Research Article

Structural Biophysics-Guided Computational Design of Semaglutide Analogues to Enhance GLP-1R Activation

Wei Li[1]

1. Contrebola Institute of Computational Interstructural Biophysics, China

CagriSema is a fixed-dose combination of cagrilintide (an amylin analogue) and semaglutide (a GLP-1 receptor agonist), and is currently an experimental obesity drug developed by Novo Nordisk. In March 2025, CagriSema underperformed expectations in a Phase III trial, achieving 15.7% weight loss instead of the anticipated 25%, raising concerns about its efficacy and clinical value. Given its chemical composition, the weight-loss efficacy of CagriSema is inextricably linked to the activations of GLP-1R and amylin receptors (AMYRs). With GLP-1R as an example target here, this study employs a structural biophysics-guided computational approach for the design of semaglutide analogues to enhance the activation of its receptor GLP-1R. To fully harness the therapeutic potential of GLP-1R activation, an experimental structural basis (PDB entry 4ZGM) of the GLP-1-GLP-1R interaction is essential for the design of semaglutide analogues, where site-specific missense mutations are engineered into its peptide backbone to establish additional stabilizing interactions with the extracellular domain (ECD) of GLP-1R. Specifically, this study puts forward an automated systemic natural amino acid scanning of the peptide backbone of semaglutide, where PDB entry 4ZGM was used as the structural template for high-throughput structural modeling by Modeller and ligand-receptor binding affinity (K_d) calculations by Prodigy. To sum up, this article reports a total of 564 computationally designed semaglutide analogues with improved GLP-1R ECD binding affinity. Moreover, this study proposes a concept of an interfacial electrostatic scaffold comprising four salt bridges at the binding interface of GLP-1R ECD and semaglutide analogues. Drawing parallels with the continued optimization in the past century of the history of insulin, this article argues that the interfacial electrostatic scaffold here constitutes a robust framework for the continued development of next-generation GLP-1R agonists, enabling more effective therapies for patients with diabetes and/or obesity.

 doi.org/10.32388/FGTLKF

Corresponding author: Wei Li, wli148@aucklanduni.ac.nz

Introduction

Over one century ago, the discovery of insulin marked a transformative milestone in diabetes treatment, leading from life-saving animal-derived therapies to recombinantly synthetic human insulin and its analogues. These continued advancements revolutionized glycemic control for patients with diabetes of both types[1][2][3][4][5]. From day one, Novo Nordisk has played a pivotal role in this evolution, including also its recent developments of glucagon-like peptide-1 receptor (GLP-1R) agonists such as semaglutide[6][7][8][9]. Since its FDA approval in December of 2017, semaglutide has facilitated once-weekly dosing for type 2 diabetes, significantly enhancing patient compliance and clinical outcomes[10][11][12][13]. Beyond type 2 diabetes, semaglutide also demonstrated efficacy in weight regulation and cardiovascular risk reduction[14]. Given this, Novo Nordisk continues to explore novel therapeutic avenues, as evidenced by the development of IcoSema[15][16][17][18][19] and CagriSema[20]. Take CagriSema, for example, which is a combination therapy comprising cagrilintide and semaglutide, designed for the treatment of type 2 diabetes and obesity. Recently, Phase III trial results revealed that CagriSema underperformed relative to expectations, achieving 15.7% weight loss instead of the anticipated 25%, raising concerns regarding its overall efficacy and clinical value.

To improve therapeutic efficacy, it is essential to consider the molecular basis of receptor activation and the structural basis of drug-target interaction. Take GLP-1R, for example, which is a class B G-protein-coupled receptor (GPCR) and plays a pivotal role in glucose homeostasis by mediating the pharmacological effects of GLP-1 agonists[21][22]. Previous studies have already highlighted the extracellular domain (ECD) of GLP-1R as a key factor in initiating ligand binding, receptor activation, and subsequent intracellular signal transduction[23][24][25][26]. As such, enhancing ligand-ECD interactions represents a reasonable structural strategy to augment the therapeutic potential of GLP-1R agonists by enhancing receptor activation[27]. Last year, for instance, Sawyer and his colleagues[27] reported a systematic structure–activity and computational modeling analysis of a series of GLP-1R agonists based on an 11-mer GLP-1 peptide, leading to a significantly enhanced optimization of GLP-1R agonist potency[27], where the primary focus[27] was a systematic analysis of both Ala- and Aib-scanning (Figure 1) of the 11-mer template, in addition to a deeper exploration of Phe6 in terms of side-chain modifications and/or -methylation[27]. Here, this study reports an automated systemic natural amino acid scanning of

the entire peptide backbone of semaglutide (Figure 1), which aims to design semaglutide analogues with site-specific missense mutations to enhance GLP-1R activation through improved ECD binding affinity[21][28].

Materials and Methods

As of 2025-04-29, there is one experimental structure determined by X-ray diffraction of the semaglutide backbone in complex with GLP-1R ECD (PDB ID: 4ZGM[24]), which is used here as the template (instead of an 11-mer GLP-1 peptide as used in[27]) for high-throughput[29] structural modeling by Modeller[30] and ligand-receptor binding affinity (K_d) calculations by Prodigy[31][32].

Specifically, this study conducted an automated systemic amino acid scanning of the peptide backbone of semaglutide (Figure 1). Here, **Modigy** is defined as an abbreviation of Modeller[30] and Prodigy[31][32] to represent an in silico high-throughput generation of structural and intermolecular binding affinity (K_d) data[29]. Here, **Modigy** follows three criteria:

1. position 8 (Aib near the black scissor in Figure 1) remains unchanged, to avoid DPP4 cleavage[33][34] of the computationally designed analogues;
2. position 26 (Lys26) of semaglutide's backbone (Figure 1) remains unchanged, to allow attachment of the fatty acid chain to the peptide backbone;
3. there is just one lysine (i.e., Lys26) in the backbone (Figure 1) of semaglutide, to ensure correct attachment of the fatty acid chain to the peptide backbone of semaglutide.

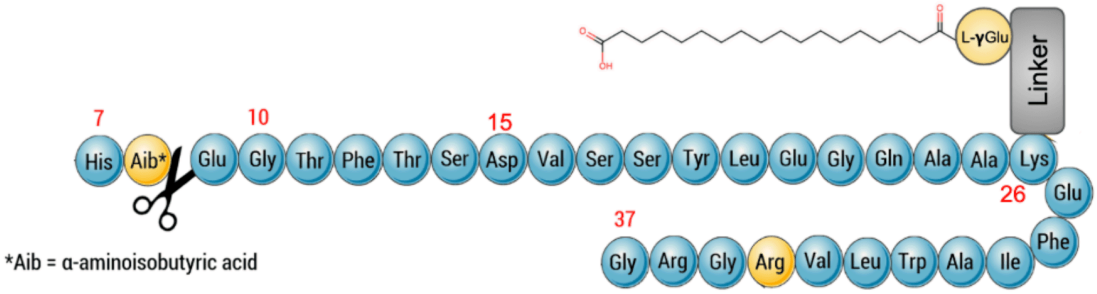


Figure 1. 2D structure of semaglutide. In this figure, the sequence of semaglutide's backbone is numbered from 7 to 37, with modifications highlighted in yellow and the dipeptidyl peptidase-4 (DPP4) cleavage site[33][34] indicated with a scissor. This figure is obtained from the PDB website (URL: https://pdb101.rcsb.org/global-health/diabetes-mellitus/drugs/incretins/drug/semaglutide/semaglutide, accessed on November 22, 2024)

Subsequently, for all Modeller-generated[30] structural models, a comprehensive structural biophysical analysis[35] was conducted to identify key electrostatic interactions at the binding interface of GLP-1R and semaglutide analogues[24][28]. Specifically, the cut-off distance is 4.0 Å for salt bridge analysis[35], while the hydrogen bond analysis employed two criteria: (a) a cutoff value of the angle formed by acceptor (A), donor (D), and hydrogen (H) ($\angle ADH$) of 30; (b) a cutoff value of donor-acceptor (D-A) distance at 3.0 Å [35]. For **Modigy**[29], homology structural modeling by Modeller[30] and K_d calculations by Prodigy[31][32] were performed on the high-performance computing (HPC) platform at the Wuxi Taihu Lake HPC center.

