Semaglutide Analogues Enhancing GLP-1R Activation: A Dynamic Structural & Electrostatic Perspective

Wei Li

Contrebola Institute of Computational Interstructural Biophysics (CICIB),
No. 88, Renaissance East Road, Nantong City 226000,
Jiangsu Province, People's Republic of China

* Corresponding email: wli148@aucklanduni.ac.nz

Abstract

On December 5, 2017, the U.S. FDA approved Novo Nordisk A/S's diabetes drug Ozempic, whose active ingredient, semaglutide, represents a major advancement in the class of glucagon-like peptide-1 (GLP-1) analogues. These analogues mimic the action of an intestinal hormone that promotes insulin secretion, thereby playing a pivotal role in metabolic regulation. Since then, semaglutide has profoundly transformed the treatment landscape for type 2 diabetes and obesity, demonstrating efficacy in glycemic control, weight management and cardiovascular risk mitigation. Yet, it remains an open question whether semaglutide is the best GLP-1 receptor (GLP-1R) agonist either as of now or in future. Given GLP-1R as the therapeutic target, a structural basis of the GLP-1-GLP-1R interaction is paramount for the design of GLP-1 analogues that fully harness the therapeutic potential of GLP-1R activation. Thus, this study employs a structural biophysical approach to enhance GLP-1R activation through the rational design of semaglutide analogues. By incorporating an electrostatic scaffold of interfacial salt bridges, site-specific missense mutations were engineered into the semaglutide backbone to establish additional stabilizing interactions with the extracellular domain (ECD) of GLP-1R, thereby enhancing ligand-receptor binding and activation. In addition to 564 semaglutide analogues, the structural biophysical analysis here also reveal the presence of an electrostatic scaffold comprising four salt bridges at the semaglutide-GLP-1R binding interface. Drawing parallels with the continued optimization of insulin analogues, this articles proposes that the electrostatic scaffold offers a robust framework for the continued development of next-generation GLP-1R agonists, enabling more precise and effective therapies for managing metabolic diseases.

Introduction

It is now more than a century since the discovery of insulin for diabetes treatment, which led to life-saving animal-derived insulin therapies [1, 2]. Afterwards, synthetic human insulin and generations of various insulin analogues revolutionized glycemic management [3–12]. Of note, the history of insulin [**?**, 13–16] is inextricably linked to Novo Nordisk, which played a pivotal role in advancing diabetes care, from continued innovations and development of insulin analogues to R&D of GLP-1 receptor (GLP-1R) agonists like semaglutide [17–21].

Approved by FDA in 2017, semaglutide's extended half-life allows weekly dosing, improving patient compliance and clinical outcomes [22–25]. Beyond type 2 diabetes,

semaglutide is also used in weight management and the treatment of cardiovascular disorders [19]. Remarkably, since the FDA nod in 2017, Novo Nordisk's market value has reached as high as 570 billion USD, surpassing the size of Denmark's entire economy (Bloomberg, accessed on May 1, 2024).

Pharmacologically, GLP-1R is a G-protein coupled receptor (GPCR) critical for glucose homeostasis, which mediates the effects of GLP-1R agonists [26, 27]. Structural understanding of GLP-1−GLP-1R interaction, especially GLP-1 interactions with the extracellular domain (ECD) of GLP-1R, is vital for optimizing GLP-1-based therapeutics and designing next-generation GLP-1R agonists with improved pharmacological profiles, as GLP-1R ECD mediates the initial step in ligand-receptor binding, downstream GLP-1R activation and transmembrane sigal transduction [28–31].

Motivation

While semaglutide's economic success as a GLP-1R agonist sets a benchmark for the market, it will not be the best GLP-1R agonist in glycemic control and weight management, and there is still need for continued innovation in the progression of peptide-based therapies for metabolic disorders, as exemplified in the history of insulin of continued design and development of analogues with enhanced pharmacological profiles [1, 32–34]. In addition, challenges like individual tolerability and competition from dual agonists, such as Eli Lilly's tirzepatide, also demand further optimization [35–37]. As a matter of fact, even Novo Nordisk is in pursuit of continued innovation in multifunctional therapeutics, exemplified by IcoSema, a combination of semaglutide and insulin icodec [38–42]. As another example, a manually designed Val27-Arg28 exchange [43] was introduced into the backbone of semaglutide to improve its GLP-1R affinity. Therefore, this study focuses on designing semaglutide analogues with site-specific missense mutations, and aims to strengthen GLP-1R activation through enhanced GLP-1R ECD binding affinity [26], contributing to the development of next-generation GLP-1R agonists with improved efficacy and specificity.

