



OPEN

Designing and computational analyzing of chimeric long-lasting GLP-1 receptor agonists for type 2 diabetes

Maryam Ehsasatvatan & Bahram Baghban Kohnehrouz

Glucagon-like peptide-1 (GLP-1) is an intestinally derived incretin that plays a vital role in engineering the biological circuit involved in treating type 2 diabetes. Exceedingly short half-life (1–2 min) of GLP-1 limits its therapeutic applicability, and the implication of its new variants is under question. Since albumin-binding DARPin as a mimetic molecule has been reported to increase the serum half-life of therapeutic compounds, the interaction of new variants of GLP-1 in fusion with DARPin needs to be examined against the GLP-1 receptor. This study was aimed to design stable and functional fusion proteins consisting of new protease-resistant GLP-1 mutants (mGLP1) genetically fused to DARPin as a critical step toward developing long-acting GLP-1 receptor agonists. The stability and solubility of the engineered fusion proteins were analyzed, and their secondary and tertiary structures were predicted and satisfactorily validated. Molecular dynamics simulation studies revealed that the predicted structures of engineered fusion proteins remained stable throughout the simulation. The relative binding affinity of the engineered fusion proteins' complex with human serum albumin and the GLP-1 receptor individually was assessed using molecular docking analyses. It revealed a higher affinity compared to the interaction of the individual GLP-1 and HSA-binding DARPin with the GLP-1 receptor and human serum albumin, respectively. The present study suggests that engineered fusion proteins can be used as a potential molecule in the treatment of type 2 diabetes, and this study provides insight into further experimental use of mimetic complexes as alternative molecules to be evaluated as new bio-breaks in the engineering of biological circuits in the treatment of type 2 diabetes.

Abbreviations

3D	Three-dimensional
DARPin	Designed ankyrin repeat protein
ECD	Extracellular domain
GLP-1	Glucagon like peptide-1
GR	Gyration radius
HSA	Human serum albumin
MD	Molecular dynamics
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
SASA	Solvent-accessible surface area
T2DM	Type 2 diabetes mellitus

Type 2 diabetes mellitus (T2DM) is a chronic, systemic metabolic disorder defined by persistent hyperglycemia and disturbances in carbohydrate, lipid, and protein metabolism due to a relative deficiency of insulin¹. The global prevalence of this disease in humans has been estimated at 415 million adults in 2017 and is predicted to rise to 629 million by 2045². Despite the availability of various T2DM treatments, such as diet, exercise, anti-diabetic medications, and subcutaneous insulin injection, glycemic control frequently fails due to inefficacy and poor patient compliance³.

Department of Plant Breeding and Biotechnology, Faculty of Agriculture, University of Tabriz, Tabriz 51666, Iran.
 email: bahramrouz@yahoo.com

Glucagon-like peptide-1 (GLP-1) and its analogs have emerged as promising medications for treating type 2 diabetes in recent years⁴. GLP-1 is a 30-residue incretin peptide hormone secreted into the bloodstream from the gut, primarily by distal ileum enteroendocrine L-cells in response to nutrient ingestion^{5,6}. GLP-1 regulates blood glucose levels through several mechanisms, including stimulation of glucose-dependent secretion of insulin, suppression of pancreatic glucagon release, satiety enhancement, delayed stomach emptying, and reduction of energy intake^{7–9}. These diverse mechanisms of action underpin the exceptional success of GLP-1 peptide analogues as diabetes and obesity treatments with little or no risk of hypoglycemia. However, the physiological half-life of GLP-1 in circulation is very short (less than 2 min), which can be attributed to both rapid proteolytic cleavage of the amide bond of alanine at position 8 (Ala8) in the N-terminus by ubiquitous dipeptidyl peptidase IV (DPP-IV), which results in the removal of the two N-terminal amino acids leading to the formation of a truncated inactive form^{10,11} and clearance by the kidney as well. As a result, significant efforts are being made to develop long-acting GLP-1R agonists, through the substitution of a susceptible amino acid in GLP-1 into a mutant to confer resistance to DPP-IV degradation¹² and the significantly prolonging of GLP-1 half-life via the virtue of the neonatal Fc receptor (FcRn) recycling mechanism through binding to IgG-Fc¹³ or human serum albumin (HSA)¹⁴.

Fusion technology, which involves linking a target protein to a more stable protein, is one of the strategies used in protein-therapeutic optimization. Fusion proteins are a type of protein that links two or more distinct protein domains into a single molecule. Over the years, researchers have used recombinant DNA technology to mimic nature's strategy for creating artificial fusion proteins, such as the use of peptide or protein tags to enable affinity chromatography techniques for one-step protein purification^{15,16}, fluorescent proteins as a biosensor for monitoring signaling molecules, or as a reporter in bioimaging¹⁷. Recently, synthetic biology applications have used artificial fusion proteins as novel protein switches¹⁸. Furthermore, synthetic antibody fragments and recombinant fusion proteins have developed as a new class of therapeutic medicines that can be used to address a wide range of issues in protein-drug development¹⁹. This approach has been proven to enhance protein-drug stability, activity, catalytic efficiency, and solubility. Furthermore, the fusion-protein method, which involves linking a target protein to a protein with a longer half-life, is employed to inhibit rapid renal clearance²⁰.

Serum albumin binding has long been used in pharmacokinetic engineering to alter the pharmacokinetic features of small proteins. Human serum albumin (HSA) is a 66 kDa plasma protein with a long half-life of roughly 19–22 days in humans²¹. It is regularly utilized in fusion technology due to its prolonged half-life. Binding to HSA causes an increase in hydrodynamic volume, which slows renal clearance^{13,14}. A further strategy to benefit from the extended plasma circulation of HSA is to derivatize GLP-1 with HSA-binding moieties. Several approaches have been described, including the use of natural albumin-binding proteins such as antibody fragments binding to serum albumin^{22–24}, albumin-binding peptides²⁵, the streptococcal protein G-derived albumin-binding domain^{26,27}, and serum albumin-binding DARPin²⁸.

Designed ankyrin repeat proteins are a promising class of small non-immunoglobulin binding proteins that bind target proteins with high affinity and specificity^{29,30}. DARPin domains have desirable biophysical features, such as high thermal and thermodynamic stability, and can be expressed in high quantities in yeast, bacteria, and tobacco chloroplast expression systems^{31–33}. Serum albumin-binding DARPin domains can be used to extend the terminal half-lives of next-generation protein therapeutics²⁸. MP0250 is a tri-specific DARPin drug that contains one hepatocyte growth factor (HGF)-binding DARPin, one vascular endothelial growth factor (VEGF)-binding DARPin, and two HSA-binding DARPins to extend its half-life³⁴. MP0250 has a half-life of about two weeks in humans³⁵. This shows that DARPins with a high affinity for HSA can be employed as a universal carrier to extend their fusion half-life.

The current study aimed to develop long-acting protease-resistant GLP-1R agonists with the potential for administration in type 2 diabetes mellitus treatment by designing and computationally analyzing the fusion proteins of the native or protease-resistant mutant GLP-1 and an HSA-binding DARPin. An earlier study with albiglutide demonstrated that the presence of albumin directly next to GLP-1 could significantly reduce its activity³⁶. Therefore, we engineered a chimeric protein by fusing GLP-1 to an HSA-binding DARPin with a high affinity for albumin that takes advantage of the extended plasma circulation of HSA to develop a long-acting variant of GLP-1. We incorporated a rigid helical linker between the GLP-1 molecule and the albumin-binding DARPin to keep protein moieties at a distance efficiently.

Since most of a protein's biological functions depend on its three-dimensional (3D) structure, appropriate folding, stability, and interaction between domains must be considered while creating multi-domain recombinant proteins. Unlike a single-domain protein, recombinant multi-domain proteins are more likely to misfold or assemble in an incorrect 3D shape because of the interaction between their distinct domains. Therefore, the *in silico* study of multi-domain proteins, followed by protein modeling and molecular dynamics (MD) simulations, is an essential stage in recombinant protein production efforts. We conducted a molecular modeling study of the engineered fusion proteins that could be used in T2DM treatment using the trRosetta protein structure prediction tool to obtain their 3D structures. The energy minimization and molecular dynamics simulation were done for the optimization of the models and monitoring their structural fluctuations. Also, the potential affinity of each moiety of the engineered fusion proteins against their specific targets was explored.

Results

Construct of the fusion proteins

Several flexible and rigid linkers were examined to select the top linker that maintains the functioning of the standard structure of the two domains of the fusion proteins (data not shown). Finally, the rigid helical (EAAAK)₃ was used as the linker to maintain the spatial distance between the domains and their independent functions (Fig. 1A)³⁷. The engineered fusion proteins became 168 amino acids in length, encoded by 504

3D structure prediction algorithm generates the five best models, with TM scores ranging from 0 to 1. The best model is the one with the highest TM score. The highest TM scores for predicted models of engineered fusion proteins were 0.928, 0.946, and 0.930 for nGLP1-DARPin, mGLP1-DARPin-1, and mGLP1-DARPin-2, respectively. The predicted models for engineered fusion proteins showed a protein with two separate domains connected through a small helical linker. The (AEEEK)₃ rigid linker keeps domains apart and ensures a rigid separation of the individual domains of the fusion proteins. The NMR analysis of GLP-1 structures revealed two α -helix between residues 13–20 and 24–35 linked by a short linker region composed of residues 21–23, while the N-terminal residues 7–13 are unstructured (Fig. 2A)³⁸. HSA-binding DARPin consists of two internal ankyrins with flanking constant N- and C-terminal capping repeats, each repeat forming a β -turn followed by two anti-parallel α -helix and an unstructured loop that binds to the β -turn of the next ankyrin repeat (Fig. 2B)²⁸. As predicted by trRosetta, the fusion of HSA-binding DARPin to GLP-1 using a rigid helical linker does not alter the tertiary structure of the fusion partner, and each moiety of the fusion proteins will most likely retain their conformational properties and relevant biological functions as the native protein. The pre-simulated 3D structure of the mGLP1-DARPin-2 fusion protein, as an example, is shown in Fig. 2C.

