating atom. Where more than one monomer is present in the asymmetric unit of the crystal structure, the overlap value listed in the table is an average over all B28-contact overlaps. Positive values indicate VDW radii overlapping. Negative values indicate the distance between VDW radii. Where more than one monomer is present in the asymmetric unit of the crystal structure, the distance value listed in the table is an average over all B28-contact overlaps.

^b Atomic interaction indicates the closest pairing of atoms involved in the B28 to other residue contact. The first atom listed is the atom from B28. The second atom listed is located on the residue listed in the other residue column.

References

- 1 Nediljko Budisa, Caroline Minks, F. Javier Medrano, Jürgen Lutz, Robert Huber, and Luis Moroder, 'Residue-Specific Bioincorporation of Non-Natural, Biologically Active Amino Acids into Proteins as Possible Drug Carriers: Structure and Stability of the Per-Thiaproline Mutant of Annexin V', *Proceedings of the National Academy of Sciences of the United States of America*, 95 (1998), 455-59.
- 2 Y. K. Kang, and H. S. Park, 'Conformational Preferences of Pseudoproline Residues', *J Phys Chem B*, 111 (2007), 12551-62.
- 3 Y. K. Kang, H. S. Park, and B. J. Byun, 'Puckering Transitions of Pseudoproline Residues', *Biopolymers*, 91 (2009), 444-55.
- 4 Y. J. Lin, C. H. Chang, and J. C. Horng, 'The Impact of 4-Thiaproline on Polyproline Conformation', *J Phys Chem B*, 118 (2014), 10813-20.
- 5 A. Choudhary, K. H. Pua, and R. T. Raines, 'Quantum Mechanical Origin of the Conformational Preferences of 4-Thiaproline and Its S-Oxides', *Amino Acids*, 41 (2011), 181-6.
- 6 Christy Catherine, Su Jin Oh, Kyung-Ho Lee, Seung-Eui Min, Jong-In Won, Hyungdon Yun, and Dong-Myung Kim, 'Engineering Thermal Properties of Elastin-Like Polypeptides by Incorporation of Unnatural Amino Acids in a Cell-Free Protein Synthesis System', *Biotechnology and Bioprocess Engineering*, 20 (2015), 417-22.
- 7 Gach K Perlikowska R, Fichna J, Toth G, Walkowiak B, do-Rego JC, Janecka A., 'Biological Activity of Endomorphin and [Dmt1]Endomorphin Analogs with Six-Membered Proline Surrogates in Position 2.', *Bioorg Med Chem.*, 17 (2009), 3789094.
- 8 S. C. Lummis, D. L. Beene, L. W. Lee, H. A. Lester, R. W. Broadhurst, and D. A. Dougherty, 'Cis-Trans Isomerization at a Proline Opens the Pore of a Neurotransmitter-Gated Ion Channel', *Nature*, 438 (2005), 248-52.
- 9 J. S. Jhon, and Y. K. Kang, 'Conformational Preferences of Proline Analogues with Different Ring Size', *J Phys Chem B*, 111 (2007), 3496-507.
- 10 C. S. Hayes, B. Bose, and R. T. Sauer, 'Proline Residues at the C Terminus of Nascent Chains Induce SsrA Tagging During Translation Termination', *J Biol Chem*, 277 (2002), 33825-32.
- 11 J. Lee, N. Joshi, R. Pasini, R. C. Dobson, J. Allison, and T. Leustek, 'Inhibition of Arabidopsis Growth by the Allelopathic Compound Azetidine-2-Carboxylate Is Due to the Low Amino Acid Specificity of Cytosolic Prolyl-Trna Synthetase', *Plant J*, 88 (2016), 236-46.
- 12 Dongxia Li, Maija Hirsilä, Peppi Koivunen, Mitchell C. Brenner, Leon Xu, Charles Yang, Kari I. Kivirikko, and Johanna Myllyharju, 'Many Amino Acid Substitutions in a Hypoxia-Inducible Transcription Factor (Hif)-1 α -Like Peptide Cause Only Minor Changes in Its Hydroxylation by the Hif Prolyl 4-Hydroxylases: Substitution of 3,4-Dehydroproline or Azetidine-2-Carboxylic Acid for the Proline Leads to a High Rate of Uncoupled 2-Oxoglutarate Decarboxylation', *Journal of Biological Chemistry*, 279 (2004), 55051-59.

