

RESEARCH ARTICLE

Structures and Interactions of Insulin-like Peptides from Cone Snail Venom

Insulin/Receptor Interactions

Biswajit Gorai¹ | Harish Vashisth^{1,*}

¹Department of Chemical Engineering,
University of New Hampshire, Durham,
NH 03824, USA.

Correspondence

*Harish Vashisth, Department of Chemical
Engineering, University of New Hampshire,
Durham, NH 03824, USA.
Email: harish.vashisth@unh.edu

Abstract

The venomous insulin-like peptides released by certain cone snails stimulate hypoglycemic shock to immobilize fish and catch the prey. Compared to human insulin (hIns), the cone snail insulins (Con-Ins) are typically monomeric and shorter in sequence, yet they exhibit moderate human insulin-like biological activity. We have modeled six variants of Con-Ins (G3, K1, K2, T1A, T1B, and T2) and carried out explicit-solvent molecular dynamics (MD) simulations of eight types of insulins, two with known structures (hIns and Con-Ins-G1) and six Con-Ins with modeled structures, to characterize key residues of each insulin that interact with the truncated human insulin receptor (μ IR). We show that each insulin/ μ IR complex is stable during explicit-solvent MD simulations and hIns interactions indicate the highest affinity for the 'site 1' of IR. The residue contact maps reveal that each insulin preferably interacts with the α CT peptide than the L1 domain of IR. Through analysis of the average non-bonded interaction energy contribution of every residue of each insulin for the μ IR, we probe the residues establishing favorable interactions with the receptor. We compared the interaction energy of each residue of every Con-Ins to the μ IR and observed that γ -carboxylated glutamate (Gla), His, Thr, Tyr, Tyr/His, and Asn in Con-Ins are favorable substitutions for GluA4, AsnA21, ValB12, LeuB15, GlyB20, and ArgB22 in hIns, respectively. The identified insulin analogs, although lacking the last eight residues of the B-chain of hIns, bind strongly to μ IR. Our findings are potentially useful in designing potent fast-acting therapeutic insulin.

KEYWORDS:

insulin, insulin analogs, cone snail, venomous peptides, molecular dynamics

INTRODUCTION

Animal venoms are a concoction of small proteins or peptides which are used as an arsenal to immobilize their prey^{1,2,3,4}. In contrast, some of them are approved for the treatment of a wide range of human pathophysiological conditions^{5,6,7,8,9,10}. Specifically, cone snails release specialized insulin-like protein toxins in water to immobilize fish and catch their prey. Recent studies have shown that the insulin from *C. geographus* (Con-Ins-G1) severely lowers the blood sugar level and induces hypoglycemic shock in the fish¹¹. Menting *et al.*¹² also reported Con-Ins-G1 as the smallest naturally available potent agonist of the human insulin receptor (hIR). Other studies have identified *C. tulipa* and *C. kinoshitai*, along with *C. geographus* species of cone snail, that release specialized insulins for fish hunting^{11,13}. The primary sequences of these insulin-like peptides are not well conserved, though the architecture of disulfide bonds of the cone snail insulin family members are very similar to the fish insulin and hIns¹³. The hIns mostly occurs in dimeric or hexameric forms but the multimeric forms of hIns needs to dissociate into the monomeric form to bind to and activate the hIR^{14,15,16,17}. The C-terminal residues of the B-chain of hIns (e.g., PheB24, PheB25, and TyrB26) play a key role in insulin's self-assembly into higher order forms. Efforts have also been made to obliterate the multimerization of hIns by trimming the last eight residues of the C-terminal of the B-chain, thereby creating the des-octapeptide(B23–B30) insulin (DOI); however this analog loses its biological activity significantly¹⁸. Intriguingly, cone snail insulins are monomeric, as they lack the B-chain C-terminal residues of the hIns (GlyB23–ThrB30), yet they show modest binding affinities for hIR¹³. Recently, Xiong *et al.*¹⁹ investigated the structure-activity relationship of the Con-Ins-G1 and reported a fully active monomeric insulin analog, mini-Ins, which was derived from Con-Ins–G1 after incorporating four mutations (HisA8, ArgA9, GluB10, and TyrB20) into DOI.

The hIns is a small protein hormone (51 residues) comprised of two chains, termed the A-chain (21 residues) and the B-chain (30 residues)²⁰. Structurally, insulin is mainly a helical protein (A-chain has two short helices and the B-chain contains a single longer helix) and consists of three disulfide bonds (CysA6-CysA11, CysA7-CysB7, and CysA20-CysB19). The mature insulin triggers a cascade of signaling reactions on binding to the extracellular domains of hIR and plays a major role in glucose homeostasis^{21,22}. Moreover, insulin binding to hIR displays negative cooperativity²³ and allosteric activation of the receptor²⁴. Insulin deficiency or insensitivity often leads to type I or type II diabetes mellitus^{25,26}. Biochemical and modeling studies have reported that insulin and hIR interact *via* two binding patches termed as 'site 1' and 'site 2'^{27,28,29,30,31,32,33}.

The 'site 1' of insulin, also known as the "classical binding surface", involves both A-chain residues (Gly1, Ile2, Val3, Glu4, Tyr19, and Asn21) as well as B-chain residues (Gly8, Ser9, Leu11, Val12, Tyr16, Phe24, Phe25, and Tyr26), and primarily interacts with specific residues of first leucine-rich (L1) domain and the C-terminal region of the α -chain (α CT) of the hIR, known as complementary 'site 1' of hIR, composed of the following residues: Asp12, Ile13, Arg14, Asn15, Gln34, Leu36, Leu37, Phe39, Glu44, Phe64, Tyr67, Phe89, Asn90, Tyr91, Phe705, Glu706, Asp707, Tyr708, Leu709, Asn711, Val712, Phe714,

Pro716 and Arg717. The 'site 2' of insulin is thought to be formed by the following residues: Thr8, Ile10, Ser12, Leu13, and Glu17 of the A-chain and His10, Glu13, and Leu17 of the B-chain, and likely interacts with the residues near the junction of the loop regions of two fibronectin type-III repeat domains (FnIII-1 and FnIII-2) of hIR consisting of residues Lys484, Leu552, Asp591, Ile602, Lys616, Asp620 and Pro621^{30,34,35}. The 'site 1' of insulin likely binds with a higher affinity with the 'site 1' of hIR and most of the 'site 1' residues of insulin are in close proximity with the α CT peptide²⁹. Specifically, the binding affinity of the recently reported mini-Ins¹⁹ for 'site 1' (referred as the "primary site" in the study) of hIR is 128 fold lower than that of the hIns. Further, molecular dynamics (MD) simulation studies revealed that the mini-Ins likely binds to the "secondary site" (binding surface formed by the loops of the FnIII-1 domain) and the "transient site" (binding surface formed by the β sheets of the FnIII-1 domain) of hIR with higher affinity and thus showed similar potency toward hIR¹⁹.

The alignment of the sequences of Con-Ins with the sequence of hIns suggests a lower sequence conservation (Figure 1) although GlyA1, ValA3, GlyB8, SerB9, and all cysteine residues of hIns are conserved within Con-Ins. Even though Con-Ins lack the eight C-terminal residues of the B-chain critical for hIns dimerization and biological activity, these peptides can notably lower the blood glucose level in zebrafish and mouse models¹³. The reported studies investigated the binding of hIns and Con-Ins-G1 to hIR and derived an equally potent surrogate of insulin from the venom of *C. geographus*^{12,13,19}. However, due to the lack of structural data on the complexes of venom insulins from *C. tulipa* or *C. kinoshitai* with hIR, the interactions between them are poorly understood. Given the need to identify a potent surrogate of hIns, structural modeling and molecular simulation techniques are useful tools to decipher the binding interactions of Con-Ins with hIR. The purpose of this study is to construct structural models of the six types of Con-Ins (G3, K1, K2, T1A, T1B, and T2) based on sequences reported by Ahorukmeyer *et al.*¹³ and study their residue-level interactions with the hIR using MD simulations. Based on structural features and interactions of Con-Ins with μ IR, we propose six residue substitutions in hIns, which are critical for the activity of monomeric Con-Ins for hIR, although Con-Ins lack several residues in the C-terminus of the B-chain that are otherwise present in hIns. We designed two insulin analogs based on our proposed substitutions and further studied their binding and interactions with μ IR, thereby revealing the enhanced binding of these analogs to hIR. These results will potentially motivate future design of potent insulin analogs.