The trRosetta-predicted 3D structures of the fusion proteins were assessed by the Ramachandran plot, ERRAT, and ProSA Web server. The Ramachandran plot showed that the proportion of residues in three engineered fusion protein models in the favored region was 94.00% and 6.00% in the allowed region. (Fig. 3A–C). The results of the Ramachandran plot indicated that the overall quality of the predicted models of the engineered fusion proteins is sufficiently good. The overall quality factors of the refined models on the ERRAT server were 99.375% for nGLP1-DARPin and mGLP1-DARPin-1, and 100% for mGLP1-DARPin-2 (Fig. S3). According to the ERRAT results, the final refined models have the highest overall quality factor. And also, the ProSA-calculated Z-scores for nGLP1-DARPin, mGLP1-DARPin-1, and mGLP1-DARPin-2 were –5.45, 5.39, and –5.46, respectively, which were inside a characteristic range for native proteins (Fig. 3D–F).

Molecular dynamics

We performed three replica molecular dynamic simulations for 500 ns to evaluate the dynamic behavior of the engineered fusion proteins as described in the methods. The simulation bridges the gap between theory and experimentation. The theory was tested through simulation using computer-generated models that provided insight into the utmost stability of complex molecules and the potential for strong interactions between molecules^{39,40}. The GROMACS program was used to assess the stability of the engineered fusion proteins, and the PyMOL package was used to visualize their structure. The MD simulation output data were evaluated based on RMSD⁴¹, RMSF, SASA, Gyration radius⁴², and the average number of H-bonds, with the results summarized in Table 2 and Fig. 4.

The average distance between the atoms that make up a protein's backbone relative to its initial structure is measured by RMSD. The RMSD plots for predicted models of the fusion proteins reached a constant value after 130 ns and flattened with a minimal deviation for the rest of the simulation time (Fig. 4A). The RMSD mean value of the nGLP1-DARPin fusion protein was 3.4 Å, which was slightly higher than for two fusion proteins containing modified GLP-1 (Fig. 4A, Table 2). This indicates that the stability of the mutant systems is relatively higher than that of the native GLP-1-containing system.

The root mean square fluctuation (RMSF) of C α atoms in engineered fusion proteins was determined using final trajectories to analyze the mean atomistic motions of distinct residues. More flexibility (more conformational fluctuation) is indicated by higher RMSF values, whereas less fluctuation in the structure is indicated by lower RMSF values. Comparing the fluctuations of the RMSF value between different fusion protein structures, as seen in Fig. 4B, we found that the GLP-1 moiety of the fusion protein is more flexible than other regions of the fusion protein. Among the most flexible regions, we observed more flexibility in modified residues of the mGLP1-DARPin-1 and mGLP1-DARPin-2 fusion proteins compared with the fusion protein containing native GLP-1.

The radius of gyration indicates the level of compaction in the structure and the folding properties of the protein. The gyration radius reached the plateau after 130 ns and remained constant over the timeframe with an average value of 2.32, 2.73, and 2.7 nm for nGLP1-DARPin, mGLP1-DARPin-1, and mGLP1-DARPin-2, respectively, which signifies that all systems remained compact across the 500 ns MD simulation (Fig. 4C). The

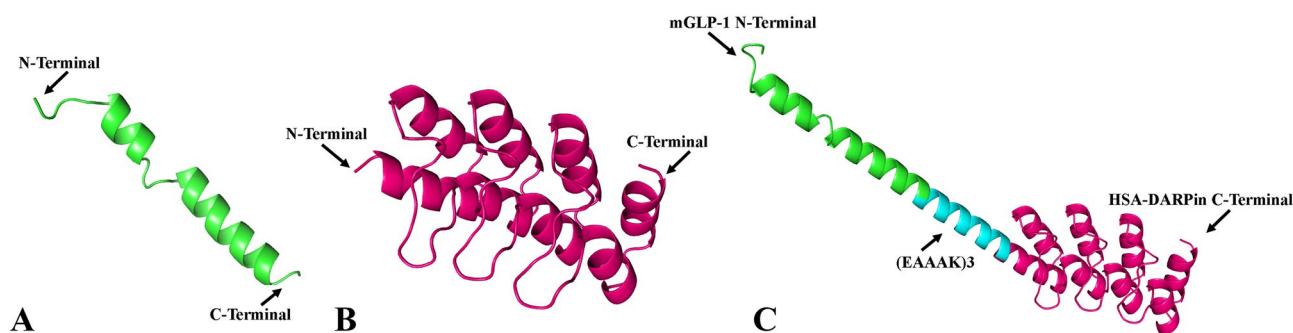


Figure 2. Three-dimensional structures of (A) GLP-1 (PDB id: 1D0R) and (B) HSA-binding DARPin (PDB id: 4GRG) from the RCSB PDB database; (C) trRosetta predicted model of the fusion protein containing GLP-1 (green), HSA-binding DARPin (magenta), and (EAAAK)₃ rigid helical linker (cyan).

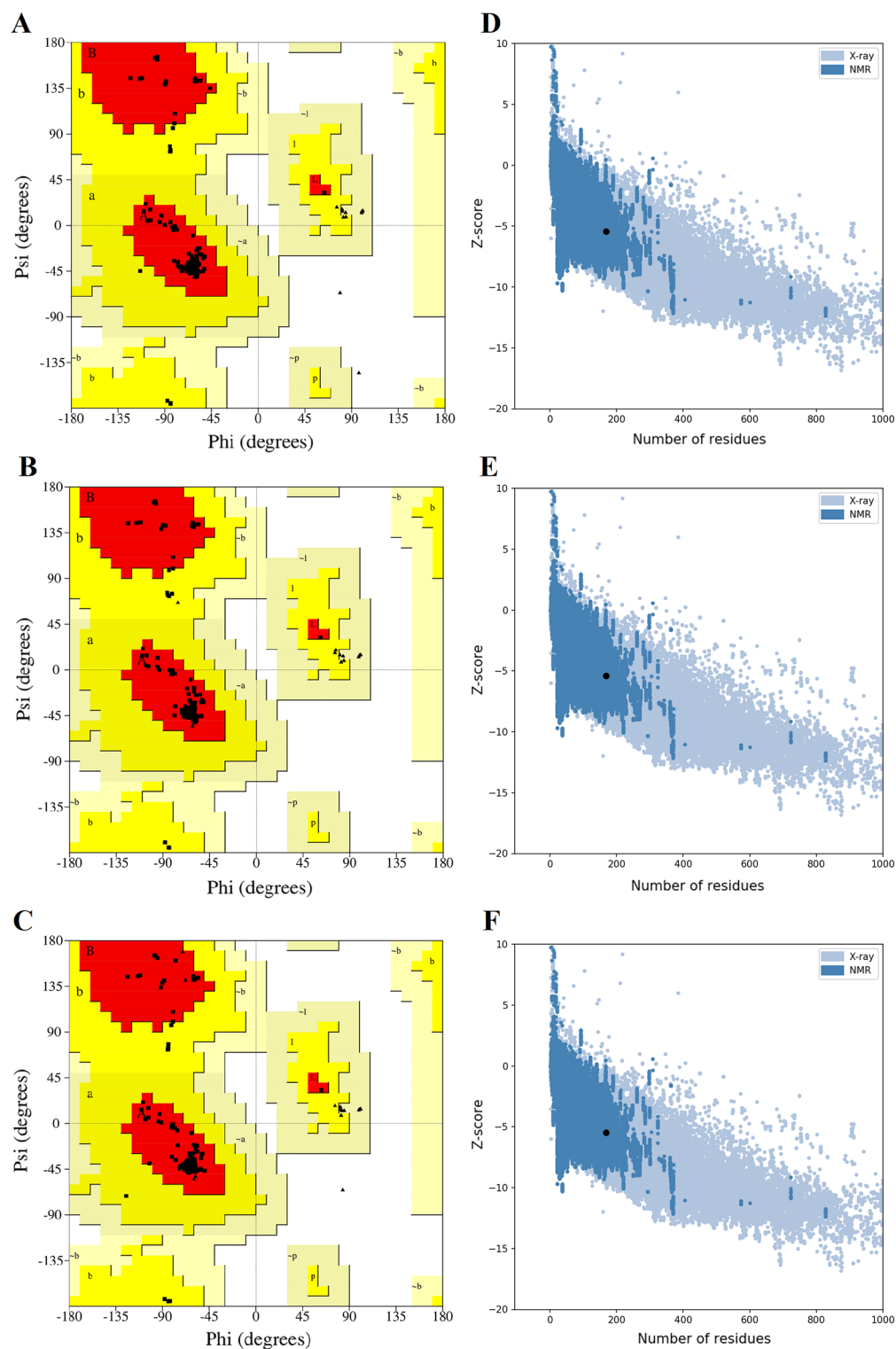


Figure 3. Ramachandran plot of the (A) nGLP1-DARPin, (B) mGLP1-DARPin-1, and (C) mGLP1-DARPin-2 fusion proteins generated by the PROCHECK program. The regions in the plots are labeled as follows: [A, B, L]: residue in most favored regions (red), [a, b, l, p]: residue in additional allowed regions (yellow), [\sim a, \sim b, \sim l, \sim p]: residue in generously allowed regions (beige), the disallowed region is in white, glycine residues were shown as triangles; The majority of residues in the core and allowed regions are due to the stereochemical quality of the protein structure; The ProSA result of the engineered fusion proteins showed a Z-score of -5.45 , -5.39 and -5.46 (black dot) for (D) nGLP1-DARPin, (E) mGLP1-DARPin-1, and (F) mGLP1-DARPin-2 fusion proteins, respectively, which was inside the range characteristic for native proteins.