Results

As mentioned above, semaglutide (a GLP-1R agonist) binding directly to the GLP-1R ECD[24] is the first step that initiates downstream GLP-1R activation and transmembrane signal transduction[25][26]. As such, the primary focus of this study is the interaction between the GLP-1R ECD and semaglutide, where PDB entry 4ZGM[36] constitutes the structural basis for the subsequent computational design of semaglutide analogues with improved GLP-1R ECD binding affinity.

A structural biophysical investigation into PDB entry 4ZGM

In 2015, a team from Novo Nordisk deposited into the Protein Data Bank[37] the first experimental structure (PDB entry 4ZGM[24][36]) of the semaglutide backbone in complex with the GLP-1R ECD, which was used as the template for **Modigy**[29][30][31][32]. Thus, PDB entry 4ZGM was first subjected to a comprehensive structural biophysical analysis[35], leading to the identification of two interfacial salt bridges (Table 1) and two interfacial hydrogen bonds (Table 2) at the binding interface of semaglutide and the GLP-1R ECD.

PDB ID	Residue A	Atom A	Residue B	Atom B	Distance (Å)
4ZGM	B_LYS_26	NZ	A_GLU_128	OE1	3.409
4ZGM	B_LYS_26	NZ	A_GLU_128	OE2	2.771

Table 1. Interfacial salt bridge analysis of PDB entry 4ZGM[24][36]). In this table, the residue naming scheme is Chain ID_Residue Name_Residue ID, and the residue ID numbering scheme is described in Table 1 of the supplementary file **supps.pdf**, and chains A and B represent the GLP-1R ECD and semaglutide, respectively.

Subsequently, the binding affinity of native semaglutide and GLP-1R ECD was calculated by Prodigy[31][32] to be 3.4 $\times$ 10^{-6} M at 37 $^{\circ}$C, with the template being PDB entry: 4ZGM[24], where a total of 42 residue pairs were found to be located at the binding interface of semaglutide and GLP-1R ECD[24][36], as listed in Table 4 of the supplementary file **supps.pdf**. According to the threshold distance of 5.0 Å used by Prodigy[31][32], among the 42, only two pairs of charged residues are in interfacial contact: (a) B_LYS_26 of semaglutide and A_GLU_128 of GLP-1R ECD; (b) B_ARG_36 of semaglutide and A_GLU_68 of GLP-1R ECD[24].

PDB ID	Acceptor (A)	Donor (D)	Hydrogen (H)	D-A (Å)	H-A (Å)	$\angle ADH(^\circ)$
4ZGM	O, B_VAL_33	NH2, A_ARG_121	HH22, A_ARG_121	2.86	2.10	23.30
4ZGM	OE2, A_GLU_128	NZ, B_LYS_26	HZ2, B_LYS_26	2.77	2.03	27.77

Table 2. Interfacial side chain hydrogen bond analysis of PDB entry 4ZGM[24][36]. In this table, the residue naming scheme is the same as in Table 1, the residue ID numbering scheme is described in Table 1 of the supplementary file **supps.pdf**, while $\angle ADH$ represents the angle formed by acceptor (A), donor (D), and hydrogen (H), and chains A and B represent GLP-1R ECD and semaglutide, respectively.

As such, the interfacial residue pair of B_LYS_26 and A_GLU_128 plays a pivotal role both in the binding of semaglutide to GLP-1 ECD and in the stabilization of the ligand-receptor complex structure[21][38][39][40][41][42], because

1. the two oppositely charged residues are in interfacial contact, according to the Prodigy[31][32] analysis of PDB entry 4ZGM[24];
2. two interfacial salt bridges (Table 1) are formed between the oppositely charged side chains of B_LYS_26 and A_GLU_128;
3. one interfacial hydrogen bond (Table 2) is formed between the oppositely charged side chains of B_LYS_26 and A_GLU_128.

PDB ID	Residue A	Atom A	Residue B	Atom B	Distance (Å)
4ZGM	B_ARG_36	NH1	A_GLU_68	OE1	4.147
4ZGM	B_ARG_36	NH1	A_GLU_68	OE2	4.835
4ZGM	B_ARG_36	NH2	A_GLU_68	OE1	4.893
4ZGM	B_ARG_36	NH2	A_GLU_68	OE2	6.269

Table 3. Interfacial residue pair distance analysis of PDB entry: 4ZGM[24]. In this table, the residue naming scheme is the same as in Table 1, and the residue ID numbering scheme is described in Table 1 of the supplementary file **supps.pdf**, and chains A and B represent GLP-1R ECD and semaglutide, respectively.

Apart from the residue pair of B_LYS_26 and A_GLU_128 (Tables 1 and 2), the Prodigy[31][32] analysis of PDB entry 4ZGM[24] identified another interfacial residue pair of B_ARG_36 of semaglutide and A_GLU_68 of GLP-1R ECD[24], for which a detailed distance analysis is included in Table 3. For B_ARG_36 and A_GLU_68, the distances between their oppositely charged side chains are well above the cut-off distance (4.0 Å) as used in[35]. As such, the electrostatic interaction between B_ARG_36 and A_GLU_68 is not categorized as a salt bridge[35] here.

Rational design of semaglutide analogues with improved GLP-1R ECD binding affinity

In 2021, I introduced a manually designed Val-Arg exchange into the peptide backbone of semaglutide to improve its GLP-1R ECD binding affinity[28]. In this study, an automated high-throughput approach was employed for a systematic amino acid scanning of semaglutide's backbone to enhance GLP-1R ECD affinity and GLP-1R activation thereafter. With the **Modigy**[29] workflow, a total of 564 amino acid sequences were computationally designed with three missense mutations according to the sequence template of the semaglutide backbone in Figure 1, where Modeller[30] was used to generate at least 5000 homology complex structural models of GLP-1R ECD-semaglutide analogue with PDB entry 4ZGM[24][36] as the structural template, and afterwards, for each structural model, Prodigy[31][32] was used for the structural biophysics-based calculations of K_d between semaglutide analogues and GLP-1R ECD.

For the 564 semaglutide analogues, the supplementary file **sema.txt** provides a summary of:

1. an identification number of the backbone sequence, i.e., from 1 to 564;
2. three site-specific missense mutations introduced into the peptide backbone;
3. three site-specific missense mutations sorted according to their residue IDs;
4. original sequence of the peptide backbone of semaglutide, as in Figure 1;
5. mutant sequence of the peptide backbone of semaglutide;
6. an amino acid sequence alignment of the original and the mutant;
7. a statistical summary of the ligand-receptor binding affinity K_d values between native semaglutide and GLP-1R ECD (PDB entry 4ZGM[24][36]);
8. a statistical summary of the ligand-receptor binding affinity K_d values between mutant semaglutide and GLP-1R ECD.

