Chapter II

INCORPORATION OF ALIPHATIC PROLINE ANALOGS INTO RECOMBINANTLY-PRODUCED INSULIN

2.1 Contributions

Chad D. Paavola, Michael P. Akers, Julie S. Moyers, and Alborz Mahdavi provided advice on proinsulin refolding. Alejandro Lopez assisted with cloning *proS* mutants, and Alex M. Chapman performed pilot ncPro incorporation studies. Janine Quijano and Cecile Donohue performed mouse experiments, with input from Hsun Teresa Ku. Analytical ultracentrifugation was performed by Amy Henrickson and Maduni Ranasinghe, with input from Borries Demeler. Jay Winkler, Songye Chen, and Mona Shahgholi assisted with circular dichroism spectroscopy, transmission electron microscopy, and mass spectrometry, respectively. Stephanie L. Breunig prepared all insulin samples and performed all other experiments.

2.2 Abstract

Non-canonical proline residues expand the chemical space about proline while maintaining its conformationally-restricting properties. Here, we demonstrate the incorporation of 4*R*-methylproline, 4*S*-methylproline, and 4-methyleneproline into recombinantly-produced insulin in *Escherichia coli* using a residue-specific replacement approach. These modest modifications to proline at position B28 affect the biophysical properties of insulin: incorporation of 4-methyleneproline at B28 accelerates fibril formation, while 4-methylation speeds dissociation from the pharmaceutically-formulated hexamer. This

work expands the scope of proline analogs amenable to incorporation into recombinant proteins, and demonstrates how non-canonical amino acid mutagenesis can engineer the therapeutically-relevant properties of protein drugs.

2.3 Introduction

Proline is unique among the canonical amino acids: the cyclic pyrrolidine side chain restricts the conformational space accessible to the residue. Replacing proline with any of the proteinogenic amino acids through standard mutagenesis approaches necessarily grants greater conformational freedom. Alternatively, non-canonical proline (ncPro) residues expand the chemical space about proline, while maintaining a pyrrolidine (or pyrrolidine-like) side chain. Because the conformational preferences of many proline analogs are known,[1,2] ncPro residues can also probe the effect that a particular proline conformation (i.e., *cis* or *trans* amide isomer) has on protein behavior. Proline analogs have demonstrated the importance of a key proline *cis-trans* isomerization event in $5HT_3$ receptor opening,[3] modified the properties of elastin-like proteins,[4–6] determined the molecular origins of collagen stability,[1,7] and probed the role of *cis-trans* isomerization in β2-microglobulin fibrillation.[8,9] To date, residue-specific (rather than site-specific) incorporation approaches remain the only method for installing proline analogs into recombinant proteins. Using *E. coli* as an expression host, this approach leads to the efficient incorporation of diverse 3- and 4-functionalized proline residues, and those with modified ring sizes and compositions.[10]

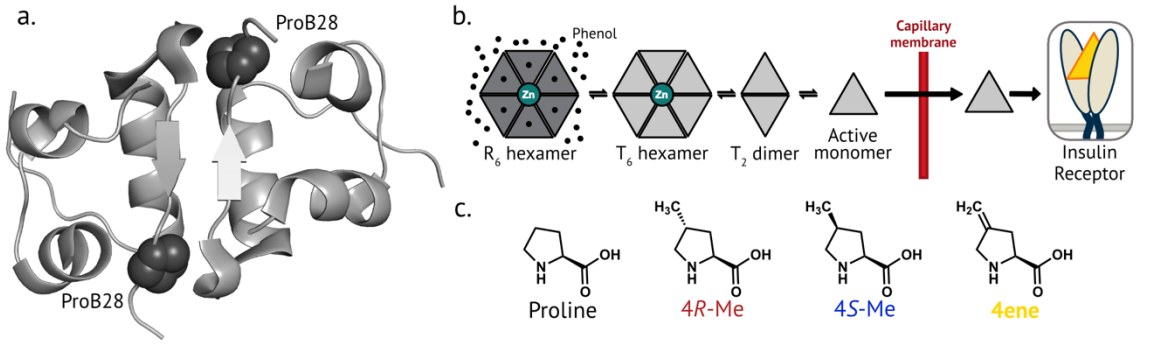

Figure 2.1. **Proline mutagenesis at position B28 of human insulin**. **a**. Crystal structure of insulin (PDB 1MSO), highlighting ProB28 located at the dimer interface. **b**. Simplified scheme of insulin dissociation after injection. Insulin exists as a hexamer in the R state in the presence of zinc and phenolic ligands, such as in the pharmaceutical formulation. After injection, insulin dissociates into lower-order oligomeric species that can more easily diffuse across the capillary membrane, enter the bloodstream, and bind to the insulin receptor. **c**. The structure of proline and the aliphatic proline analogs used in this study.

Insulin (Figure 2.1a) is a 5.8 kDa peptide hormone normally released from the pancreatic β cells in response to elevated levels of blood glucose. Its binding to the insulin receptor induces several intracellular responses that ultimately lower blood glucose concentrations.[11] Diabetes mellitus is a result of dysfunctional insulin signaling, either through an impaired ability to produce and secrete insulin (type 1), or by insulin resistance (type 2). Subcutaneous injection of exogenous insulin is a common strategy in diabetes treatment, especially for individuals with type 1 diabetes. The recombinant production of insulin[12] has significantly advanced diabetes treatment by enabling both its production at scale, and the creation of modified insulins with desirable properties through standard mutagenesis approaches.[13–16]

To more closely mimic the insulin-action profile of a healthy pancreas, two broad classes of insulin analogs have been developed: long-acting (or basal) insulins, and fast-acting insulins (FAIs).[17] Long-acting variants recapitulate the lower levels of insulin secretion that maintain metabolism in an anabolic state. Conversely, FAIs aim to mimic the transient increases in insulin concentration stimulated by elevated blood glucose, such as after a meal. Typical insulin replacement therapy relies on a combination of regular basal insulin treatments, and FAI injections before meals.

Many FDA-approved insulin variants contain minimal alterations to insulin's amino acid sequence that lead to pronounced pharmacokinetic effects.[18] Notably, insulin aspart[14,19] (NovoLog, marketed by Novo Nordisk) and insulin lispro[15,20] (Humalog, Eli Lilly) both involve changes to ProB28, a key dimer-stabilizing residue[15] near the C-terminus of the B-chain (Figure 2.1a). Insulin aspart is achieved by the single point mutation ProB28Asp, while insulin lispro contains an inversion of ProB28 and LysB29; both changes destabilize oligomer formation. Because the rate-limiting step for insulin absorption into the bloodstream is dissociation of oligomer to monomer[21] (Figure 2.1b), these changes accelerate insulin's onset of action. Insulins are also prone to chemical and physical denaturation,[22–24] processes slowed by the formation of protective oligomers. As a result, insulin production and distribution require a cold chain,[25] and its long-term storage in continuous subcutaneous insulin infusion pumps is challenging.[26]

Intrigued by the role that ProB28 plays in insulin biophysics, we recently introduced ncPro residues at position B28 of human insulin.[27,28] Because mature insulin contains one proline,

a residue-specific replacement approach results in site-specific proline replacement without the need for an orthogonal aminoacyl-tRNA synthetase/tRNA pair. These efforts, which focused on proline analogs known to incorporate well into recombinant proteins in *E. coli*,[5] illustrated how proline mutagenesis of insulin can tune its biophysical characteristics.[27,28]

Here, we demonstrate the efficient incorporation of three new aliphatic proline residues (4*R*-methylproline, 4*R*-Me; 4*S*-methylproline, 4*S*-Me; and 4-methyleneproline, 4ene; Figure 2.1c) at position B28 of recombinantly-produced insulin; these insulin variants will be referred to as ins-4*R*-Me, ins-4*S*-Me, and ins-4ene, respectively. We find that these modifications alter insulin behavior: replacement of ProB28 with 4ene speeds fibril formation, while 4-methylation accelerates hexamer dissociation. This work expands the range of proline analogs that can be incorporated into recombinant proteins in *E. coli*. It also demonstrates how small molecular changes introduced through non-canonical amino acid mutagenesis can be used to probe and engineer the therapeutically-relevant properties of protein drugs.

2.4 Results and discussion

2.4.1 Aliphatic proline residues are accepted by the E. coli translational machinery

To identify an expanded set of ncPro residues accepted by the *E. coli* translational machinery, we expressed proinsulin (a precursor to insulin) under conditions that favor ncPro incorporation.[10] We monitored ncPro replacement by proinsulin expression and mass spectrometry, and noted a range of incorporation efficiencies for the 15

commercially-available proline analogs tested (Figure 2.S1, Table 2.S1). Notably, the aliphatic proline residues 4*R*-Me, 4*S*-Me, and 4ene led to high levels of proinsulin expression and good (~90%) incorporation efficiencies under optimized conditions (Figure 2.2a-d, Table 2.S2, Table 2.S3). We chose to replace ProB28 of mature human insulin with each of these three aliphatic proline analogs (Figure 2.2e-h), and determined the effect that aliphatic proline replacement has on insulin behavior.

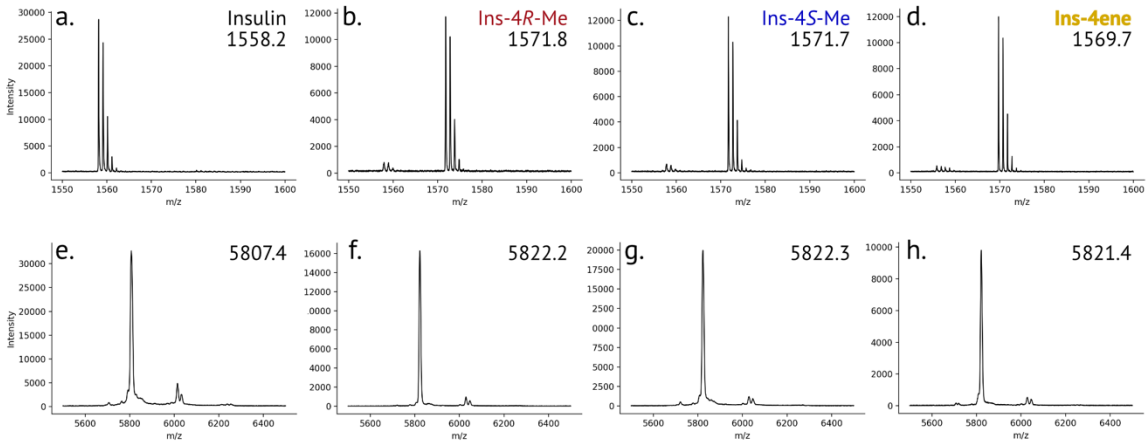

Figure 2.2. Mass spectrometry characterization of insulin variants. a-d. Characterization of proline analog incorporation. The solubilized inclusion body (containing proinsulin) after expression in the presence of proline (a), 4*R*-Me (b), 4*S*-Me (c), 4ene (d) was digested with Glu-C and analyzed by MALDI-TOF MS. The peptide that contains position B28 of mature insulin is ^{50}RGFFYT**P**KTRRE (expected m/z = 1557.8). **e-h**. MALDI-TOF characterization of mature and purified insulin variants: human insulin (e), Ins-4*R*-Me (f), Ins-4*S*-Me (g), and Ins-4ene (h). The larger molecular weight peaks (m/z ~6050) present in these spectra correspond to adducts of the sinapic acid matrix.