MATERIALS AND METHODS

Structural modeling of cone snail insulins

We obtained the Con-Ins sequences of *C. geographus* (G3), *C. kinoshitai* (K1 and K2), and *C. tulipa* (T1A, T1B, and T2) from the previous work by Ahorukmeyer *et al.*¹³. We modeled the tertiary structure of these six venom insulins using MODELLER v9.10³⁶. We used the structure of the cone snail insulin from *C. geographus* (G1, PDB ID: 5JYQ) as a template during

model building using the homology modeling approach³⁷. We initially transformed the post-translationally modified (PTM) unnatural residues present in the sequences of Con-Ins to their nearest natural amino acid homolog. For example, we considered the residues γ -carboxylated glutamate (Gla) and hydroxyproline (Hyp) of the cone snail insulins as glutamic acid (Glu) and proline (Pro), respectively. We used the multi-chain modeling approach to predict the structures of cone snail insulins with the A- and B-chains and preserved the disulfide bonds present in the template during model generation. We generated 200 models of each insulin using the MODELLER and the best model was selected based on the lowest discrete optimized protein energy (DOPE) score³⁸. We then used the Vienna-PTM 2.0 online portal³⁹ to revise the transformed residues, which were modified from unnatural to natural amino acids during the structure prediction, to their wild-type PTM form in the modeled cone snail structures. Before performing MD simulations, we further used the PROPKA online portal⁴⁰ to assign correct protonation states to the side-chains of residues at a pH value of 7.

Structural modeling of insulin/ μ IR complexes

We obtained the initial coordinates of the crystal structures of hIns (PDB ID: 6VEP) and Con-Ins-G1 (PDB ID: 6VEQ) bound to the truncated hIR (μ IR) from the Protein Data Bank (PDB). Specifically, the μ IR construct is comprised of residues 1–154 of the L1 domain and a short C-terminal peptide of the α -chain composed of residues 704–719 (α CT). We modeled the missing residues in the crystal structure of hIns and Con-Ins-G1 bound to μ IR using MODELLERv9.10. Using the PyMOL software⁴¹, we then performed an all-atom structural alignment of each modeled cone snail insulin structure on the Con-Ins-G1 structure (PDB ID: 6VEQ) to obtain their complexes with the μ IR. Overall, we obtained eight insulin- μ IR complexes, two among which were experimentally determined structures (each for the hIns and Con-Ins-G1 in complex with the μ IR) and the remaining six were modeled Con-Ins structures in complex with the μ IR.

Setup for all-atom MD simulations

To simulate the Con-Ins structures with PTM residues, we performed classical all-atom MD simulations of various insulins and their complexes with μ IR using the GROMACsv2020.4 software package⁴² combined with the modified GROMOS ffG54a8 force-field⁴³ with extended parameters from the Vienna-PTM server³⁹. Each insulin or insulin- μ IR complex was placed at the center of a dodecahedron box extending a minimum distance of 12 Å from any atom of the protein and solvated by the SPC explicit water model⁴⁴. Next, the requisite sodium and chloride ions were added to neutralize and maintain an ionic concentration of 140 mM of salt in each system. Each system was minimized using the steepest-descent algorithm⁴⁵. The minimized systems were equilibrated at 300 K for 500 ps using the Berendsen thermostat with a coupling time of 0.1 ps. Further, the systems were equilibrated using the Berendsen barostat at 1 atm pressure for 10 ns. The heavy atoms of each peptide or protein during the initial minimization and equilibration steps were positionally restrained. We used periodic boundary conditions, a non-bonded

cut-off of 14 Å, and calculated the long-range electrostatic interactions using the particle-mesh Ewald method⁴⁶. The bond-lengths were constrained using the LINCS algorithm⁴⁷ and time-step of 2 fs was used in all simulations. After equilibration, the restraints were removed and longer runs for 500 ns were performed in the NPT ensemble. The V-rescale thermostat with a coupling time of 1.0 ps was used during the longer timescale MD simulations⁴⁸. The coordinates from simulation trajectories were saved at every 20 ps. The analyses of MD trajectories were performed using the tools in GROMACS⁴⁹ and VMD⁵⁰. Each set of simulation was performed in triplicate for each insulin as well as the insulin- μ IR complex. The details of all the MD simulations are summarized in Table S1.

RESULTS

Structural models and all-atom MD simulations of unbound Con-Ins structures

Using the comparative modeling approach, we generated the tertiary structures of six different variants of cone snail insulins: G3, K1, K2, T1A, T1B, and T2 (Figure 2A). We preserved the secondary structure components (3 helices) and three disulfide bonds during model building and retained all unnatural PTM residues. The all-atom root mean squared deviation (RMSD) obtained after superimposing each model on the Con-Ins-G1 template are less than 1 Å which implies that the predicted initial structures are consistent with the template used.

To study the conformational stability and variability of the modeled Con-Ins structures, we conducted three independent conventional MD simulations (each 500 ns long) of each Con-Ins in an aqueous environment. We applied the Gromos clustering algorithm⁵¹ using a RMSD cut-off of 2 Å on the ensemble of insulin structures generated from the last 400 ns of each simulation trajectory. We selected the central structure of the dominant cluster as the representative conformer of each insulin after the MD simulation. We then superimposed each representative insulin conformer on their starting structure (Figure 2B) to observe any major structural differences. Among the superimposed structures, Con-Ins-K1 and Con-Ins-K2 showed the highest deviations (RMSD > 3.0 Å) from the initial structure due to additional residues at the C- and N-termini of the A- and B-chains, respectively. Those residues are flexible given their location in the unstructured regions which significantly contributes to the higher RMSD of the conformers. The models with comparatively smaller sequence length (Con-Ins-G3, Con-Ins-T1A, Con-Ins-T1B, and Con-Ins-T2) showed lower RMSD values (Figure 2B).

All-atom MD simulations of insulin- μ IR complexes

Using conventional MD simulations, we equilibrated each insulin- μ IR complex in an aqueous environment for 500 ns. To assess the stability of the complexes, we calculated the time evolution of the center of mass (COM) separation between each insulin and the μ IR, averaged over three independent simulations (Figure S1). We found that the complexes are stable and the ligand-receptor

pairs maintained contact throughout each simulation. To examine the effect on the structure of each insulin after binding to the μ IR, we also tallied the distributions of the RMSD from receptor-bound insulin simulations with the unbound MD simulations of each insulin (Figure 3). We observed that the peak of the RMSD histograms of bound insulins are at lower RMSD values in comparison to the RMSD peak for respective unbound insulins, which implies that the native like structure of the bound insulins were more preserved than those of unbound insulins.

To further delineate the influence of binding of insulin to μ IR on the fluctuations of residues, we calculated the change in the root mean squared fluctuation (Δ RMSF) per residue of each bound insulin relative to each unbound insulin (Figure S2). The Δ RMSF data revealed that the conformational fluctuations in residues of each insulin decrease on binding to μ IR. The fluctuations of the residues of hIns in the helical regions decreased more than the unstructured regions (formed by loops and turns). Additionally, the fluctuations of the C-terminal residues of the B-chain of hIns significantly decreased on binding to μ IR. We also observed a significant reduction in the flexibility of the Con-Ins residues in the bound state with μ IR. These observations have implications on the interaction of insulins with μ IR because those residues for which conformational fluctuations decreased most likely participate in favorable interactions with the receptor.

Insulin- μ IR interactions

To elucidate the residues having favorable contacts, we computed the residue-residue distance correlations from the simulation trajectories and mapped them on a contact map for each insulin- μ IR complex. Specifically, we calculated the distances between the center of mass of each residue of the insulin and μ IR averaged from three simulation runs and projected on 2D contour maps (Figure S3). Notably, the C-terminus of each insulin chain shows residue contacts with the N-terminal residues of the L1 domain and the α CT peptide of μ IR. The B-chain of hIns, with a longer C-terminal segment (B23-B30), shows favorable contacts with the N-terminal residues of the L1 domain and the N-terminal residues of the α CT peptide of μ IR. However, Con-Ins having shorter B-chains lack these interactions with the N-terminal residues of the L1 domain and the α CT peptide, which is likely a reason for their lower binding affinity. The contact maps reveal that each insulin preferably interacts with the α CT peptide of μ IR, which is consistent with the previous work²⁹.

We also assessed the average non-bonded interaction energy (cumulative van der Waals and electrostatic interaction energies) between all residues of each insulin and all residues of μ IR from three independent simulations (Figure S4). The non-bonded interaction energy of hIns to μ IR is the lowest (-938 ± 182 kJ/mol) followed by T1A (-654 ± 141 kJ/mol), K1 (-626 ± 145 kJ/mol), T2 (-623 ± 117 kJ/mol), T1B (-604 ± 130 kJ/mol), G3 (-566 ± 122 kJ/mol), G1 (-561 ± 113 kJ/mol), and K2 (-522 ± 101 kJ/mol). The negative non-bonded energies suggest favorable interactions between each insulin and μ IR. As the interactions are stable, we sought to identify key residues of each insulin variant that can be attributed to their reduced or enhanced biological activity for μ IR.

