Computational Insights into the Allosteric Behavior of Mini-Insulin in Receptor Binding Driven by C-Peptide Mobility

Esra Ayan[1,2,3*]

[1] Center for Life Sciences and Technologies, Bogazici University, 34684, Istanbul, Türkiye

[2] Research Institute for Health Sciences and Technologies, Istanbul Medipol University, 34674, Istanbul, Türkiye

[3] Omica Research Innovation Consulting Testing Service Industry and Trade Ltd. Co., Istanbul, Turkey

* To whom correspondence should be addressed. Email: esrayan91@gmail.com

Abstract

Background/aim: The production of recombinant insulin remains challenging, particularly in enhancing refolding efficiency and bioactivity. Mini-proinsulin analogs, which involve reducing the length of the C-peptide, offer potential improvements in insulin production. This study aims to evaluate mini-proinsulin analogs' design and receptor binding dynamics to optimize recombinant insulin production in *E. coli*.

Materials and methods: Mini-proinsulin analogs were engineered by replacing the 33-residue C-peptide with a pentapeptide sequence to improve refolding. The three-dimensional structure of mini-proinsulin was predicted using AlphaFold and performed docking analysis of mini-analogs to the insulin receptor using AutoDock Tools, with comparisons made to previously available NMR-determined analog and the native insulin-insulin receptor complex. Normal Mode Analyses (GNM and ANM) were performed in detail to assess binding dynamics.

Results: *In silico* analyses revealed that mini-proinsulin analogs closely replicate the structural features of native insulin and display receptor binding dynamics similar to native insulin, though they follow distinct receptor interaction paths.

Conclusion: All analysis suggests that C-peptide mobility may contribute to the allosteric behavior observed in mini-proinsulin analogs during receptor interaction.

Key words: Mini-proinsulin analogs; Recombinant insulin production; Allosteric behavior.

1. Introduction

Insulin, a polypeptide hormone secreted by pancreatic β-cells, comprises A and B chains. The chains are connected by two interchain and an intrachain disulfide bond (Ayan et al., 2022). The hormone is initially produced as a single-chain precursor, preproinsulin, which undergoes proteolytic cleavage in the pancreas to forge its active form. In the human body, preproinsulin has signal peptide, B-peptide, C-peptide and A-peptide. However, in the recombinant production, preproinsulin has fusion tag, B-peptide, C-peptide and A-peptide. In the absence of signal peptide in human body or without fusion tag in recombinant production, insulin is termed as proinsulin (Ayan et al., 2022)

Escherichia coli (*E. coli*) is typically employed in the production of recombinant human insulin, where preproinsulin is synthesized in the bacterial cytoplasm and accumulates as inclusion bodies within a fusion protein. Following the CNBr administration to the fusion product and sulfonation of the released proinsulin, the refolding process of S-sulfonated proinsulin can proceed with relatively high efficiency (Siew et al., 2021). After tryptic digestion of proinsulin, mature insulin, which consists of B-peptide and A-peptide alone, is produced, releasing the C-peptide, whose length varies between 26 and 38 amino acid residues across different animal species (Gross et al., 1989). Mature insulin is vital for regulating blood glucose levels, while the C-peptide, co-secreted with insulin, exhibits biological activity, including its ability to repair the diminished Na^+-K^+ dependent ATPase activity observed in diabetic conditions (Ido et al., 1997; Steiner et al., 1997).

Proinsulin displays only 2% of the receptor binding capacity compared to mature insulin (Chang et al., 1998). Conversely, an insulin analog that was directly cross-linked

between the C-terminus of the B chain and the N-terminus of the A chain exhibited a near-total loss of bioactivity despite its crystal structure being nearly indistinct from that of native insulin (Derewenda et al., 1991; Hua et al., 1998). It indicates that the length of the cross-link is critical in determining insulin's bioactivity. Still, a study conducted by Brems et al. (1991) systematically investigated the effect of cross-linking between the B and A chains, revealing that a specific cross-link length is critical for maintaining bioactivity (Brems et al., 1991). Additionally, site-directed mutagenesis research has indicated that keeping secondary structural elements is crucial for insulin’s biological function (Pittman IV et al., 1997). The conclusions reveal that the residues responsible for receptor binding are localized on both the B and A chains. Insulin has activity only when it possesses sufficient flexible C-peptide to adopt an optimal orientation, allowing the interacting residues to align correctly with the receptor.

The three-dimensional structure of insulin indicates that a peptide linker could connect the A and B chains significantly shorter than the 33-residue C-peptide (Cho et al., 2000). However, it remains novel whether reducing the length of the C-peptide would enable the resulting mini-proinsulin to exhibit receptor interaction dynamics comparable to those of mature insulin. In related research, mini proinsulin was engineered to enhance the refolding efficacy of the insulin precursor and was also facilitated recombinant production by increasing productivity in *E. coli* (Chang et al., 1998). These parameters are still challenging for recombinant insulin production in industrial applications. The findings revealed a 20–40% improvement of mini proinsulin use in refolding efficiency compared to standard proinsulin. Furthermore, the mini proinsulin exhibited substantial receptor binding activity, at least 50% compared to native insulin (Chang et al., 1998; Cho et al., 2000).

Here, the 33-residue C-peptide is replaced with a pentapeptide sequence designed to form a short turn, thereby promoting refolding. The novel designer mini proinsulin, P28B, is substituted by D28B to minimize aggregate formation *in solution.* Additionally, the substitution of N60A for G60A is performed to enhance stability in acidic conditions. The three-dimensional structure of mini proinsulin predicted using AlphaFold (AF), has been perform of docking with the insulin receptor. Similarly, the mini proinsulin, whose structure had been previously determined through an NMR study in the literature (PDB ID: 1EFE), has been also performed of docking with the receptor. Both docking structures are compared with the experimental insulin-insulin receptor complex (PDB ID: 6VEP) (Xiong et al., 2020) using *in-silico* analyses. Previously available experimental results had demonstrated a high ratio of insulin moiety in the fusion protein as well as enhanced the refolding capacity of mini proinsulin (Chang et al., 1998). The *in-silico* analysis suggests that despite mini proinsulin analogs having nearly identical intrinsic dynamics post-binding mechanism to that of native insulin, they display allosteric behavior in receptor binding due to probably C-peptide mobility. Collectively, further well-optimized mini-proinsulin analogs might be used as final products to produce recombinant insulin in E. coli.

2. Materials and methods

2.1. Protein Structures

The AlphaFold prediction of novel designer mini-proinsulin structure (FVNQHLCGSHLVEALYLVCGEKGFFYTDKTRRYPGDVKRGIVEQCCTSICSLY QLENYCG) has been used for this study along with the mini-proinsulin structure from the Protein Data Bank (PDB) under the 1EFE accession code. Additionally, insulin-

insulin receptor structure from PDB (ID: 6VEP) has been used for this study to perform docking analyses between mini-proinsulin and insulin receptor as well as to compare receptor interaction dynamics between genuine insulin-insulin receptor and mini-proinsulin-insulin receptor. An unbound native insulin structure has been used to compare dynamicity in isolation between native insulin from PDB (ID: 3I40) and mini-proinsulin analogs.