2.4.2 Proline analogs do not affect insulin secondary structure or bioactivity

The secondary structure of each insulin variant was assessed by circular dichroism spectroscopy (Figure 2.3a-c; Table 2.S4). CD can assess insulin oligomerization: for

monomeric insulins, the ratio of negative ellipticities at 208 and 222 nm is increased compared to that of the insulin dimer.[29] The far-UV CD spectrum for each variant closely matched that of human insulin at 60 µM, suggesting that proline replacement has not significantly perturbed secondary structure or dimer formation under these conditions.

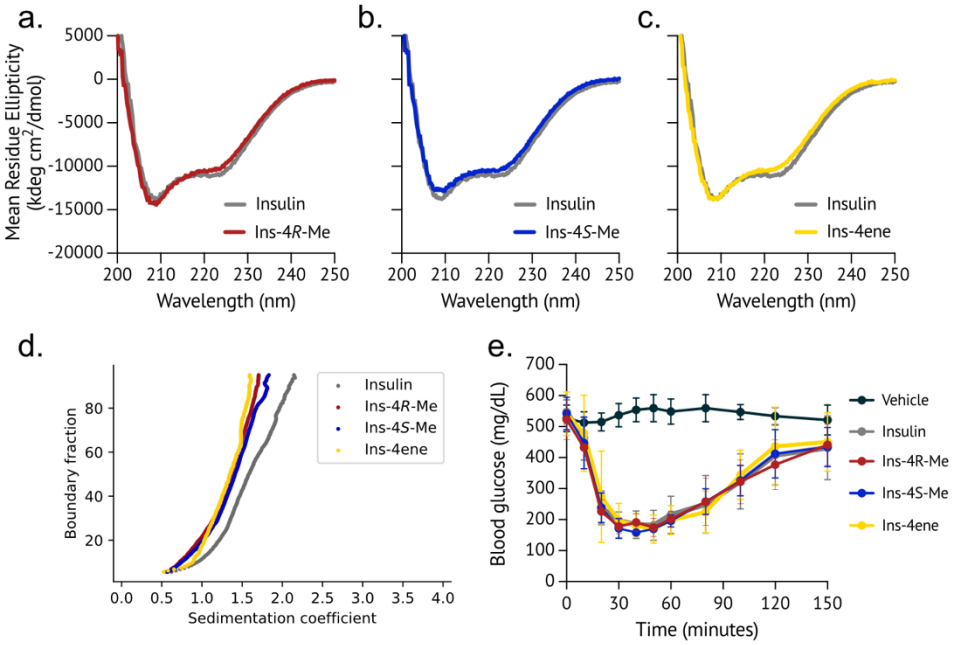

Figure 2.3. Circular dichroism spectroscopy, sedimentation, and bioactivity of insulin variants. **a-c**. Far-UV circular dichroism spectra of insulin and insulin variants (60 µM in 100 mM phosphate buffer, pH 8.0). The spectrum of each insulin variant is overlaid with that of human insulin (grey). **d**. Velocity sedimentation of insulin and insulin variants (60 µM in 100 mM phosphate buffer, pH 8.0). **e**. Insulins were injected subcutaneously into diabetic mice, and blood glucose was measured over time after injection.

Insulin oligomerization was studied further by velocity sedimentation analytical ultracentrifugation (Figure 2.3d, Table 2.S4). The broad sedimentation profiles of each insulin under these conditions (60 µM insulin, 100 mM phosphate, pH 8.0) are suggestive of a mixture of oligomeric states (here, predominantly monomers and dimers). The

sedimentation coefficients of all insulin variants (1.3-1.4S) were modestly reduced compared to that of human insulin (1.6S). These smaller sedimentation coefficients might be attributed to increased monomer-dimer dissociation constants. The K_d of Ins-4ene was determined from these data (25 μM, 95% CI [22, 28]), and is indeed moderately increased compared to the reported monomer-dimer K_d of human insulin (9 μM).[30] However, it was difficult to accurately determine the dissociation constants for the other insulins due to the existence of higher-order oligomers in the samples.

To validate biological activity *in vivo*, insulins were formulated with zinc and phenolic ligands, and injected subcutaneously into diabetic mice; blood glucose was monitored over 2.5 h. Rodent models can assess insulin activity, but cannot distinguish differences in onset of action between human insulin and fast-acting analogs.[31] Since the B-chain C-terminus does not interact with the insulin receptor,[32,33] we did not expect modification of ProB28 to hinder insulin's bioactivity. Indeed, all insulin variants reduced blood glucose in diabetic mice (Figure 2.3e).

2.4.3 4-methylene at position B28 accelerates fibril formation

To assess stability against physical denaturation, insulins were subjected to vigorous shaking at 37°C, and fibril formation was monitored over time with the fibril-specific dye thioflavin T (ThT; Figure 2.4a; Table 2.S4). The introduction of either 4*R*-Me or 4*S*-Me at insulin B28 did not significantly affect the lag time to fibril formation under these experimental conditions, compared to human insulin (lag time = 16.6 ± 4.1 h). However, ins-4ene (8.2 ± 4.0 h) formed fibrils more rapidly than the other insulins assessed here. All

aggregates were fibrillar in nature when assessed by transmission electron microscopy (TEM). Compared to the micrometer-long fibrils described in previous reports,[34–37] those observed here were relatively short (tens to hundreds of nanometers). Under these conditions, ins-4ene formed shorter fibrils compared to the other insulins (Figure 2.4b-e).

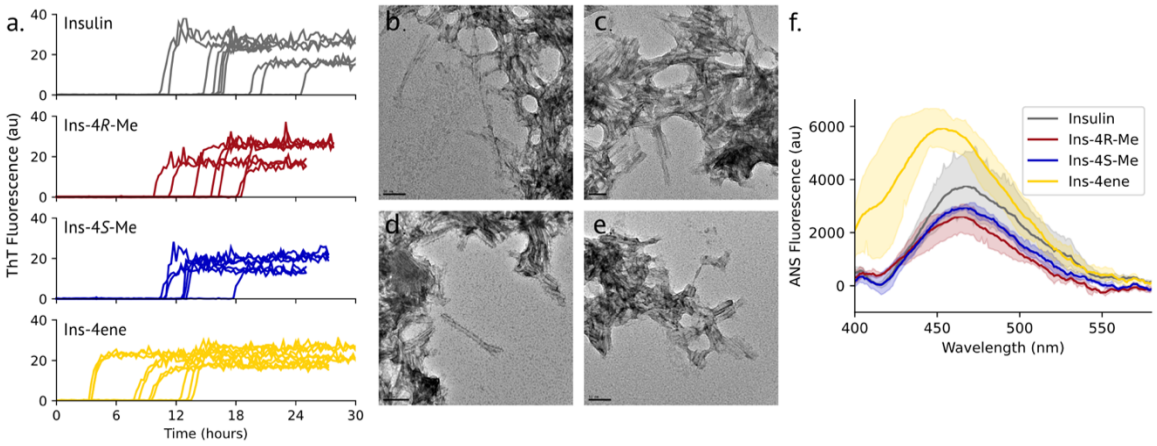

Figure 2.4. **Fibrillation of insulin variants a.** Insulin variants (60 µM in 100 mM phosphate buffer, pH 8.0) were incubated at 37°C with vigorous shaking, and fibril formation was monitored by ThT fluorescence over time. **b-e**. Representative TEM images of insulin (b; scale bar: 50 µm), ins-4R-Me (c; 50 µm), ins-4S-Me (d; 100 µm), and ins-4ene (e; 50 µm) aggregates. **f**. ANS emission spectra of insulin variants (1 µM insulin variant labeled with 5 µM ANS in 100 mM phosphate buffer, pH 8.0).

We sought to better understand the molecular mechanism for the decreased stability of ins-4ene. An early step in the mechanism of insulin fibril formation is thought to be detachment of the C-terminus of the insulin B-chain from the core of the molecule, exposing hydrophobic residues.[37] We probed insulin disorder with the dye 8-anilino-1-naphthalenesulfonic acid (ANS).[38] Compared to the other insulins measured, ins-4ene exhibits a blue-shift in the emission maximum and increase in fluorescence intensity upon labeling with ANS (Figure 2.4f; Table 2.S4), suggesting greater disorder. We analyzed

DMSO-solubilized ins-4ene fibrils by mass spectrometry, but did not observe chemical modification of the exocyclic alkene (data not shown).

2.4.4 4-methylation of ProB28 speeds hexamer dissociation

Increased negative ellipticity at 222 nm is indicative of insulin oligomerization.[29,39] Inspired by a previous report,[40] we measured the rate of dissociation to the monomer state using CD spectroscopy. Insulins were formulated under conditions that mimic the pharmaceutical formulation (600 μM insulin, 25 mM *m*-cresol, 250 μM $ZnCl_2$) and favor the R_6 hexamer state.[41] Dissociation was monitored by tracking mean residue ellipticity at 222 nm over time after 150-fold dilution into ligand-free buffer (Figure 2.5a). CD spectra after dilution are distinct from that of chemically-denatured insulin (Figure 2.S3), suggesting that changes in CD signal are not a result of denaturation. While ins-4ene dissociated similarly to human insulin ($t_{1/2}$ = 17.0 ± 2.3 s and 18.4 ± 2.8 s, respectively), the dissociation kinetics of ins-4*R*-Me ($t_{1/2}$ = 9.8 ± 1.8 s) and ins-4*S*-Me ($t_{1/2}$ = 9.9 ± 2.5 s) were accelerated under these conditions (Figure 2.5b-f, Table 2.S4).

Steric effects might play a role in accelerating the dissociation of ins-4*R*-Me and ins-4*S*-Me. A 4*R*- or 4*S*-methyl substituent installed at ProB28 in published structures of the R_6 insulin hexamer are in close proximity (2.2 and 2.3 Å, respectively) to backbone carbonyl oxygen atoms of the adjacent monomer (Figure 2.S4a); these distances are within the van der Waals radii of the respective atoms. Similar interactions are observed in the T_6 hexamer (Figure 2.S4b). The resulting steric clashes perhaps destabilize the hexamer and accelerate dissociation.

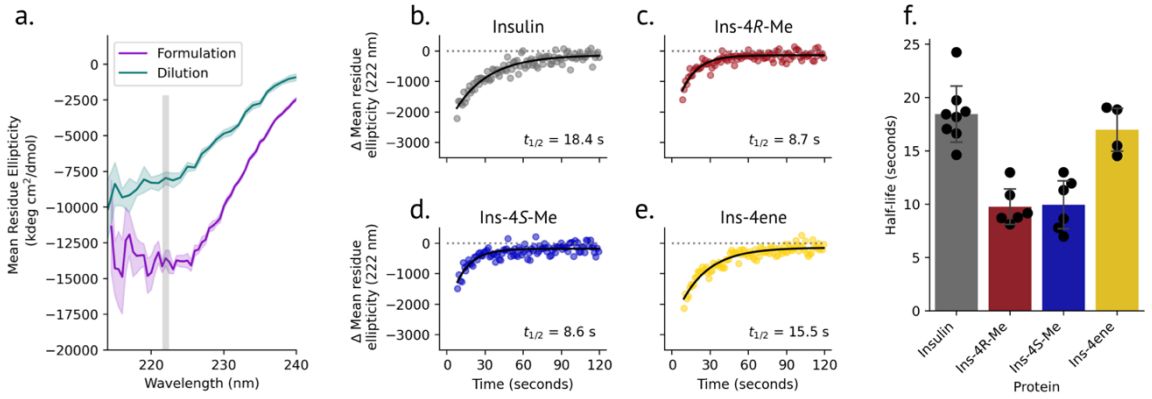

Figure 2.5. Hexamer dissociation kinetics of insulin variants a. Equilibrium CD spectra of insulin before and after dilution. To measure dissociation kinetics, the decrease in negative ellipticity at 222 nm was monitored over time after dilution. **b-e**. Representative dissociation kinetics for insulin (**b**), ins-4R-Me (**c**), ins-4S-Me (**d**), and ins-4ene (**e**). Each dilution experiment was fit to a mono-exponential function, and the half-life for each displayed replicate is indicated. **f**. Summary of dissociation half-life values.