- 13 Kyrylo Bessonov, Kenrick A. Vassall, and George Harauz, 'Parameterization of the Proline Analogue Aze (Azetidine-2-Carboxylic Acid) for Molecular Dynamics Simulations and Evaluation of Its Effect on Homo-Pentapeptide Conformations', *Journal of Molecular Graphics and Modelling*, 39 (2013), 118-25.
- 14 A. Zagari, K. A. Palmer, K. D. Gibson, G. Nemethy, and H. A. Scheraga, 'The Effect of the L-Azetidine-2-Carboxylic Acid Residue on Protein Conformation. Iv. Local Substitutions in the Collagen Triple Helix', *Biopolymers*, 34 (1994), 51-60.
- 15 Berenfeld L Trotter EW, Krause SA, Petsko GA, Gray JV., 'Protein Misfolding and Temperature up-Shift Cause G1 Arrest Via a Common Mechanism Dependent on Heat Shock Factor in *Saccharomyces cerevisiae*.', *Proc Natl Acad Sci U S A.*, 98 (2001), 7313-8.
- 16 Y. K. Kang, and H. S. Park, 'Conformational Preferences and Cis-Trans Isomerization of L-3,4-Dehydroproline Residue', *Biopolymers*, 92 (2009), 387-98.
- 17 A. Jain, V. Ramanathan, and R. Sankararamakrishnan, 'Lone Pair ··· H Interactions between Water Oxygens and Aromatic Residues: Quantum Chemical Studies Based on High-Resolution Protein Structures and Model Compounds', *Protein Sci*, 18 (2009), 595-605.
- 18 O. Carugo, and P. Argos, 'Accessibility to Internal Cavities and Ligand Binding Sites Monitored by Protein Crystallographic Thermal Factors', *Proteins*, 31 (1998), 201-13.
- 19 V. E. Angarica, and J. Sancho, 'Protein Dynamics Governed by Interfaces of High Polarity and Low Packing Density', *PLoS One*, 7 (2012), e48212.
- 20 S. J. Hubbard, K. H. Gross, and P. Argos, 'Intramolecular Cavities in Globular Proteins', *Protein Eng*, 7 (1994), 613-26.
- 21 Shrihari Sonavane, and Pinak Chakrabarti, 'Cavities and Atomic Packing in Protein Structures and Interfaces', *PLOS Computational Biology*, 4 (2008), e1000188.

CHAPTER V

Future avenues for insulin engineering

Discussion

The work described herein is, to the author's knowledge, the most thorough analysis of proline analogs in a globular protein to date. Nine proline analogs were incorporated into insulin at position B28. Detailed structural, biochemical and biophysical data was collected for each of these analogs and can be found in this thesis as well as the complementary work in the thesis of Katharine Fang (Caltech, 2017). Insulin is a model protein system for computational chemistry and molecular dynamics simulations due to its small size (5.8 kDa) and wealth of model protein interactions (zinc coordination, disulfides, hydrophobic core). Few proline analogs have been crystalized in globular proteins and in no case is systematic data available for the same position with more than two analogs. The structures and data herein should provide for a robust starting point for any future computational studies on proline analogs and their effects on protein structure and function.

We can also draw several conclusions about the C-terminus of the B-chain in insulin from our collective set of proline analogs. First, the efficiency of protein hydrophobic packing at C-terminus of the B-chain can be improved. Second, the rates of fibrillation, hexamer dissociation and dimer binding affinity can be uncoupled. Third, the importance of CH/ π packing between ProB28 and TyrB26 may have been historically underestimated. Finally, one can speculate that systematic application of ncAA mutagenesis to most positions in insulin (or more broadly any therapeutic protein) has the potential to discover new and improved variants.