2.2. Structure Prediction

The novel designer mini-proinsulin has been calculated as a monomer using AlphaFold (AF) with MMseqs2 v1.5.5. The structure prediction process followed the standard approach outlined in the AlphaFold paper (Jumper et al., 2021), which involves five key steps: multiple sequence alignment (MSA) construction, template search, inference with five models, ranking of models based on the mean predicted local distance difference test (pLDDT), and constrained relaxation of the predicted structures. For the novel mini-insulin, AF has been executed with the model preset set to "monomer" using the following command:

```
alphafold --fasta_paths=<fasta file path> --model_preset=<monomer> --output_dir=<output directory> --db_preset=full_dbs --use_gpu_relax --max_template_date=<max template date>
```

The structure with the highest prediction accuracy has been selected for further analysis. Molecular images have been generated by PyMol version 2.3.0 (Schrödinger, LLC, New York, NY, USA).

2.3. Docking analysis

AutoDock and AutoDock Tools (Huey et al., 2012) were used to optimize docking parameters in precision. The necessary coordinate files and related information were converted from PDB format to PDBQT format, which includes atom and bond types,

partial atomic charges, and other relevant data. After loading the insulin receptor structure, the target file was prepared by adding hydrogen atoms, correcting bond orders, and renumbering residues. Kollman charges, derived from quantum mechanical calculations, were applied to assign appropriate partial charges to the receptor. Similarly, the mini-proinsulin ligand's coordinate files were processed in PDBQT format, with hydrogen atoms added, bond orders corrected, and residues renumbered. Gasteiger charges were applied to the ligand to facilitate the docking procedure. A grid parameter file was prepared to calculate grid maps for interaction energies across various atom types, essential for calibrating the docking procedure. A grid box was generated to define the binding site coordinates for the search engine. A docking parameter file was created to execute the docking using AutoDock Tools. The following commands were used to run AutoGrid and AutoDock:

```
autogrid4.exe -p dock.gpf -l dock.glg &
autodock4.exe -p dock.dpf -l dock.dlg &
```

The resulting dlg file contained detailed information about the docking simulations, including the calculated binding energy in kcal/mol. The best conformations had binding energies of −6.35 kcal/mol for the nM2PI-IR co-complex and −4.46 kcal/mol for the M2PI-IR co-complex. Additionally, the root-mean-square deviations (RMSD) of atomic positions of nM2PI-IR and M2PI-IR co-complexes from the reference co-complex (PDB ID: 6VEP) were 0.329 Å and 0.312 Å, respectively.

2.4. Gaussian Network Model (GNM)

The Gaussian Network Model (GNM) analysis was performed using ProDy (Bakan et al., 2011). GNM analysis was completed on both mini-insulin variants and native insulin in isolation, as well as on docked co-complexes (nM2PI-IR and M2PI-IR) and

the native insulin receptor-insulin (IR-INS) co-complexes. For all structures, Cα atoms were operated to construct contact maps. The cutoff distance was set to 8.00 Å, and N − 1 nonzero normal modes were obtained, where N represents the number of Cα atoms. Cross-correlations between residue fluctuations were calculated based on the cumulative 10 slowest GNM modes. The squared fluctuations for each structure were calculated across the weighted 10 fastest modes to compare their local motions and dynamic behavior. Graphics illustrating the normalized results were generated using Matplotlib.

2.5. Anisotropic Network Model (ANM)

ANM (Anisotropic Network Model) analysis was conducted on the docking structures of nM2PI-IR, M2PI-IR, and the native INS-IR co-complex (PDB ID: 6VEP). A cutoff distance of 12.0 Å and the default spring constant of 1.0 were used. The ANM modes were focused on individual mode 5. The resulting protein motions were visualized using the NMWiz tool within the VMD software.

3. Results

3.1. Novel mini-insulin mimics the structural landscape of native insulin

The novel mini-proinsulin (hereafter referred to as nMPI2) was designed by reducing the C-peptide from 33 residues to a 5-residue mini-C peptide, along with substituting Arginine with Lysine at position 22 of Chain B (hereafter referred to as R22B, K22B) and Proline with Aspartic acid at position 28 (hereafter referred to as P28B, D28B). Additionally, Asparagine at position 60 of Chain A was replaced by Glycine (hereafter referred to as N60A, G60A). To obtain the three-dimensional structure of nMPI2 (Fig. 1A), an AF prediction was conducted, as described in the Methods section. The

experimentally determined mini proinsulin structure (hereafter referred to as MPI2) (PDB ID: 1EFE) (Fig. 1B), resolved by NMR techniques, was used to compare with the predicted nMPI2 structure, indicating a perfect superposition with a root-mean-square deviation (RMSD) of up to 1.2 Å. Furthermore, the structure of human insulin (hereafter referred to as INS) (PDB ID: 3I40), determined through X-ray crystallography, was used to compare the dynamic deviations between the mini proinsulin and native INS in isolation. A highly consistent RMSD matching up to 0.4 Å was observed between the mini-analogs and native INS (Fig. 1C).

Docking analysis, as described in the Methods section, was conducted to evaluate if the mini proinsulins could effectively bind to the insulin receptor. This analysis included a comparison with the experimental structure of the insulin-insulin receptor co-complex (hereafter referred to as IR-INS) (PDB ID: 6VEP) (Fig. 1D-F). According to the molecular docking results, nMPI2 exhibited a relatively higher affinity (-6.35) for the αCT domain of the insulin receptor (IR), whereas MPI2 showed a lower affinity (-4.46). Both mini-analogs demonstrated ideal alignment with the experimental IR-INS structure, with RMSD of 0.319 Å for the IR-nMPI2 co-complex and 0.312 Å for the IR-MPI2 co-complex. Electrostatic forces can give required information about protein-protein interactions and protein stability (Zhou et al., 2018). The insulin-binding region for each co-complex was surrounded by an acidic and neutral region concentrated with stable negatively charged and uncharged residues in the presence even of the charge substitutions in the nM2PI conformer. The electrostatic distribution here highlights areas where charged interactions, such as hydrogen bonds or ionic interactions, might occur between the nM2PI and the IR (Fig. 1G). The charge distribution might be crucial for receptor binding and overall affinity. However, the electrostatic distribution, while

similar, may show slight differences in charge regions compared to the IR-nM2PI co-complex, reflecting lower