Parameters	nGLP1-DARPin	mGLP1-DARPin-1	mGLP1-DARPin-2
Protein backbone RMSD	0.34 ± 0.022	0.29 ± 0.017	0.33 ± 0.024
Protein backbone RMSF	0.264 ± 0.121	0.265 ± 0.133	0.268 ± 0.159
Gyration radius graph	2.32 ± 0.025	2.73 ± 0.023	2.70 ± 0.027
Solvent Accessible Surface Area	106.56 ± 1.54	105.93 ± 1.75	107.64 ± 2.02
Hydrogen bond formation/deformation	122.53 ± 5.73	124.84 ± 5.85	123.79 ± 5.80

Table 2. Results of the last 500 ns molecular dynamic simulation for the fusion proteins. Data represent mean ± SD. *RMSD* root mean square deviation, *RMSF* root mean square fluctuation; *Rg* radius of gyration.

higher *Rg* values of the mGLP1-DARPin-1 and mGLP1-DARPin-2 fusion proteins than the nGLP1-DARPin indicate loose packing of the protein structure in the mutant systems, which means more flexible conformation. In theory, changes in protein accessibility to solvents can be determined by computing the solvent-accessible surface area (SASA). The SASA plots of all engineered fusion proteins are shown in Fig. 4D, and their average quantities are depicted in Table 2. During the simulations, the SASA of the proteins would naturally increase as hydration of the hydrophobic core occurred during unfolding, disrupting hydrophobic interactions among non-polar residues. The fusion protein's SASA plot revealed slightly diminishing trends, indicating reduced exposure of the hydrophobic core to solvation as unfolding progressed, making the proteins more stable (Fig. 4D). The average number of H-bonds in each frame over time corresponds to the number of hydrogen bonds formed or broken throughout the molecular simulation. Figure 4E displayed a stable fluctuation in the pattern of the number of hydrogen bonds formed in all systems, and this number was practically constant during the 500 ns of MD simulation in all systems, suggesting the stability of the molecule structures. The accuracy of our engineered fusion protein structures was validated by the fusion protein simulation results analysis.

The calculation of the free energy landscape is based on the potential of mean force (PMF), which is defined by $\Delta G = -k_B T \ln p(x, y)$ where temperature is represented by *T* and k_B is the Boltzmann constant. Similarly, *x* and *y* represent *Rg* and RMSD, respectively. In Fig. 5, 2D-PMF profiles are displayed. Our findings showed that for the nGLP1-DARPin, mGLP1-DARPin-1, and mGLP1-DARPin-2, the most stable configurations appeared approximately at (RMSD = 0.35 nm and *Rg* = 2.35 nm), (RMSD = 0.25 nm and *Rg* = 2.74 nm), and (RMSD = 0.25 nm and *Rg* = 2.75 nm), respectively. Based on the results of the MF profile computation, it can be concluded that the changed residues of the mGLP1-DARPin-1 and mGLP1-DARPin-2 fusion proteins are more stable than those of the native protein.

Protein–protein interaction of fusion proteins

MD results indicated that all engineered fusion proteins may retain their biological activity. Using molecular docking, we aimed to examine whether each moiety of the engineered fusion proteins could bind to its receptor. The frame with the lowest RMSD value with the average structure from the last 50 ns of each trajectory from all replicas of MD simulations of fusion proteins was selected for further docking steps. We used the crystallography structures of the extracellular domain of the GLP-1 receptor (PDB ID: 3IOL) and human serum albumin (PDB ID: 1A06) to enter the docking steps. We used the ClusPro 2.0 online server for protein–protein docking analysis and the PRODIGY server to calculate binding free energy for each engineered fusion protein. As a control, GLP-1 (PDB ID: 1D0R) with the GLP-1 receptor docking analyses, and the HSA-binding DARPin (trRosetta predicted structure) with human serum albumin have been conducted.

We used the docking energy score as the selection criterion for the best-docked complex, so that the lower the binding energy scores, the better the binding affinity between the two proteins. The results indicated that all engineered fusion proteins showed a relatively high docking score, suggesting that they may have lower energy constraints for binding (Table 3). Regarding binding affinity, docking analysis of engineered fusion proteins with HSA showed higher quantities than the GLP-1 receptor, suggesting their highest potential for binding to HSA (Table 3). This can be supported by the highest number of hydrogen bonds in the DARPin moiety of fusion proteins established with HSA (Fig. 6). In addition, among the fusion proteins, nGLP1-DARPin had a higher binding affinity to HSA ($\Delta G = -9.2$ kcal mol⁻¹) compared to the mGLP1-DARPin-1 ($\Delta G = -8.1$ kcal mol⁻¹) and mGLP1-DARPin-2 ($\Delta G = -8.3$ kcal mol⁻¹) fusion proteins, which is comparable with the individual DARPin complexed with HSA ($\Delta G = -9.6$ kcal mol⁻¹). In the docking analysis with the GLP-1 receptor, it can be seen that the nGLP1-DARPin fusion protein had a higher binding affinity to the GLP-1 receptor ($\Delta G = -6.1$ kcal mol⁻¹) compared with the mGLP1-DARPin-1 ($\Delta G = -5.5$ kcal mol⁻¹) and mGLP1-DARPin-2 ($\Delta G = -5.8$ kcal mol⁻¹) fusion proteins (Table 3, Fig. 7). Regarding the modified GLP-1 containing fusion proteins, mGLP1-DARPin-2 with Aln8Gly, Lys26Gn, and Lys34-Asp substitutions had a higher binding affinity to the HSA and GLP-1 receptor than GLP1-DARPin-1 with an Aln8Gly substitution. The findings demonstrated that the interactions between GLP-1/GLP-1R and HSA-DARPin/HSA persisted even after the amino acid substitution in GLP-1 and the formation of the fusion protein.

The complexes with a higher docking score were selected and molecular dynamically simulated using GROMACS 2019.2 software. The RMSD, representing structural variations during the simulation period compared to the time zero state, was obtained. The RMSD plot of the fusion proteins complex with HSA demonstrates that all three systems reached a stable state after 50 ns of the MD simulation. No significant fluctuations were observed for the rest of the simulation period, with average RMSD values of 0.47 ± 0.06 , 0.58 ± 0.04 , and 0.48 ± 0.12 for the nGLP1-DARPin, mGLP1-DARPin-1, and mGLP1-DARPin-2 complexed with HSA,

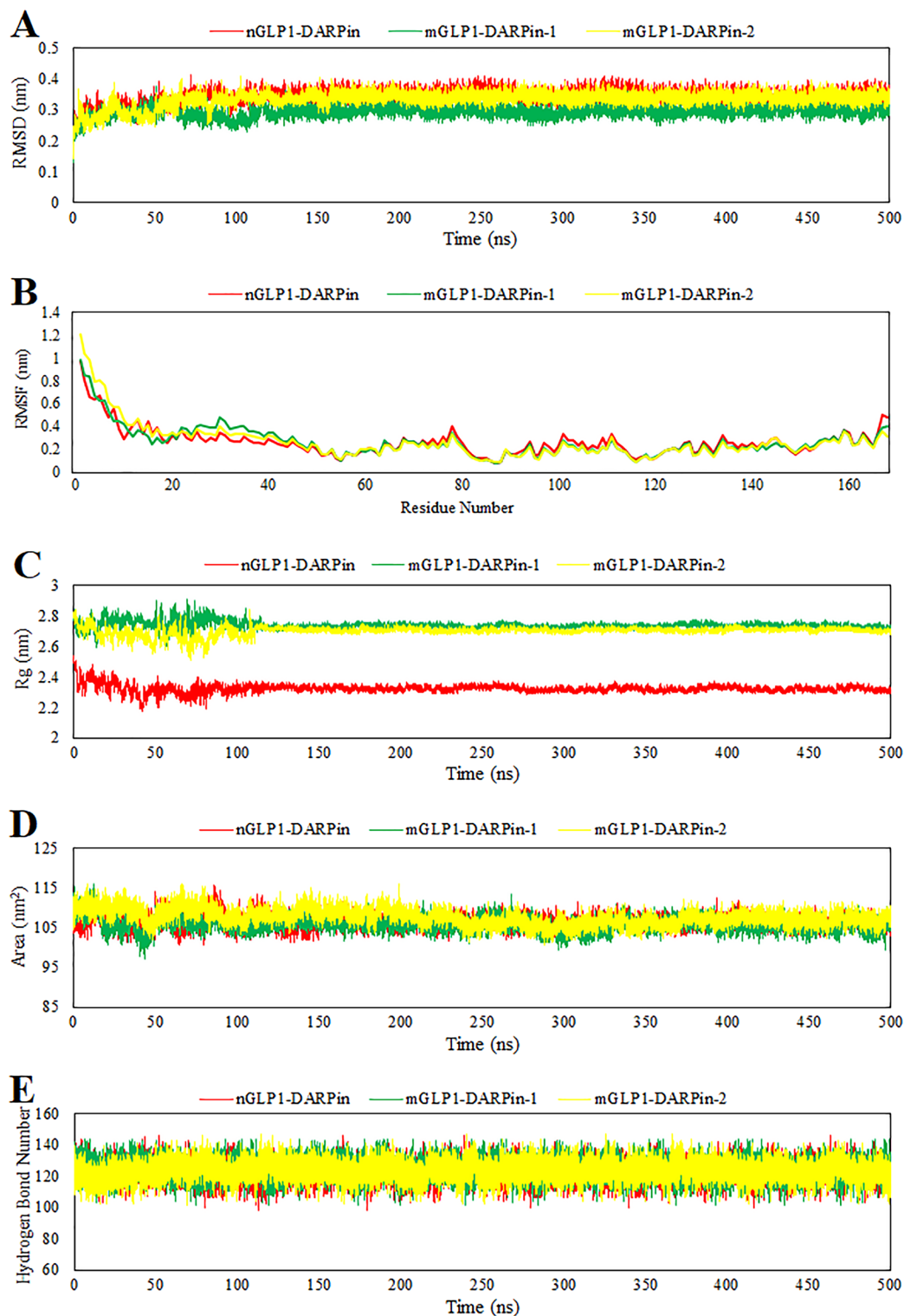


Figure 4. (A) Root mean square deviation, (B) root mean square fluctuation, (C) gyration radius graph, (D) solvent accessible surface area (SASA), and (E) hydrogen bond formation and deformation for MD simulation of the predicted structure of the nGLP1-DARPin (red), mGLP1-DARPin-1 (green), and mGLP1-DARPin-2 (yellow) fusion proteins.

respectively (Fig. 8A). The number of H-bonds formed between fusion proteins and the receptors at each step of the simulation was counted to ensure that the binding interaction was stable throughout the MD simulation.