We examined the oligomerization of insulin and ins-4*S*-Me in more detail by sedimentation velocity analytical ultracentrifugation. At 300 µM and in the presence of zinc and *m*-cresol, the sedimentation coefficients of both insulin (3.3S) and ins-4*S*Me (3.4S) correspond to the hexamer[42] (Figure 2.6a, Table 2.S5). After dilution, both insulins behave as monomers (Figure 2.6b, Table 2.S5), further validating that the dilution kinetics measured above indeed describe hexamer dissociation. Interestingly, we found that 4*S*-methylation of ProB28 affects insulin oligomerization in the absence of ligands. Compared to the broad sedimentation profile of insulin, which is indicative of insulin's heterogeneous oligomerization behavior,[43] the sedimentation of ins-4*S*-Me (3.1S) is nearly unchanged in the absence of zinc and *m*-cresol (Figure 2.6b).

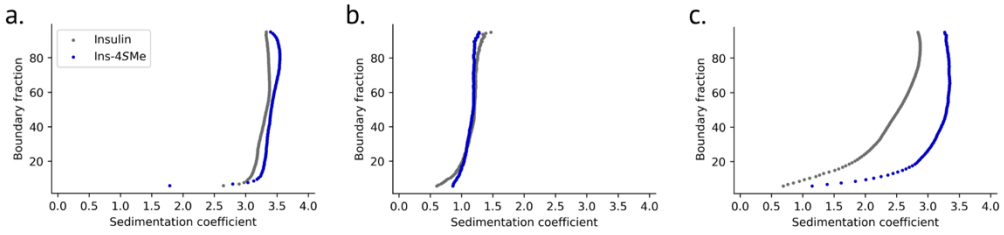

Figure 2.6. Sedimentation velocity of insulin and ins-4S-Me. Insulins were formulated under the following conditions: **a**. 300 µM insulin, 12.5 mM m-cresol, 125 µM Zn. **b**. 4 µM insulin, 167 µM m-cresol, 1.67 µM Zn. **c**. 300 µM insulin. All samples were in 25 mM tris buffer, pH 8.0.

2.5 Conclusion

Here, we have expanded the set of ncPro residues that can be incorporated into proteins in *E. coli*, and demonstrated their utility in modifying the properties of recombinantly-produced insulin. We found that the proline analogs 4*R*-Me, 4*S*-Me, and 4ene could be incorporated with high efficiencies into recombinant proinsulin. During the completion of this work, the incorporation of 4*R*-Me was also reported in the recombinant expression of thioredoxin.[44] In that case, incorporation of the 4*S* diastereomer could not be detected, perhaps due to its interference with thioredoxin folding. Conversely, insoluble proinsulin is purified from the inclusion body fraction, so any influence of ncPro replacement on protein stability during expression is likely reduced. We were also able to detect low to modest levels of incorporation of other ncPro residues, such as the diazirine-containing variant photo-proline, and 3-hydroxy and 4-oxo analogs (Figure 2.S1, Table 2.S1). Engineering the prolyl-tRNA synthetase or other components of the *E. coli* translational machinery might lead to improved incorporation of these analogs.

Besides expanding the size and hydrophobic surface area at proline residues, the three aliphatic ncPro residues studied here complement other translationally-active proline

analogs used to probe proline conformation. For instance, 4*R*-Me and 4*S*-Me have opposing conformational preferences relative to their more commonly-used 4-fluoroproline counterparts,[1,2,45] and together have been used to validate the stereoelectronic origin of collagen stability.[7,45] The amplitude of the pyrrolidine ring pucker of 4ene is expected to be attenuated compared to proline (Figure 2.S5), a conformation approaching that of the planar 3,4-dehydroproline.[46] Furthermore, the olefin present in 4ene might be used as a chemical handle for protein modification at proline residues.[47,48]

This work also demonstrates how small molecular changes can have oversized effects on protein behavior. For instance, introducing an exocyclic alkene at B28 of insulin shortened its lag time to fibril formation. This finding, together with increased disorder of ins-4ene as evidenced by its ANS emission spectrum, supports current models for the mechanism of insulin fibrillation.[37] Furthermore, installing a simple methyl group at ProB28 of human insulin significantly accelerates dissociation from the insulin hexamer, behavior associated with achieving a rapid onset-of-action for FAIs.[49] Future work is needed to more fully understand the effect that 4-methylation has on insulin oligomerization, and to determine if the faster dissociation kinetics observed for ins-4*R*-Me and ins-4*S*-Me *in vitro* result in a faster onset-of-action *in vivo*.

2.6 Materials and methods

2.6.1 Chemicals

All chemicals were purchased from MilliporeSigma unless otherwise indicated. 4-methyleneproline (4ene) was purchased as the N-boc protected version from Acros

Organics, and deprotected with trifluoroacetic acid (TFA) in dichloromethane. 4ene was extracted with H_2O and lyophilized; complete deprotection and >95% purity was verified by ^{1}H NMR. All other proline analogs were used as received: 4*R*-methylproline (4*R*-Me) was purchased from Advanced Chemblocks as the hydrochloride salt; ^{1}H NMR analysis indicated the presence of 5-10% of the 4*S*-methylproline diastereomer. 2-methylproline (2-Me) was purchased from Advanced ChemBlocks as the hydrochloride salt. 4*S*-methylproline (4*S*-Me) was purchased from AstaTech as the hydrochloride salt. L-*cis*-pyrrolidine-2,4-dicarboxylic acid (4*S*-COOH) was purchased from Boc Sciences. L-*trans*-pyrrolidine-2,4-dicarboxylic acid (4*R*-COOH) was purchased from Tocris Biosciences. L-pyroglutamic acid (5-oxo) was purchased from Aldrich. 4,4-dimethylproline (44-diMe) was purchased from J&W Pharmlab. Piperazine-2-carboxylic acid (Pip-Az) was purchased from Ark Pharm as the dihydrochloride salt. (2*S*,5*S*)-5-hydroxypiperidine-2-carboxylic acid (Pip-OH) was purchased from Ark Pharm. 4*S*-aminoproline (4*S*-NH_2) was purchased from Toronto Research Chemicals as the dihydrochloride salt. 3*R*-hydroxyproline (3*R*-OH) was purchased from Combi-Blocks. 3*S*-hydroxyproline (3*S*-OH) was purchased from Ark Pharm. 4-oxoproline (4-oxo) was purchased as the hydrobromide salt from MilliporeSigma.

2.6.2 Enzymes

Restriction enzymes, kinases, and ligases were purchased from New England Biolabs. Trypsin was purchased from MilliporeSigma. Carboxypeptidase B was purchased from Worthington Biochemical. Glu-C peptidase was purchased from Promega.

2.6.3 Strains and plasmids

The proline-auxotrophic *E. coli* strain CAG18515 was obtained from the Coli Genetic Stock Center (CGSC) at Yale University. Strain DH10B was used for standard cloning operations; electrocompetent CAG18515 were transformed with purified plasmid products.

The plasmid pQE80_H27R-PI_proS contains an IPTG-inducible proinsulin gene and the *E. coli* prolyl-tRNA synthetase gene controlled by its endogenous promoter. Proinsulin is translationally fused to an N-terminal leader peptide (H27R) that increases expression yields,[50] and a 10x-his tag to facilitate proinsulin enrichment after refolding. The gene for H27R-PI was ordered as a g-Block gene fragment from Integrated DNA Technologies (IDT) after codon optimization of the N-terminal leader peptide. A restriction enzyme cloning approach (XhoI and BamHI restriction enzyme cut sites) was used to replace the hexahistidine-tagged proinsulin gene in the plasmid pQE80PI-proS, which was described previously.[28] Correct installation of the gene of interest was verified by Sanger sequencing.

A blunt-end ligation approach was used to install the C443G and M157Q mutations. The *proS*-containing plasmid was amplified with primers AL01004_fwd & AL01004_rev (C443G), or AL01005_fwd & AL01005_rev2 (M157Q). The linear PCR product was phosphorylated (T4 PNK) and circularized (T4 DNA ligase). Correct installation of the point mutation was verified by Sanger sequencing.

2.6.4 Primers

DNA oligos were purchased from Integrated DNA Technologies (IDT). Nucleotides responsible for installing the M157Q and C443G mutations are underlined.

AL01004_fwd: ATACCGTAGCCACCCATCGTCAGG
AL01004_rev: CGGGGTAACGCGTGTGGT
AL01005_fwd: AGCATCTTTCTGCAGGAATTCGC
AL01005_rev2: TACTCTTTCCATACTTCTCAGGAATCC

2.6.5 Nucleotide and amino acid sequences

H27R-PI: The g-Block gene fragment was purchased from IDT. The coding sequence is in UPPERCASE, XhoI and BamHI cut sites are underlined.

gcccttt cgtcttcacctcgagaaatcataaaaaatttatttgctttgtgagcggataacaattataatagattcaattgtgagcggata
acaatttcacacagaattcattaaagaggagaaattaactATGACAATGATCACTAATTCACCCGAGA
TTTCCCACCATCATCATCATCATCACCACCACCATCAGTTGATCTCGGAGGCC
CGTTTTGTGAACCAGCACCTGTGCGGTAGCCACCTGGTGGAAGCTCTGTACCT
GGTTTGCGGTGAGCGTGGTTTCTTCTACACGCCAAAGACCCGCCGTGAAGCT
GAAGATCTGCAGGTGGGCCAGGTAGAACTGGGCGGTGGTCCGGGTGCCGGCT
CTCTGCAACCGCTGGCACTGGAAGGTTCCCTGCAAGCGCGTGGTATCGTAGA
GCAGTGCTGTACTTCTATCTGCTCCCTGTACCAGCTGGAGAACTACTGTAATT
AAggatccgcatgcgagc

The sequence of the H27R leader peptide is underlined, proinsulin is in **bold**. The A-chain and B-chain in mature insulin are colored **red** and **blue**, respectively.