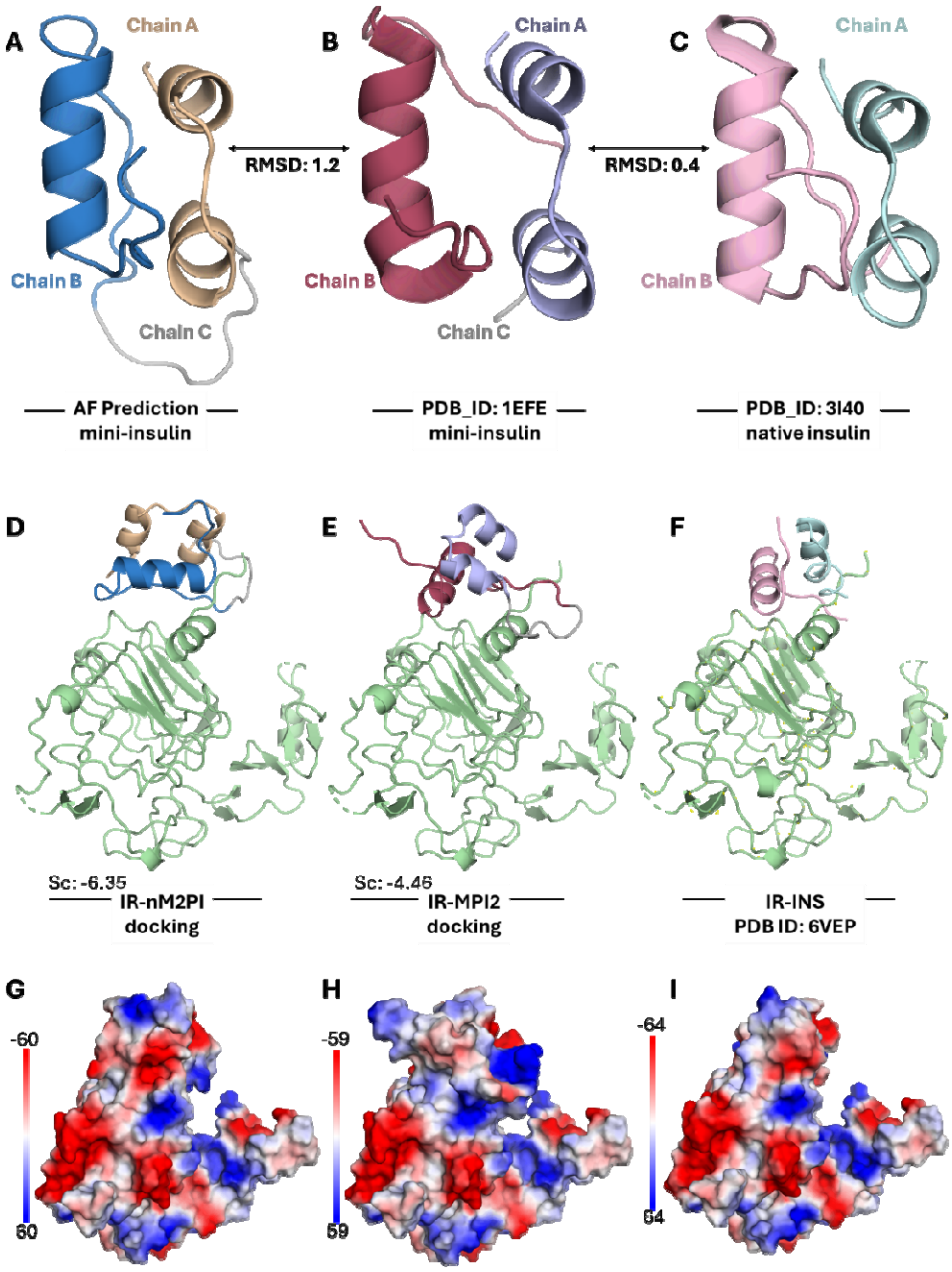

Figure 1. Structural representation of mini-analogs for calculated and experimental co-complexes. (A) AlphaFold prediction of the novel designer mini-proinsulin (nMPI2). (B) Experimental structure of mini-proinsulin from the Protein Data Bank (PDB ID: 1EFE), referred

to as MPI2. (C) Experimental structure of native insulin from the PDB (ID: 3I40). RMSD values between the structures are indicated with bidirectional arrows. (D) Docking co-complex of the experimental insulin receptor (IR, shown in green) and nMPI2, indicating a relatively higher affinity score. (E) Docking co-complex of the experimental IR (green) and MPI2, indicating a relatively lower affinity score. (F) Experimental IR-INS Cryo-EM co-complex from the PDB (ID: 6VEP). (G, H, I) Poisson-Boltzmann electrostatics of the docking and experimental co-complexes, illustrating the charge-smoothed protein contact potential. Red regions represent areas of negative electrostatic potential, blue regions indicate areas of positive electrostatic potential, and white (or neutral) regions represent areas with near-neutral electrostatic potential.

affinity or less optimized electrostatic interaction between IR and M2PI co-complex (Fig. 1H). Notably, certain blue and red areas seem less extensive or less aligned with the receptor, potentially explaining the lower binding affinity, which is highly compatible with the docking score of the IR-M2PI co-complex (Fig. 1E). The electrostatic potential map of the experimental IR-INS co-complex demonstrates their natural, highly optimized charge interaction (Fig 1I). The alignment and balance of positive and negative regions are more uniform and organized, likely indicating stronger, more stable interactions than the docked mini-analogs. Still, Panel G in Figure 1 illustrates that the IR- nM2PI co-complex attempts to mimic the natural electrostatic landscape of native insulin (Fig. 1I) but with varying degrees of success. It is essential that there are highly similar electrostatic charge distributions between IR-nM2PI and experimental co-complexes (Fig G and Fig. I), whereas potential surface charge redistributions may force shape perturbations of a protein by changing hydrogen bonds and salt bridges.

3.2. Mini proinsulins exhibit enhanced stability compared to the unbound state

Protein binding is determined by setting novel noncovalent interactions between specific residues of two proteins. These interactions can impact residues located at distant sites by