Therefore, we calculated the average non-bonded interaction energy contribution of each residue of each insulin for the μ IR (Figure 4A) and identified the interactions common in at least two representative structures, the central conformer of the most populated cluster, from three independent MD simulations to delineate the crucial insulin residues engaged in the interaction with the receptor. We observed that the N-terminal residues (Gly1-Glu4) of the A-chain of hIns interact with the His710, Asn711, and Phe714 of the α CT peptide (Figure 4B). The α CT residue His710 also exhibits hydrophobic interactions with the residues ValA3, GlyB8, and ValB12 of hIns (Figure 4C). The residue Phe714 of α CT is oriented toward the hydrophobic pocket formed by GlyA1, IleA2, ValA3, TyrA19, LeuB11, and LeuB15. The C-terminal residues (AsnA18 and TyrA19) of the hIns's A-chain establish favorable interactions with the Phe714, Pro716, and Arg717 of the α CT peptide (Figure 4D). The side chain of PheB25 interacts with the Pro716, Arg717, and Pro718 of the α CT peptide (Figure 4E). The side chain of TyrB16 and ValB12 pairs with the residues Phe39 and Phe64 of the L1 domain, respectively (Figure 4F-G). The side chain of GluB13 forms a salt-bridge with the side-chain of Arg65 of the L1 domain (Figure 4H). The observed interacting residues are in agreement with the 'site 1' binding surface of hIns reported in the literature^{12,13,27,28}.

The Con-Ins structures lack nine residues that are present at the C-terminus of the B-chain of hIns. The Con-Ins also exist in the monomeric form and are known to activate IR¹³. Xiong *et al.*¹⁹ have interpreted crucial interactions between the Con-Ins-G1 and μ IR based on the crystal structure. They elucidated that Con-Ins-G1 lacks the aromatic triplet PheB24-PheB25-TyrB26 of hIns, yet it binds to μ IR. We have carried out MD simulations of the experimental structure of the Con-Ins-G1/ μ IR complex and computed the average per residue energy contribution which illustrates that Con-Ins-G1 mostly interacts with the N- and C-terminal residues of the A-chain and with the residues in the C-terminus of the B-chain (Figure 4A). Among the PTM residues, GlaA4 showed significant non-bonded interaction by establishing a stable hydrogen bonding interaction with the Asn711 of the α CT peptide (Figure 4I), which is consistent with the previous work¹⁹. We also observed that the average energy contribution of the Con-Ins-G1's GlaA4 is substantially lower than the hIns's GluA4, thus Gla4 is a potential substitution to enhance the activity of a mutant insulin. We observed that TyrB17 of Con-Ins-G1 occupies the space inhabited by PheB24 of hIns (Figure 4J), which is in agreement with the earlier reports^{12,19}. Similarly, we suggest residues ThrB14 and TyrB22 of Con-Ins-G1 as possible substitutions for the residues ValB12 and GlyB20 of hIns, respectively, as they exhibit relatively lower non-bonded energies. The residues ThrB14 and TyrB22 of hIns form hydrogen bonds with the side-chains of residues Arg65 and Asn15, respectively, of the L1 domain (Figure 4K-L).

The pattern of per residue binding energy of Con-Ins-G3 is similar to the Con-Ins-G1 (Figure 4A), with marginal differences. The Con-Ins-G3 PTM residue, GlaA4, also forms a hydrogen bond with the residue Asn711 of α CT, as also observed in Con-Ins-G1. The TyrB17 residue of cone snail insulins, also known as a surrogate for hIns's PheB24¹², occupies the hydrophobic pocket formed by Asn15, Leu37, and Phe39 of the L1 domain and Phe714 of the α CT peptide (Figure 4M). The C-terminal

residue of Con-Ins-G3, HisB22 shows a higher non-bonded energy, formed favorable interactions with Phe39 and Arg42 of the L1 domain (Figure 4N), and may be selected as a probable replacement for GlyB20 of hIns.

The sequence alignment of Con-Ins with hIns (Figure 1) reveals that the C-terminus of the A-chain in both Con-Ins-K1 and Con-Ins-K2 is longer compared to hIns, while the N-terminus of the B-chain in both Con-Ins-K1 and Con-Ins-K2 is longer compared to hIns. The N-terminal residues of the A-chain of the Con-Ins-K1, GlyA1 and IleA2, interact with the Phe714 residue of the α CT peptide (Figure 4O). The main-chain of GluA18 interacts by establishing a hydrogen bond with the side-chain of Arg717 (Figure 4P). The non-bonded binding energy contribution of HisA21 of Con-Ins-K1 (-49.41 kJ/mol) is significantly lower than the residue Asn21 of hIns (-18.22 kJ/mol), thus the His residue is more likely to fit at the 21st position of hIns. The side-chain of the HisA21 of Con-Ins-K1 interacts with the residues Arg14 and Asn15 of the L1 domain of μ IR (Figure 4Q). However, the per residue energy plot suggests that Con-Ins-K2 interacts moderately with μ IR (Figure 4A), except that the end residue AsnB27 shows the lowest non-bonded interaction energy. The side-chain of AsnB27 residue of Con-Ins-K2 forms a hydrogen bond with the main-chain of Asn15 of the L1 domain (Figure 4R). Thus, the Asn residue may be a potential substitution for ArgB22 of hIns.

The Con-Ins-T1A, Con-Ins-T1B, and Con-Ins-T2 sequences are very similar (Figure 1) and they exhibit comparable non-bonded binding energy with the residues of μ IR (Figure S4). The residues ProB14 and IleB18 occupy a pocket on the surface of the L1 domain formed by residues Phe39, Arg65, and Tyr67 residues (Figure 4S). The side chain of TyrB17 of T1A, also known as a surrogate for hIns's PheB24, interacts with the residues Phe39 (L1) and Phe714 (α CT) (Figure 4T). The residue PheB20 interacts with the residues Lys40 and Arg42 of the L1 domain (Figure 4U). We noticed that the per residue non-bonded interaction energy (Figure 4A), and the interaction pattern of Con-Ins-T1B and Con-Ins-T2 with μ IR are very similar to Con-Ins-T1 (Figure S7). The residues GluA4 and TyrB17 of Con-Ins-T1A, Con-Ins-T1B, and Con-Ins-T2 exhibit significantly lower interaction energies for μ IR than the residues GluA4 and LeuB15 of hIns. These two residues (GluA4 and TyrB17) are thus crucial and they may be potential substitutes in a modified insulin with enhanced biological activity.

Thus, we assessed the non-bonded interaction energy contribution with the μ IR for each residue of hIns with each residue at the equivalent position in seven variants of Con-Ins and observed that certain Con-Ins residues demonstrate noticeable lower interaction energies (i.e. stronger interactions) in comparison to the hIns residues at equivalent positions (Figure 5A). Specifically, Con-Ins residues Glu (Figure 5B), His (Figure 5C), Thr (Figure 5D), Tyr (Figure 5E), Tyr/His (Figure 5F/G), and Asn (Figure 5H) exhibit lower interaction energy with the μ IR in comparison to the hIns residues GluA4, AsnA21, ValB12, LeuB15, GlyB20, and ArgB22, respectively.

Based on these proposed substitutions, we designed the structural models of two insulin analogs (Figure 6A). The analogs are shorter in length, as they lack the last eight residues of the B-chain of hIns, and they differ from each other only by one residue at position 20 of the B-chain; analog 1 and analog 2 consist of residues Tyr and His at position 20, respectively. Each

analog in complex with the μ IR was subjected to triplicate all-atom MD simulations (Table S1). We calculated the average non-bonded interaction energy between residues of these analogs and μ IR. We observed that the average interaction energy of analog 1 (-963 ± 141 kJ/mol) for μ IR is lower than that of the analog 2 (-771 ± 149 kJ/mol), and significantly lower than each Con-Ins/ μ IR complex ($> -654 \pm 141$ kJ/mol), and most importantly, even lower than the interaction energy of hIns with the μ IR (-938 ± 182 kJ/mol). To further scrutinize the significance of these substitutions, we estimated the difference of the non-bonded interaction energy (ΔE) of each residue of analog 1 (Figure 6B) and analog 2 (Figure 6C) from the non-bonded interaction energy of respective residues of hIns. The substituted residues: GluA4 (Figure 6D), HisA21 (Figure 6E), ThrB12 (Figure 5F), TyrB15 (Figure 6G), Tyr/HisB20 (Figures 6H and 6I), and AsnB22 (Figure 6J) significantly enhance the binding energy of analogs to μ IR. Among these six substitutions, TyrB15 and AsnB22 significantly boost the affinity of the analogs for μ IR. The higher energy contribution by residues TyrB15 and AsnB22 of analog 1 in comparison to the equivalent residues in analog 2 explains the better binding energy of the former. Except the residues GlyA1, GluA17, AsnA18, TyrA19, GlyB8, SerB9, and HisB10 of the analogs, other residues favor the enhanced binding to μ IR.