In supplementary file **sema.txt**, a statistical summary includes the average, the standard deviation, the maximum, and the minimum of the ligand-receptor binding affinity K_d values, where the average K_d was used to rank the 564 semaglutide analogues, with the average K_d of native semaglutide as the baseline. Of note, for three site-specific missense mutations in supplementary file **sema.txt**, the residue ID numbering scheme for semaglutide's peptide backbone is from 1 to 28, as included in Table 1 of supplementary file **supps.pdf**, instead of from 7 to 37 as in Figure 1.

Rational design of semaglutide analogues: a biophysical perspective

In supplementary file **sema.txt**, semaglutide mutant 524 is ranked first due to its highest average GLP-1R ECD binding affinity (K_d). Chemically, semaglutide mutant 524 is computationally designed as the peptide backbone of semaglutide (Figure 1) with three missense mutations, i.e., I20B_Q, L23B_R, and V24B_N (Table 1 of supplementary file **supps.pdf**). To further test if the three missense mutations do lead to a significantly increased GLP-1R ECD binding affinity, the **Modigy**[29] workflow was conducted 10,000 times for both native semaglutide and semaglutide mutant 524, generating 10,000 structural models built by Modeller[30] each for native semaglutide and semaglutide mutant 524. Afterwards, the 20,000 structural models were subjected to ligand-receptor K_d calculations by Prodigy[31][32].

Of the two sets of K_d data, a statistical analysis indicates a clear difference between them, where neither dataset follows a normal distribution (p-values ≈ 0). Given the non-normality of the data (Figure 2), a Wilcoxon test was used, yielding a final p-value of approximately 0, strongly suggesting that the two datasets are extremely different from each other. Taken together, as shown also in Figure 2, compared

with that (3.278×10^{-6} M) of native semaglutide (PDB entry 4ZGM[24]), semaglutide mutant 524 possesses a significantly higher (at least one degree of order) GLP-1R ECD binding affinity (1.462×10^{-7} M) to enhance GLP-1R activation.

PDB file	Residue A	Atom A	Residue B	Atom B	Distance (Å)
4452	B_LYS_117	NZ	A_GLU_100	OE2	3.505
4452	B_ARG_123	NH1	A_GLU_40	OE1	3.417
4452	B_ARG_127	NH1	A_GLU_40	OE1	3.122
4452	B_ARG_127	NH1	A_GLU_40	OE2	4.023

Table 4. Interfacial salt bridge analysis[35] of the 4452^{th} structural model (supplementary file **4452.pdb**) of semaglutide mutant 524, the first in supplementary file **sema.txt**. In this table, the residue naming scheme is the same as in Table 1, and the residue ID numbering scheme is described in Table 1 of supplementary file **supps.pdf**. Chains A and B represent GLP-1R ECD and semaglutide, respectively, while the one row with a yellow background represents the interfacial salt bridge also identified in PDB entry 4ZGM[24][36].

To further investigate how the three missense mutations (I20B_Q, L23B_R, and V24B_N, Figure 2) led to a significantly increased GLP-1R ECD binding affinity, a comprehensive structural biophysical analysis[35] was conducted for the 10000 structural models of semaglutide mutant 524 built by Modeller[30]. As mentioned above, the role of charge-charge interactions is only as important as 4.76% (2/42, Table 4 of supplementary file **supps.pdf**) in the stabilization of the complex structure of GLP-1R ECD and semaglutide[24], including only two interfacial salt bridges (Table 1) and two interfacial hydrogen bonds (Table 2). Hence, the comprehensive structural biophysical analysis[35] here focuses primarily on the electrostatic interactions at the binding interface between GLP-1R ECD and semaglutide or its analogue (mutant 524).

Specifically, one structural model stood out, i.e., the 4452^{th} structural model (supplementary file **4452.pdb**) of semaglutide mutant 524, where four salt bridges were identified at the binding interface of GLP-1R ECD and semaglutide mutant 524, as listed in Table 4. Among the four, three new interfacial salt

bridges (Figures 3 and 4) were found to be formed due to the missense mutations introduced into semaglutide's backbone (including in particular L23B_R), whereas only one interfacial salt bridge (yellow row in Table 4) was found to be formed and was the same as the salt bridge (Table 1) identified in PDB entry 4ZGM[24][36].

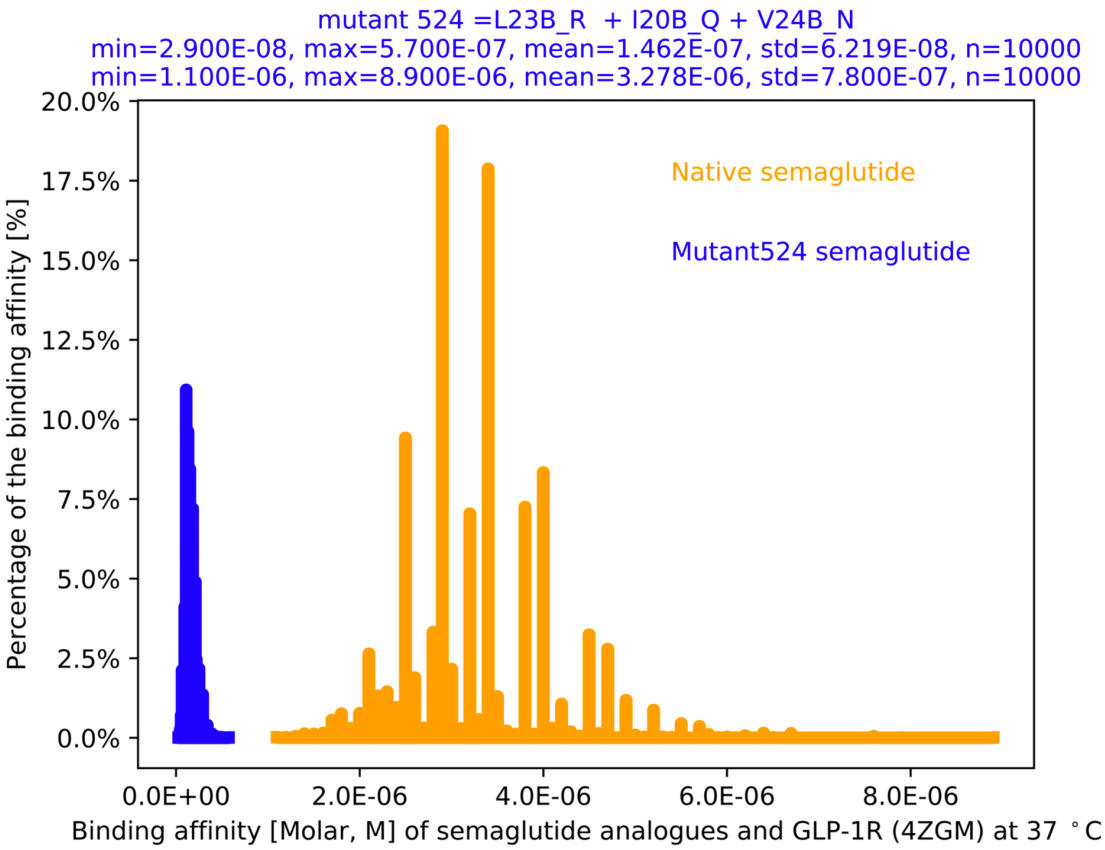


Figure 2. Distributions of the GLP-1R ECD binding affinities of native (the orange histogram) semaglutide (PDB ID: 4ZGM[24]) and semaglutide mutant 524 (the blue histogram). In the title (top) of this figure, L23B R, I20B Q, and V24B R represent three positions where site-specific missense mutations are introduced to the peptide backbone of semaglutide to improve its GLP-1R ECD binding affinity.