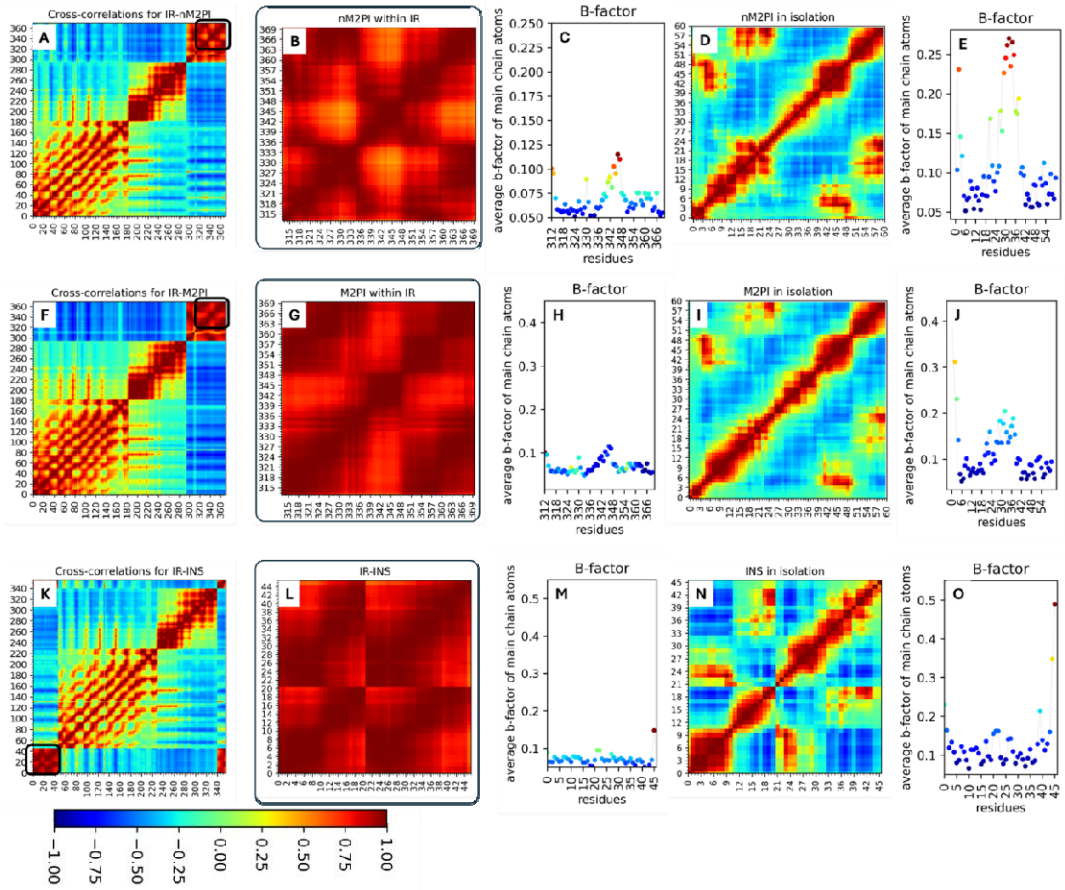

Figure 2. Cross-correlation and calculated B-factor by GNM comparisons between calculated and experimental co-complexes. (A) Cross-correlation for IR-nM2PI. Highlight square indicated docking insulin monomer' cross-correlation. Residue pairs exhibiting correlated motion, moving in the same direction, are depicted in red regions ($C_{ijorient} > 0$), while those exhibiting anticorrelated motion, moving in opposite directions, are represented in blue regions ($C_{ijorient} < 0$). Uncorrelated residue pairs are shown in green ($C_{ijorient} = 0.0$), as indicated

by the color bar on lower left corner. (B) Cross-correlation for nM2PI monomer within IR-nM2PI co-complex. (C) B-factor of nM2PI monomer within IR-nM2PI co-complex. (D) Cross-correlation for nM2PI monomer in isolation. (E) B-factor of nM2PI monomer in isolation. (F) Cross-correlation for IR-M2PI. (G) Cross-correlation for M2PI monomer within IR-M2PI co-complex. (H) B-factor of M2PI monomer within IR-M2PI co-complex. (I) Cross-correlation for M2PI monomer in isolation. (J) B-factor of M2PI monomer in isolation. (K) Cross-correlation for IR-INS. (L) Cross-correlation for INS monomer within IR-INS co-complex. (M) B-factor of INS monomer within IR-INS co-complex. (N) Cross-correlation for INS monomer in isolation. (O) B-factor of INS monomer in isolation.

disrupting current communication paths and creating new ones. The reorganization has the potential to significantly change the conventional residue communication network (Eren et al., 2021). Such modifications can be effectively identified and analyzed using the GNM analysis. Here we performed cross-correlations between residue fluctuations of weighted 10 slowest GNM modes for each case (IR-nM2PI; IR-M2PI; IR-INS) (Fig. 2). Cross-correlation for calculated IR-nM2PI co-complex indicated nM2PI almost behaves as a rigid block while binding to IR, all residues practically moving along the same mode axis (cross-correlations close to 1 (red)) (Fig. 2A-B), which is highly compatible with diminished B-factor (Fig. 2C) of sliced nM2PI (Fig. 2B) compared to greater mobile B-factor (Fig. 2E) of nM2PI in isolation (Fig. 2D). Namely, nM2PI in isolation reveal greater variations in the movement directions of various residues or groups of residues, with some moving in opposite directions and being separated by hinge regions (cross-correlations close to 1 (blue)) (Fig. 2D). Cross-correlation for calculated IR-M2PI co-complex also shows the similar characteristic with the calculated IR-nM2PI co-complex (Fig. 2F-J). However, experimental IR-INS co-complex

indicated more optimized and strongest correlation compared to Panel A and Panel F in Figure 2 (Fig. 2K-L). Additionally, native insulin exhibits the lowest mobility once binding to the IR, suggesting that the interaction is characterized by maximal affinity and structural stability (Fig. 2M) compared to unbound conformer (Fig. 2N-O).

Collectively, cross-correlation analyses show that native insulin provides the most robust and conformationally stable form while binding to IR (Fig. 2L-M). While both nM2PI and M2PI exhibit comparable correlation profiles, native insulin presents a more refined optimization of structural integrity. These observations are verified by B-factor analyses, where native insulin exhibits the lowest dynamic mobility post-receptor binding (Fig. 2M). Although the novel and alternative mini-insulin analogs have relatively higher residue mobility after receptor docking, they display markedly improved structural rigidity compared to their unbound conformer (Fig. 2B-D-G-I). Thus, native insulin displays the highest binding affinity and structural fidelity, while the mini-insulin analogs, although demonstrating a similar binding paradigm, do not perform the same level of stabilization and optimization, confirming mini-insulin demonstrates receptor binding efficacy at approximately 50% of the potency exhibited by native insulin (Chang et al., 1998).

3.3. ANM modes support nM2PI shows post-binding stability

The ANM analysis can illustrate that intrinsic residue fluctuations in C-peptide mobility drive the allosteric behavior of mini-proinsulin variants post-binding to IR. ANM provides insight into residue motion directions by calculating theoretical displacements of residues. The analysis utilized the global 5 slow modes, likely to drive these conformational changes. The direction of motions, derived from the eigenvectors of

Mode 5, was mapped onto the IR-INS, IR-M2PI, and IR-nM2PI co-complexes, with a color gradient representing the amplitude of the residue fluctuations. In the native IR-INS interaction, the correlation map reveals a well-coordinated residue motion pattern with distinct regions of coherent dynamics (Fig. 3A). The movements suggest that INS