MTMITNSPEISHHHHHHHHHHQLISEAR**FVNQHLCGSHLVEALYLVCGERGFF**
YTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEGSLQARGIVEQCCTSICS
LYQLENYCN

proS: the endogenous *proS* promoter is underlined, the coding sequence is in UPPERCASE.

attcacgccttctcttttgacatttcttttgcactggtaaactaaatcactttttttttgtcccaggctcgccttgagcctgttctaccttcc
aactggaaccgtaacaacATGCGTACTAGCCAATACCTGCTCTCCACTCTCAAGGAGAC
ACCTGCCGACGCCGAGGTGATCAGCCATCAGCTGATGCTGCGCGCCGGGATG
ATCCGCAAGCTGGCCTCCGGGTTATATACCTGGCTGCCGACCGGCGTGCGCG
TTCTGAAAAAAGTCGAAAACATCGTGCGTGAAGAGATGAACAACGCCGGTGC
GATCGAGGTGTCGATGCCGGTGGTTCAGCCAGCCGATTTGTGGCAAGAGAGT
GGTCGTTGGGAACAGTACGGTCCGGAACTGCTGCGTTTTGTTGACCGTGGCG
AGCGTCCGTTCGTACTCGGCCCAACTCATGAAGAAGTTATCACTGACCTGATT
CGTAACGAGCTTAGCTCTTACAAACAGCTGCCGCTGAACTTCTATCAGATCCA
GACCAAGTTCCGCGACGAAGTGCGTCCGCGTTTCGGCGTCATGCGTTCCCGC
GAATTCCTGATGAAAGATGCTTACTCTTTCCATACTTCTCAGGAATCCCTGCA
GGAAACCTACGATGCAATGTATGCGGCCTACAGCAAAATCTTCAGCCGCATG
GGGCTGGATTTCCGCGCCGTACAAGCCGACACCGGTTCTATCGGCGGCAGCG
CCTCTCACGAATTCCAGGTGCTGGCGCAGAGCGGTGAAGACGATGTGGTCTT
CTCCGACACCTCTGACTATGCAGCGAACATTGAACTGGCAGAAGCTATCGCG
CCGAAAGAACCGCGCGCTGCTGCTACCCAGGAAATGACGCTGGTTGATACGC
CGAACGCGAAAACCATCGCGGAACTGGTTGAACAGTTCAATCTGCCGATTGA
GAAAACGGTTAAGACTCTGCTGGTTAAAGCGGTTGAAGGCAGCAGCTTCCCG
CAGGTTGCGCTGCTGGTGCGCGGTGATCACGAGCTGAACGAAGTTAAAGCAG
AAAAACTGCCGCAGGTTGCAAGCCCGCTGACTTTCGCGACCGAAGAAGAAAT
TCGTGCCGTGGTTAAAGCCGGTCCGGGTTCACTGGGTCCGGTAAACATGCCG
ATTCCGGTGGTGATTGACCGTACCGTTGCGGCGATGAGTGATTTCGCTGCTGG
TGCTAACATCGATGGTAAACACTACTTCGGCATCAACTGGGATCGCGATGTC
GCTACCCCGGAAGTTGCAGATATCCGTAACGTGGTGGCTGGCGATCCAAGCC
CGGATGGCCAGGGTAGGCTGCTGATCAAACGTGGTATCGAAGTTGGTCACAT
CTTCCAGCTGGGTACCAAGTACTCCGAAGCACTGAAAGCCTCCGTACAGGGT
GAAGATGGCCGTAACCAAATCCTGACGATGGGTTGCTACGGTATCGGGGTAA
CGCGTGTGGTAGCTGCGGCGATTGAGCAGAACTACGACGAACGAGGCATCGT

ATGGCCTGACGCTATCGCGCCGTTCCAGGTGGCGATTCTGCCGATGAACATG
CACAAATCCTTCCGCGTACAAGAGCTTGCTGAGAAACTGTACAGCGAACTGC
GTGCACAAGGTATCGAAGTGCTGCTGGATGACCGCAAAGAGCGTCCGGGCGT
GATGTTTGCTGATATGGAACTGATCGGTATTCCGCACACTATTGTGCTGGGCG
ACCGTAACCTCGACAACGACGATATCGAATATAAATATCGTCGCAACGGCGA
GAAACAGTTAATTAAGACTGGTGACATCGTCGAATATCTGGTGAAACAGATT
AAAGGCTGA

MRTSQYLLSTLKETPADAEVISHQLMLRAGMIRKLASGLYTWLPTGVRVLKKVE
NIVREEMNNAGAIEVSMPVVQPADLWQESGRWEQYGPELLRFVDRGERPFVLG
PTHEEVITDLIRNELSSYKQLPLNFYQIQTKFRDEVRPRFGVMRSREFLMKDAYSF
HTSQESLQETYDAMYAAYSKIFSRMGLDFRAVQADTGSIGGSASHEFQVLAQSG
EDDVVFSDTSDYAANIELAEAIAPKEPRAAATQEMTLVDTPNAKTIAELVEQFNL
PIEKTVKTLLVKAVEGSSFPQVALLVRGDHELNEVKAEKLPQVASPLTFATEEEIR
AVVKAGPGSLGPVNMPIPVVIDRTVAAMSDFAAGANIDGKHYFGINWDRDVAT
PEVADIRNVVAGDPSPDGQGRLLIKRGIEVGHIFQLGTKYSEALKASVQGEDGRN
QILTMGCYGIGVTRVVAAAIEQNYDERGIVWPDAIAPFQVAILPMNMHKSFRVQ
ELAEKLYSELRAQGIEVLLDDRKERPGVMFADMELIGIPHTIVLGDRNLDNDDIE
YKYRRNGEKQLIKTGDIVEYLVKQIKG

2.6.6 Screening ncPro incorporation

A single colony of *E. coli* strain CAG18515 harboring plasmid pQE80_H27R-PI_proS was used to inoculate a culture of Luria Bertani (LB) medium supplemented with ampicillin. The culture was grown overnight at 37°C until stationary phase was reached, then diluted 1:100 into 100 mL of 1x M9 medium, supplemented with all twenty amino acids. The medium composition of M9 is as follows: 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^{-1} $FeSO_4$, 1 µg L^{-1}

trace metals [Cu^{2+}, Mn^{2+}, Zn^{2+}, MoO_4^{2-}], 35 mg L^{-1} thiamine HCl, 10 mg L^{-1} biotin, 20 mM D-glucose, 100 mg L^{-1} ampicillin, 50 mg L^{-1} of each L-amino acid.

The culture was grown at 37°C until it reached OD ~0.8, after which it was subjected to a medium shift: cells were pelleted via centrifugation (5 kg, 5 min, 4°C) and washed twice with 10 mL ice-cold 0.9% NaCl. Washed cells were resuspended in 80 mL of 1.25x M9 –Pro, a 1.25x concentrated form of M9 that omits proline. The culture was split into 4 mL aliquots, and incubated for 30 min at 37°C to deplete residual proline. A 1 mL solution containing 2.5 mM ncPro and 1.5 M NaCl was added (0.5 mM ncPro and 0.3 M NaCl working concentrations). After 30 min of incubation at 37°C to allow for ncPro uptake, proinsulin expression was induced by the addition of 1 mM IPTG. Cultures were incubated for 2.5 h at 37°C, after which cells were harvested via centrifugation and stored at -80°C until further processing.

Cell pellets were thawed and lysed with B-PER Complete (Thermo Fisher Scientific) for 1 h at room temperature with shaking, then centrifuged (20 kg, 10 min) and the supernatant discarded. The pellet (containing insoluble proinsulin) was washed once with Triton wash buffer (2 M urea, 20 mM Tris, 1% Triton X-100, pH 8.0), and twice with ddH_2O. The pellet was resuspended in solubilization buffer (8 M urea, 300 mM NaCl, 50 mM NaH_2PO_4, pH 8.0), and proinsulin was allowed to dissolve for 1 h at room temperature with shaking. Samples were centrifuged, and the supernatant removed for analysis by SDS-PAGE and MALDI-TOF (described in section 2.6.10 below).

2.6.7 Proinsulin expression

A single colony of *E. coli* strain CAG18515 harboring plasmid pEQ80_H27R-PI_proS (or the corresponding plasmid with a point mutation in the *proS* gene; see Table 4.S3) was used to inoculate 70 mL of LB medium containing ampicillin, and the culture was grown overnight at 37°C to stationary phase. The overnight culture was used to inoculate 5 L (as 4x1.25 L cultures) of 1x Andrew's Magical Medium (AMM),[51] a defined medium containing all 20 proteinogenic amino acids, in 2.8 L Fernbach flasks. The composition of AMM was the following: 3.60 g L^{-1} glucose, 3.5 g L^{-1} KH_2PO_4, 6.56 g L^{-1} $K_2HPO_4 \bullet 3H_2O$, 3.5 g L^{-1} $(NH_4)_2HPO_4$, 8.37 g L^{-1} MOPS, 0.72 g L^{-1} tricine, 2.92 g L^{-1} NaCl, 0.51 g L^{-1} NH_4Cl, 0.26 g L^{-1} $MgCl_2 \bullet 7H_2O$, 50 mg L^{-1} K_2SO_4, 0.246 mg L^{-1} $MgSO_4 \bullet 7H_2O$, 12.3 mg L^{-1} $CaCl_2$,$\bullet 2H_2O$, 2.8 mg L^{-1} $FeSO_4 \bullet 7H_2O$, 0.5 mg L^{-1} thiamine, 24 µg L^{-1} boric acid, 1 µg L^{-1} trace metals (Cu^{2+}, Mn^{2+}, Zn^{2+}, MoO_4^{2-}), and 50 mg L^{-1} each amino acid.

When growth reached mid-exponential phase (OD_{600} ~0.8), the culture was subjected to a medium shift: cells were pelleted via centrifugation (5 kg, 5 min, 4°C) and washed twice with 100 mL ice-cold 0.9% NaCl. Washed cells were resuspended in 1 L of 1.25x AMM –Pro, a 1.25x concentrated form of AMM that omits proline. Cells were incubated for 30 min at 37°C to deplete residual proline, after which 250 mL of a solution containing 2.5-5.0 mM ncPro (see Table 4.S3) and 2.5 M NaCl was added (0.5-1.0 mM ncPro and 0.5 M NaCl working concentrations). After 30 min of incubation at 37°C to allow for ncPro uptake, proinsulin expression was induced by the addition of 1 mM IPTG. Cultures were incubated overnight at 37°C, after which cells were harvested via centrifugation and stored at -80°C until further processing.

Proline-containing proinsulin was expressed using strain CAG18515 harboring plasmid pQE80-H27R-PI_proS in 7.5 L (as 6 x 1.25 L cultures) of Terrific Broth (TB). 1 mM IPTG was added at mid-log phase (OD_{600} ~0.8) to induce proinsulin expression. Cultures were incubated at 37°C for 3 h, after which cells were harvested via centrifugation and stored at -80°C until further processing.

2.6.8 Proinsulin refolding

Cell pellets were warmed from -80°C to room temperature and resuspended in 5 mL IB buffer (50 mM tris, 100 mM NaCl, 1 mM EDTA, pH 8.0) per gram cell pellet. 1 mg L^{-1} lysozyme and 1 mM PMSF were added, and the slurry sat on ice for 30 min before cells were lysed via sonication. The lysate was centrifuged (14 kg, 30 min, 4°C) and the soluble fraction was discarded. The pellet was washed twice with IB buffer + 1% Triton X-100, once with IB buffer, and once with water; this final step required centrifugation for 45 min. The washed inclusion body pellet was resuspended in a minimal amount of water, and the mass of proinsulin in the inclusion body pellet was estimated by SDS-PAGE.

In preparation for proinsulin refolding, the inclusion body was resuspended in 3 M urea and 10 mM cysteine in water, such that the proinsulin concentration was 1 mg proinsulin per L total slurry. To dissolve proinsulin, the pH was adjusted to 12 and sample stirred for 1 h at room temperature. At this stage, ncPro incorporation was assessed by MALDI-TOF, which is described in section 2.6.10 below. The solubilized proinsulin solution was diluted ten-fold into refold buffer (10 mM CAPS, pH 10.6) that had been pre-cooled to 4°C. The pH of the refold solution was adjusted to 10.7 and the sample stored at 4°C; care was taken

to ensure that the solution pH remained between 10.6 and 10.8 throughout the refolding process. Proinsulin refolding progress was monitored by reverse-phase HPLC, and usually reached completion within 50 h.