From Figures 3 and 4, it is quite clear that there is a cluster of three charged residues, i.e., Arg123 and Arg127 of semaglutide's backbone, and Glu40 of GLP-1R ECD, sitting at the binding interface of the semaglutide-GLP-1R ECD complex structure. Here, both Arg123 and Arg127 of semaglutide's backbone (cyan cartoon, Figures 3 and 4) formed interfacial salt bridges with Glu40 of GLP-1R ECD (green cartoon, Figures 3 and 4), the details of which are included in Table 4 and shown in Figure 4.

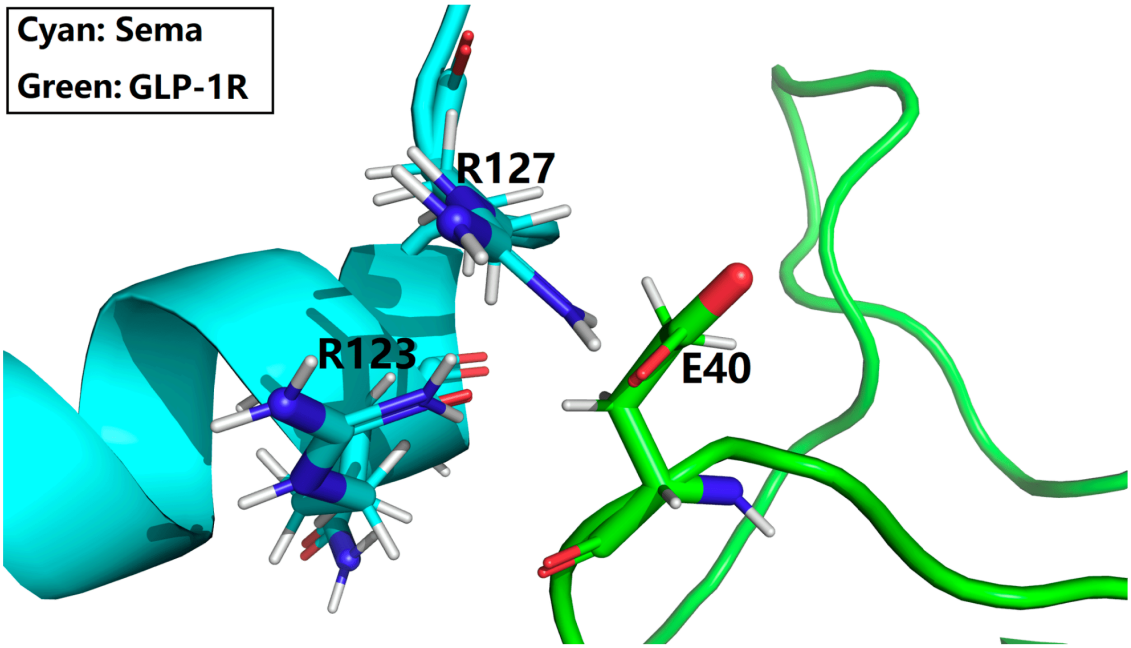


Figure 3. A zoomed-in view of the binding interface of GLP-1R and semaglutide mutant 4452 (Figure 2). This figure is prepared with PyMol[43] with supplementary file 4452.pdb, and the details of the interfacial salt bridges are included in Table 4.

doi.org/10.32388/FGTLKF

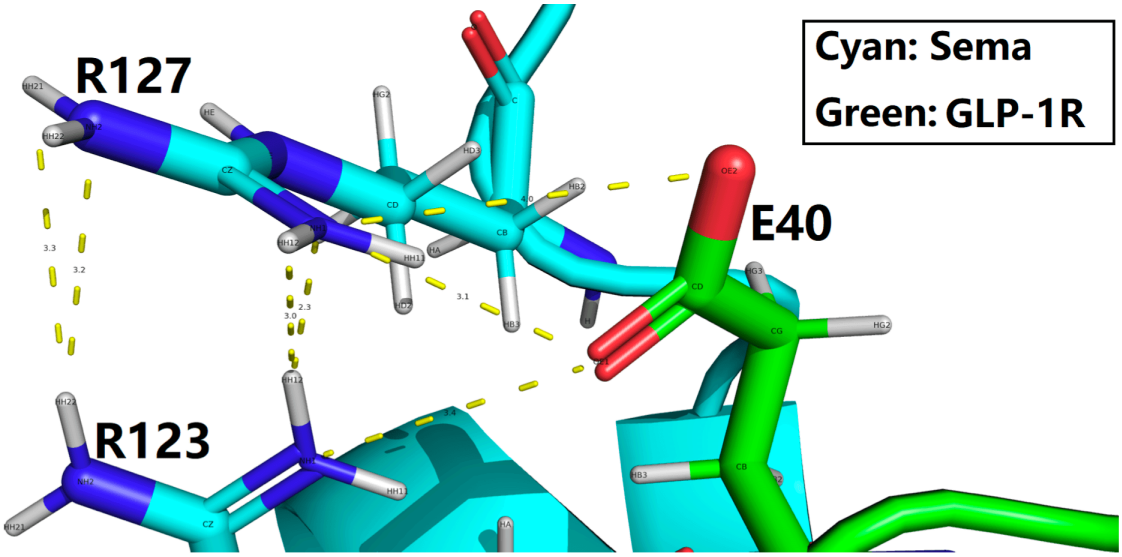


Figure 4. Salt bridges at the binding interface of GLP-1R and semaglutide mutant 4452 (Figure 2). This figure is prepared with PyMol[43] with supplementary file 4452.pdb, the details of the interfacial salt bridges (yellow dotted lines) are included in Table 4, and the details of other H-bond-like electrostatic interactions are included in Table 5.

Here, of note, as described in Tables 2 and 3 of supplementary file **supps.pdf**,

1. B_LYS_117 of semaglutide mutant 524 (supplementary file **4452.pdb**) corresponds to B_LYS_26 of native semaglutide (PDB entry 4ZGM[24][36]);
2. B_ARG_123 of semaglutide mutant 524 (supplementary file **4452.pdb**) corresponds to B_LEU_32 of native semaglutide (PDB entry 4ZGM[24][36]);
3. B_ARG_127 of semaglutide mutant 524 (supplementary file **4452.pdb**) corresponds to B_ARG_36 of native semaglutide (PDB entry 4ZGM[24][36]);
4. A_GLU_100 of semaglutide mutant 524 (supplementary file **4452.pdb**) corresponds to A_GLU_128 of native semaglutide (PDB entry 4ZGM[24][36]);
5. A_GLU_40 of semaglutide mutant 524 (supplementary file **4452.pdb**) corresponds to A_GLU_68 of native semaglutide (PDB entry 4ZGM[24][36]).