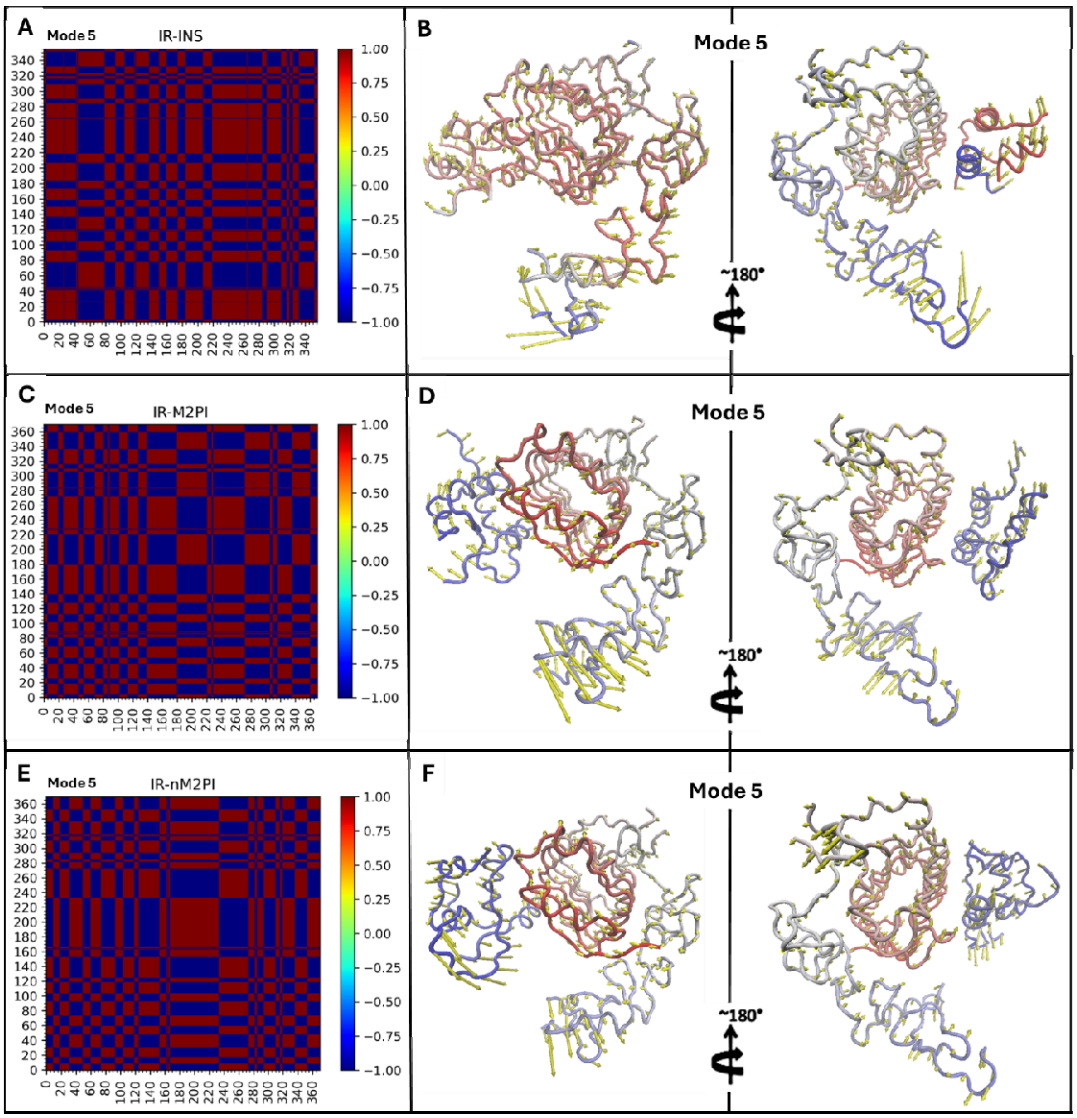

Figure 3. Comparison of individual 5 slow ANM mode direction of motions in INS, M2PI, and nM2PI bound to the IR. (A) Cross-correlations of IR-INS map by ANM. Residue pairs exhibiting correlated motion, moving in the same direction, are depicted in red regions ($C_{ijorient}$ > 0), while those exhibiting anticorrelated motion, moving in opposite directions, are represented

in blue regions ($C_{ijorient} < 0$) (B) Mode 5 of IR-INS is drawn on the structure. Ribbon representation was rendered blue to gray and red coloring, representing high to low mobility (C) Cross-correlations of IR-M2PI map by ANM. (D) Mode 5 of IR-M2PI is drawn on the structure. (E) Cross-correlations of IR-nM2PI map by ANM. (F) Mode 5 of IR-nM2PI is drawn on the structure.

displays highly optimized collective motions with the receptor, indicative of a well-tuned allosteric communication network captured by the ANM analysis. Instead, the IR-M2PI appears more disordered and less coordinated than the INS conformer (Fig. 3C). Yet, the correlation map of IR-nM2PI unveils relatively concerted and robust correlations than M2PI (Fig. 3E). The red and blue regions are more coherent, indicating relatively coordinated collective motions and a well-optimized dynamic interaction with the receptor, suggesting nM2PI favors a dynamic modulation closely similar to INS. The INS direction of motion within the IR-INS co-complex are more dominant in Mode 5 (Fig. 3B) compared to the IR-M2PI and IR-nM2PI co-complexes (Fig. 3D-F). Notably, Chain B in the M2PI and nM2PI analogs exhibits increased mobility, indicating less restricted dynamics in the region. Still, M2PI exhibits relatively weaker collective dynamics and more irregular motion patterns (Fig. 3D), suggesting reduced post-binding stability compared to INS conformer.

3.4. Local motions reveal unprecedented hotspot residues in mini proinsulins

GNM yield precise estimations of localized dynamics within the fast vibrational modes. Residues exhibiting peak fluctuations in the fast modes are potential hot loci and are structurally constrained and play pivotal roles in folding and ligand binding processes. Variations in residue motion profiles offer valuable insights into the protein's functional mechanisms and allosteric modulation. To assess and compare local dynamics, squared

fluctuations across the top 10 weighted fastest GNM modes were computed for both predicted co-complexes compared experimental IR-INS interactions (Fig. 4). Additionally, to assess collective motion between only αCT domain of IR, which is