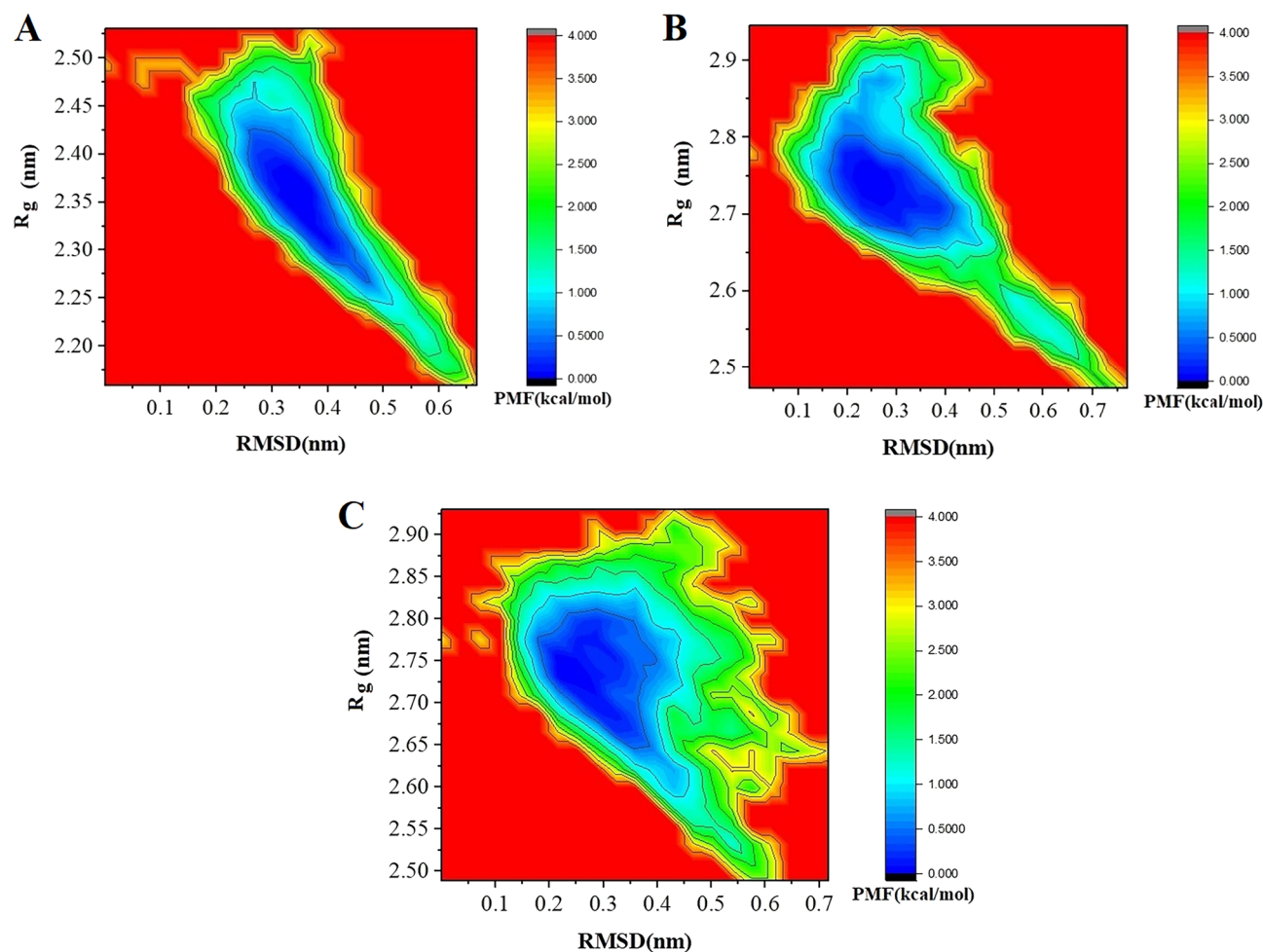


Figure 5. 2D-PMF profiles regarding the RMSD and R_g of (A) nGLP1-DARPin, (B) mGLP1-DARPin-1, and (C) mGLP1-DARPin-2.

Protein-Protein Complex	Docking score	ΔG (kcal mol ⁻¹)	Dissociation constant (M)
DARPin::HAS	-732.3	-9.6	1.7e-07
nGLP1-DARPin::HAS	-752.3	-9.2	1.8e-07
mGLP1-DARPin-1::HAS	-746.2	-8.1	3.9e-06
mGLP1-DARPin-2::HAS	-810.6	-8.6	1.7e-07
GLP-1::GLP-1 receptor	-820.6	-6.8	1.6e-05
nGLP1-DARPin::GLP-1 receptor	-897.4	-6.1	3.2e-06
mGLP1-DARPin-1::GLP-1 receptor	-821.9	-5.5	3.4e-05
mGLP1-DARPin-2::GLP-1 receptor	-852.3	-5.8	1.2e-05

Table 3. Docking scores and ΔG of protein-protein complexes obtained from the Cluspro and PRODIGY servers.

The total number of H-bonds forming in all three fusion proteins complexed with HSA was about 600, which remained consistent over the simulation time (Fig. 8B).

The RMSD plot analysis of fusion proteins complexed with GLP-1 receptor during 200 ns of the MD simulation showed that all three docked complexes stabilized after 50 ns and remained stable throughout the rest of the simulation time with RMSD values of 0.8 ± 0.08 , 0.9 ± 0.09 , and 0.81 ± 0.1 for nGLP1-DARPin, mGLP1-DARPin-1, and mGLP1-DARPin-2 complex with GLP-1R, respectively (Fig. 8C). Regarding H-bonds, internal hydrogen bonds in all three fusion proteins complexed with the GLP-1 receptor were about 195 and remained stable throughout the simulation with no obvious fluctuations (Fig. 8D).

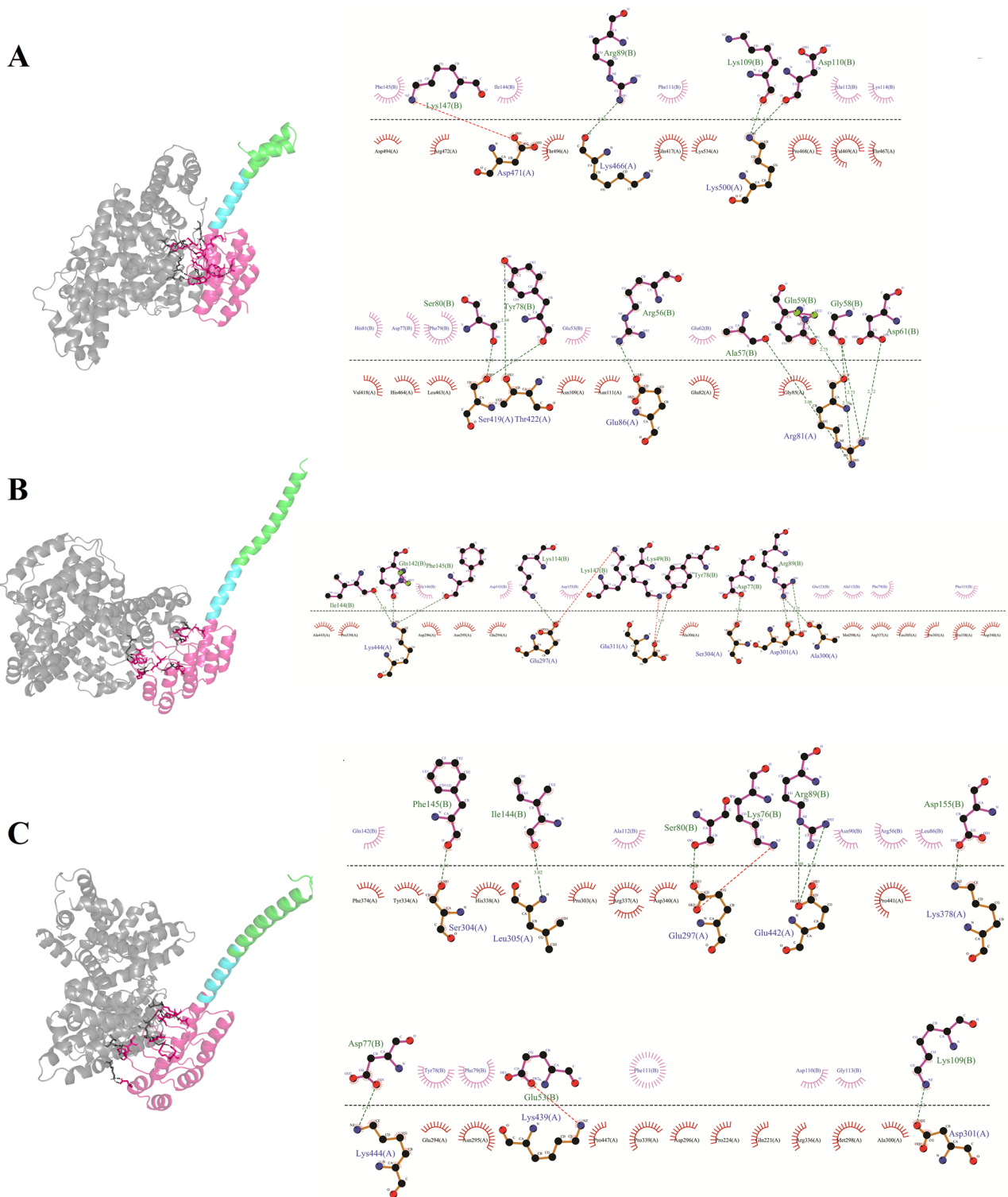


Figure 6. Interaction between (A) nGLP1-DARPin, (B) mGLP1-DARPin-1, and mGLP1-DARPin-2 fusion proteins with human serum albumin (gray). GLP-1 is depicted in green, while the linker is shown in cyan. HSA-binding DARPin is colored hot pink.