Materials and Methods

To date, there is one experimental structure determined by X-ray diffraction of the semaglutide backbone in complex with the extracellular domain (ECD) of GLP-1R (PDB ID: 4ZGM [29]), which is used here as the template for structural modeling by Modeller [44] and ligand-receptor binding affinity (K_d) calculations by Prodigy [45, 46]. In short, this article employs an automated systemic scanning the amino acids of the backbone of semaglutide as shown in Figure 1, where the semaglutide-GLP-1R complex structure (PDB ID: 4ZGM [29]), an in silico high-throughput generation [47] of structural and intermolecular binding affinity (K_d) data follows three criteria:

1. position 8 (scissor in Figure 1) remains unchanged;
2. position 26 (lysine) of semaglutide's backbone remains unchanged;
3. there is only one lysine (Lys26) in the backbone of semaglutide.

Subsequently, for all Modeller-generated [44] structural models of semaglutide analogues in complex with GLP-1R ECD, a comprehensive structural biophysical analysis [50] was conducted for the identification of key amino acid residues at the binding interface [29, 43]. Specifically, the cut-off distance is 4.0 Å for salt bridge analysis [50], 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

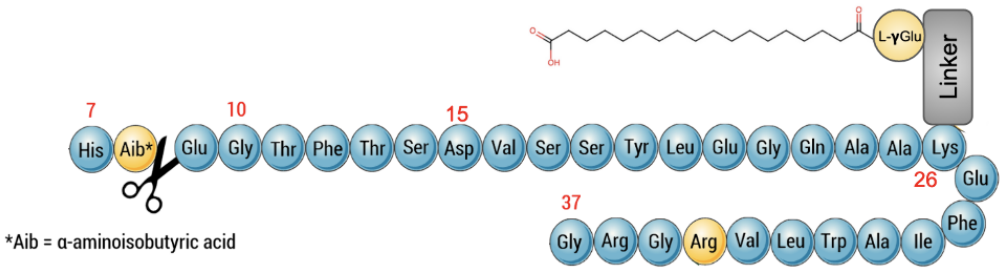

Fig 1. 2D structure of semaglutide. In this figure, the sequence of semaglutide's backbone is numbered from 7 to 37, with modifications highlighted in yellow and the dipeptidyl peptidase-4 (DPP4) cleave site [48, 49] indicated with a scissor. This figure is obtainted from the PDB website (URL: https://pdb101.rcsb.org/global-health/diabetes-mellitus/drugs/incretins/drug/semaglutide/semaglutide, accessed on November 22, 2024).

cutoff value of donor-acceptor (D-A) distance at 3.0 Å [50]. All structural modeling (Modeller [44]) and K_d calculations (Prodigy [45, 46]) were performed with the high performance computing facility at Wuxi Taihu Lake HPC center.

Results

Semaglutide binding to GLP-1R ECD [29] is the first step that initiates the ligand-receptor binding and downstream GLP-1R activation and transmembrane sigal transduction [28, 30, 31]. As a result, this section focuses on direct binding and interaction of semaglutide to GLP-1R ECD to ensure that the semaglutide analogue reaches GLP-1R ECD faster, binds to GLP-1R ECD stronger (K_d) and faster (K_{on}), and dissociates from GLP-1R ECD slower (K_{off}), compared to native semaglutide reaching and binding to GLP-1R ECD [29].

A comprehensive structural biophysical analysis of semaglutide backbone in complex with the GLP-1 receptor ECD

On September 9, 2015, a team of Novo Nordisk deposited into Protein Data Bank [51] the first crystal structure (PDB ID: 4ZGM) of semaglutide backbone in complex with the GLP-1R ECD, PDB entry 4ZGM [29, 52].

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 bridging analysis of semaglutide backbone in complex with the GLP-1 receptor ECD (PDB ID: 4ZGM [29, 52]). In this table, the residue naming scheme is the same as Table 1, and the amino acids of semaglutide's backbone is numbered from 7 to 37 according to Figure 1.

As described above, this experimental structure (PDB ID: 4ZGM [29, 52]) was used as the template for structural modeling by Modeller [44] and ligand-receptor binding affinity calculations by Prodigy [45, 46], as described in [47]. Before that, however, it was first subjected to a comprehensive structural biophysical analysis [50] of the binding interface of semaglutide and GLP-1R ECD, which led to the identification of two salt bridges between B_LYS_26 of semaglutide and A_GLU_128 of GLP-1R ECD (Table 1).