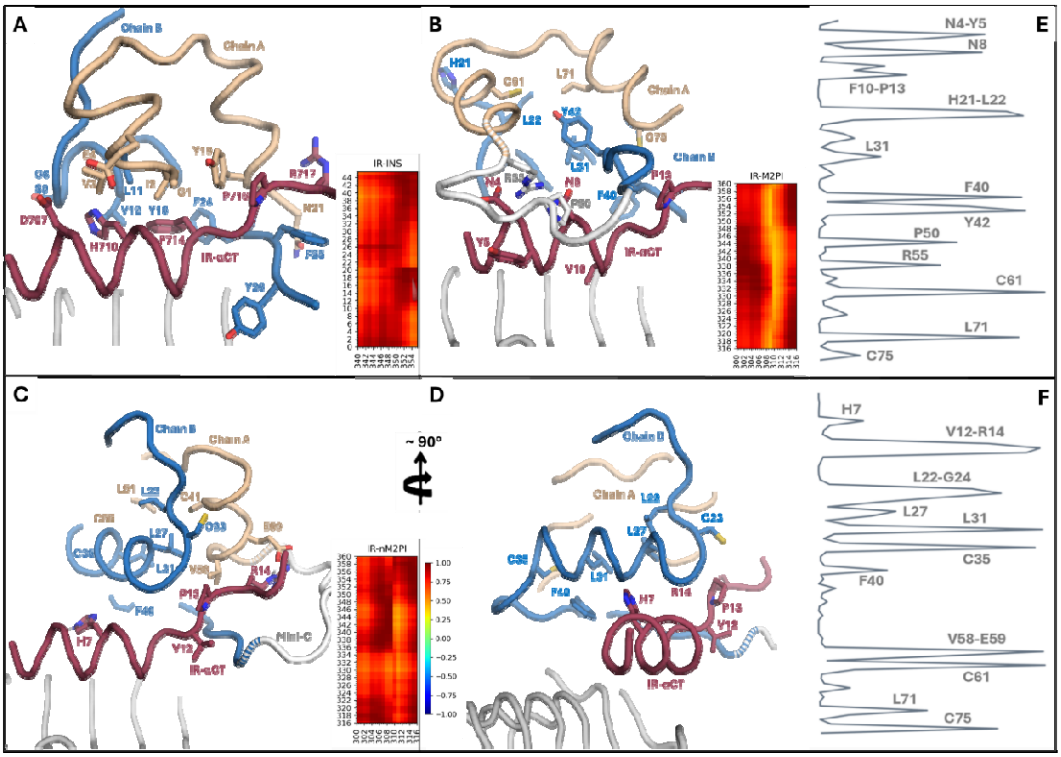

Figure 4. Structural localization of potential receptor interaction residues and their GNM analysis. (A) Native αCT-INS interaction residues adapted from PDB structure under the 6VEP accession code. Also, cross-correlation analysis between αCT domain and INS to emphasize their strong collective motion. (B) αCT-M2PI interaction residues determined by 10 fastest GNM modes and cross-correlation analysis between αCT domain and M2PI to emphasize their relatively strong collective motion. (C) αCT-nM2PI interaction residues determined by 10 fastest GNM modes and cross-correlation analysis between αCT domain and nM2PI to emphasize their strong collective motion. (D) Turn panel C to 90 degrees for inspecting the detailed interaction between the αCT- nM2PI complex. Residue pairs exhibiting correlated motion, moving in the same direction, are depicted in red regions ($C_{ijorient} > 0$), while those exhibiting anticorrelated motion, moving in opposite directions, are represented in blue regions

($C_{ijorient}$ < 0). Uncorrelated residue pairs are shown in green ($C_{ijorient}$ = 0.0), as indicated by the color bar in those panels. (E, F) Ten fastest GNM modes of αCT-mini analogs interactions. Potential hot loci have been labelled with relevant residues.

critical part of the insulin binding, and insulin variants (M2PI, nM2PI, and INS), performed cross-correlation over the 10 slowest GNM modes.

Compared to the experimental IR-INS interaction (Fig. 4A), the local dynamics between the αCT domain of the IR and the mini-analogs (Fig. 4B-F) reveal distinct variations within the binding region. Despite their overall similarity, notable differences are observed between the interaction processes of αCT-M2PI and αCT-nM2PI. Specifically, peaks at N4-Y5, F10-P13, H21, Y42, P50, and R55 are only to the αCT-M2PI complex (Fig. 4E), while peaks at H7, V12, R14, G24, L27, C35, and V58-E59 are present at the αCT-nM2PI interaction alone (Fig. 4F). Shared hotspots are identified at residues L22, L31, F40, C61, L71, and C75 in both mini-complexes, suggesting mini-proinsulin variants engage in allosteric behavior to provide a distinct kinetic stabilization strategy with IR, likely driven by the flexible mini-C peptide. Additionally, cross-correlation analysis was conducted to investigate the collective motion between the αCT domain of IR and the mini-analogs. Interestingly, the αCT domain exhibits strong correlations with the mini-proinsulin variants (Fig. 4B-C), closely similar to the native αCT-INS complex (Fig. 4A), suggesting that despite the mini-analogs adopting receptor interaction strategies distinct from that of native insulin, they still manage to obtain favorable and identical structural stability (Fig. 5). Regarding the inspection of the lowest frequency collective motions between the αCT domain of IR and the C-peptides of mini-analogs alone, coherent features intrinsically endowed by the rigid blocks within this complex are observed, especially between the aCT and the C-peptide of nM2PI more pronounced

(Fig. 5B). This allosteric regulation, again, due to probably the mobility of the C-peptides, might be highlighting the receptor interaction ability of mini-analogs to stabilize effectively, even when their methods of achieving such stability diverge from the native form.

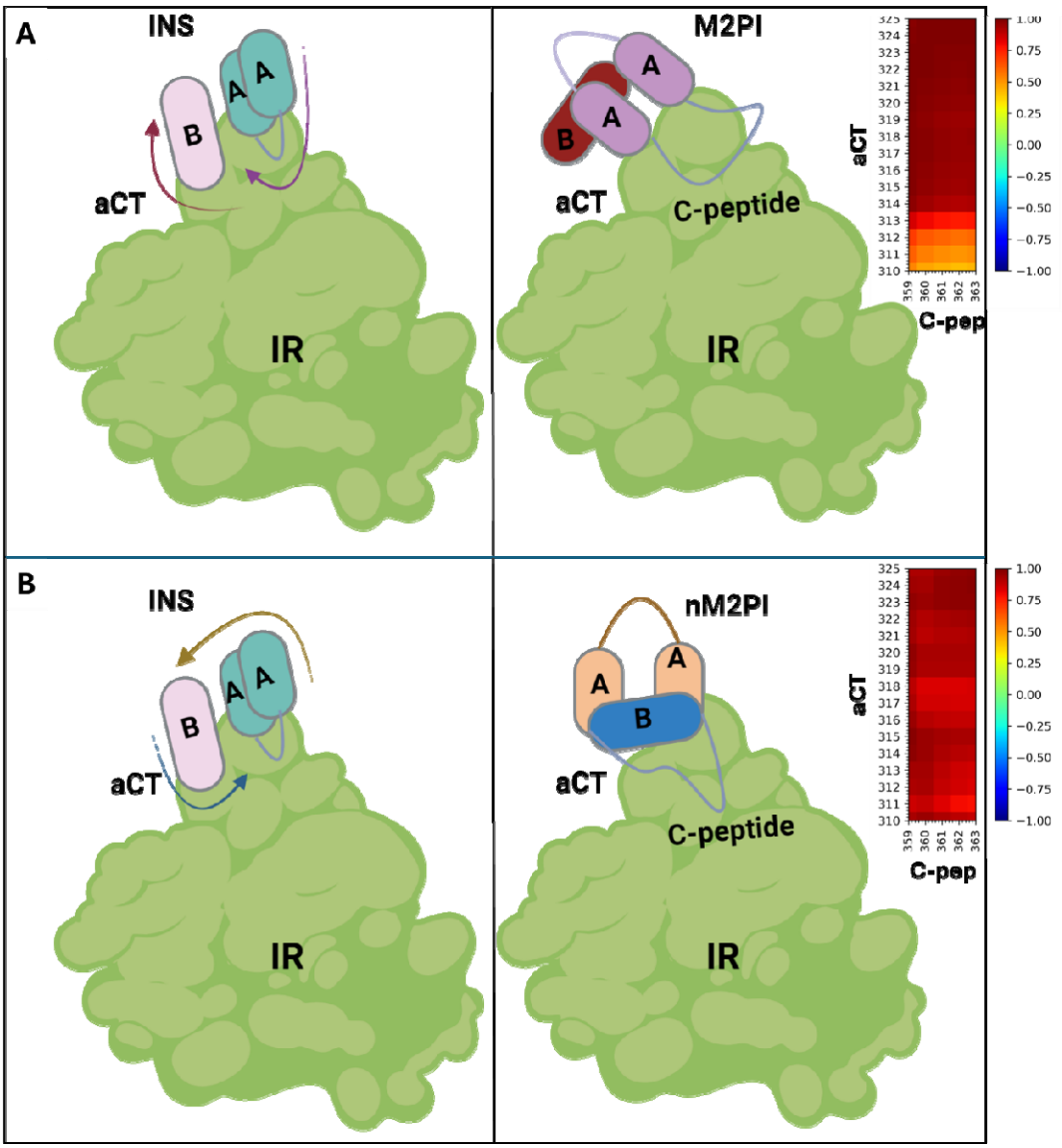

Figure 5. Pictorial representation the allosteric behavior of M2PI and nM2PI compared INS in receptor binding driven by C-peptide mobility. (A) Comparison of how the M2PI position alters its conformation after binding to the αCT receptor domain compared to the INS