From Figures 3 and 4, in particular, it is obvious that the newly formed Arg123-Glu40 salt bridge is due to the missense mutation L23B_R (Figure 2), i.e., the substitution of leucine with a positively charged arginine at position 32 (instead of 23, Table 1 of supplementary file **supps.pdf**) of semaglutide's backbone

(Figure 2). Moreover, the newly formed Arg127-Glu40 salt bridge is at least partly due to the newly formed Arg123-Glu40 salt bridge, i.e., the presence of a positively charged arginine at position 32 of semaglutide's backbone (Figure 2), because in PDB entry 4ZGM[24][36], the distances (Table 3) between the oppositely charged side chains of A_GLU_68 and B_ARG_36 are far beyond the cut-off distance (4.0 Å) for salt bridge analysis as used in[35]. While in the case of semaglutide mutant 524 (supplementary file **4452.pdb**) here, one clear-cut interfacial salt bridge was formed at 3.122 Å (Table 4) between Arg127 in semaglutide[24][36]) and Glu40 of GLP-1R ECD, and another one also categorized here as an interfacial salt bridge at 4.023 Å, as the inter-actomic distance (4.023 Å, Table 4) equals almost the cut-off distance (4.0 Å) for salt bridge analysis as used in[35].

PDB	Acceptor (A)	Donor (D)	Hydrogen (H)	D-A (Å)	H-A (Å)	$\angle ADH(^\circ)$
4452	NH1, B_ARG_127	NH1, B_ARG_123	HH12, B_ARG_123	3.23	2.29	16.94
4452	NE, B_ARG_127	NH1, B_ARG_123	HH12, B_ARG_123	3.19	2.36	28.90

Table 5. Interfacial hydrogen bond analysis of the 4452th structural model (supplementary file **4452.pdb**) of semaglutide mutant 524. In this table, the residue naming scheme is the same as in Table 1, and the residue ID numbering scheme is described in Table 1 of supplementary file **supps.pdf**, and chains A and B represent GLP-1R ECD and semaglutide, respectively, while $\angle ADH$ represents the angle formed by acceptor (A), donor (D), and hydrogen (H).

In Figures 3 and 4, the cluster of three charged residues consists of Arg123, Arg127, and Glu40, where the two arginines possess positively charged side chains, while the glutamate (Glu40) possesses a negatively charged side chain. Energetically, it is not favorable for Arg123 and Arg127 to be too close to each other due to charge-charge repulsion, despite the presence of a negatively charged Glu40 which attracts them. Of interest is the existence of two interfacial hydrogen bonds (Table 5) between the side chains of Arg123 and Arg127, which constitute an additional set of attractive forces for the structural stability of the cluster of three charged residues (Figures 3 and 4) consisting of Arg123, Arg127, and Glu40. As such, these two interfacial hydrogen bonds (Table 5), along with the three interfacial salt bridges (Figures 3 and 4) between Arg123, Arg127, and Glu40, led to the establishment of a delicate structural electrostatic balance,

while maintaining the structural stabilization of the complex of GLP-R and semaglutide analogues, where such a delicate structural balance in hindsight entails a paradigm shift from the manual design approach in 2011[28] to an automated systemic approach as described in this manuscript, as it does not appear so straightforward for a manual design approach to uncover such a delicate structural balance at a ligand-receptor binding interface[24].

Overall, from a molecular energetic point of view, the formation of these three new interfacial salt bridges (Figures 3 and 4) acts like three electrostatic clips to further stabilize the ligand-receptor complex structure and contributes favorably to the binding affinity of semaglutide mutant 542 to GLP-1R ECD, with the potential of the semaglutide analogue binding stronger and thereby leading to stronger GLP-1R activation[24]. Yet, it remains an open question whether there is still room for optimization of the structural energetic (both electrostatics and hydrophobics) balance at the binding interface of GLP-1R ECD and semaglutide analogues, because after all, only three missense mutations were introduced into the peptide backbone of semaglutide with the Modigy[29] workflow here.

An electrostatic scaffold for the design of newer GLP-1R agonists

As mentioned above, the role of charge-charge interactions is only as important as 4.76% (2/42, Table 4 of the supplementary file **supps.pdf**) in the stabilization of the complex structure of GLP-1R ECD and semaglutide[24], including only two interfacial salt bridges (Table 1) and two interfacial hydrogen bonds (Table 2). As such, with 564 semaglutide analogues (supplementary file **sema.txt**) in place, a new set of Modigy[29] workflow was conducted with four missense mutations introduced into the peptide backbone of semaglutide, leading to the identification of a structural electrostatic scaffold at the binding interface of GLP-1R ECD and semaglutide analogue A747 (supplementary file **A747.pdb**) with four site-specific missense mutations, namely L23B_R, I20B_Q, V24B_R, and R25B_A.

Here, of note, this new Modigy[29] workflow with four missense mutations was only carried out for those key residue positions in semaglutide's backbone and is incomplete, because a complete Modigy[29] workflow here would require a total of 3,276,000,000 ($\frac{28!}{4!(24)!}$) times of structural modeling by Modeller[30] and ligand-receptor binding affinity (K_d) calculations by Prodigy[31][32], where the size of the computational task is too huge even for a traditional HPC facility[29]. In addition, as described in Tables 1 and 3 of the supplementary file **supps.pdf**,

1. I20B_Q represents a substitution of B_ILE_29 of native semaglutide (PDB entry 4ZGM[24][36]) with B_GLN_120 in the supplementary file **A747.pdb**;
2. L23B_R represents a substitution of B_LEU_32 of native semaglutide (PDB entry 4ZGM[24][36]) with B_ARG_123 in the supplementary file **A747.pdb**;
3. V24B_R represents a substitution of B_VAL_33 of native semaglutide (PDB entry 4ZGM[24][36]) with B_ARG_124 in the supplementary file **A747.pdb**;
4. R25B_A represents a substitution of B_ARG_34 of native semaglutide (PDB entry 4ZGM[24][36]) with B_ALA_125 in the supplementary file **A747.pdb**. Interestingly, this mutation along with V24B_R is equivalent to the manually designed Val-Arg exchange plus a Val-Ala substitution at position 34 of native semaglutide (PDB entry 4ZGM[24][36]).

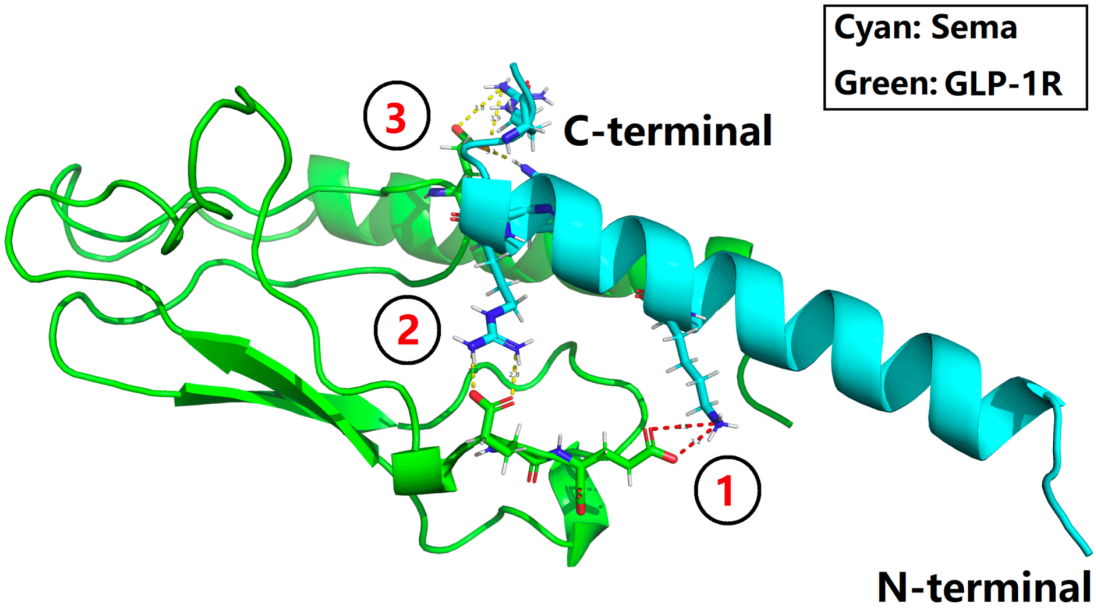


Figure 5. An overview of the complex structure of GLP-1R and semaglutide mutant A747. In this figure, the three red numbers (1, 2, and 3) within black circles represent the three sets of interfacial salt bridges (red and yellow dotted lines) of GLP-1R and semaglutide mutant A747, which constitute an electrostatic scaffold for the design of next-generation GLP-1R agonists with improved GLP-1R ECD affinity. This figure is prepared with PyMol[43] with the supplementary file A747.pdb, with details of the three sets of interfacial salt bridges included in Table 6.