Proinsulin was enriched from the refold solution after adjusting the pH to 8.0 and incubating the sample overnight with Ni-NTA resin and 10 mM imidazole. The resin was washed with wash buffer (25 mM imidazole in PBS, pH 8.0), and proinsulin was eluted with elution buffer (250 mM imidazole in PBS, pH 8.0). Fractions containing proinsulin were combined and extensively dialyzed against 10 mM sodium phosphate, pH 8.0.

2.6.9 Insulin maturation and purification

Refolded proinsulin was warmed to 37°C and digested with trypsin (20 U mL^{-1}) and carboxypeptidase-B (10 U mL^{-1}) at 37°C for 90 min to remove the N-terminal tag and C chain. Digestion was halted by adjusting the pH to ~3 with 6 N HCl.

Insulins were immediately purified after proteolysis by reverse-phase HPLC on a C_4 column (Penomenex Jupiter 5 µm particle size, 300 Å pore size, 250x10 mm) using 0.1% TFA in water (solvent A) and 0.1% TFA in acetonitrile (solvent B) as mobile phases. A gradient of 25-32% solvent B was applied over 65 min, and fractions containing insulin were collected. Samples for purity analysis were removed at this stage; the remaining portion of the fraction was lyophilized. Each insulin fraction was analyzed by analytical reverse-phase HPLC, MALDI-TOF MS (Figure 2.2e-h), and SDS-PAGE (Figure 2.S2) to verify sample quality and ensure ≥95% purity for all downstream analyses. Lyophilized insulin powders were stored at -20°C until further use.

2.6.10 MALDI-TOF MS

To assess levels of incorporation of ncPros into the corresponding proinsulins, samples were digested with Glu-C, which results in a peptide fragment containing ProB28 (^{50}RGFFYT**P**KTRRE). A 20 µL aliquot of proinsulin was subjected to cysteine reduction (5 mM DTT, 55°C for 20 min) and alkylation (15 mM iodoacetimide, RT for 15 min in the dark), prior to 10-fold dilution into 100 mM NH_4HCO_3, pH 8.0 (100 µL final volume). Digestion was started with addition of 0.6 µL Glu-C (0.5 µg μL^{-1} in ddH_20) at 37°C for 2.5 h. The digestion reaction was quenched by adding 10 µL of 5% TFA. Peptides were desalted using ZipTip C_{18} columns (MilliporeSigma) according to the manufacturer's protocol. Desalted peptides (in 50% acetonitrile, 0.1% TFA) were diluted 3:1 into the matrix solution (α-cyanohydroxycinnamic acid in 50% ACN, 0.1% TFA) and analyzed by MALDI-TOF MS. Analog incorporation was calculated by comparing the area under the curve (AUC) of the ncPro form of the peptide (m/z = 1572 for 4*R*-Me and 4*S*-Me, and 1570 for 4ene) with the AUC of the canonical proline peptide (m/z = 1558).

HPLC-purified insulins were analyzed as full-length, mature proteins. Aliquots directly from HPLC purification (~30% ACN, 0.1% TFA) were mixed 1:1 with matrix solution (sinapic acid in 30% ACN, 0.1% TFA) before analysis by MALDI-TOF MS.

2.6.11 Circular dichroism spectroscopy

Equilibrium measurements: The circular dichroism spectra of insulin samples (60 µM in 100 mM sodium phosphate, pH 8.0) were measured at 25°C in 1 mm quartz cuvettes on an Aviv Model 430 Circular Dichroism Spectrophotometer using a step size of 0.5 nm and

averaging time of 1 s. A reference buffer spectrum was subtracted from each sample spectrum.

Kinetic measurements: Insulin samples in 100 mM sodium phosphate buffer pH 8.0 were dialyzed overnight against 28.6 mM tris buffer, pH 8.0 (Slide-A-Lyzer dialysis cassettes, 3.5 kDa MWCO, ThermoFisher). Insulins were formulated as the following: 600 μM insulin, 250 μM $ZnCl_2$, 25 mM *m*-cresol, 25 mM tris buffer, pH 8. To a stirred buffer solution containing 2.98 mL of 25 mM tris, pH 8.0 in a 10 mm quartz cuvette was injected 20 μL of the insulin formulation (150-fold dilution). Ellipticity was monitored at 222 nm over 120 s (1 s kinetic interval, 0.5 s time constant, 10 nm bandwidth) at 25°C. A typical run led to a rapid drop in CD signal as mixing occurred (~5 s), then a gradual rise to an equilibrium ellipticity representative of an insulin monomer. Data preceding the timepoint with the greatest negative ellipticity (representing the mixing time) were omitted from further analysis. Runs were discarded if the maximum change in mean residue ellipticity from equilibrium did not exceed 750 deg cm^2 $dmol^{-1}$, which indicated poor mixing. The remaining data were fit to a mono-exponential function using Scipy (Python). The data presented here are from at least two separate insulin HPLC fractions, measured on two different days.

For quality control, an equilibrium spectrum for each protein was obtained after dilution as described above; all spectra resembled that of the insulin monomer[29] (data not shown). The CD spectrum of human insulin under pre-dilution formulation conditions was obtained

using a 0.1 mm quartz cuvette. In each case, a blank spectrum containing all buffers and ligands was subtracted from the sample spectrum.

2.6.12 Reduction of blood glucose in diabetic mice

NODscid mice (NOD.CB17-Prkdcscid/J) were obtained from Jax Mice (Bar Harbor, ME). Mice were maintained under specific pathogen-free conditions, and experiments were conducted according to procedures approved by the Institutional Animal Care and Use Committee (IACUC) at the City of Hope. Adult (8-12 week old) male NODscid mice were injected intraperitoneally (45 mg kg^{-1} day^{-1} for 3 days) with freshly prepared streptozotocin (STZ) in 50 mM 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 (200-600 mg dL^{-1}) as measured by a glucomonitor (Freestyle, Abbott Diabetes Care, Alameda, CA) in blood sampled from the lateral tail vein. Insulin analogs were diluted to 100 µg mL^{-1} in formulation buffer (1.6 mg mL^{-1} *m*-cresol, 0.65 mg mL^{-1} phenol, 3.8 mg mL^{-1} sodium phosphate pH 7.4, 16 mg mL^{-1} glycerol, 0.8 µg mL^{-1} $ZnCl_2$). Insulin analogs were injected (35 µg kg^{-1}) subcutaneously at the scruff and blood glucose was measured at 0, 10, 20, 30, 40, 50, 60, 80, 100, 120, and 150 min.

2.6.13 Fibrillation

Insulin samples (60 µM in 100 mM sodium phosphate, pH 8.0) were centrifuged at 22,000 g for 1 h at 4°C, prior to the addition of 1 µM thioflavin T (ThT). Each insulin (200 µL) was added to a 96-well, black, clear bottom plate (Greiner Bio-One) and sealed. Samples were shaken continuously at 960 rpm on a Varioskan multimode plate reader at 37°C, and

fluorescence readings were recorded every 15 min (444 nm excitation, 485 nm emission). Fibrillation runs were performed on at least two separate HPLC fractions, each in triplicate or quadruplicate, on at least two different days. The growth phase of each fibrillation replicate was fit to a linear function, and fibrillation lag times were reported as the x-intercept of this fit. Fibril samples were stored at 4°C until analysis by TEM.

2.6.14 Transmission electron microscopy

Insulin fibrils were centrifuged (5 kg, 1 min), then washed twice and resuspended in ddH_2O. Fibrils were stained with 2% uranyl acetate on a 300-mesh formvar/carbon coated copper grid (Electron Microscopy Sciences) and imaged on a Tecnai T12 LaB6 120 eV transmission electron microscope.

2.6.15 ANS fluorescence

1 μM insulin or insulin variant was mixed with 5 μM ANS in 100 mM phosphate buffer, pH 8.0. Fluorescence emission spectra were measured in 1 cm quartz cuvettes at ambient temperature using a PTI QuantaMaster fluorescence spectrofluorometer. A 350 nm excitation wavelength and scan rate of 2 nm s^{-1} were used. Measurements for each insulin were performed in triplicate from three separate HPLC fractions. Spectra were smoothed before plotting and determining the emission maxima.

2.6.16 Analytical ultracentrifugation

Insulins were dissolved in 100 mM phosphate buffer, pH 8.0 and formulated at 60 μM. Alternatively, insulins were dialyzed against 28.6 mM tris buffer, pH 8.0, and formulated at 300 μM insulin, 12.5 mM *m*-cresol, and 125 μM $ZnCl_2$. Ligand-free insulins in tris buffer

were formulated from the same dialysis sample. Velocity sedimentation experiments were performed by Amy Henrickson and Maduni Ranasinghe at the Canadian Center for Hydrodynamics at the University of Lethbridge, with input from Borries Demeler. Samples containing 12.5 mM *m*-cresol were measured by interference optics, due to the high background absorbance of *m*-cresol. Ligand-free samples were measured using absorbance optics. All data were analyzed with UltraScan III version 4.0 release 6606.[52] Sedimentation coefficients were determined from the enhanced van Holde-Weischhet analysis[53] to generate diffusion-corrected sedimentation coefficient distributions.

2.6.18 Calculations of proline and proline analog conformation

The equilibrium geometry conformations of the N-methyl, O-methyl ester protected versions of proline, 4-methyleneproline, and 3,4-dehydroproline in water were calculated using Spartan Student (Wavefunction) at the B3LYP/G-31+G** level of theory. Pseudorotation parameters were calculated from the dihedral angles about the pyrrolidine ring, as previously reported.[54]

2.6.19 Models of ins-4R-Me and ins-4S-Me hexamers

Crystal structures of the T_6 (PDB: 1MSO) and R_6 (1EV6) insulin hexamers were downloaded from the Protein Data Bank and visualized with Pymol. The hydrogen atoms at the C^{γ} position of ProB28 were replaced with methyl groups; no additional energy minimization was used.

2.7 Supplementary figures and tables

Figure 2.S1. SDS-PAGE analysis of proinsulin expression in media supplemented with non-canonical proline analogs. Proinsulin (12.7 kDa) was expressed after a medium shift to ncPro-containing medium. The inclusion body fraction was isolated, solubilized, and analyzed by SDS-PAGE.