DISCUSSION

The transition of hIns from a hexamer to a monomer impedes its bioavailability and delays the pharmacological activity^{12,19,29}. Certain cone snail venom insulin-like peptides have inherited the ability to bind to and activate the hIR, though they are monomeric and have a shorter sequence length than hIns. In this work, we have investigated the key residues of cone snail venom insulins interacting with the 'site 1' of μ IR using all-atom MD simulations with the aim to understand the characteristic features of a bioactive insulin based on interactions of Con-Ins with the receptor. We studied two experimentally available insulin structures (hIns and Con-Ins-G1) and six modeled structures of cone snail venom insulin like proteins (Con-Ins-G3, Con-Ins-K1, Con-Ins-K2, Con-Ins-T1A, Con-Ins-T1B, and Con-Ins-T2). The alignment of the primary sequences of all insulin proteins suggests that the sequence identity between the hIns and Con-Ins sequences are lower ($< 30\%$). Contrary to the disparity among their sequences, the Con-Ins show notable hIns like molecular mechanism¹³. Among the Con-Ins, the sequence of the A-chain is more conserved than the B-chain (Figure 1).

First, we modeled the tertiary structures of six cone snail insulin like peptides and then performed conventional MD simulations of each model to examine their stabilities in an aqueous environment. We observed that the structures of the modeled insulins are maintained throughout simulations with marginal flexibilities for residues in the loop regions. Next, we simulated each insulin docked to the 'site 1' of μ IR and we observed that the RMSD and per residue RMSF of the insulins bound to μ IR are lower than that of the unbound insulins. The total non-bonded interaction energy between each insulin and μ IR suggests that the hIns has a higher affinity for the receptor than Con-Ins. Additionally, we estimated the per residue contribution of each insulin based on the total non-bonded interaction energy to delineate key residues interacting with the μ IR. Interestingly, the

last eight residues of the B-chain of hIns contribute ~40% to the total non-bonded binding energy, thus explaining the notable loss of potency of DOI for the hIR¹⁸. We have also observed that the first four and the last five residues of the A-chain (GlyA1, IleA2, ValA3, GluA4, GluA17, AsnA18, TyrA19, CysA20, and AsnA21), and the B-chain residues (SerB9, ValB12, TyrB16, PheB24, PheB25, TyrB26, ThrB27, LysB29, and ThrB30) of hIns are key residues interacting with the 'site 1' of μ IR. Similarly, the N- and C-terminal residues of the A-chain, and certain residues of the B-chain of cone snail insulins (starting from the 10th position) are primarily involved in interactions with the μ IR.

Among the key interacting residues of the A-chain of hIns and Con-Ins-G1, the GluA4 residue of the venom insulin forms a stable hydrogen bond with the side-chain of the residue Asn711 of the α CT peptide and exhibits significantly lower non-bonded interaction energy. The residues ThrB14 and TyrB22 of Con-Ins-G1 show stronger interactions than the residues ValB12 and GlyB20 at the equivalent positions of hIns (Figure 1). Therefore, we recommend residues Glu, Thr, and Tyr from Con-Ins-G1 as probable replacements at the 4th position of the A-chain and the 12th and 20th positions of B-chain of hIns, respectively. The per residue interaction pattern of Con-Ins-G3 and Con-Ins-G1 are very similar, except that the residue HisB22 of Con-Ins-G3 shows a relatively enhanced binding affinity for the μ IR. Hence, we propose that replacing the GlyB20 of hIns by a His residue may enhance the binding affinity of an insulin analog for the receptor.

We compared the per residue non-bonded interaction energy of Con-Ins-K1 and Con-Ins-K2 variants with the hIns. Both venom insulins include a longer fragment (3-5 residues) at the C- and N-termini of the A- and B-chains, respectively. The C-terminal residues of Con-Ins-K1 exhibit lower non-bonded interaction energies than the C-terminal residues of hIns. We observed that the residues HisA21, ThrA22, LeuA23, and GlnA24 of Con-Ins-K1 have significant per-residue binding energies with the residues of the receptor. Based on these binding energies, we propose that replacing AsnA21 of hIns with HisA21 may be potentially useful in enhancing the binding affinity of the obtained insulin analog, which will have the same length of sequence as the A-chain of hIns. Both Con-Ins-K1 and Con-Ins-K2 have Asn as the end residue (Figure 1), and the calculated non-bonded energy obtained for them are significantly lower than the equivalent residue of hIns at the 22nd position, i.e., ArgB22 (Figure 4A). Hence, replacing ArgB22 of hIns by Asn will likely increase the binding affinity of a modified insulin analog.

The sequences of Con-Ins-T1A, Con-Ins-T1B, and Con-Ins-T2 are mostly conserved within the studied proteins and they show very similar per residue non-bonded binding energies. We noted that the side chains of TyrB17 of Con-Ins-G1, Con-Ins-G3, Con-Ins-T1A, Con-Ins-T1B, and Con-Ins-T2 interact with the residue Phe714 of the L1 domain. The residue TyrB17 of Con-Ins is also known as a surrogate of the residue PheB24 of hIns and is one of the key residues responsible for the activity of cone snail insulins instead of the last nine residues of the B-chain of hIns^{12,13,19,27}. Thus, we recommend Tyr at the 15th position of the hIns as a potential substitution.

Based on our proposed substitutions, we designed two insulin analogs and report their binding energies with the μ IR. These analogs are eight residues shorter along the C-terminal end of the B-chain of hIns. The analogs reveal better binding energies

with the μ IR than any of the Con-Ins. Notably, analog 1 exhibits a better binding energy with the receptor than the wild type hIns. The substitutions derived from the interaction study between Con-Ins and μ IR augment the binding energies of the analogs. We also show that the B-chain substitutions, TyrB15 and AsnB22, are the most promising mutations responsible for the enhanced binding energy of analog 1 for the μ IR.

In summary, certain residues of Con-Ins exhibit substantially lower non-bonded interaction energies in comparison to the equivalent hIns residues complexed with μ IR. Our results suggest that an insulin analog with Glu, His, Thr, Tyr, Tyr/His, and Asn substitutions for the residues GluA4, AsnA21, ValB12, LeuB15, GlyB20, and ArgB22, respectively, of hIns will likely enhance the bioactivity in comparison to hIns. The recently reported fully-active smallest monomeric insulin analog, mini-Ins, derived only from the Con-Ins-G1 binds to the “secondary site” of hIR have His, Arg, Glu, and Tyr substitutions at ThrA8, SerA9, HisB10, and GlyB20, respectively, of hIns¹⁹. Moreover, we extracted specific structural features from seven variants of cone snail insulin like peptides interacting with the ‘site 1’ of hIR and derived two shorter-length insulin analogs. We noted that the binding energy of the analog 1 for μ IR is higher than that of the hIns. Thus, the findings from this study may potentially contribute to future design of fast-acting monomeric insulin analogs with enhanced affinity for hIR.

CONCLUSION

In this work, we initially modeled the structures of six Con-Ins subtypes (G3, K1, K2, T1A, T1B, and T2) using the experimental structure of the Con-Ins-G1/ μ IR complex as a template. We studied the complexes of seven Con-Ins with μ IR and investigated their key residues interacting with the ‘site 1’ of μ IR using all-atom MD simulations. We compared the binding energy contributions with the receptor for each residue of each cone snail insulin as well as of hIns to extract inherent structural features in Con-Ins for potential substitutions in hIns. Our study suggests that GluA4, AsnA21, ValB12, LeuB15, GlyB20, and ArgB22 of hIns may be substituted by Glu, His, Thr, Tyr, Tyr/His, and Asn residues, respectively, to design an active monomeric insulin analog with superior pharmacological potential.