With this partial Modigy[29] workflow, a total of 9 interfacial salt bridges (Table 6) were identified at the binding interface of GLP-1R ECD and semaglutide analogue A747 (supplementary file **A747.pdb**), in comparison with 2 interfacial salt bridges (Table 1) for PDB entry 4ZGM[24][36]), and in comparison with 7 interfacial salt bridges (Table 4) for supplementary file **4452.pdb**. Specifically, three are a total of 7 newly formed interfacial salt bridges (Table 6) which are due to the four site-specific missense mutations L23B_R, I20B_Q, V24B_R, and R25B_A introduced into the peptide backbone (Figure 1) of semaglutide.

From a structural point of view, the nine interfacial salt bridges (Table 6) constitute a structural electrostatic scaffold (Figure 5) at the binding interface of GLP-1R ECD and semaglutide analogue A747, which consists of three sets of interfacial salt bridges, as designated by three red numbers (1, 2, and 3) within black circles in Figure 5, and also in Figures 1-5 in supplementary file **supps.pdf**. Moreover, the role of charge-charge interactions is as important as 18.18% (8/44, Table 5 in supplementary file **supps.pdf**) for semaglutide analogue A747 to GLP-1R ECD, compared with 4.76% (2/42, Table 4 in supplementary file **supps.pdf**) for native semaglutide to GLP-1R ECD, in the stabilization of the semaglutide backbone in complex with GLP-1R ECD (PDB ID: 4ZGM[24]).

PDB file	Residue A	Atom A	Residue B	Atom B	Distance (Å)
A747	B_LYS_117	NZ	A_GLU_100	OE1	3.770
A747	B_LYS_117	NZ	A_GLU_100	OE2	3.153
A747	B_ARG_123	NH2	A_GLU_40	OE1	2.661
A747	B_ARG_124	NH1	A_GLU_99	OE1	2.751
A747	B_ARG_124	NH1	A_GLU_99	OE2	3.489
A747	B_ARG_124	NH2	A_GLU_99	OE1	3.557
A747	B_ARG_124	NH2	A_GLU_99	OE2	3.026
A747	B_ARG_127	NH2	A_GLU_40	OE1	3.889
A747	B_ARG_127	NH2	A_GLU_40	OE2	3.764

Table 6. Interfacial salt bridging network analysis of semaglutide analogue A747 in complex with GLP-1R ECD (supplementary file **A747.pdb**. In this table, the residue naming scheme is the same as Table 1, and the residue ID numbering scheme is described in Table 1 of supplementary file **supps.pdf**, and chains A and B represent GLP-1R ECD and semaglutide, respectively, while rows with yellow, green, and cyan backgrounds represent the first (1 in Figure 5), the second (2 in Figure 5), and the third (3 in Figure 5) sets of interfacial salt bridges, as shown in Figure 5 and also in the Graphical Abstract of this manuscript.

Conclusion

In summary, this article reports a structural biophysics-driven computational framework, i.e., the **Modigy** workflow[29], for the design of semaglutide analogues (Figure 6), emphasizing structural and electrostatic optimization to enhance GLP-1R activation, to address the concern of the market with respect to the weight loss efficacy of CargiSema developed by Novo Nordisk. In addition to 564 semaglutide analogues (supplementary file **sema.txt**) with site-specific missense mutations, this study delves into the sequence space of GLP-1R ECD and semaglutide, and uncovers an electrostatic scaffold of intermolecular salt bridges at the ligand-receptor binding interface, which help stabilize the complex structure to enhance GLP-1R activation, providing a robust framework for the development of next-generation GLP-1R agonists with improved efficacy[44][45][46][47].

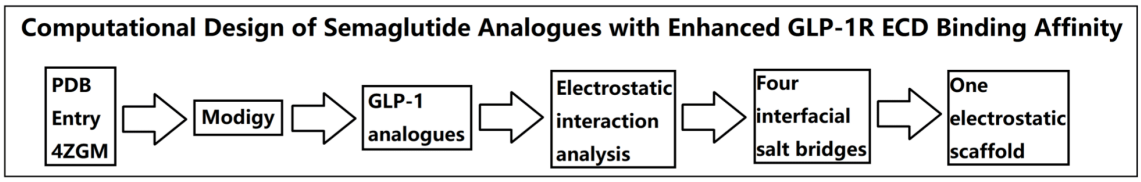


Figure 6. A flowchart overview of the materials, methods, results, data analysis, and conclusion of this study. In this figure, Modigy[29] is an abbreviation of Modeller[30] and Prodigy[31][32].

Discussion

Originally coined in 2024[29], **Modigy** is defined as an abbreviation of Modeller[30] and Prodigy[31][32] to represent an in silico high-throughput generation of structural and intermolecular binding affinity (K_d) data[29]. Since its inception, the **Modigy** workflow has aimed to contribute to structure-based and biophysics-driven biomolecular discovery and design, making use of the experimental complex structures deposited in the Protein Data Bank, e.g., ligand-receptor, antibody-antigen, etc. As such, for the continued development of next-generation GLP-1R agonists (e.g., CagriSema[20] and IcoSema[15]) with improved efficacy[44][45][46][47], the **Modigy** workflow is applicable not just for GLP-1R and GLP-1 (analogues), but also for amylin[48][49], glucose-dependent insulinotropic polypeptide receptor (GIPR)[50], glucagon receptor (GCGR)[51], and beyond, so long as an experimental complex structure of the interacting partners is already deposited in the Protein Data Bank[37] [52].

Since the regulatory approval of semaglutide in 2017, the obesity and diabetes drug market has become highly competitive, led by Novo Nordisk and Eli Lilly. For instance, Novo Nordisk's semaglutide-based drugs (Ozempic, Wegovy) face competition from Eli Lilly's tirzepatide-based Mounjaro and Zepbound, which have demonstrated superior weight loss efficacy, while Novo Nordisk's CagriSema has recently shown promising but not market-leading results, further intensifying industry competition. Both companies are expanding production and clinical trials to solidify their positions, as, for instance, on December 6, 2024, the European Commission approved Novo Holdings' acquisition of Catalent and related manufacturing sites for weight-loss drug production. While both market competition and patient needs are commercially relevant for pharmaceutical companies, long-term commercial success depends on delivering continuously effective, safe, and innovative treatments for patients. As such, given that the R&D process of a drug is itself a continued multi-parameter optimization process, this study here does

not focus primarily on market/product competition. Instead, this study here focuses primarily on continued exploration of the unexplored sequence/chemical space of GLP-1R agonists, not just a systematic analysis of both Ala- and Aib-scanning (Figure 1), for the development of next-generation GLP-1R agonists with improved efficacy[44][45].