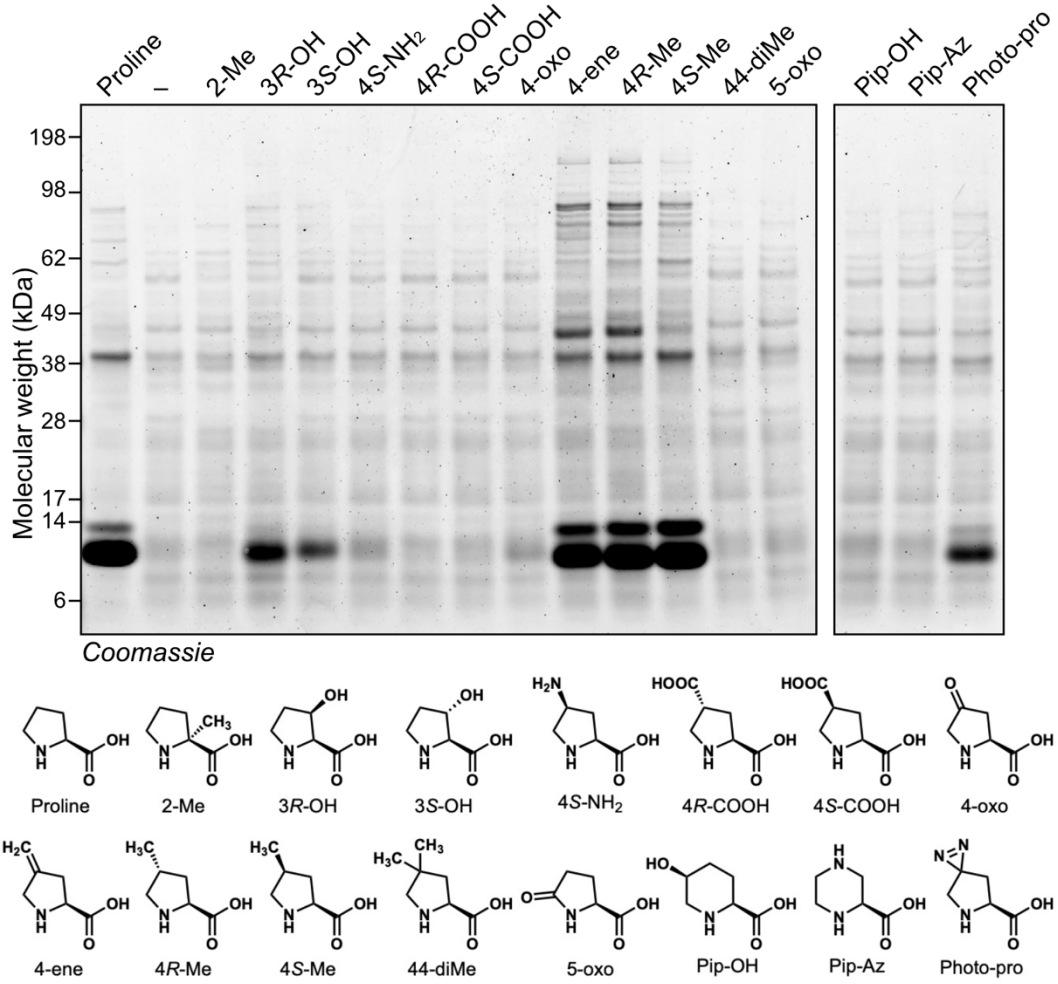

Figure 2.S2. Purity of insulin samples assessed by SDS-PAGE. HPLC-purified insulin (a), Ins-4*R*-Me (b), Ins-4*S*-Me (c), and Ins-4ene were analyzed by SDS-PAGE to validate purity; shown are representative lanes corresponding to individual HPLC fractions.

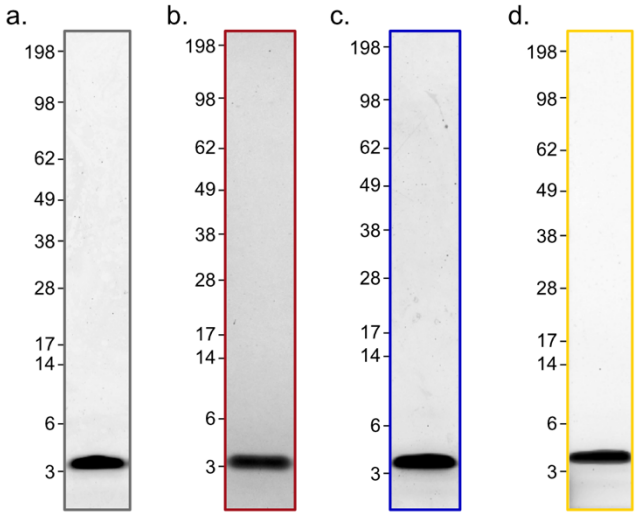

Figure 2.S3. Circular dichroism controls. At 60 µM, insulin is expected to exist as a dimer at pH 8, monomer in 20% ethanol, and denatured in the presence of 8 M guanidinium chloride. These spectra are overlaid with equilibrium spectra collected before and after dilution for kinetic CD measurements. Spectra below 210-215 nm were omitted for some samples due to high levels of buffer absorbance at these wavelengths.

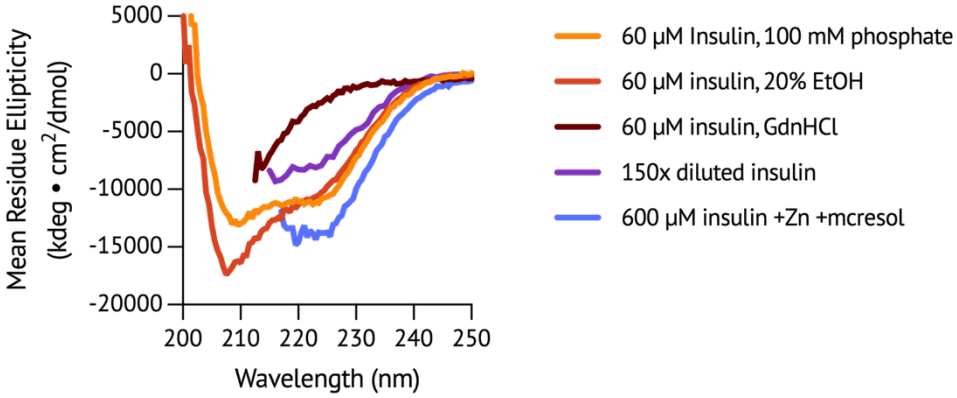

Figure 2.S4. Models of ins-4*R*-Me and ins-4*S*-Me hexamers. 4*S*-Me and 4*R*-Me were modeled in the structures of the R_6 (a) and T_6 (b) insulin hexamers (PDB ID: 1EV3 & 1MSO, respectively). Atoms near to each methyl substituent are indicated; distance measurements are in Å.

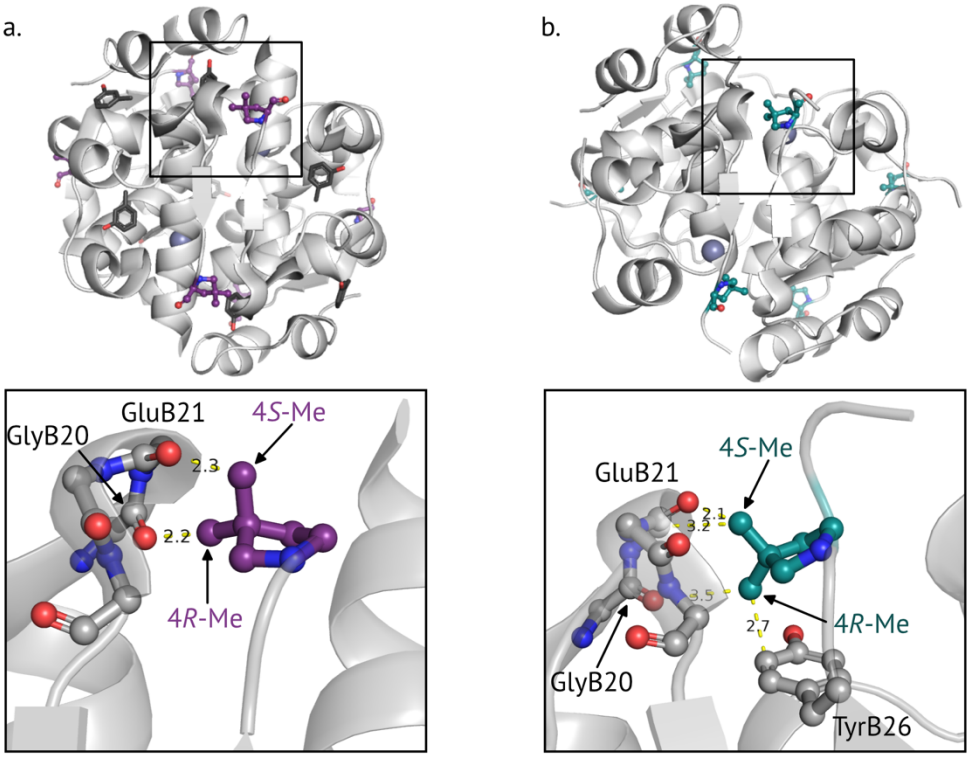

Figure 2.S5. Conformations of proline, 4-methyleneproline, and 3,4-dehydroproline. The equilibrium geometry conformations of protected versions of proline (a), 4ene (b), and 3,4-dehydroproline (c) were calculated (B3LYP/G-31+G**). The pseudorotation parameters amplitude (A) and phase angle (P)[55] are indicated for each structure. Amplitude corresponds to the degree of puckering for each proline, and phase angle represents puckering geometry. The *endo* (P~198°) and *exo* (P~18°) ring puckers of proline are nearly isoenergetic and rapidly interconvert.[54] More notable in this case is the puckering amplitude, which decreases with the addition of sp^2 hybridized carbon atoms.

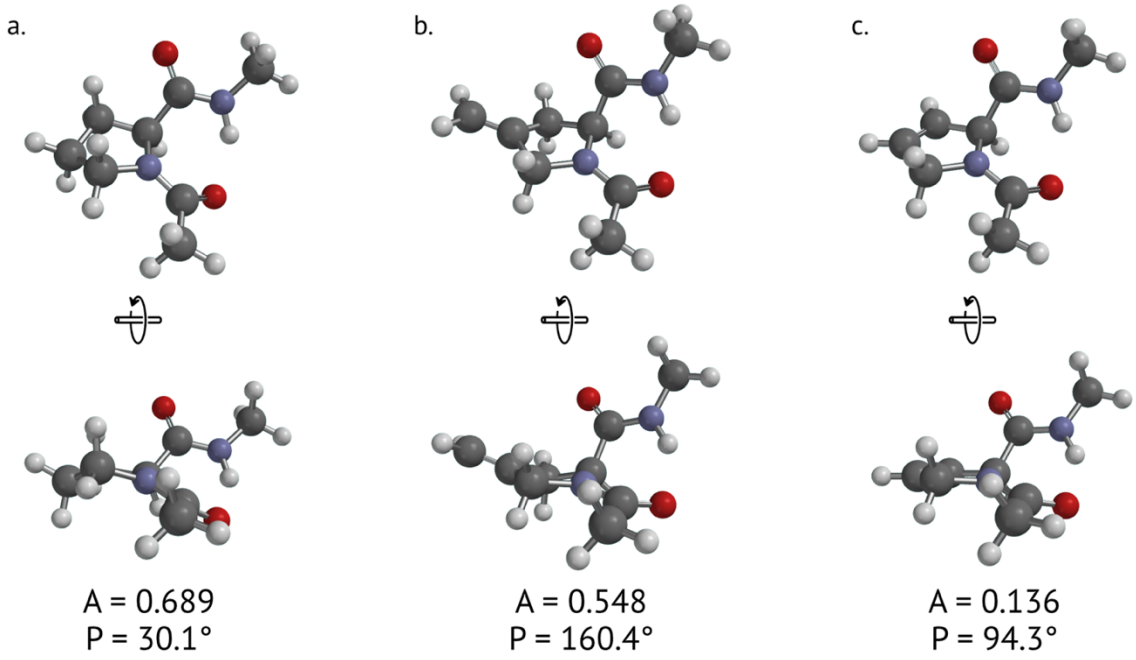

Table 2.S1. Incorporation of non-canonical proline residues

ncPro	Expected m/z	Observed m/z[‡]	Expected Δ m/z[#]	Observed Δ m/z[‡#]	Incorporation efficiency[‡]
Proline	1557.90	1557.19 ± 0.02	–	–	–
2-Me	1571.91	n.d.[*]	14.01	–	–
3*R*-OH	1573.88	1573.184 ± 0.006	15.98	15.5 ± 0.7	0.672 ± 0.004
3*S*-OH	1573.88	1573.184 ± 0.008	15.98	15.990 ± 0.005	0.537 ± 0.002
4*S*-NH_2	1572.90	1572.191 ± 0.001	15.00	14.990 ± 0.006	0.17 ± 0.03
4*R*-COOH	1601.89	n.d.	43.99	–	–
4*S*-COOH	1601.89	n.d.	43.99	–	–
4-oxo	1571.87	1571.182 ± 0.009	13.97	13.974 ± 0.009	0.15 ± 0.01
4-ene	1569.89	1569.218 ± 0.008	11.99	11.971 ± 0.001	0.926 ± 0.002
4*R*-Me	1571.91	1571.241 ± 0.006	14.01	13.90 ± 0.02	0.853 ± 0.004
4*S*-Me	1571.91	1571.26 ± 0.01	14.01	14.017 ± 0.004	0.775 ± 0.007
44-diMe	1585.93	n.d.	28.03	–	–
5-oxo	1571.87	n.d.	13.97	–	–
Pip-OH	1587.91	n.d.	30.01	–	–
Pip-Az	1572.90	n.d.	15.00	–	–
Photo-pro[%]	528.61	528.61	8.66	8.67	0.373