Discussion

Our research aimed to design and perform computational analysis of the fusion proteins with high affinity and specificity for human serum albumin and a protease-resistant GLP-1 peptide to produce potent GLP-1 agonists with an extended plasma half-life for the treatment of type 2 diabetes. The rapid enzymatic degradation and short plasma half-life of GLP-1 (about 2 min) limit its therapeutic use. To address this limitation, research has concentrated on many strategies, including identifying and developing DPP-IV inhibitors^{43,44}, altering the

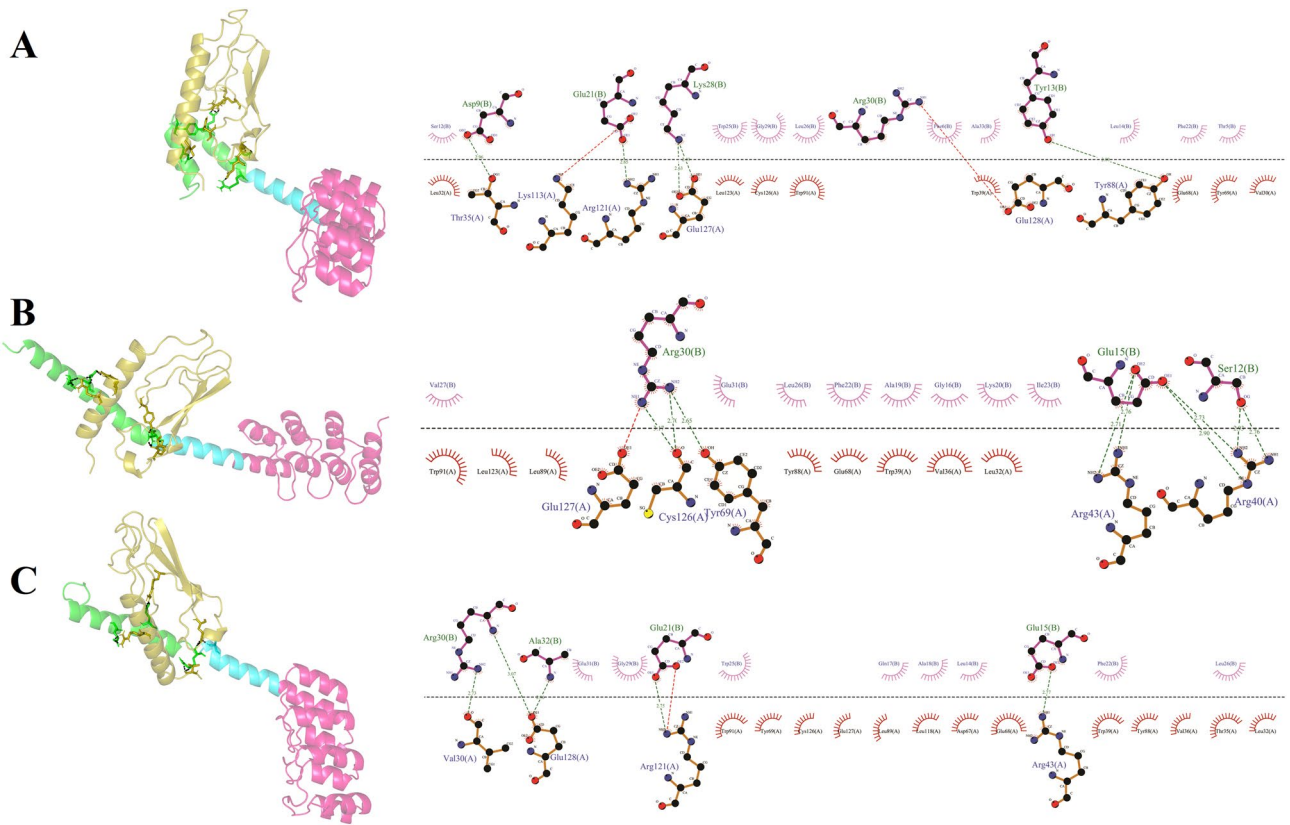


Figure 7. Interaction between (A) nGLP1-DARPin, (B) mGLP1-DARPin-1, and mGLP1-DARPin-2 fusion proteins with the extracellular domain of the GLP-1 receptor (olive). GLP-1 is depicted in green, while the linker is shown in cyan. HSA-binding DARPin is colored hot pink.

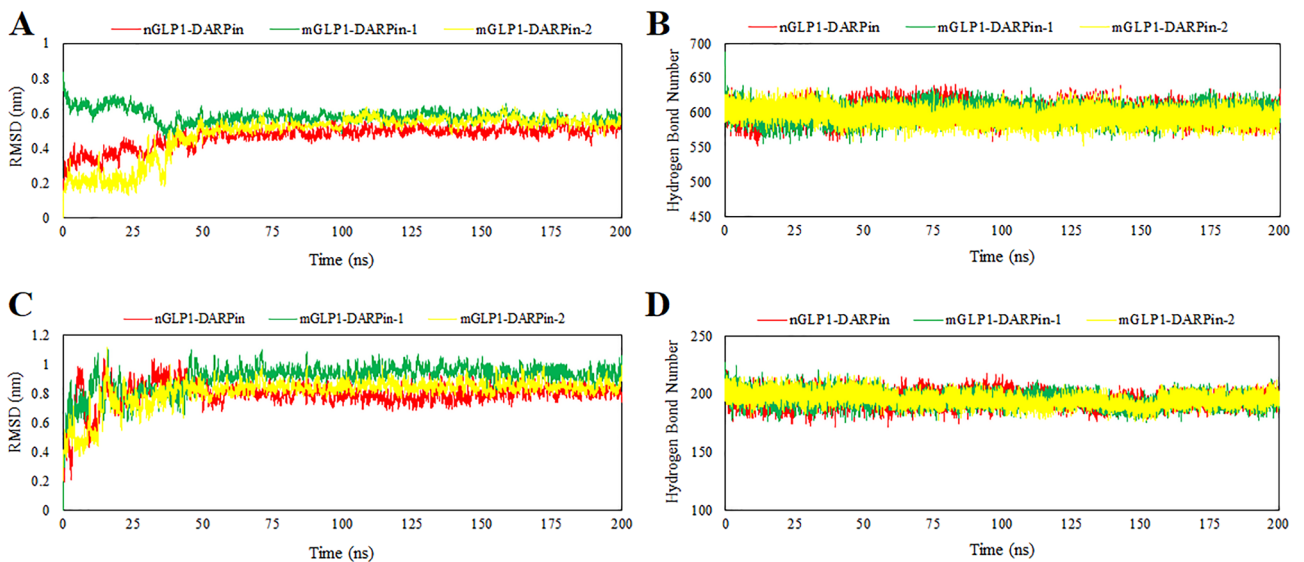


Figure 8. Carbon backbone RMSD profiles of the fusion proteins in complex with (A) human serum albumin and (B) the extracellular domain of the GLP-1 receptor; and the average number of H-bonds formed by each fusion protein with (C) human serum albumin and (D) the ECD of the GLP-1R during 200 ns MD simulations.

structural makeup of GLP-1, searching for structural analogues that are more resistant to DPP-IV cleavage⁴⁵, and increasing its hydrodynamic radius (R_h) above the glomerular filtration cutoff by conjugating it to proteins such as transferrin⁴⁶, albumin⁴⁷, and elastin-like polypeptides⁴⁸. Since the fusion of an albumin-binding moiety to proteins instead of directly fusing biologics to albumin has several advantages in terms of production cost and safety⁴⁹, the use of a binding protein known as DARPins has been decided upon in this work. DARPins can selectively bind any given target protein with high affinity³⁰, and HSA can be specifically targeted by albumin-binding

DARPin too. As demonstrated by MP0250²⁸, the therapeutic protein's half-lives can be significantly extended by genetically fusing non-HSA-binding DARPins or other therapeutic proteins to albumin-binding DARPins. The many benefits that DARPins possess make them the perfect scaffold for the development of protein drugs, including high thermal stability, with denaturation midpoints from 66 to 95 °C, and high solubility in solution, up to 100 mg/mL^{30,50}. In incubation with human serum (ex vivo), no proteolytic digestion was observed, indicating high stability for potential drug use. *E. coli*, yeast, and plants can all express DARPins at high levels in the soluble form^{31,33,34}, and based on chemical characterization, DARPins are predicted to be non-immunogenic⁵¹. The albumin-binding DARPin domain demonstrated nanomolar affinities for humans, cynomolgus monkeys, mice, and rats' serum albumin at pH 7.4 or 6.0. Because of this broad species selectivity, half-life data can be extrapolated from mice to humans²⁸. Due to these characteristics, albumin-binding DARPins are a potential substitute scaffold for extending the half-life of therapeutic proteins and peptides.

Based on these studies and the functional form of native GLP-1 (7–36), chimeric proteins as potent GLP-1 agonists with an extended plasma half-life were engineered using a native or two modified human GLP-1 (mGLP-1) that were resistant to DPP-IV or both DPP-IV and trypsin cleavage and fused to a DARPin with high affinity and specificity for HSA. To create the modified GLP-1, the enzyme-sensitive Ala at position 8 and Lys at positions 26 and 34 were substituted by Gly, Gln, and Asp, respectively. It has been demonstrated that these structural alterations exert complete resistance to GLP-1 cleavage caused by DPP-IV and trypsin degradation^{52–55}. In a recent study, Tan and coworkers engineered and expressed two GLP-1 analogues by genetically fusing a modified GLP-1 (Ala8Gly) to one or two tandem HSA-binding DARPins through a (GGGS)₃ flexible linker. Their findings revealed that GLP-2DARPin, which binds two HSA molecules, had a half-life roughly three times longer than GLP-DARPin (52.3 h versus 18.0 h). The bioactivity results, however, showed that GLP-DARPin had a stronger blood glucose-lowering effect than GLP-2DARPin. In addition, according to the oral glucose tolerance tests, GLP-DARPin significantly lowered blood glucose levels for at least 48 h, whereas GLP-2DARPin did so for only 24 h⁵⁶. In another recent study, Xia et al. designed and expressed four injectable long-acting GLP-1R agonists in a bacterial expression system by genetically fusing Exendin-4 to the N-terminus of HSA-binding DARPin via linkers of different types and lengths. Their results demonstrated that Ex-DARPin fusion proteins were substantially stable, resulting in incomplete denaturation even at 80 °C. The in vitro bioactivity results showed that Ex-DARPin fusion proteins could bind to HSA, activate GLP-1R, and have a longer half-life than native Ex⁵⁷.