position within the IR. Cross-correlation analysis by GNM between aCT and C-peptide within the IR-M2PI co-complex. N-terminal of aCT shows at least 50% correlation with C-peptide of IR-M2PI co-complex. (B) Comparison of how the nM2PI position alters its conformation after binding to the αCT receptor domain compared to the INS position within the IR. Cross-correlation analysis by GNM between aCT and C-peptide within the IR-nM2PI co-complex. aCT shows strong correlation with C-peptide of IR-nM2PI co-complex. Residue pairs exhibiting correlated motion, moving in the same direction, are depicted in red regions ($C_{ijorient} > 0$), while those exhibiting anticorrelated motion, moving in opposite directions, are represented in blue regions ($C_{ijorient} < 0$). Uncorrelated residue pairs are shown in green ($C_{ijorient} = 0.0$), as indicated by the color bar in right panels.

4. Discussion

The present study computationally models the dynamic interactions between mini-insulin analogs and the insulin receptor. Two key rationales underpin the cleaving of the C-peptide in conventional proinsulin. First, optimize the early and late stages of protein folding pathways (Matouschek et al., 1990). Second, predictions indicate that most dibasic processing sites are located in a β-turn adjacent to regions of α-helix or β-sheet (Rholam et al., 1986). Hence, introducing a β-turn between these dibasic sites is proposed to facilitate the enzymatic cleavage process required to release the mature insulin, potentially enhancing post-processing efficiency. β-turns are essential structural elements on the protein surface, with their backbone conformation predominantly governed by critical interactions. In the shortening of the C-peptide, the YPGDV sequence was selected due to its notable advantages in promoting β-turn formation. Specifically, studies on β-turns have shown that the presence of proline at the 2nd position and glycine at the 3rd position results in the highest β-turn population.

Additionally, the inclusion of deprotonated aspartic acid at the 4th position significantly enhances the stability of the β-turn. The YPGDV sequence has been demonstrated to exhibit a high frequency of β-turn conformations in isolation (Dyson et al., 1988).

Experimental validations reveal that mini-proinsulin incorporating the YPGDV pentapeptide demonstrates a substantially higher refolding yield than the conventional proinsulin isoform (Chang et al., 1998; Cho et al., 2000). Previous attempts to construct mini-proinsulin using the RRGSKR linker sequence showed reduced purification efficiency, with unresolved confirmation of correct disulfide bond formation (Wetzel et al., 1981). Conversely, mini-proinsulin with the YPGDV sequence exhibited optimal refolding at a pH range of 11–11.5, which is more alkaline than the conditions typically favored for proinsulin refolding (Chang et al., 1998). These results suggest that electrostatic repulsion between the dibasic residues, RR and KR, plays a more pronounced role in mini-proinsulin folding kinetics than full-length proinsulin. Additionally, the differential refolding efficiencies between the YPGDV-based mini-proinsulin and other mini-proinsulin variants with distinct C-peptide sequences become increasingly evident at higher protein concentrations, indicating that YPGDV-based design strategy could serve as a robust alternative for scalable recombinant human insulin production.

Here, we utilized two forms of mini-proinsulin that have YPGDV sequences in C-peptide: the 1st is a previously characterized M2PI structure determined by NMR studies (PDB ID: 1EFE), and the 2nd is a novel designer M2PI variant. The novel variant has been calculated by AF and incorporates specific amino acid substitutions, including R22B to K22B to reduce susceptibility to trypsin degradation, P28B to D28B to

mitigate potential aggregation in solution, and N60A to G60A to enhance stabilization of nM2PI in acidic environments (Fig 1).

To assess the proper binding of mini-analogs to the IR, docking studies were performed using the experimental receptor structure (PDB ID: 6VEP). The results indicate that nM2PI displays higher binding affinity to IR relative to the M2PI structure, likely attributed to the positioning of the C-peptide on the αCT domain (Fig. 1). The binding conformations of the mini-analogs diverge substantially from each other and the native IR-INS complex. Intra-monomer dynamics reveal that both mini-analogs exhibit enhanced cooperativity when bound to IR, while increased residual mobility is observed when the analogs are in isolation. Despite these conformational differences, both mini-analogs show collective motions that align with those of native INS within the IR, suggesting notable interaction patterns between the mini-analogs and IR (Fig. 2). Additionally, these findings confirm that mini-proinsulin exhibits significant receptor-binding activity, with at least 50% of the binding potency of native INS (Cho et al., 2000).

The global 5 slow ANM modes and local 10 fastest GNM modes reveal that the mini-analogs interact with the IR, adopting similar intrinsic dynamics to native INS, yet exhibiting conformational alterations that disrupt conventional interaction pathways and create new ones (Fig. 3 and Fig. 4). In particular, the local 10 fastest GNM modes identify novel potential hot spots that differ significantly from those observed in the native receptor interaction (Fig. 4), suggesting the mobility of the C-peptide controls the allosteric behavior of mini-proinsulin analogs in receptor binding (Fig 5).

Upon comparing the receptor binding activity of mini-analogs with other reported mini-proinsulin variants, it was observed that some designs displayed receptor interaction

strengths as low as 0.5% of that of native insulin (Kobayashi et al., 1989). For example, insulin cross-linked between A1 and B29 with 2,7-diaminosuberic acid showed substantially reduced receptor binding, similar to proinsulin (Cutfield et al., 1981), while a mini-proinsulin with a direct peptide bond between B29 and A1 was effectively inactive (Derewenda et al., 1991). In contrast, the high receptor binding activity exhibited by the experimental mini-proinsulin (M2PI) in previously available study was supported by our computational analyses. Notably, besides the novel identified potential hot spots, the RR and KR enzyme processing regions were found to play a significant role in receptor interaction. The literature further suggests that these basic residues impact the receptor binding conformation (Chang et al., 1998). In M2PI and nM2PI, these residues induce conformational changes within the receptor binding region, potentially contributing to the allosteric modulation of those mini-proinsulins. That is why, in proinsulin, the larger C-peptide likely blocks critical active surfaces on the insulin molecule from interacting with the receptor (Marques et al., 2004), which may explain the reduced receptor binding affinity observed in proinsulin.