[‡]Average ± standard deviation of two technical replicates

[#]Mass shift compared to the proline-containing peptide present in the spectrum

[*]n.d., not detected

[%]Digested photo-proline peptide was analyzed by LC-ESI-MS, due to diazirine photolysis during MALDI-TOF analysis. We quantified the $[M+3H]^{+3}$ ion for the proline and ncPro-containing peptides. We also note the presence (27%; m/z = 519.3) of a 3,4-dehydroproline-containing peptide that likely results from photo-pro degradation within the culture, followed by co-translational incorporation (see Ch. 4). The incorporation efficiency for photo-pro reported here is calculated with respect to the proline and dehydroproline ions.

Table 2.S2. Mass spectrometry characterization of insulin variants

	Digested peptide			Mature insulin	
Protein	**Expected m/z**	**Observed m/z‡**	**Incorporation efficiency‡**	**Expected m/z**	**Observed m/z‡**
Insulin	1557.78	1557.43 ± 0.06	–	5808.6	5807.7 ± 0.4
Ins-4*R*-Me	1571.86	1571.83 ± 0.03	0.901 ± 0.010	5822.6	5822.2 ± 0.3
Ins-4*S*-Me	1571.86	1571.70 ± 0.03	0.896 ± 0.002	5822.6	5822.34 ± 0.01
Ins-4ene	1569.84	1569.56 ± 0.02	0.9417 ± 0.0005	5820.6	5821.2 ± 0.3

‡Average ± standard deviation of two technical replicates

Table 2.S3. Expression conditions and insulin yields

Protein	**ProRS variant**	**[ncPro] (mM)**	**[NaCl] (M)**	**Approx. proinsulin yield (mg L^{-1})‡**	**Approx. mature insulin yield (mg L^{-1})**
Insulin*	wt	–	–	35	8.0
Ins-4*R*-Me	C443G	0.5	0.5	29	3.9
Ins-4*S*-Me	wt	0.5	0.5	53	7.9
Ins-4ene	M157Q	1.0	0.5	23	1.5

‡Yields determined by measuring absorbance (280 nm) after proinsulin refolding and Ni-NTA enrichment.

*Expressed in terrific broth (TB)

Table 2.S4. Summary of insulin variant characterization

Protein	**Ellipticity ratio (208/222 nm)‡**	**Sedimentation coefficient‡**	**Fibrillation lag time (h)‡**	**Hexamer dissociation $t_{1/2}$ (s)**	**ANS emission maximum (nm)***
Insulin	1.24 ± 0.03	1.6S	16.6 ± 4.1	18.4 ± 2.8	470 ± 5
Ins-4*R*-Me	1.37 ± 0.05	1.3S	15.0 ± 3.5	9.8 ± 1.8	465 ± 3
Ins-4*S*-Me	1.19 ± 0.06	1.4S	12.5 ± 2.5	9.9 ± 2.5	466 ± 6
Ine-4ene	1.32 ± 0.05	1.3S	8.2 ± 4.0	17.0 ± 2.3	450 ± 11

‡60 μM insulin, 100 mM phosphate buffer, pH 8.0

*1 μM insulin, 5 μM ANS, 100 mM phosphate buffer, pH 8.0

Table 2.S5. Sedimentation coefficients of insulin and ins-4*S*-Me

Protein	Formulation[*]	Dilution[‡]	Ligand-free[%]
Insulin	3.3S	1.1S	2.4S
Ins-4*S*-Me	3.4S	1.1S	3.1S

[*]300 μM insulin, 125 μM $ZnCl_2$, 12.5 mM *m*-cresol, 25 mM tris, pH 8.0

[‡]4 μM insulin, 1.7 μM $ZnCl_2$, 167 μM *m*-cresol, 25 mM tris, pH 8.0

[%]300 μM insulin, 25 mM tris, pH 8.0

2.8 References

(1) Shoulders, M. D.; Raines, R. T. Collagen Structure and Stability. *Annu Rev Biochem* **2009**, *78*, 929–958. https://doi.org/10.1146/annurev.biochem.77.032207.120833.

(2) Ganguly, H. K.; Basu, G. Conformational Landscape of Substituted Prolines. *Biophys Rev* **2020**, *12*, 25–39. https://doi.org/10.1007/s12551-020-00621-8.

(3) Lummis, S. C. R.; Beene, D. L.; Lee, L. W.; Lester, H. A.; Broadhurst, R. W.; Dougherty, D. A. *Cis–Trans* Isomerization at a Proline Opens the Pore of a Neurotransmitter-Gated Ion Channel. *Nature* **2005**, *438*, 248–252. https://doi.org/10.1038/nature04130.

(4) Kim, W.; Hardcastle, K. I.; Conticello, V. P. Fluoroproline Flip-Flop: Regiochemical Reversal of a Stereoelectronic Effect on Peptide and Protein Structures. *Angew Chem Int Ed* **2006**, *45*, 1841–8145. https://doi.org/10.1002/anie.200603227.

(5) 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* **2004**, *5* (7), 928–936. https://doi.org/10.1002/cbic.200400052.

(6) Kim, W.; McMillan, R. A.; Snyder, J. P.; Conticello, V. P. A Stereoelectronic Effect on Turn Formation Due to Proline Substitution in Elastin-Mimetic Polypeptides. *J Am Chem Soc* **2005**, *127* (51), 18121–18132. https://doi.org/10.1021/ja054105j.

(7) Holmgren, S. K.; Taylor, K. M.; Bretscher, L. E.; Raines, R. T. Code for Collagen's Stability Deciphered. *Nature* **1998**, *392*, 666–667. https://doi.org/10.1038/33573.

(8) Torbeev, V. Y.; Hilvert, D. Both the *Cis-Trans* Equilibrium and Isomerization Dynamics of a Single Proline Amide Modulate β2-Microglobulin Amyloid Assembly. *Proc Natl Acad Sci* **2013**, *110* (50), 20051–20056. https://doi.org/10.1073/pnas.1310414110.

(9) Torbeev, V.; Ebert, M. O.; Dolenc, J.; Hilvert, D. Substitution of Proline32 by α-Methylproline Preorganizes β2-Microglobulin for Oligomerization but Not for Aggregation into Amyloids. *J Am Chem Soc* **2015**, *137* (7), 2524–2535. https://doi.org/10.1021/ja510109p.

(10) Breunig, S. L.; Tirrell, D. A. Incorporation of Proline Analogs into Recombinant Proteins Expressed in *Escherichia coli*. *Methods Enzymol* **2021**, *656*, 545–571. https://doi.org/10.1016/BS.MIE.2021.05.008.

(11) Haeusler, R. A.; McGraw, T. E.; Accili, D. Biochemical and Cellular Properties of Insulin Receptor Signalling. *Nat Rev Mol Cell Biol* **2018**, *19*, 31–44. https://doi.org/10.1038/nrm.2017.89.

(12) Goeddel, D. V.; Kleid, D. G.; Bolivar, F.; Heyneker, H. L.; Yansura, D. G.; Crea, R.; Hirose, T.; Kraszewski, A.; Itakura, K.; Riggs, A. D. Expression in *Escherichia coli* of Chemically Synthesized Genes for Human Insulin. *Proc Natl Acad Sci* **1979**, *76* (1), 106–110. https://doi.org/10.1073/pnas.76.1.106.

(13) Rakatzi, I.; Ramrath, S.; Ledwig, D.; Dransfeld, O.; Bartels, T.; Seipke, G.; Eckel, J. A Novel Insulin Analog With Unique Properties: LysB3 ,GluB29 Insulin Induces

Prominent Activation of Insulin Receptor Substrate 2, but Marginal Phosphorylation of Insulin Receptor Substrate 1. *Diabetes* **2003**, *52*, 2227–2238. https://doi.org/10.2337/diabetes.52.9.2227.

(14) Brange, J.; Ribel, U.; Hansen, J. F.; Dodson, G.; Hanse, M. T.; Havelund, S.; Melberg, S. G.; Norris, F.; Norris, K.; Snel, L.; Sørensen, A. R.; Voigt, H. O. Monomeric Insulins Obtained by Protein Engineering and Their Medical Implications. *Nature* **1988**, *333*, 679–682. https://doi.org/10.1038/333679a0.

(15) Brems, D. N.; Alter, L. A.; Beckage, M. J.; Chance, R. E.; Dimarchi, R. D.; Green, L. K.; Long, H. B.; Pekar, A. H.; Shields, J. E.; Frank, B. H. Altering the Association Properties of Insulin by Amino Acid Replacement. *Protein Eng* **1992**, *5* (6), 527–533. https://doi.org/10.1093/protein/5.6.527.

(16) Lepore, M.; Pampanelli, S.; Fanelli, C.; Porcellati, F.; Bartocci, L.; Di Vincenzo, A.; Cordoni, C.; Costa, E.; Brunetti, P.; Bolli, G. B. Pharmacokinetics and Pharmacodynamics of Subcutaneous Injection of Long-Acting Human Insulin Analog Glargine, NPH Insulin, and Ultralente Human Insulin and Continuous Subcutaneous Infusion of Insulin Lispro. *Diabetes* **2000**, *49*, 2142–2148. https://doi.org/10.2337/diabetes.49.12.2142.

(17) Mathieu, C.; Gillard, P.; Benhalima, K. Insulin Analogues in Type 1 Diabetes Mellitus: Getting Better All the Time. *Nat Rev Endocrinol* **2017**, *13*, 385–399. https://doi.org/10.1038/nrendo.2017.39.

(18) Bolli, G. B.; Di Marchi, R. D.; Park, G. D.; Pramming, S.; Koivisto, V. A. Insulin Analogues and Their Potential in the Management of Diabetes Mellitus. *Diabetologia* **1999**, *42* (10), 1151–1167. https://doi.org/10.1007/s001250051286.

(19) Owens, D.; Vora, J. Insulin Aspart: A Review. *Expert Opin Drug Metab Toxicol* **2006**, *2* (5), 793–804. https://doi.org/10.1517/17425255.2.5.793.

(20) Holleman, F.; Hoekstra, J. B. L. Insulin Lispro. *N Engl J Med* **1997**, *337* (3), 176–183. https://doi.org/10.1056/NEJM199707173370307.