We engineered our chimeric proteins with a rigid helical linker because previous research with albiglutide demonstrated that albumin's proximity to GLP-1 can significantly reduce its activity. That's why albiglutide contains two copies of GLP1 in tandem, with the second copy fused to HSA, which serves as a steric spacer³⁶. The selection of a linker is a crucial step in the design process of a fusion protein, especially in preserving the biological activity of the domains. We chose (EAAAK)₃ rigid helical linkers to work because we believe their rigidity allows for the separation of the protein domains and improves the functionality of individual domains^{58,59} when compared to flexible linkers, which appear to be suitable for linking normally connected domains in a single genetic construct, such as antibody domains in scFvs⁶⁰. With a GLP-1 fusion protein containing an albumin binding domain, Yousefpour and colleagues⁶¹ showed that the longer length and rigidity of the helical linker effectively kept the albumin, which is bound to the albumin binding domain, far enough from GLP1 to prevent it from interfering with its binding to the GLP1-R. In contrast, the shorter and more flexible linker does not completely remove the steric hindrance, leading to decreased receptor binding.

Bioinformatics approaches can result in significant time, cost, and failure reductions in experimental attempts. Using various bioinformatics tools and servers, we investigated the structural, evolutionary, and physicochemical characteristics of the engineered fusion proteins. The engineered constructs were 168 amino acids in length, with native or modified GLP-1 ranging from 1 to 30. The linker sequence was amino acids 31–45 (15 aa), and the HSA-binding DARPin was 46–168 (123 aa). Because the N-terminus of GLP-1 is critical to its biological function, we engineered the fusion proteins to leave the N-terminus of GLP-1 free. ProtParam was used to determine the primary structural features of the engineered fusion proteins. The high extinction coefficient, low instability index, and high aliphatic index of engineered fusion proteins indicate they are stable over a wide temperature range. SOPMA and PORTER, which perform a probabilistic assessment for all amino acids together, assisted us in making our secondary structure prediction. Alpha-helices make up a large portion of our fusion proteins. Furthermore, Protein-sol revealed that our engineered fusion constructs were more soluble than the soluble *E. coli* proteins in the experimental dataset. Following that, in silico analysis was performed to ensure that each domain in the engineered fusion proteins was folded correctly. The details of the 3D structure of our engineered fusion proteins can be used to study protein function, dynamics, and interaction with receptors. TrRosetta was used for the 3D modeling of our fusion constructs, and structure validation was performed using the Ramachandran plot, ProSA-web, and ERRAT tools to detect errors in the 3D modeled structures of the fusion constructs. Further validation of the engineered fusion protein's modeled tertiary structure was performed using the GROMACS MD simulation package for 500 ns in a water environment. After simulation, molecular docking studies were used to study the stable interactions of the engineered fusion proteins against the GLP-1 receptor and human serum albumin. More importantly, docking results demonstrated that the engineered fusion proteins can recognize and bind both GLP-1R and HSA with high affinity.

Conclusion

In this study, we designed long-acting chimeric proteins as GLP-1 receptor agonists for the treatment of type 2 diabetes mellitus and evaluated them through bioinformatics approaches. The mGLP1-DARPin-1 fusion protein that was more resistant to DPP-IV cleavage can be used as a long-lasting injectable form of GLP-1, and the mGLP1-DARPin-2 fusion protein that was resistant to both DPP-IV and trypsin cleavage can be used as a candidate for oral delivery of GLP-1 bioencapsulated in plant cells. According to the in silico results, the engineered

fusion proteins can identify and bind to the two target proteins (GLP-1R and HSA). Our study's methodology facilitates rapid analysis of engineered chimeric constructs prior to starting laboratory experiments with recombinant fusion proteins. Since all software and servers are available for free, it should go without saying that this process is quick, affordable, and easy—especially for novice users in this field. In conclusion, as we demonstrate here, the fusion of peptide therapeutics to an albumin-binding DARPIn, if engineered properly, is a straightforward and modular protein engineering strategy to enhance the therapeutic efficacy of short-circulating biologics. Although the present computational study provides three potential candidates as long-acting GLP-1 agonists, an experimental process needs to be established, a matter which is the theme of our ongoing study.

Methods

The current study involves the fusion of a native or mutant GLP-1 (30 residues) with an HSA-binding DARPIn (123 residues) via a rigid helical linker (15 residues) to create fusion proteins (168 residues) resistant to DPP-IV and, or trypsin with extended half-lives using structural modeling, which are subsequently characterized using MD simulation and molecular docking analyses.

Sequences retrieval and fusion proteins design

The amino acid sequence for the DARPIn with a high affinity for human serum albumin was obtained from US Patent 2016/9458211 B1⁶². The GLP-1 molecule, which originally consisted of 30 amino acids from human GLP-1 (7–36), was modified with the substitution of alanine (Aln) by glycine (Gly) at position 8 to prevent DPP-IV recognition and degradation and with glutamine (Gln) and aspartic acid (Asp) instead of lysine (Lys) at positions 26 and 34, respectively, to inhibit trypsin digestion. Based on previous studies, the N-terminal of GLP-1 is important for its biological function; therefore, we engineered the fusion constructs to free the N-terminal of GLP-1. The native or two modified GLP-1 (mGLP-1) with Ala8Gly substitution or Ala8Gly, Lys26Gln, and Lys34Asp substitutions were genetically fused to the N-terminus of HSA-binding DARPIn through a rigid helical linker (EAAAK)₃ to create the fusion proteins, denoted as nGLP1-DARPIn, mGLP1-DARPIn-1, and mGLP1-DARPIn-2, respectively. Rigid linkers are used when there is a requirement to preserve the spatial distance between the domains and their independent functions.

Primary structure properties of fusion proteins

Physical and chemical characteristics of engineered fusion proteins, including all molecular weight, molecular formula, theoretical isoelectric point (pI), the total number of positive and negatively charged residues in the sequence, extinction coefficient (E), instability index (II), aliphatic index (AI), and grand average hydropathy (GRAVY), were attained using the ExPASy ProtParam server⁶³. The instability index estimates a protein's in vitro stability, so the instability index must be below 40 for a protein to be considered stable. A protein aliphatic index is described as the relative volume occupied by aliphatic side chains (leucine, isoleucine, valine, and alanine) and is viewed as a favorable factor for the rise of thermostability in globular proteins. The GRAVY score is calculated by adding all of the amino acid hydropathy scores and dividing them by the number of residues in the sequence.

Prediction of secondary structure

The engineered fusion proteins' secondary structures, including the regions lacking regular shape, sections with low complexity, the proportion of random coils, extended strands, and alpha helices, were predicted by the SOPMA⁶⁴ and PORTER⁶⁵ servers.

Three-dimensional model prediction

The 3D structure of the engineered fusion proteins was modeled using the trRosetta web server⁶⁶. The trRosetta algorithm predicts protein structures quickly and precisely. The raw amino acid sequences of the engineered fusion proteins were submitted to the trRosetta server in FASTA format. It builds the 3D structure using a restricted Rosetta and direct energy minimization. When the conditions of Confidence > 0.6, with an E-value of 0.001, and Coverage > 0.3 are met, a template is employed for subsequent prediction. The TM-score of the predicted models is calculated using the probability of the top anticipated distance and the convergence of the top models. The TM-score is a number between 0 and 1, with a value greater than 0.5, indicating that the model's topology was predicted correctly. The trRosetta results produced five top models for each entry, of which the one with the greatest estimated TM-score (e-TM) represented the best model and was selected for this study. The predicted 3D structures were visualized by PyMOL Version 2.3.2_81 (the PyMOL Molecular Graphics System).

Tertiary structure validation

Procheck Ramachandran Plot v.3.5.4⁶⁷ and ERRAT⁶⁸ of the SAVES V6.0 web tool and ProSA-Web⁶⁹ were used for structural evaluation and stereochemical analyses. The Ramachandran plot showed the depiction and distribution of the residues in favored, allowed, and outlier regions and was used to determine the validity and quality of protein models. In the predicted structure, the ERRAT server evaluated the statistics of non-bonded interactions among various atom types by comparison with statistics from highly refined structures, and its higher scores indicate higher quality. The ProSA Z-score estimates the total energy of the structure from an energy distribution produced from random conformations and represents the overall model quality.

Protein solubility prediction

The Protein-Sol web server⁷⁰ was used to estimate the solubility of engineered fusion proteins. It analyzes amino acid sequences and computes anticipated solubility and other attributes. For experimental solubility, since the

population average for the experimental dataset (PopAvrSol) is 0.45, any scaled solubility value greater than 0.45 is predicted to be more soluble than the average soluble proteins of *E. coli*⁷¹. In contrast, any protein with a lower scaled solubility value is predicted to be less soluble. The toxic potential of engineered fusion proteins was predicted using the ToxDL⁷² web server. The ToxDL server is devised for deep learning prediction of toxic domains in protein structures.