Statements and Declarations

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the author used OpenAI's ChatGPT for the choice of words and for the organization of a few sentences in order to improve the readability of the manuscript. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for this manuscript.

References

1. ^*Emily K. Sims, Alice L. J. Carr, Richard A. Oram, Linda A. DiMeglio, Carmella Evans-Molina. (2021). 100 years of insulin: Celebrating the past, present and future of diabetes therapy. Nature Medicine. 27(7):1154–1164. doi:10.1038/s41591-021-01418-2*
2. ^*Ignazio Vecchio, Cristina Tornali, Nicola Luigi Bragazzi, Mariano Martini. (2018). The discovery of insulin: An important milestone in the history of medicine. Frontiers in Endocrinology. 9. doi:10.3389/fendo.2018.00613*
3. ^*Irl B. Hirsch. (2005). Insulin analogues. New England Journal of Medicine. 352(2):174–183.*
4. ^*G. Scapin, V. P. Dandey, Z. Zhang, C. Strickland, C. S. Potter, et al. Insulin Receptor ectodomain in complex with one insulin molecule. Worldwide Protein Data Bank 2018.*
5. ^*Wei Li. (2020). Delving deep into the structural aspects of the BPro28-BLys29 exchange in insulin lispro: A structural biophysical lesson.*
6. ^*Lotte Bjerre Knudsen, Jesper Lau. (2019). The discovery and development of liraglutide and semaglutide. Frontiers in Endocrinology. 10:1–32.*
7. ^*John D. Bucheit, Lauren G. Pamulapati, Nicole Carter, Kevin Malloy, Dave L. Dixon, et al. (2020). Oral semaglutide: A review of the first oral glucagon-like peptide 1 receptor agonist. Diabetes Technology & Therapeutics. 22(1):10–18.*

8. ^Helena W. Rodbard, Julio Rosenstock, Luis H. Canani, Chaicharn Deerochanawong, Janusz Gumprecht, et a l. (2019). Oral semaglutide versus empagliflozin in patients with type 2 diabetes uncontrolled on metformi n: The PIONEER 2 trial. Diabetes Care. 42(12):2272–2281.
9. ^Ying Xiao, Lin Sun. (2019). Semaglutide in weight management. The Lancet. 394(10205):1226.
10. ^Sarah L. Anderson, Trevor R. Beutel, Jennifer M. Trujillo. (2020). Oral semaglutide in type 2 diabetes. Journ al of Diabetes and its Complications. 34(4):107520.
11. ^Michael E. Røder. (2019). Clinical potential of treatment with semaglutide in type 2 diabetes patients. Drug s in Context. 8:1–11.
12. ^Vanita R. Aroda, Matthew S. Capehorn, Louis Chaykin, Juan P. Frias, Nanna L. Lausvig, et al. (2019). Impact of baseline characteristics and beta-cell function on the efficacy and safety of subcutaneous once-weekly se maglutide: A patient-level, pooled analysis of the SUSTAIN 1-5 trials. Diabetes, Obesity and Metabolism. 22 (3):303–314.
13. ^Richard Pratley, Aslam Amod, Søren Tetens Hoff, Takashi Kadowaki, Ildiko Lingvay, et al. (2019). Oral sem aglutide versus subcutaneous liraglutide and placebo in type 2 diabetes (PIONEER 4): A randomised, double -blind, phase 3a trial. The Lancet. 394(10192):39–50.
14. ^Mansoor Husain, Andreas L. Birkenfeld, Morten Donsmark, Kathleen Dungan, Freddy G. Eliaschewitz, et a l. (2019). Oral semaglutide and cardiovascular outcomes in patients with type 2 diabetes. New England Jour nal of Medicine. 381(9):841–851.
15. a, bSanjay Kalra, Saptarshi Bhattacharya, Nitin Kapoor. (2021). Contemporary classification of glucagon-lik e peptide 1 receptor agonists (GLP1ras). Diabetes Therapy. 12(8):2133–2147.
16. ^Rajmohan Seetharaman. (2024). IcoSema's leap forward: New data from COMBINE 3 paves the way. Jour nal of Basic and Clinical Physiology and Pharmacology.
17. ^Lisbet Westergaard, Lene Alifrangis, Stephen T. Buckley, Hans Veit Coester, Thomas Klitgaard, et al. (2024). Pharmacokinetic properties of a once-weekly fixed-ratio combination of insulin icodec and semaglutide co mpared with separate administration of each component in individuals with type 2 diabetes mellitus. Clin Drug Investig. 44(11):849–861.
18. ^Ildiko Lingvay, Marisse Asong, Cyrus Desouza, Pierre Gourdy, Soumitra Kar, et al. (2023). Once-weekly ins ulin icodec vs once-daily insulin degludec in adults with insulin-naive type 2 diabetes: The ONWARDS 3 ran domized clinical trial. JAMA : the journal of the American Medical Association. 330(3):228–237.
19. ^Athena Philis-Tsimikas, Marisse Asong, Edward Franek, Ting Jia, Julio Rosenstock, et al. (2023). Switching t o once-weekly insulin icodec versus once-daily insulin degludec in individuals with basal insulin-treated ty

pe 2 diabetes (ONWARDS 2): A phase 3a, randomised, open label, multicentre, treat-to-target trial. Lancet Di abetes and Endocrinology. 11(6):414–425.

20. a, b *Caroline M. Apovian, Marie E. McDonnell. (2023). CagriSema and the link between obesity and type 2 di abetes. The Lancet. 402(10403):671–673. doi:10.1016/s0140-6736(23)01291-6*

21. a, b, c *Chelsea K. Longwell, Stephanie Hanna, Nina Hartrampf, R. Andres Parra Sperberg, Po-Ssu Huang, et a l. (2020). Identification of n-terminally diversified GLP-1R agonists using saturation mutagenesis and chem ical design. ACS Chemical Biology. 16(1):58–66. doi:10.1021/acschembio.0c00722*

22. ^ *Haibo Zhang, Tianxiao Wu, Yong Wu, Yuran Peng, Xian Wei, et al. (2024). Binding sites and design strategi es for small molecule GLP-1R agonists. European Journal of Medicinal Chemistry. 275:116632. doi:10.1016/j.ej mech.2024.116632*

23. ^ *Xin Zhang, Matthew J. Belousoff, Yi-Lynn Liang, Radostin Danev, Patrick M. Sexton, et al. (2021). Structure and dynamics of semaglutide- and taspoglutide-bound GLP-1R-Gs complexes. Cell Reports. 36(2):109374. d oi:10.1016/j.celrep.2021.109374*

24. a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, s, t, u, v, w, x, y, z, ^, ^, ^, ^, ^, ^, ^, ^, ^, ^, ^, ^ *Jesper Lau, Paw Bloch, Lauge Sch äffer, Ingrid Pettersson, Jane Spetzler, et al. (2015). Discovery of the once-weekly glucagon-like peptide-1 (GL P-1) analogue semaglutide. Journal of Medicinal Chemistry. 58(18):7370–7380.*

25. a, b *Ting Dang, Jie Yu, Zhihe Cao, Bingjie Zhang, Shanshan Li, et al. (2024). Endogenous cell membrane inter actome mapping for the GLP-1 receptor in different cell types. Nature Chemical Biology. doi:10.1038/s41589-024-01714-1*