Moreover, this comprehensive structural biophysical analysis [50] also identified two hydrogen bonds the binding interface of semaglutide and GLP-1R ECD [29, 52], as listed in Table 2. Of note, one of the interfacial hydrogen bond is formed between the side chains of B_LYS_26 of semaglutide and A_GLU_128 of GLP-1R ECD.

PDB ID	Acceptor (A)	Donor (D)	Hydrogen (H)	D-A (Å)	H-A (Å)	$\angle ADH(°)$
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. 4ZGM-specific interfacial side chain and main chain hydrogen bonding analysis. In this table, the residue naming scheme is the same as Table 1, $\angle ADH$ represents the angle formed by acceptor (A), donor (D) and hydrogen (H).

Subsequently, the binding affinity of native semaglutide and GLP-1R ECD was calculated by Prodigy [45, 46] to be 3.4×10^{-6} M at 37 °C, with the template being PDB entry: 4ZGM [29], where a total of 42 residue pairs were found to be located at the binding interface of semaglutide and GLP-1R ECD [29, 52], as listed in supplementary file **4zgm.ic**. Among the 42, there are only two pairs of charged residues: (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, as listed in Tables 1 and 3. As a result, this article focuses on the interaction between semaglutide and GLP-1R ECD with a structural electrostatic perspective in particular, including salt bridging and hydrogen bonding analysis [50] of the ligand-receptor binding interface [26, 53–57].

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

Table 3. Interfacial residue pair analysis of PDB entry: 4ZGM [29]. In this table, the residue naming scheme is the same as Table 1. In light of the cut-off distance (4.0 Å), the electrostatic interactions listed in this table are not categorized as salt bridges [50].

As an example of the structural electrostatic perspective, in 2021, a manually designed structural modification involving a Val27-Arg28 (Val33-Arg34 in PDB entry 4ZGM) exchange (V-R exchange) [43] was introduced into the backbone of semaglutide towards the enhancement of semaglutide-GLP-1R binding affinity (K_d) from 3.4×10^{-6} M at 37 °C to 1.1×10^{-6} M at 37 °C, where a newly established structure-stabilizing interfacial salt bridge was formed between Glu104 (A_Glu_127 in PDB entry 4ZGM) of GLP-1R ECD and Arg27 (B_Val_33 in PDB entry 4ZGM) of semaglutide, which is entirely due to the manually designed V-R exchange [43]. Here, an amino acid residue numbering issue arises due to different methods of experimental structural determinantion (X-ray diffraction [29, 52]) or computational structural modeling (SWISS-MODEL [58, 59] or Modeller [44]). To address this residue numbering issue, a set of three numbering schemes are summarized as below:

1. the amino acid residue numbering is from 7 to 37, the same as Figure 1 for the experimental structure determined by X-ray diffraction of the semaglutide backbone in complex with GLP-1R ECD (PDB ID: 4ZGM [29]);
2. the amino acid sequence of semaglutide's backbone is numbered from 1 to 31 for the SWISS-MODEL [58, 59] homology structural model of the semaglutide backbone in complex with GLP-1R ECD;
3. the amino acid sequence of semaglutide's backbone is numbered from 98 to 128 for the Modeller [44] homology structural model of the semaglutide backbone in complex with GLP-1R ECD.

Furthermore, in another study [47, 60], one semaglutide analogue demonstrated a binding affinity to GLP-1R that is more than two orders of magnitude (113.3 times) higher than native semaglutide, achieving a K_d of 3.0×10^{-8} M compared to the K_d of 3.4×10^{-6} M for native semaglutide. Nonetheless, with respect to the two examples, only one homology structural model was build for the semaglutide mutant, and no structural dynamics was taken into account, which is the focus of next section.

Design of semaglutide analogues with improved GLP-1R ECD affinity: a dynamic structural electrostatic perspective

First, PDB entry 4ZGM [29, 52] was used as the template for structural modeling by Modeller [44] and ligand-receptor binding affinity calculations by Prodigy [45–47], where a total of 10000 ligand-receptor K_d values were generated with a statistical analysis in Figure 2 as the baseline for the design of semaglutide analogues with improved GLP-1R ECD affinity.