Molecular dynamics simulation

The GROMACS package, version 2019.2, was used to simulate the molecular dynamics of engineered fusion proteins⁷³. For each fusion protein, the best model obtained from molecular modeling was subjected to MD simulation. The all-atom Charmm27 forcefield was selected for making topology files, and the structure was placed in a triclinic unit cell 1.0 nm from the box edge. The system was first solvated using a simple point charge (SPC) water model before being ionized and neutralized by Na⁺ and Cl⁻ ions with a salt concentration of 0.15 M. Following that, a conjugate gradient algorithm with Steepest Descent (SD) minimization was used to minimize energy on a model that had already been predicted. Then the position restraint simulation was conducted under an NVT (constant number, volume, and temperature) ensemble heated to 300 °K at 100 ps and an NPT (constant number, pressure, and temperature) equilibrated to the 1 bar pressure at 100 ps. Finally, an unrestrained MD simulation was run for 500 ns on three replicas with coupled temperature (300 °K) and pressure (1 bar). Microsoft Office Excel was used to plot graphs and perform a comparative analysis of structural deviations based on extracting parameters from the MD trajectories such as Root Mean Square Deviation (RMSD), Root Mean Square Fluctuation (RMSF), Solvent-Accessible Surface Area (SASA), Radius of Gyration (Rg), and H-bond formation or deformation. The potential mean force (PMF) of the studied systems was computed using the MolAICal package^{74,75}.

Protein–protein interaction

Following molecular dynamics simulations, the fusion proteins were subjected to molecular docking simulations. The ClusPro 2.0 online server, which is a fully automated server engineered for protein–protein docking simulations, was used to undertake a molecular docking analysis of the fusion proteins against human serum albumin (HSA) and the extracellular domain of the GLP-1 receptor (ECD-GLP-1R)⁷⁶. Protein–protein interactions are computed in three steps: rigid body docking using the Fast Fourier Transform (FFT) correlation method; clustering of generated structures based on RMSD to find the largest cluster that represents the likely models of the complex; and refinement of selected structures. ClusPro 2.0 requires PDB files of the proteins as input and provides four types of output models based on the scoring algorithms designated as balanced, electrostatic-favored, hydrophobic-favored, and Van der Waals electrostatic.

In our docking, we designated ECD-GLP-1R and HSA as the receptors and engineered fusion proteins as the ligands. To validate the docking studies, GLP-1 and HSA-binding DARPIn were docked against the GLP-1 receptor and human serum albumin, respectively. The crystal structure of GLP-1 in complex with ECD-GLP-1R (PDB ID: 3IOL) was obtained from the RCSB data bank. To our knowledge, the crystallographic structure of HSA-binding DARPIn has not been reported in the protein database. Therefore, we carried out computational modeling of HSA-binding DARPIn by the MODELLER software 15.9⁷⁷ based on the most similar available conformation (PDB ID: 4GRG). The complexes obtained after docking were used as input for further MD simulations to increase the accuracy of the resulting interactions. We carried out three replicas of MD simulations for 200 ns using the same procedure in the MD section described above, to study the stability of these docking complexes. MD output data were analyzed regarding RMSD and total numbers of H-bond formation or deformation. Subsequently, MD simulation trajectories were submitted to the PRODIGY server (PROtein binDing energy prediction) to obtain free energy (ΔG)⁷⁸, and studied for interaction analysis using LigPlot + v.2.2⁷⁹. The PyMOL package was utilized for visual depiction.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Received: 19 June 2023; Accepted: 17 October 2023

Published online: 18 October 2023

References

- Piero, M., Nzaro, G. & Njagi, J. Diabetes mellitus—A devastating metabolic disorder. *Asian J. Biomed. Pharmaceut. Sci.* **5**(40), 1 (2015).
- Laakso, M. Biomarkers for type 2 diabetes. *Mol. Metab.* **27**(Suppl 1), S139–S146 (2019).
- Talchai, C., Xuan, S., Lin, H. V., Sussel, L. & Accili, D. Pancreatic β cell dedifferentiation as a mechanism of diabetic β cell failure. *Cell* **150**(6), 1223–1234 (2012).
- Gupta, V. Glucagon-like peptide-1 analogues: An overview. *Indian J. Endocrinol. Metab.* **17**(3), 413–421 (2013).
- Müller, T. D. *et al.* Glucagon-like peptide 1 (GLP-1). *Mol. Metab.* **30**, 72–130 (2019).
- Vahl, T. P., Paty, B. W., Fuller, B. D., Prigeon, R. L. & D'Alessio, D. A. Effects of GLP-1-(7–36) NH₂, GLP-1-(7–37), and GLP-1-(9–36) NH₂ on intravenous glucose tolerance and glucose-induced insulin secretion in healthy humans. *J. Clin. Endocrinol. Metab.* **88**(4), 1772–1779 (2003).
- Baggio, L. L. & Drucker, D. J. Biology of incretins: GLP-1 and GIP. *Gastroenterology.* **132**(6), 2131–2157 (2007).
- Zhang, Y. *et al.* GLP-1 receptor in pancreatic α -cells regulates glucagon secretion in a glucose-dependent bidirectional manner. *Diabetes.* **68**(1), 34–44 (2019).
- López-Ferreras, L. *et al.* Lateral hypothalamic GLP-1 receptors are critical for the control of food reinforcement, ingestive behavior and body weight. *Mol. Psychiatry* **23**(5), 1157–1168 (2018).

10. Kieffer, T. J., McIntosh, C. H. & Pederson, R. A. Degradation of glucose-dependent insulinotropic polypeptide and truncated glucagon-like peptide 1 in vitro and in vivo by dipeptidyl peptidase IV. *Endocrinology*. **136**(8), 3585–3596 (1995).
11. Hui, H., Farilla, L., Merkel, P. & Perfetti, R. The short half-life of glucagon-like peptide-1 in plasma does not reflect its long-lasting beneficial effects. *Eur. J. Endocrinol.* **146**(6), 863–869 (2002).
12. Green, B. D. *et al.* Novel dipeptidyl peptidase IV resistant analogues of glucagon-like peptide-1(7–36)amide have preserved biological activities in vitro conferring improved glucose-lowering action in vivo. *J. Mol. Endocrinol.* **31**(3), 529–540 (2003).
13. Kontermann, R. E. Half-life extended biotherapeutics. *Expert Opin. Biol. Ther.* **16**(7), 903–915 (2016).
14. Strohl, W. R. Fusion proteins for half-life extension of biologics as a strategy to make biobetters. *BioDrugs* **29**(4), 215–239 (2015).
15. Bell, M. R., Engleka, M. J., Malik, A. & Strickler, J. E. To fuse or not to fuse: What is your purpose?. *Protein Sci.* **22**(11), 1466–1477 (2013).
16. Terpe, K. Overview of tag protein fusions: From molecular and biochemical fundamentals to commercial systems. *Appl. Microbiol. Biotechnol.* **60**(5), 523–533 (2003).
17. Pédelacq, J.-D., Cabantous, S., Tran, T., Terwilliger, T. C. & Waldo, G. S. Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.* **24**(1), 79–88 (2006).
18. Grünberg, R. & Serrano, L. Strategies for protein synthetic biology. *Nucleic Acids Res.* **38**(8), 2663–2675 (2010).
19. Schmidt, S. R. *Fusion Protein Technologies for Biopharmaceuticals: Applications and Challenges.* (2013).
20. Yang, H., Liu, L. & Xu, F. The promises and challenges of fusion constructs in protein biochemistry and enzymology. *Appl. Microbiol. Biotechnol.* **100**(19), 8273–8281 (2016).
21. Peters, T. Jr. Serum albumin. *Adv Protein Chem.* **37**, 161–245 (1985).
22. Holt, L. J. *et al.* Anti-serum albumin domain antibodies for extending the half-lives of short lived drugs. *Protein Eng. Des. Sel.* **21**(5), 283–288 (2008).
23. Müller, M. R. *et al.* Improving the pharmacokinetic properties of biologics by fusion to an anti-HSA shark VNAR domain. *MAbs.* **4**(6), 673–685 (2012).
24. Van Roy, M. *et al.* The preclinical pharmacology of the high affinity anti-IL-6R Nanobody* ALX-0061 supports its clinical development in rheumatoid arthritis. *Arthritis Res. Ther.* **17**(1), 135 (2015).
25. Dennis, M. S. *et al.* Albumin binding as a general strategy for improving the pharmacokinetics of proteins. *J. Biol. Chem.* **277**(38), 35035–35043 (2002).
26. Linhult, M., Binz, H. K., Uhlén, M. & Hober, S. Mutational analysis of the interaction between albumin-binding domain from streptococcal protein G and human serum albumin. *Protein Sci.* **11**(2), 206–213 (2002).
27. Jonsson, A., Dogan, J., Herne, N., Abrahmsén, L. & Nygren, P. A. Engineering of a femtomolar affinity binding protein to human serum albumin. *Protein Eng. Des. Sel.* **21**(8), 515–527 (2008).
28. Steiner, D. *et al.* Half-life extension using serum albumin-binding DARPin* domains. *Protein Eng. Des. Sel.* **30**(9), 583–591 (2017).
29. Binz, H. K. *et al.* High-affinity binders selected from designed ankyrin repeat protein libraries. *Nat. Biotechnol.* **22**(5), 575–582 (2004).
30. Stumpp, M. T., Binz, H. K. & Amstutz, P. DARPins: A new generation of protein therapeutics. *Drug Discov. Today* **13**(15–16), 695–701 (2008).
31. Ehsasatvatan, M., Kohnehrouz, B. B., Gholizadeh, A., Ofoghi, H. & Shanebandi, D. The production of the first functional antibody mimetic in higher plants: The chloroplast makes the DARPin G3 for HER2 imaging in oncology. *Biol. Res.* **55**(1), 1–18 (2022).
32. Plückthun, A. Designed ankyrin repeat proteins (DARPins): Binding proteins for research, diagnostics, and therapy. *Annu. Rev. Pharmacol. Toxicol.* **55**, 489–511 (2015).
33. Goldstein, R. *et al.* Development of the designed ankyrin repeat protein (DARPin) G3 for HER2 molecular imaging. *Eur. J. Nucl. Med. Mol. Imaging* **42**(2), 288–301 (2015).
34. Binz, H. K., Bakker, T. R., Phillips, D. J., Cornelius, A., Zitt, C., Göttler, T. *et al.* *Design and Characterization of MP0250, a Tri-Specific Anti-HGF/Anti-VEGF DARPin* Drug Candidate.* In *MAbs.* (Taylor & Francis, 2017).
35. Stumpp, M. T., Dawson, K. M. & Binz, H. K. Beyond antibodies: The DARPin(*) drug platform. *BioDrugs* **34**(4), 423–433 (2020).
36. Baggio, L. L., Huang, Q., Brown, T. J. & Drucker, D. J. A recombinant human glucagon-like peptide (GLP)-1-albumin protein (albugon) mimics peptidergic activation of GLP-1 receptor-dependent pathways coupled with satiety, gastrointestinal motility, and glucose homeostasis. *Diabetes.* **53**(9), 2492–2500 (2004).
37. Ehsasatvatan, M. & Kohnehrouz, B. Effect of linker's length and sequence on the structure and stability of mGLP-1-DARPin fusion protein for treatment of type 2 diabetes: A computational study. *Yafteh* **25**(2), 26–47 (2023).
38. Thornton, K. & Gorenstein, D. G. Structure of glucagon-like peptide (7–36) amide in a dodecylphosphocholine micelle as determined by 2D NMR. *Biochemistry.* **33**(12), 3532–3539 (1994).
39. Wriggers, W., Chakravarty, S. & Jennings, P. A. Control of protein functional dynamics by peptide linkers. *Biopolymers* **80**(6), 736–746 (2005).
40. Astuti, A. & Mutiara, A. Performance analysis on molecular dynamics simulation of protein using GROMACS. *arXiv preprint arXiv:0912.0893* (2009).
41. Maiorov, V. N. & Crippen, G. M. Significance of root-mean-square deviation in comparing three-dimensional structures of globular proteins. *J. Mol. Biol.* **235**(2), 625–634 (1994).
42. Lobanov, M., Bogatyreva, N. S. & Galzitskaia, O. V. Radius of gyration is indicator of compactness of protein structure. *Mol. Biol. (Mosk)* **42**(4), 701–706 (2008).
43. Chen, P. *et al.* Imidazopiperidine amides as dipeptidyl peptidase IV inhibitors for the treatment of diabetes. *Bioorg. Med. Chem. Lett.* **17**(21), 5853–5857 (2007).
44. Kushwaha, R. N. *et al.* Design, synthesis, biological screening, and molecular docking studies of piperazine-derived constrained inhibitors of DPP-IV for the treatment of type 2 diabetes. *Chem. Biol. Drug Des.* **85**(4), 439–446 (2015).
45. Johnson, L. M. *et al.* A potent α/β -peptide analogue of GLP-1 with prolonged action in vivo. *J. Am. Chem. Soc.* **136**(37), 12848–12851 (2014).
46. Kim, B.-J. *et al.* Transferrin fusion technology: A novel approach to prolonging biological half-life of insulinotropic peptides. *J. Pharmacol. Exp. Ther.* **334**(3), 682–692 (2010).
47. St Onge, E. L. & Miller, S. A. Albiglutide: A new GLP-1 analog for the treatment of type 2 diabetes. *Expert Opin. Biol. Ther.* **10**(5), 801–806 (2010).
48. Luginbuhl, K. M. *et al.* One-week glucose control via zero-order release kinetics from an injectable depot of glucagon-like peptide-1 fused to a thermosensitive biopolymer. *Nat. Biomed. Eng.* **1**(6), 1–14 (2017).
49. Fleer, R., Fournier, A., Guittou, J.-D., Jung, G. & Yeh, P. *Fusion Polypeptides of Human Serum Albumin and a Therapeutically Active Polypeptide* (Google Patents, 2004).
50. Binz, H. K., Stumpp, M. T., Forrer, P., Amstutz, P. & Plückthun, A. Designing repeat proteins: Well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins. *J. Mol. Biol.* **332**(2), 489–503 (2003).
51. Caputi, A. P. & Navarra, P. Beyond antibodies: Ankyrins and DARPins: From basic research to drug approval. *Curr. Opin. Pharmacol.* **51**, 93–101 (2020).
52. Hou, J. *et al.* Oral administration of a fusion protein containing eight GLP-1 analogues produced in *Escherichia coli* BL21 (DE3) in streptozotocin-induced diabetic rats. *Biotechnol. Lett.* **29**(10), 1439–1446 (2007).