NOTES

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We acknowledge financial support from the National Institutes of Health (NIH) through grant R35GM138217 (H.V). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. We thank computational support through the following resources: Premise, a central shared high-performance computing cluster at the University of New

Hampshire supported by the Research Computing Center; and BioMade, a heterogeneous CPU/GPU cluster supported by the NSF EPSCoR award (OIA-1757371).

References

1. Fry BG, Roelants K, Champagne DE, et al. The toxicogenomic multiverse: convergent recruitment of proteins into animal venoms. *Annu. Rev. Genomics Hum. Genet.* 2009; 10: 483–511.
2. Utkin YN. Animal venom studies: Current benefits and future developments. *World J. Biol. Chem.* 2015; 6(2): 28.
3. Schendel V, Rash LD, Jenner RA, Undheim EA. The diversity of venom: the importance of behavior and venom system morphology in understanding its ecology and evolution. *Toxins* 2019; 11(11): 666.
4. Walker AA, Robinson SD, Hamilton BF, Undheim EA, King GF. Deadly proteomes: a practical guide to proteotranscriptomics of animal venoms. *Proteomics* 2020; 20(17-18): 1900324.
5. Herzig V, Cristofori-Armstrong B, Israel MR, Nixon SA, Vetter I, King GF. Animal toxins—Nature’s evolutionary-refined toolkit for basic research and drug discovery. *Biochem. Pharmacol.* 2020; 181: 114096.
6. Yang X, Wang Y, Wu C, Ling EA. Animal venom peptides as a treasure trove for new therapeutics against neurodegenerative disorders. *Curr. Med. Chem.* 2019; 26(25): 4749–4774.
7. Souza dJM, Goncalves BD, Gomez MV, Vieira LB, Ribeiro FM. Animal toxins as therapeutic tools to treat neurodegenerative diseases. *Front. Pharmacol.* 2018; 9: 145.
8. Mohamed Abd El-Aziz T, Soares AG, Stockand JD. Snake venoms in drug discovery: valuable therapeutic tools for life saving. *Toxins* 2019; 11(10): 564.
9. Zambelli V, Pasqualoto KFM, Picolo G, Chudzinski-Tavassi AM, Cury Y. Harnessing the knowledge of animal toxins to generate drugs. *Pharmacol. Res.* 2016; 112: 30–36.
10. Robinson SD, Li Q, Bandyopadhyay PK, et al. Hormone-like peptides in the venoms of marine cone snails. *Gen. Comp. Endocrinol.* 2017; 244: 11–18.
11. Safavi-Hemami H, Gajewiak J, Karanth S, et al. Specialized insulin is used for chemical warfare by fish-hunting cone snails. *Proc. Natl. Acad. Sci. U.S.A.* 2015; 112(6): 1743–1748.
12. Menting JG, Gajewiak J, MacRaild CA, et al. A minimized human insulin-receptor-binding motif revealed in a *Conus geographus* venom insulin. *Nat. Struct. Mol. Biol.* 2016; 23(10): 916–920.

13. Ahorukomeye P, Disotuar MM, Gajewiak J, et al. Fish-hunting cone snail venoms are a rich source of minimized ligands of the vertebrate insulin receptor. *Elife* 2019; 8: e41574.
14. Derewenda U, Derewenda Z, Dodson G, Hubbard R, Korber F. Molecular structure of insulin: the insulin monomer and its assembly. *Br. Med. Bull.* 1989; 45(1): 4–18.
15. Dong J, Wan Z, Popov M, Carey PR, Weiss MA. Insulin assembly damps conformational fluctuations: Raman analysis of amide I linewidths in native states and fibrils. *J. Mol. Biol.* 2003; 330(2): 431–442.
16. Palivec V, Viola CM, Kozak M, et al. Computational and structural evidence for neurotransmitter-mediated modulation of the oligomeric states of human insulin in storage granules. *J. Biol. Chem.* 2017; 292(20): 8342–8355.
17. Weiss MA, Lawrence MC. A thing of beauty: structure and function of insulin’s “aromatic triplet”. *Diabetes Obes. Metab.* 2018; 20: 51–63.
18. Bao Sj, Xie Dl, Zhang Jp, Chang Wr, Liang Dc. Crystal structure of desheptapeptide (B24–B30) insulin at 1.6 Å resolution: Implications for receptor binding. *Proc. Natl. Acad. Sci. U.S.A.* 1997; 94(7): 2975–2980.
19. Xiong X, Menting JG, Disotuar MM, et al. A structurally minimized yet fully active insulin based on cone-snail venom insulin principles. *Nat. Struct. Mol. Biol.* 2020; 27(7): 615–624.
20. Adams MJ, Blundell TL, Dodson EJ, et al. Structure of rhombohedral 2 zinc insulin crystals. *Nature* 1969; 224(5218): 491–495.
21. Kahn CR. The molecular mechanism of insulin action. *Annu. Rev. Med.* 1985; 36(1): 429–451.
22. Petersen MC, Shulman GI. Mechanisms of insulin action and insulin resistance. *Physiol. Rev.* 2018; 98(4): 2133–2223.
23. De Meyts P, Roth J, Neville Jr DM, Gavin III JR, Lesniak MA. Insulin interactions with its receptors: experimental evidence for negative cooperativity. *Biochem. Biophys. Res. Commun.* 1973; 55(1): 154–161.
24. De Meyts P. Insulin/receptor binding: the last piece of the puzzle? What recent progress on the structure of the insulin/receptor complex tells us (or not) about negative cooperativity and activation. *Bioessays* 2015; 37(4): 389–397.
25. Roep BO, Thomaïdou S, Tienhoven vR, Zaldumbide A. Type 1 diabetes mellitus as a disease of the β -cell (do not blame the immune system?). *Nat. Rev. Endocrinol.* 2021; 17(3): 150–161.
26. DeFronzo RA, Ferrannini E, Groop L, et al. Type 2 diabetes mellitus. *Nat. Rev. Dis. Primers* 2015; 1(1): 1–22.
27. Ward CW, Menting JG, Lawrence MC. The insulin receptor changes conformation in unforeseen ways on ligand binding: sharpening the picture of insulin receptor activation. *Bioessays* 2013; 35(11): 945–954.

28. Menting JG, Ward CW, Margetts MB, Lawrence MC. A thermodynamic study of ligand binding to the first three domains of the human insulin receptor: Relationship between the receptor α -chain C-terminal peptide and the Site 1 insulin mimetic peptides. *Biochemistry* 2009; 48(23): 5492–5500.
29. Menting JG, Whittaker J, Margetts MB, et al. How insulin engages its primary binding site on the insulin receptor. *Nature* 2013; 493(7431): 241–245.
30. Vashisth H, Abrams CF. Docking of insulin to a structurally equilibrated insulin receptor ectodomain. *Proteins* 2010; 78(6): 1531–1543.
31. Vashisth H, Abrams CF. All-atom structural models of insulin binding to the insulin receptor in the presence of a tandem hormone-binding element. *Proteins* 2013; 81(6): 1017–1030.
32. Vashisth H. Flexibility in the insulin receptor ectodomain enables docking of insulin in crystallographic conformation observed in a hormone-bound microreceptor. *Membranes* 2014; 4(4): 730–746.
33. Vashisth H. Theoretical and computational studies of peptides and receptors of the insulin family. *Membranes* 2015; 5(1): 48–83.
34. Whittaker L, Hao C, Fu W, Whittaker J. High-affinity insulin binding: insulin interacts with two receptor ligand binding sites. *Biochemistry* 2008; 47(48): 12900–12909.
35. Rentería ME, Gandhi NS, Vinuesa P, Helmerhorst E, Mancera RL. A comparative structural bioinformatics analysis of the insulin receptor family ectodomain based on phylogenetic information. *PloS One* 2008; 3(11): e3667.
36. Šali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* 1993; 234(3): 779–815.
37. Martí-Renom MA, Stuart AC, Fiser A, Sánchez R, Melo F, Šali A. Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomol. Struct.* 2000; 29(1): 291–325.
38. Shen My, Sali A. Statistical potential for assessment and prediction of protein structures. *Protein Sci.* 2006; 15(11): 2507–2524.
39. Margreitter C, Petrov D, Zagrovic B. Vienna-PTM web server: a toolkit for MD simulations of protein post-translational modifications. *Nucleic Acids Res.* 2013; 41(W1): W422–W426.
40. Li H, Robertson AD, Jensen JH. Very fast empirical prediction and rationalization of protein pKa values. *Proteins: Struct., Funct., Bioinf.* 2005; 61(4): 704–721.