(21) Bakaysa, D. L.; Radziuk, J.; Havel, H. A.; Brader, M. L.; Li, S.; Dodd, S. W.; Beals, J. M.; Pekar, A. H.; Brems, D. N. Physiochemical Basis for the Rapid Time-Action of LysB28ProB29-Insulin: Dissociation of a Protein-Ligand Complex. *Protein Sci* **1996**, *5* (12), 2521–2531. https://doi.org/10.1002/pro.5560051215.

(22) Brange, J.; Langkjaer, L. Insulin Structure and Stability. In *Stability and Characterization of Protein and Peptide Drugs: Case Histories*; Wang, Y. J., Pearlman, R., Eds.; Plenum Press: New York, 1993; pp 315–350. https://doi.org/10.1007/978-1-4899-1236-7_11.

(23) Brange, J.; Havelund, S.; Hougaard, P. Chemical Stability of Insulin. 2. Formation of Higher Molecular Weight Transformation Products During Storage of Pharmaceutical Preparations. *Pharm Res* **1992**, *9* (6), 727–734. https://doi.org/10.1023/a:1015887001987.

(24) Brange, J.; Langkjær, L.; Havelund, S.; Vølund, A. Chemical Stability of Insulin. 1. Hydrolytic Degradation During Storage of Pharmaceutical Preparations. *Pharm Res* **1992**, *9* (6), 715–726. https://doi.org/10.1023/a:1015835017916.

(25) Weiss, M. A. Design of Ultra-Stable Insulin Analogues for the Developing World. *J Health Specialties* **2013**, *1* (2), 59–70. https://doi.org/10.4103/1658-600X.114683.

(26) Kerr, D.; Morton, J.; Whately-Smith, C.; Everett, J.; Begley, J. P. Laboratory-Based Non-Clinical Comparison of Occlusion Rates Using Three Rapid-Acting Insulin Analogs in Continuous Subcutaneous Insulin Infusion Catheters Using Low Flow Rates. *J Diabetes Sci Technol* **2008**, *2* (3), 450–455. https://doi.org/10.1177/193229680800200314.

(27) Fang, K. Y.; Lieblich, S. A.; Tirrell, D. A. Replacement of ProB28 by Pipecolic Acid Protects Insulin against Fibrillation and Slows Hexamer Dissociation. *J Polym Sci A Polym Chem* **2019**, *57* (3), 264–267. https://doi.org/10.1002/pola.29225.

(28) Lieblich, S. A.; Fang, K. Y.; Cahn, J. K. B.; Rawson, J.; LeBon, J.; Teresa Ku, H.; Tirrell, D. A. 4*S*-Hydroxylation of Insulin at ProB28 Accelerates Hexamer Dissociation and Delays Fibrillation. *J Am Chem Soc* **2017**, *139*, 8384–8387. https://doi.org/10.1021/jacs.7b00794.

(29) Pocker, Y.; Biswas, S. B. Conformational Dynamics of Insulin in Solution. Circular Dichroic Studies. *Biochemistry* **1980**, *19* (22), 5043–5049. https://doi.org/10.1021/bi00563a017.

(30) Antolíková, E.; Žáková, L.; Turkenburg, J. P.; Watson, C. J.; Hančlová, I.; Šanda, M.; Cooper, A.; Kraus, T.; Brzozowski, A. M.; Jiráček, J. Non-Equivalent Role of Inter- and Intramolecular Hydrogen Bonds in the Insulin Dimer Interface. *J Biol Chem* **2011**, *286* (42), 36968–36977. https://doi.org/10.1074/jbc.M111.265249.

(31) Plum, A.; Agersø, H.; Andersen, L. Pharmacokinetics of the Rapid-Acting Insulin Analog, Insulin Aspart, in Rats, Dogs, Pigs, and Pharcamodynamics of Insulin Aspart in Pigs. *Drug Metab Dispos* **2000**, *28*, 155–160.

(32) Scapin, G.; Dandey, V. P.; Zhang, Z.; Prosise, W.; Hruza, A.; Kelly, T.; Mayhood, T.; Strickland, C.; Potter, C. S.; Carragher, B. Structure of the Insulin Receptor-Insulin Complex by Single-Particle Cryo-EM Analysis. *Nature* **2018**, *556* (7699), 122–125. https://doi.org/10.1038/nature26153.

(33) Menting, J. G.; Whittaker, J.; Margets, M. B.; Whittaker, L. J.; Kong, G. K.-W.; Smith, B. J.; Watson, C. J.; Žáková, L.; Kletvíková, E.; Jiracek, J.; Chan, S. J.; Steiner, D. F.; Dodson, G. G.; Brzozowski, A. M.; Weiss, M. A.; Ward, C. W.; Lawrence, M. C. How Insulin Engages Its Primary Binding Site on the Insulin Receptor. *Nature* **2013**, *493*, 241–245. https://doi.org/10.1038/nature11781.

(34) Jansen, R.; Dzwolak, W.; Winter, R. Amyloidogenic Self-Assembly of Insulin Aggregates Probed by High Resolution Atomic Force Microscopy. *Biophys J* **2005**, *88*, 1344–1353. https://doi.org/10.1529/biophysj.104.048843.

(35) Nielsen, L.; Frokjaer, S.; Brange, J.; Uversky, V. N.; Fink, A. L. Probing the Mechanism of Insulin Fibril Formation with Insulin Mutants. *Biochemistry* **2001**, *40* (28), 8397–8409. https://doi.org/10.1021/bi0105983.

(36) Jiménez, J. L.; Nettleton, E. J.; Bouchard, M.; Robinson, C. V; Dobson, C. M.; Saibil, H. R.; Caspar, D. L. D. The Protofilament Structure of Insulin Amyloid Fibrils. *Proc Natl Acad Sci* **2002**, *99* (14), 9196–9201. https://doi.org/10.1073/pnas.1424599399.

(37) Akbarian, M.; Yousefi, R.; Farjadian, F.; Uversky, V. N. Insulin Fibrillation: Toward Strategies for Attenuating the Process. *Chem Commun* **2020**, *56* (77), 11354–11373. https://doi.org/10.1039/d0cc05171c.

(38) Semisotnov, G. V.; Rodionova, N. A.; Razgulyaev, O. I.; Uversky, V. N.; Gripas', A. F.; Gilmanshin, R. I. Study of the "Molten Globule" Intermediate State in Protein Folding by a Hydrophobic Fluorescent Probe. *Biopolymers* **1991**, *31* (1), 119–128. https://doi.org/10.1002/BIP.360310111.

(39) Goldman, J.; Carpenter, F. H. Zinc Binding, Circular Dichroism, and Equilibrium Sedimentation Studies on Insulin (Bovine) and Several of Its Derivatives. *Biochemistry* **1963**, *13* (22), 4566–4574. https://doi.org/10.1021/bi00719a015.

(40) Gast, K.; Schüler, A.; Wolff, M.; Thalhammer, A.; Berchtold, H.; Nagel, N.; Lenherr, G.; Hauck, G.; Seckler, R. Rapid-Acting and Human Insulins: Hexamer Dissociation Kinetics upon Dilution of the Pharmaceutical Formulation. *Pharm Res* **2017**, *34* (11), 2270–2286. https://doi.org/10.1007/s11095-017-2233-0.

(41) Smith, G. D.; Ciszak, E.; Magrum, L. A.; Pangborn, W. A.; Blessing, R. H. R_6 Hexameric Insulin Complexed with *m*-Cresol or Resorcinol. *Acta Cryst* **2000**, *56* (12), 1541–1548. https://doi.org/10.1107/S0907444900012749.

(42) Richards, J. P.; Stickelmeyer, M. P.; Flora, D. B.; Chance, R. E.; Frank, B. H.; DeFelippis, M. R. Self-Association Properties of Monomeric Insulin Analogs under Formulation Conditions. *Pharm Res* **1998**, *15* (9), 1434–1441. https://doi.org/10.1023/a:1011961923870.

(43) Pekar, A. H.; Frank, B. H. Conformation of Proinsulin. A Comparison of Insulin and Proinsulin Self-Association at Neutral PH. *Biochemistry* **1972**, *11* (22), 4013–4016. https://doi.org/10.1021/bi00772a001.

(44) Caporale, A.; Loughlin, J. O.; Ortin, Y.; Rubini, M. A Convenient Synthetic Route to (2*S*,4*S*)-Methylproline and Its Exploration for Protein Engineering of Thioredoxin. *Org Biomol Chem* **2022**. https://doi.org/10.1039/D2OB01011A.

(45) Shoulders, M. D.; Hodges, J. A.; Raines, R. T. Reciprocity of Steric and Stereoelectronic Effects in the Collagen Triple Helix. *J Am Chem Soc* **2006**, *128*, 8112–8113. https://doi.org/10.1021/ja061793d.

(46) Kang, Y. K.; Park, H. S. Conformational Preferences and *Cis–Trans* Isomerization of L-3,4-Dehydroproline Residue. *Peptide Sci* **2009**, *92* (5), 387–398. https://doi.org/10.1002/BIP.21203.

(47) Loosli, S.; Foletti, C.; Papmeyer, M.; Zürich, E. Synthesis of 4-(Arylmethyl)Proline Derivatives. *Synlett* **2019**, *30*, 508–510. https://doi.org/10.1055/s-0037-1611672.

(48) Deming, T. J.; Fournier, M. J.; Mason, T. L.; Tirrell, D. A. Biosynthetic Incorporation and Chemical Modification of Alkene Functionality in Genetically Engineered

Polymers. *J Macromol Sci A* **1997**, *34* (10), 2143–2150. https://doi.org/10.1080/10601329708010331.

(49) Bakaysa, D. L.; Radziuk, J.; Havel, H. A.; Brader, M. L.; Li, S.; Dodd, S. W.; Beals, J. M.; Pekar, A. H.; Brems, D. N. Physicochemical Basis for the Rapid Time-Action of LysB28ProB29-Insulin: Dissociation of a Protein-Ligand Complex. *Protein Sci* **1996**, *5*, 2521–2531.

(50) 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* **2011**, *151* (4), 350–356. https://doi.org/10.1016/j.jbiotec.2010.12.023.

(51) He, W.; Fu, L.; Li, G.; Andrew Jones, J.; Linhardt, R. J.; Koffas, M. Production of Chondroitin in Metabolically Engineered *E. coli*. *Metab Eng* **2015**, *27*, 92–100. https://doi.org/10.1016/J.YMBEN.2014.11.003.

(52) Demeler, B.; Gorbet, G. E. Analytical Ultracentrifugation Data Analysis with Ultrascan-III. *Analytical Ultracentrifugation: Instrumentation, Software, and Applications* **2016**, 119–143. https://doi.org/10.1007/978-4-431-55985-6_8.

(53) Demeler, B.; Van Holde, K. E. Sedimentation Velocity Analysis of Highly Heterogeneous Systems. *Anal Biochem* **2004**, *335* (2), 279–288. https://doi.org/10.1016/J.AB.2004.08.039.

(54) Park, H. S.; Kang, Y. K. Puckering Transition of the Proline Residue along the Pseudorotational Path: Revisited. *New J Chem* **2021**, *45* (22), 9780–9793. https://doi.org/10.1039/D1NJ01361K.

(55) Hudáky, I.; Baldoni, H. A.; Perczel, A. Peptide Models XXXVIII. Proline Conformers from X-Ray Crystallographic Database and from Ab Initio Computations. *J Mol Struct Theochem* **2002**, *582* (1–3), 233–249. https://doi.org/10.1016/S0166-1280(01)00796-5.