53. Ma, B. *et al.* Expression and purification of optimized rolGLP-1, a novel GLP-1 analog, in *Escherichia coli* BL21 (DE3) and its good glucoregulatory effect on type 2 diabetic mice. *Curr. Pharmaceut. Biotechnol.* **14**(11), 985–994 (2013).
54. Dou, W.-F. *et al.* Expression, purification, and characterization of recombinant human serum albumin fusion protein with two human glucagon-like peptide-1 mutants in *Pichia pastoris*. *Protein Exp. Purif.* **61**(1), 45–49 (2008).
55. Qian, K. *et al.* Efficient expression of glucagon-like peptide-1 analogue with human serum albumin fusion protein in *Pichia pastoris* using the glyceraldehyde-3-phosphate dehydrogenase promoter. *Biotechnol. Bioprocess Eng.* **20**(4), 694–700 (2015).
56. Tan, H. *et al.* Generation of novel long-acting GLP-1R agonists using DARPins as a scaffold. *Int. J. Pharmaceut.* **607**, 121043 (2021).
57. Xia, J., Gao, G., Zhang, C., Ying, J. & Li, J. Albumin-binding DARPins as scaffold improve the hypoglycemic and anti-obesity effects of exendin-4 in vivo. *Eur. J. Pharmaceut. Sci.* **185**, 106422 (2023).
58. Amet, N., Lee, H. F. & Shen, W. C. Insertion of the designed helical linker led to increased expression of tf-based fusion proteins. *Pharm. Res.* **26**(3), 523–528 (2009).
59. Arai, R. Design of helical linkers for fusion proteins and protein-based nanostructures. *Methods Enzymol.* **647**, 209–230 (2021).
60. Chen, X., Zaro, J. L. & Shen, W.-C. Fusion protein linkers: Property, design and functionality. *Adv. Drug Deliv. Rev.* **65**(10), 1357–1369 (2013).
61. Yousefpour, P., Varanko, A., Subrahmanyam, R. & Chilkoti, A. Recombinant fusion of glucagon-like peptide-1 and an albumin binding domain provides glycemic control for a week in diabetic mice. *Adv. Ther.* **3**(10), 2000073 (2020).
62. Bakker, T., Stumpp, M. T., Binz, H. K., Phillips, D., Dolado, I., Forrer, P. *et al.* Recombinant Proteins that Simultaneously Bind HGF, VEGF-A and Serum Albumin, Comprising Ankyrin Repeat Domains. (Google Patents, 2016).
63. Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M. R., Appel, R. D. & Bairoch, A. Protein identification and analysis tools on the ExPASy server. In *The Proteomics Protocols Handbook*. 571–607 (2005).
64. Combet, C., Blanchet, C., Geourjon, C. & Deléage, G. NPS@: Network protein sequence analysis. *Trends Biochem. Sci.* **25**(3), 147–150 (2000).
65. Pollastri, G. & McLysaght, A. Porter: A new, accurate server for protein secondary structure prediction. *Bioinformatics* **21**(8), 1719–1720 (2005).
66. Du, Z. *et al.* The trRosetta server for fast and accurate protein structure prediction. *Nat. Protoc.* **16**(12), 5634–5651 (2021).
67. Laskowski, R., MacArthur, M. & Thornton, J. PROCHECK: Validation of Protein-Structure Coordinates. (2006).
68. Colovos, C. & Yeates, T. ERRAT: An empirical atom-based method for validating protein structures. *Protein Sci.* **2**(9), 1511–1519 (1993).
69. Wiederstein, M. & Sippl, M. J. ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* **35**(Web Server issue), W407–W410 (2007).
70. Hebditch, M., Carballo-Amador, M. A., Charonis, S., Curtis, R. & Warwicker, J. Protein-Sol: A web tool for predicting protein solubility from sequence. *Bioinformatics* **33**(19), 3098–3100 (2017).
71. Niwa, T. *et al.* Bimodal protein solubility distribution revealed by an aggregation analysis of the entire ensemble of *Escherichia coli* proteins. *Proc. Natl. Acad. Sci.* **106**(11), 4201–4206 (2009).
72. Pan, X. *et al.* ToxDL: Deep learning using primary structure and domain embeddings for assessing protein toxicity. *Bioinformatics* **36**(21), 5159–5168 (2021).
73. Van Der Spoel, D. *et al.* GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **26**(16), 1701–1718 (2005).
74. Bai, Q. *et al.* MolAICal: A soft tool for 3D drug design of protein targets by artificial intelligence and classical algorithm. *Brief. Bioinform.* **22**(3), 161 (2021).
75. Bai, Q., Research and development of MolAICal for drug design via deep learning and classical programming. *arXiv preprint arXiv:2006.09747*. (2020).
76. Kozakov, D. *et al.* The ClusPro web server for protein-protein docking. *Nat. Protoc.* **12**(2), 255–278 (2017).
77. Fiser, A. & Sali, A. Modeller: Generation and refinement of homology-based protein structure models. *Methods Enzymol.* **374**, 461–491 (2003).
78. Xue, L. C., Rodrigues, J. P., Kastriitis, P. L., Bonvin, A. M. & Vangone, A. PRODIGY: A web server for predicting the binding affinity of protein-protein complexes. *Bioinformatics* **32**(23), 3676–3678 (2016).
79. Wallace, A. C., Laskowski, R. A. & Thornton, J. M. LIGPLOT: A program to generate schematic diagrams of protein–ligand interactions. *Protein Eng.* **8**(2), 127–134 (1995).

Acknowledgements

This research is supported by the Research Grant of the University of Tabriz (number S/4829-14020128). The authors would like to thank Dr. Jaber Jahanbin Sardroodi and Dr. Parisa Jahanbakhsh-Bonab, Azarbaijan Shahid Madani University for their help in PMF calculation.

Author contributions

M.E., designed the construct, performed analyses, and wrote the manuscript. B.B.K., supervised the study, provided specialized scientific and revised the paper. All authors have agreed with the manuscript and provided their consent for publication.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-45185-1>.

Correspondence and requests for materials should be addressed to B.B.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023