Substitution of proline with several analogs (FzpI, DfpI, PipI, HypI, HzpI, Thial) of increased bulk did not seem to destabilize the protein. Indeed, out of all the proline analogs tested here, the bulkier replacements represent the insulins with the greatest enhancement to fibrillation resistance. One must draw the conclusion that insulin can readily accept at least small increases in bulk at B28 and so must have either the unfilled

cavity space or reorganization capacities to accommodate the additional atomic space. One can easily glean from the alignments of the high-resolution structures that no systematic or obvious reorganizations occurred. Therefore a substantial amount of the increased bulk must have been accommodated by unfilled and pre-existing void volume. It seems likely that filling this void volume improves hydrophobic packing and thus tends to lead to enhanced protection against fibrillation/aggregation by stabilizing the folded native state. It is not always the case that bulk aids against fibrillation. HypI and FypI both add bulk, but, for either electrostatic or steric reasons, the sum total of their interactions does not protect against fibrillation.

Previous attempts to stabilize insulin against fibrillation have run into the roadblock that stabilized insulins could not be made fast-acting. It seemed that the only way to protect against fibrillation was to alter the distribution of oligomeric states present in solution by proportionally increasing the amount of hexamer or stability of the hexamer. The work herein shows conclusively that the binding affinity of the dimer and rates of kinetic disassociation as well as rates of fibrillation need not be linked. The canonical form of insulin might be thought of as Pareto optimal by balancing the rates of dissociation with the rate of fibrillation. By using ncAA prolines we can break through the canonical amino acid Pareto limit. We observed insulins with fast dissociation and fast fibrillation (AzeI, FypI) and others with slow dissociation and slow fibrillation (PipI), which are typical for other canonical mutants. But we also observed monomerized insulins with WT-like fibrillation and kinetics (HypI), WT-like dimerizing insulins with slow fibrillation and fast dissociation (HzpI, dFpI) and WT-like insulins with slow fibrillation (ThiaI). Single atomic differences, even as small as the presence of a single or double bond (DhpI), are sufficient to alter the association and aggregation behavior of insulin. We can only speculate that future studies, through further ncAA mutagenesis at B28 or other positions in insulin, will unlock insulins with further enhanced characteristics. Perhaps replacement at B28 with an azepane, an azocane or methylated piperidines would result in further resistance to fibrillation. Or perhaps strengthening the hydrogen bond between HzpI and Glu21', by ncAA replacement with an aminoproline or

several hydroxyl groups would result in yet faster dissociating and fibrillation resistant insulins. The Tirrell lab has already begun explorations amongst some of these derivatives. Diabetics worldwide and I await the results with baited breath.

It has long been known that ProB28 sits in a hydrophobic pocket, but the appreciation that TyrB26 and ProB28 may interact via a specific CH/ π interaction above and beyond the normal hydrophobic interaction has not been previously addressed. Any substantial disruption to this interaction (FypI, AzeI) leads to dramatically more fibrillation prone variants. Conversely this interaction was preferentially maintained over other contacts even as substitutions lead to perturbations in the constellation of interactions around B28 (DhpI, dFpI, FzpI). A CH/ π interaction would also help to explain the observation that non-aromatic mutations at B26 lead to faster fibrillation.

Finally the work described here is a landmark as protein-engineering techniques continue to improve. Small molecules are regularly subjected to atom-by-atom testing in order to achieve the best ligand for medicinal purposes. Proteins are merely large molecules, and so in theory atom-by-atom testing would also provide us with the best protein molecule for medicinal purposes. Nature has provided scientists with rapid means to screen a wide range of chemical species through the application of directed evolution and related techniques (pioneered by Frances Arnold and Pim Stemmer amongst others). However, existing directed evolution and protein engineering methods are limited by the chemical diversity present in the canonical 20 amino acids. The work here demonstrates that even for a protein like insulin, which is heavily studied and optimized, non-canonical amino acids can provide for further improvements through atom-by-atom testing. One can only hope that protein engineers will discover rapid methods to screen ncAA protein variants. Nine mutants were screened in this thesis over the course of several years, a rate that might be compared to the rate of single site mutagenesis techniques used at the dawn of protein engineering in the 1980's. If similar advancements in screening occur as

occurred for protein engineering, we will be screening hundreds or thousands of ncAAs by the 2050s.