41. DeLano WL, others . Pymol: An open-source molecular graphics tool. *CCP4 Newsletter on Protein Crystallography* 2002; 40(1): 82–92.
42. Abraham MJ, Murtola T, Schulz R, et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 2015; 1: 19–25.
43. Reif MM, Winger M, Oostenbrink C. Testing of the GROMOS force-field parameter set 54A8: structural properties of electrolyte solutions, lipid bilayers, and proteins. *J. Chem. Theory Comput.* 2013; 9(2): 1247–1264.
44. Berendsen HJ, Postma JP, Gunsteren vWF, Hermans J. Interaction models for water in relation to protein hydration. In: Springer. 1981 (pp. 331–342).
45. Chester C, Friedman B, Ursell F. An extension of the method of steepest descents. In: . 53. Cambridge University Press. ; 1957: 599–611.
46. Darden T, York D, Pedersen L. Particle mesh Ewald: An N log (N) method for Ewald sums in large systems. *J. Chem. Phys.* 1993; 98(12): 10089–10092.
47. Hess B, Bekker H, Berendsen HJ, Fraaije JG. LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* 1997; 18(12): 1463–1472.
48. Bussi G, Zykova-Timan T, Parrinello M. Isothermal-isobaric molecular dynamics using stochastic velocity rescaling. *J. Chem. Phys.* 2009; 130(7): 074101.
49. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ. GROMACS: fast, flexible, and free. *J. Comput. Chem.* 2005; 26(16): 1701–1718.
50. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J. Mol. Graph.* 1996; 14(1): 33–38.
51. Daura X, Gademann K, Jaun B, Seebach D, Van Gunsteren WF, Mark AE. Peptide folding: when simulation meets experiment. *Angew. Chem. Int.* 1999; 38(1-2): 236–240.

FIGURE LEGENDS

FIGURE 1

Sequence alignment among human and cone snail venom insulins. The residue numbering at the top corresponds to the hIns sequence. The conserved cysteine residues are enclosed within orange boxes. The residues in cone snail insulins those are similar to hIns are shown in green. The post-translationally modified residues: γ -carboxylated glutamate (γ) and hydroxyproline (O) are shown in blue. The C-terminal amidation is represented by an asterisk (*).

FIGURE 2

Structural models of Con-Ins. (A) The cartoon representations of models of Con-Ins variants are depicted. The A- and B-chains of each cone snail insulin are depicted in red and cyan, respectively. The side-chains of cysteine residues forming disulfide bonds are depicted in yellow sticks. (B) Shown are the superimposed structures of the representative Con-Ins structure of the largest cluster from each simulation trajectory (darker colors) on the initial structure (lighter colors). The RMSD values (along with the standard deviations) computed relative to the initial structure in each MD simulation are depicted below the superimposed structure of each Con-Ins.

FIGURE 3

RMSD distributions from unbound and receptor-bound MD simulations of each insulin. Shown are the normalized frequency distributions of the RMSD for the unbound insulins (blue) and insulins bound to μ IR (pink). Data represent the averaged RMSD, computed for the backbone atoms relative to the initial conformation, from three independent MD simulations.

FIGURE 4

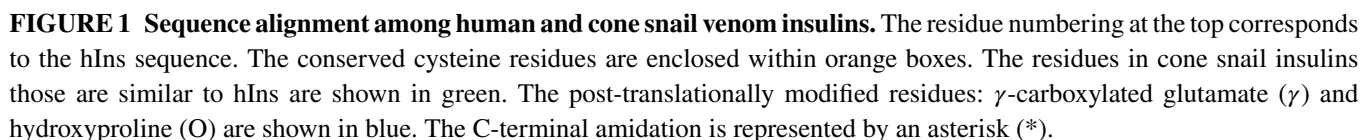
Per-residue non-bonded binding energy contribution and major structural interactions between insulins and μ IR. (A) The total non-bonded (Lennard-Jones and Coulomb interactions) energy of each residue of insulin with the μ IR are depicted. A more negative value for a given residue indicates a stronger interaction. The residues belonging to the A-chain and B-chain of insulins are depicted in cyan and magenta bars, respectively. (B-H) Major interactions between hIns, (I-L) Con-Ins-G1, (M-N) Con-Ins-G3, (O-Q) Con-Ins-K1, (R) Con-Ins-K2, and (S-U) Con-Ins-T1A with the μ IR are shown. The carbon atom of interacting residues of the insulins, L1, and α CT are labeled and depicted in pink, white, and orange stick representations, respectively. Shown are the common interactions in at least two representative structures of the dominant cluster from three independent MD simulations of each insulin- μ IR complex. See also Figures S5 and S6 for detailed interaction patterns.

FIGURE 5

Potential residue substitutions. The comparison of non-bonded energy and the interaction pattern of key Con-Ins residues is shown. (A) The non-bonded energy of hIns residues (cyan), Con-Ins residues (blue), and the difference of energy (ΔE) of hIns residues from the energy of Con-Ins residues (dark blue) is shown. The position and the likely substitution of hIns residues are depicted along the x-axis. For example, the ($A4^{Glu}_{Gla}$) signifies Glu residue at the fourth position of the A-chain of hIns can be substituted by the Gla residue derived from Con-Ins. The representative structure of hIns (represented by transparent cyan cartoon model) and Con-Ins (represented by transparent red cartoon model) from the simulation trajectory are superimposed over each other to identify the difference in the interactions of hIns residues and the equivalent residues of Con-Ins at (B) position 4 of the A-chain, (C) position 21 of the A-chain, (D) position 12 of the B-chain, (E) position 17 of the B-chain, (F-G) position 20 of the B-chain, and (H) position 22 of the B-chain. The interacting residues of hIns and Con-Ins are represented in blue and red sticks, respectively.

FIGURE 6

Sequence alignment and per residue energy contribution of insulin analogs. (A) Sequence alignment of hIns, analog 1, and analog 2 are shown. The substituted residues are shown in blue. The single residue difference within analog 1 and analog 2 sequences is highlighted in orange. The total non-bonded energy difference (ΔE) of each residue of (B) analog 1 and (C) analog 2 with the μIR from the respective residues of hIns are depicted. The substitutions belonging to the A-chain and B-chain of insulins are depicted in blue and purple bars, respectively. The representative structure of hIns (represented by transparent cyan cartoon model) and Con-Ins (represented by a transparent red cartoon) from the simulation trajectory are superimposed over each other to identify the differences in the interactions of hIns residues and the equivalent residues of the analogs at (D) position 4 of the A-chain (P_{A4}), (E) position 21 of the A-chain (P_{A21}), (F) position 12 of the B-chain (P_{B12}), (G) position 17 of the B-chain (P_{B17}), (H-I) position 20 of the B-chain (P_{B20}), and (H) position 22 of the B-chain (P_{B22}). The interacting residues of hIns and analogs are represented in blue and red sticks, respectively.