26. a, b *Shane C. Wright, Aikaterini Motso, Stefania Koutsilieri, Christian M. Beusch, Pierre Sabatier, et al. (2023). GLP-1R signaling neighborhoods associate with the susceptibility to adverse drug reactions of incretin mim etics. Nature Communications. 14(1). doi:10.1038/s41467-023-41893-4*

27. a, b, c, d, e, f *Jonathon R. Sawyer, Joseph A. Audie, Jon Swanson, David Diller, Solimar Santiago, et al. (2024). D esign, structure–activity relationships, and computational modeling studies of a series of α-Helix biased, ul tra-short glucagon-like peptide-1 receptor agonists. Molecules (Basel, Switzerland). 30(1):12. doi:10.3390/mo lecules30010012*

28. a, b, c, d *Wei Li. (2021). Strengthening semaglutide-GLP-1R binding affinity via a Val27-Arg28 exchange in th e peptide backbone of semaglutide: A computational structural approach. Journal of Computational Biophy sics and Chemistry. 20(05):495–499.*

29. a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p *Wei Li. (2024). In silico generation of structural and intermolecular binding af finity data with reasonable accuracy: Expanding horizons in drug discovery and design. doi:10.20944/prepr*

ints202405.1739.v1

30. a, b, c, d, e, f, g, h, i, j, k*Benjamin Webb, Andrej Sali. Protein structure modeling with MODELLER. In: Methods i n molecular biology.: Springer US 2020. pp. 239–255.*

31. a, b, c, d, e, f, g, h, i, j, k, l, m*Anna Vangone, Alexandre MJJ Bonvin. (2015). Contacts-based prediction of binding affinity in protein-protein complexes. eLife. 4.*

32. a, b, c, d, e, f, g, h, i, j, k, l, m*Li C. Xue, João Pglm Rodrigues, Panagiotis L. Kastritis, Alexandre Mjj Bonvin, Anna Vangone. (2016). PRODIGY: A web server for predicting the binding affinity of protein-protein complexes. Bi oinformatics (Oxford, England). :btw514.*

33. a, b*Matthew P. Gilbert, Richard E. Pratley. (2020). GLP-1 analogs and DPP-4 inhibitors in type 2 diabetes the rapy: Review of head-to-head clinical trials. Frontiers in Endocrinology. 11. doi:10.3389/fendo.2020.00178*

34. a, b*Carolyn F. Deacon. (2020). Dipeptidyl peptidase 4 inhibitors in the treatment of type 2 diabetes mellitus. Nature Reviews Endocrinology. 16(11):642–653. doi:10.1038/s41574-020-0399-8*

35. a, b, c, d, e, f, g, h, i, j, k*Wei Li. (2017). How do SMA-linked mutations of SMN1 lead to structural/functional defi ciency of the SMA protein? Cedric Raouleditor. PLOS ONE. 12(6):e0178519.*

36. a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, s, t, u, v*S. Reedtz-Runge. Crystal structure of Semaglutide peptide backbo ne in complex with the GLP-1 receptor extracellular domain. Worldwide Protein Data Bank 2015.*

37. a, b*Helen Berman, Kim Henrick, Haruki Nakamura. (2003). Announcing the worldwide protein data bank. N ature Structural & Molecular Biology. 10(12):980–980.*

38. ^*Di Wu, Jing Sun, Tianlei Xu, Shuning Wang, Guoqing Li, et al. (2010). Stacking and energetic contribution of aromatic islands at the binding interface of antibody proteins. Immunome Research. 6(Suppl 1):S1.*

39. ^*F. P. Davis, A. Sali. (2005). PIBASE: A comprehensive database of structurally defined protein interfaces. Bi oinformatics (Oxford, England). 21(9):1901–1907.*

40. ^*Harry C. Jubb, Arun P. Pandurangan, Meghan A. Turner, Bernardo Ochoa-Montaño, Tom L. Blundell, et al. (2017). Mutations at protein-protein interfaces: Small changes over big surfaces have large impacts on hum an health. Progress in Biophysics and Molecular Biology. 128:3–13.*

41. ^*James A. Wells, Christopher L. McClendon. (2007). Reaching for high-hanging fruit in drug discovery at pr otein-protein interfaces. Nature. 450(7172):1001–1009.*

42. ^*Emília Antolíková, Lenka Žáková, Johan P. Turkenburg, Christopher J. Watson, Ivona Hančlová, et al. (201 1). Non-equivalent role of inter- and intramolecular hydrogen bonds in the insulin dimer interface. Journal of Biological Chemistry. 286(42):36968–36977.*

43. a, b, cDeLano WL. Pymol: An open-source molecular graphics tool. CCP4 Newsletter On Protein Crystallography. 2002;40:82–92.

44. a, b, cDileep Francis, Aksa Mariyam Chacko, Anagha Anoop, Subramani Nadimuthu, Vaishnavi Venugopal. Evolution of biosynthetic human insulin and its analogues for diabetes management. In: Hormones.: Elsevier 2024. pp. 191–256. ISBN 978-0-443-19346-0

45. a, b, cJulio Rosenstock, Rattan Juneja, John M. Beals, Julie S. Moyers, Liza Ilag, et al. (2024). The basis for weekly insulin therapy: Evolving evidence with insulin icodec and insulin efsitora alfa. Endocrine Reviews. 45 (3):379–413. doi:10.1210/endrev/bnad037

46. a, bU. Derewenda, Z. Derewenda, G. G. Dodson, R. E. Hubbard, F. Korber. (1989). Molecular Structure of insulin: The insulin monomer and its assembly. British Medical Bulletin. 45(1):4–18.

47. a, bL. J. Slieker, G. S. Brooke, R. D. DiMarchi, D. B. Flora, L. K. Green, et al. (1997). Modifications in the B10 and B26-30 regions of the B chain of human insulin alter affinity for the human IGF-I receptor more than for the insulin receptor. Diabetologia. 40(14):S54–S61.

48. ^Thomas Kruse, Jakob Lerche Hansen, Kirsten Dahl, Lauge Schäffer, Ulrich Sensfuss, et al. (2021). Development of cagrilintide, a long-acting amylin analogue. Journal of Medicinal Chemistry. 64(15):11183–11194. doi:10.1021/acs.jmedchem.1c00565

49. ^Wei Li. (2024). Identification of electrostatic hotspots at the binding interface of amylin and insulin-degrading enzyme: A structural and biophysical investigation. doi:10.20944/preprints202402.0063.v1

50. ^Wenzhuo Li, Qingtong Zhou, Zhaotong Cong, Qingning Yuan, Wenxin Li, et al. (2024). Structural insights into the triple agonism at GLP-1R, GIPR and GCGR manifested by retatrutide. Cell Discovery. 10(1). doi:10.1038/s41421-024-00700-0

51. ^Julio Rosenstock, Juan Frias, Ania M. Jastreboff, Yu Du, Jitong Lou, et al. (2023). Retatrutide, a GIP, GLP-1 and glucagon receptor agonist, for people with type 2 diabetes: A randomised, double-blind, placebo and active-controlled, parallel-group, phase 2 trial conducted in the USA. The Lancet.

52. ^Wei Li, Gary Vottevor. (2023). Towards a truly general intermolecular binding affinity calculator for drug discovery & design. doi:10.20944/preprints202208.0213.v2

Supplementary data: available at https://doi.org/10.32388/FGTLKF

Declarations

Funding: No specific funding was received for this work.

Potential competing interests: No potential competing interests to declare.