Collectively, despite the ongoing challenges in insulin recombinant production, which remains monopolized by certain companies, this study suggests that with further experimental validation and optimization, mini-proinsulin analogs hold great potential as viable end products for efficient recombinant insulin production in *E. coli*.

Competing Interests

The author declare no competing interests.

References

Ayan, E., & DeMirci, H. (2022). A Brief Atlas of Insulin. *Current Diabetes Reviews*, *19*. doi: 10.2174/1573399819666220610150342

Bakan, A., Meireles, L. M., & Bahar, I. (2011). ProDy: Protein dynamics inferred from theory and experiments. *Bioinformatics*. doi: 10.1093/bioinformatics/btr168

Brems, D. N., Brown, P. L., Nakagawa, S. H., & Tager, H. S. (1991). The conformational stability and flexibility of insulin with an additional intramolecular cross-link. *Journal of Biological Chemistry*. doi: 10.1016/s0021-9258(18)52338-3

Chang, S. G., Kim, D. Y., Choi, K. D., Shin, J. M., & Shin, H. C. (1998). Human insulin production from a novel mini-proinsulin which has high receptor-binding activity. *Biochemical Journal*. doi: 10.1042/bj3290631

Cho, Y., Chang, S. G., Choi, K. D., Shin, H., Ahn, B., & Kim, K. S. (2000). Solution Structure of an Active Mini-Proinsulin, M2PI: Inter-chain Flexibility is Crucial for Insulin Activity. *Journal of Biochemistry and Molecular Biology*.

Cutfield, J., Cutfield, S., Dodson, E., Dodson, G., Hodgkin, D., & Reynolds, C. (1981). Evidence concerning insulin activity from the structure of a cross-linked derivative. *Hoppe-Seyler's Zeitschrift Fur Physiologische Chemie*. doi: 10.1515/bchm2.1981.362.1.755

Derewenda, U., Derewenda, Z., Dodson, E. J., Dodson, G. G., Bing, X., & Markussen, J. (1991). X-ray analysis of the single chain B29-A1 peptide-linked insulin molecule. A completely inactive analogue. *Journal of Molecular Biology*. doi: 10.1016/0022-2836(91)90022-X

Dyson, H. J., Rance, M., Houghten, R. A., Lerner, R. A., & Wright, P. E. (1988). Folding of immunogenic peptide fragments of proteins in water solution. I. Sequence requirements for the formation of a reverse turn. *Journal of Molecular Biology*. doi: 10.1016/0022-2836(88)90446-9

Eren, M., Tuncbag, N., Jang, H., Nussinov, R., Gursoy, A., & Keskin, O. (2021). Normal Mode Analysis of KRas4B Reveals Partner Specific Dynamics. *Journal of Physical Chemistry B*. doi: 10.1021/acs.jpcb.1c00891

Gross, D. J., Villa-Komaroff, L., Kahn, C. R., Weir, G. C., & Halban, P. A. (1989). Deletion of a highly conserved tetrapeptide sequence of the proinsulin connecting peptide (C-peptide) inhibits proinsulin to insulin conversion by transfected pituitary corticotroph (AtT20) cells. *Journal of Biological Chemistry*. doi: 10.1016/s0021-9258(20)88210-6

Hua, Q. X., Hu, S. Q., Jia, W., Chu, Y. C., Burke, G. T., Wang, S. H., Wang, R. Y., Katsoyannis, P. G., & Weiss, M. A. (1998). Mini-proinsulin and Mini-IGF-I: Homologous protein sequences encoding non-homologous structures. *Journal of Molecular Biology*. doi: 10.1006/jmbi.1997.1574

Huey, R., Morris, G. M., & Forli, S. (2012). Using AutoDock 4 and AutoDock Vina with AutoDockTools: A Tutorial. In The Scripps Research Institute Molecular.

Ido, Y., Vindigni, A., Chang, K., Stramm, L., Chance, R., Heath, W. F., DiMarchi, R. D., Di Cera, E., & Williamson, J. R. (1997). Prevention of vascular and neural dysfunction in diabetic rats by C- peptide. *Science*. doi: 10.1126/science.277.5325.563

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*. doi: 10.1038/s41586-021-03819-2

Kobayashi, M., Sasaoka, T., Sugibayashi, M., Iwanishi, M., & Shigeta, Y. (1989). Receptor binding and biologic activity of biosynthetic human insulin and mini-proinsulin produced by recombinant gene technology. *Diabetes Research and Clinical Practice*. doi: 10.1016/0168-8227(89)90040-5

Marques, R. G., Fontaine, M. J., & Rogers, J. (2004). C-peptide: Much more than a byproduct of insulin biosynthesis. In Pancreas. doi: 10.1097/00006676-200410000-00009

Matouschek, A., Kellis, J. T., Serrano, L., Bycroft, M., & Fersht, A. R. (1990). Transient folding intermediates characterized by protein engineering. *Nature*. doi: 10.1038/346440a0

Pittman IV, I., Nakagawa, S. H., Tager, H. S., & Steiner, D. F. (1997). Maintenance of the B-chain β-turn in [Gly(B24)]insulin mutants: A steady-state fluorescence anisotropy study. *Biochemistry*. doi: 10.1021/bi9624144

Rholam, M., Nicolas, P., & Cohen, P. (1986). Precursors for peptide hormones share common secondary structures forming features at the proteolytic processing sites. *FEBS Letters*. doi: 10.1016/0014-5793(86)80002-3

Siew, Y. Y., & Zhang, W. (2021). Downstream processing of recombinant human

insulin and its analogues production from E. coli inclusion bodies. In Bioresources and Bioprocessing. doi: 10.1186/s40643-021-00419-w

Steiner, D. F., & Rubenstein, A. H. (1997). Proinsulin C-peptide-biological activity? In Science. doi: 10.1126/science.277.5325.531

Wetzel, R., Kleid, D. G., Crea, R., Heyneker, H. L., Yansura, D. G., Tadaaki, H., Kraszewski, A., Riggs, A. D., Keiichi, I., & Goeddel, D. V. (1981). Expression in escherichia coli of a chemically synthesized gene for a “mini-c” analog of human proinsulin. *Gene*. doi: 10.1016/0378-1119(81)90061-5

Xiong, X., Menting, J. G., Disotuar, M. M., Smith, N. A., Delaine, C. A., Ghabash, G., Agrawal, R., Wang, X., He, X., Fisher, S. J., MacRaild, C. A., Norton, R. S., Gajewiak, J., Forbes, B. E., Smith, B. J., Safavi-Hemami, H., Olivera, B., Lawrence, M. C., & Chou, D. H. C. (2020). A structurally minimized yet fully active insulin based on cone-snail venom insulin principles. *Nature Structural and Molecular Biology*. doi: 10.1038/s41594-020-0430-8

Zhou, H. X., & Pang, X. (2018). Electrostatic Interactions in Protein Structure, Folding, Binding, and Condensation. In Chemical Reviews. doi: 10.1021/acs.chemrev.7b00305