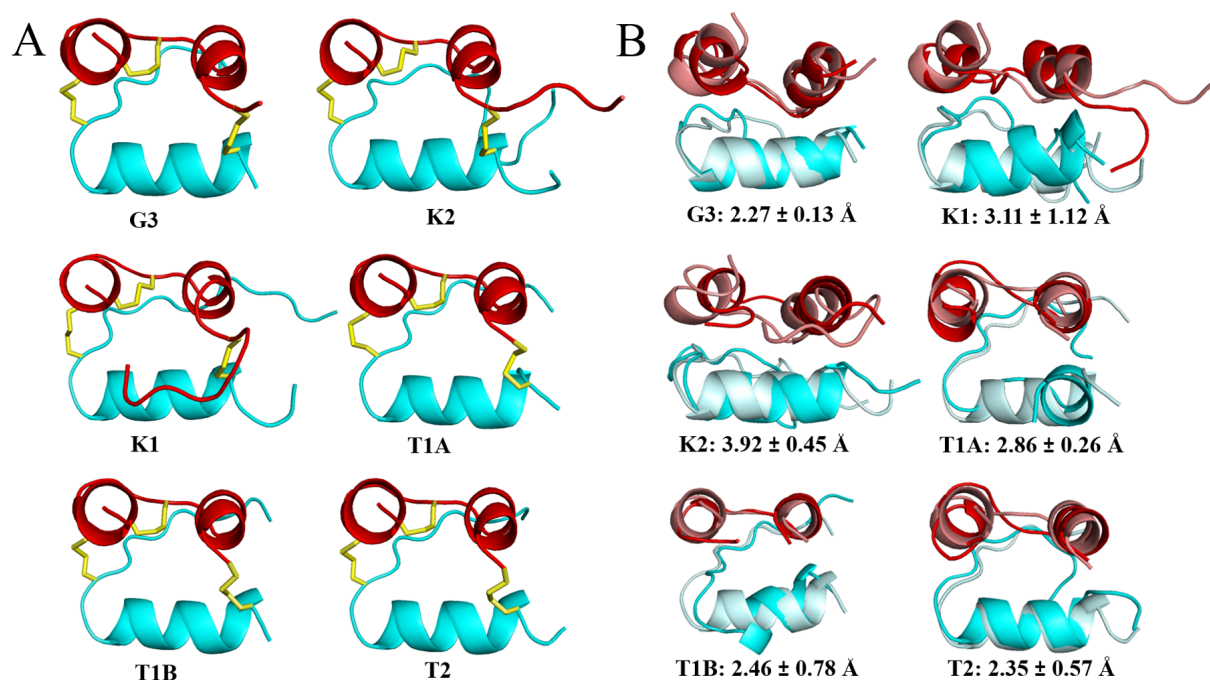


FIGURE 2 Structural models of Con-Ins. (A) The cartoon representations of models of Con-Ins variants are depicted. The A- and B-chains of each cone snail insulin are depicted in red and cyan, respectively. The side-chains of cysteine residues forming disulfide bonds are depicted in yellow sticks. (B) Shown are the superimposed structures of the representative Con-Ins structure of the largest cluster from each simulation trajectory (darker colors) on the initial structure (lighter colors). The RMSD values (along with the standard deviations) computed relative to the initial structure in each MD simulation are depicted below the superimposed structure of each Con-Ins.

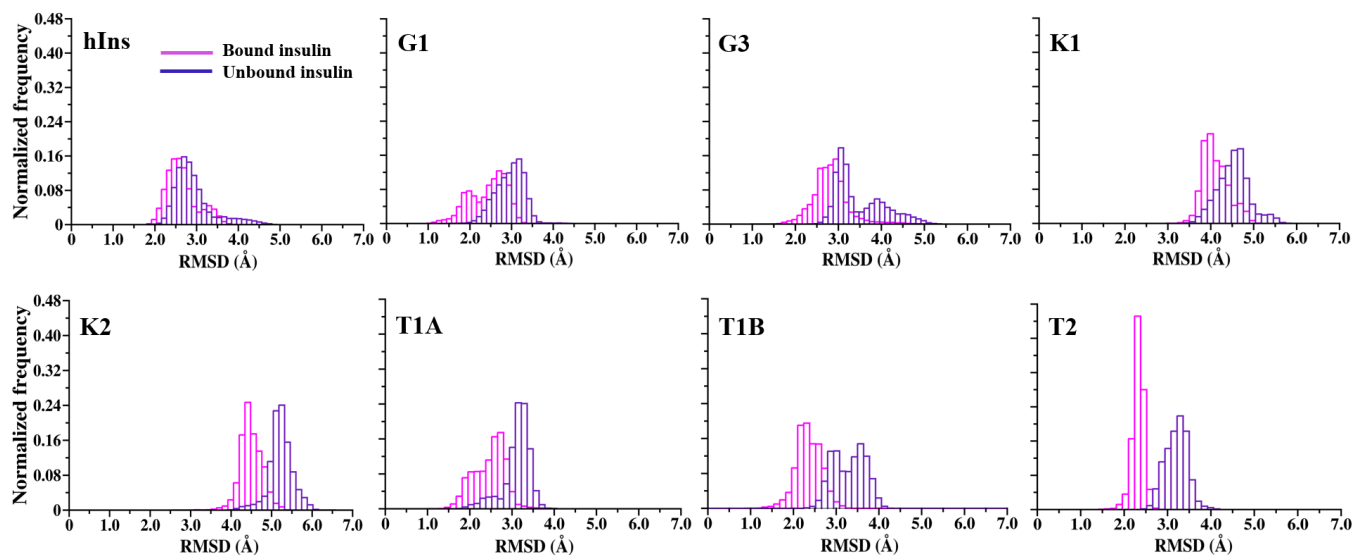


FIGURE 3 RMSD distributions from unbound and receptor-bound MD simulations of each insulin. Shown are the normalized frequency distributions of the RMSD for the unbound insulins (blue) and insulins bound to μ IR (pink). Data represent the averaged RMSD, computed for the backbone atoms relative to the initial conformation, from three independent MD simulations.

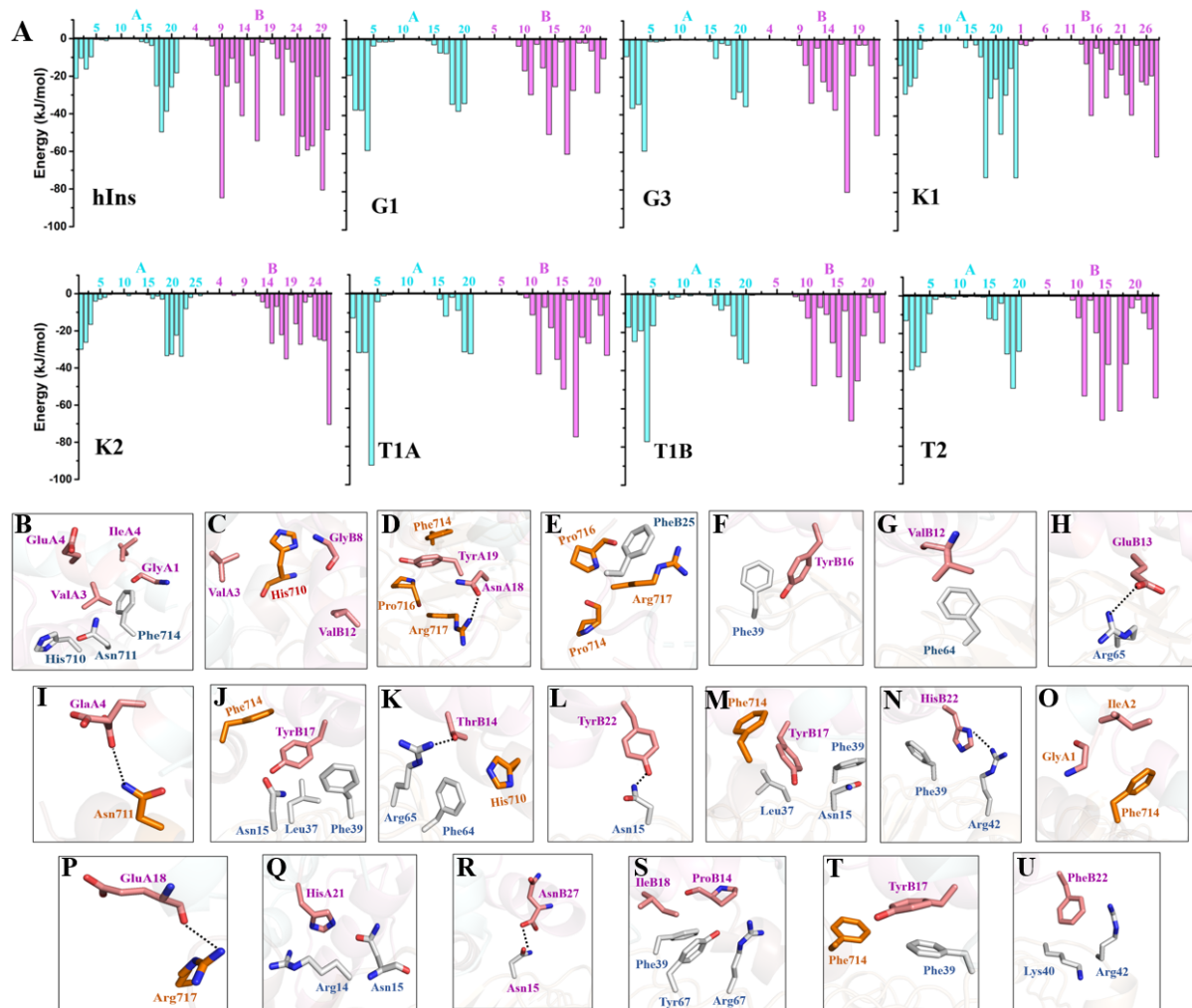


FIGURE 4 Per-residue non-bonded binding energy contribution and major structural interactions between insulins and μ IR. (A) The total non-bonded (Lennard-Jones and Coulomb interactions) energy of each residue of insulin with the μ IR are depicted. A more negative value for a given residue indicates a stronger interaction. The residues belonging to the A-chain and B-chain of insulins are depicted in cyan and magenta bars, respectively. (B-H) Major interactions between hIns, (I-L) Con-Ins-G1, (M-N) Con-Ins-G3, (O-Q) Con-Ins-K1, (R) Con-Ins-K2, and (S-U) Con-Ins-T1A with the μ IR are shown. The carbon atom of interacting residues of the insulins, L1, and α CT are labeled and depicted in pink, white, and orange stick representations, respectively. Shown are the common interactions in at least two representative structures of the dominant cluster from three independent MD simulations of each insulin- μ IR complex. See also Figures S5 and S6 for detailed interaction patterns.

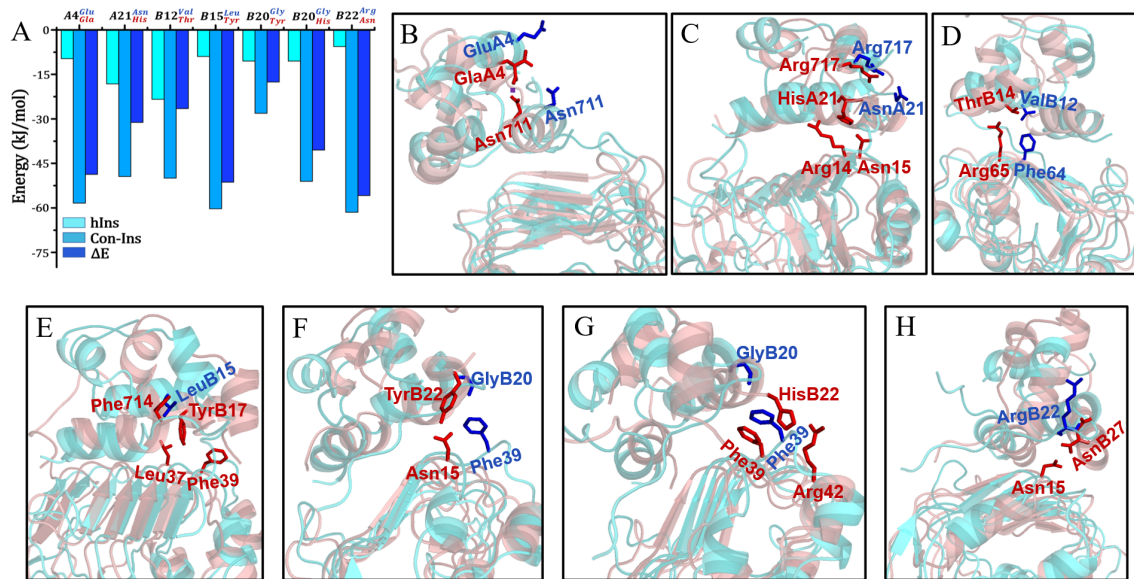


FIGURE 5 Potential residue substitutions. The comparison of non-bonded energy and the interaction pattern of key Con-Ins residues is shown. (A) The non-bonded energy of hIns residues (cyan), Con-Ins residues (blue), and the difference of energy (ΔE) of hIns residues from the energy of Con-Ins residues (dark blue) is shown. The position and the likely substitution of hIns residues are depicted along the x-axis. For example, the (A4^{Glu}) signifies Glu residue at the fourth position of the A-chain of hIns can be substituted by the Glu residue derived from Con-Ins. The representative structure of hIns (represented by transparent cyan cartoon model) and Con-Ins (represented by transparent red cartoon model) from the simulation trajectory are superimposed over each other to identify the difference in the interactions of hIns residues and the equivalent residues of Con-Ins at (B) position 4 of the A-chain, (C) position 21 of the A-chain, (D) position 12 of the B-chain, (E) position 17 of the B-chain, (F-G) position 20 of the B-chain, and (H) position 22 of the B-chain. The interacting residues of hIns and Con-Ins are represented in blue and red sticks, respectively.

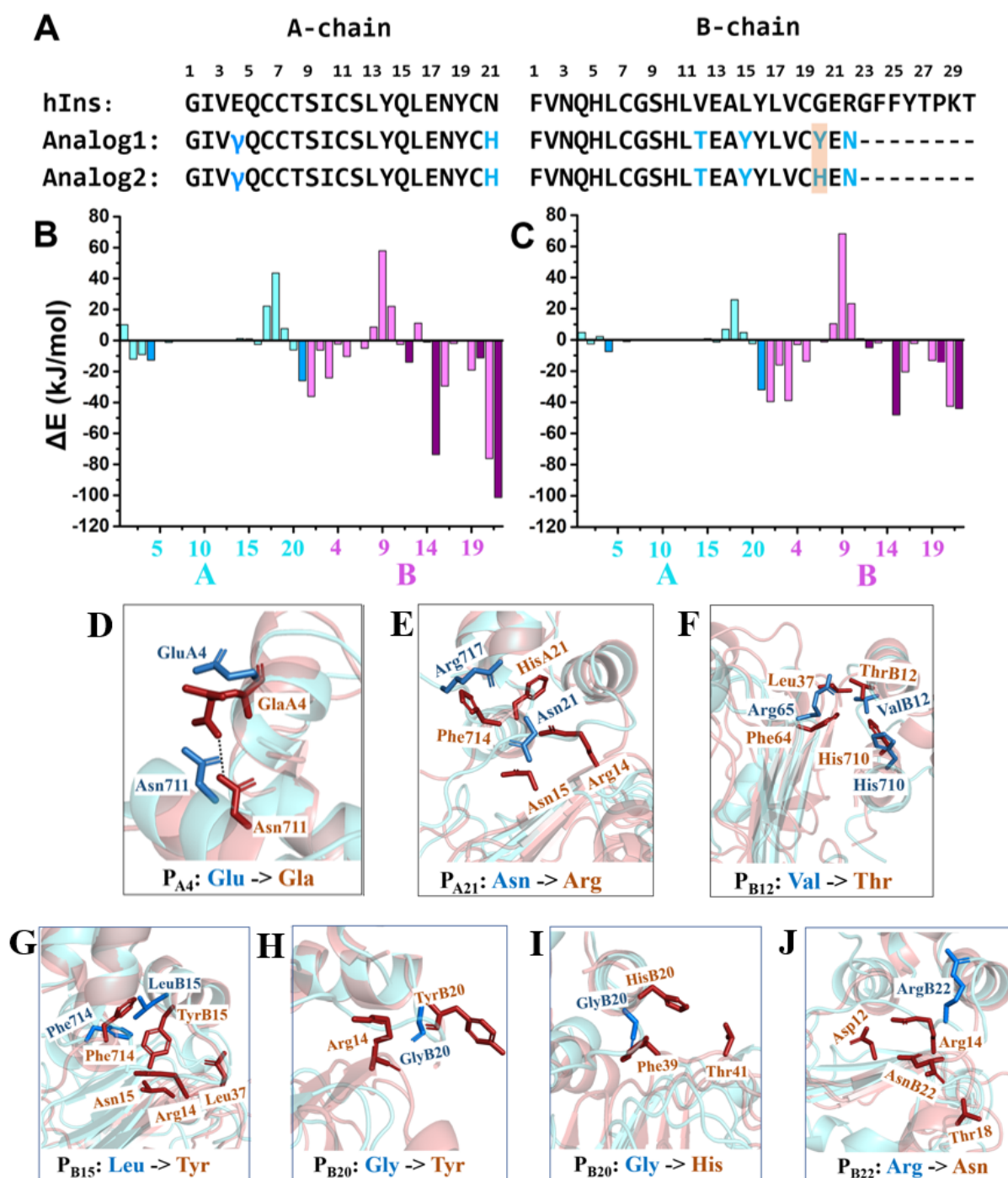


FIGURE 6 Sequence alignment and per residue energy contribution of insulin analogs. (A) Sequence alignment of hIns, analog 1, and analog 2 are shown. The substituted residues are shown in blue. The single residue difference within analog 1 and analog 2 sequences is highlighted in orange. The total non-bonded energy difference (ΔE) of each residue of (B) analog 1 and (C) analog 2 with the μ IR from the respective residues of hIns are depicted. The substitutions belonging to the A-chain and B-chain of insulins are depicted in blue and purple bars, respectively. The representative structure of hIns (represented by transparent cyan cartoon model) and Con-Ins (represented by a transparent red cartoon) from the simulation trajectory are superimposed over each other to identify the differences in the interactions of hIns residues and the equivalent residues of the analogs at (D) position 4 of the A-chain (P_{A4}), (E) position 21 of the A-chain (P_{A21}), (F) position 12 of the B-chain (P_{B12}), (G) position 17 of the B-chain (P_{B17}), (H-I) position 20 of the B-chain (P_{B20}), and (H) position 22 of the B-chain (P_{B22}). The interacting residues of hIns and analogs are represented in blue and red sticks, respectively.

.