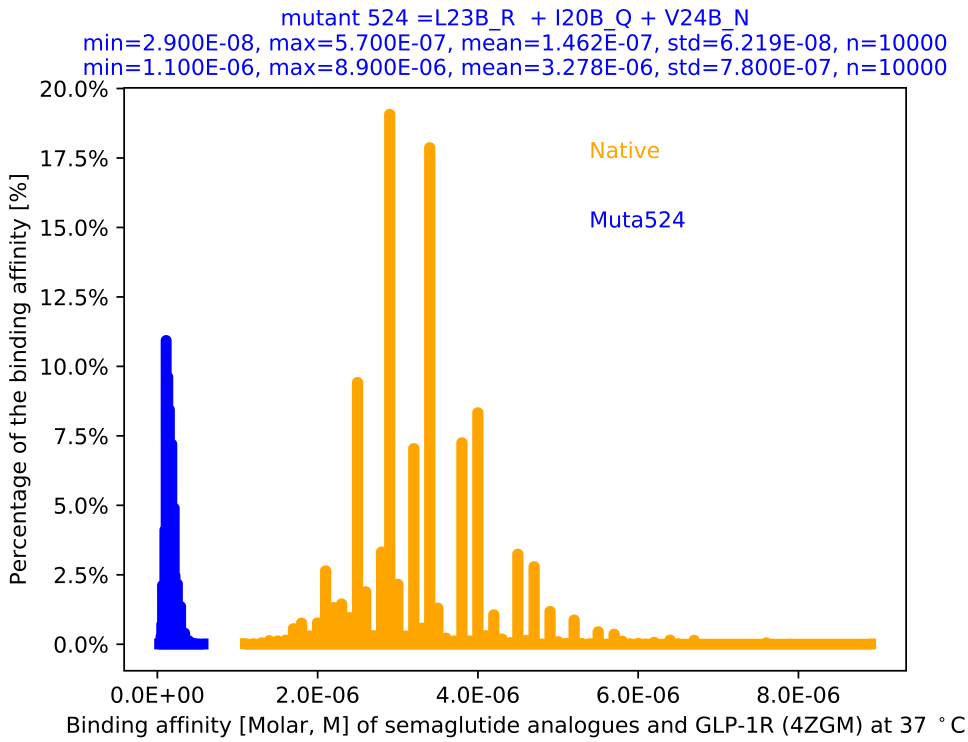

Fig 2. Distribution of the GLP-1R binding affinities of native semaglutide (PDB ID: 4ZGM [29]) and semaglutide mutant 524 as modeled by Modeller [44] and calculated by Prodigy [45, 46] for 10000 times. In the title (top) of this figure, L23B_R, I20B_Q and V24B_R represent L123B_R, I120B_Q and V124B_R in supplementary file **4452.pdb**, respectively, and the second and the third lines of the blue texts represents the statistical summaries of the GLP-1R ECD's binding affinity K_d values of the backbones of mutant (blue) and native (orange) semaglutide, respectively.

With PDB entry 4ZGM [29, 52] as the structural template, a set of key amino acid residues (supplementary file **4zgm.ic**) were identified by Prodigy [45, 46] and a comprehensive structural biophysical analysis [50]. Afterwards, a total of 564 amino acid sequences (supplementary file **rewriteEverything.out**) of semaglutide backbone were designed according to Figure 1. Subsequently, for each semaglutide backbone

sequence, Modeller [44] was used to generate at least 6000 homology structural models of GLP-1R-semaglutide analogue with PDB entry 4ZGM [29, 52] as template. Afterwards, for each structural model, Prodigy [45, 46] was used for the structural biophysics-based calculations of K_d between semaglutide analogues and GLP-1R ECD. Finally, for all 564 amino acid sequences of the backbone of semaglutide analogues, supplementary file **rewriteEverything.out** provides a summary of:

1. an identification number of the backbone sequence, i.e., from 1 to 564, e.g., semaglutide mutant 524 (Figure 2);
2. three site-specific missense mutations introduced into the backbone sequence;
3. original sequence of the backbone of semaglutide (Figure 1);
4. mutant sequence of the backbone of semaglutide;
5. alignment of the original and mutant sequences;
6. a statistical summary of the ligand-receptor binding affinity K_d values between native semaglutide and GLP-1R ECD;
7. a statistical summary of the ligand-receptor binding affinity K_d values between mutant semaglutide and GLP-1R ECD.

Of note, a statistical summary (supplementary file **rewriteEverything.out**) includes the average, the standard deviation, the maximum and the minimum of the ligand-receptor binding affinity K_d values, where average K_d was used to rank the structural biophysical analysis results of the 564 amino acid sequences of the backbone of semaglutide analogues, with native semaglutide as the baseline for the design of semaglutide analogues with improved GLP-1R ECD affinity. In supplementary file **rewriteEverything.out**, semaglutide mutant 524 was ranked the first with the highest average GLP-1R ECD affinity with three missense mutations, i.e., I20B_Q, L23B_R and V24B_N. To further investigate why these three missense mutations led to a significantly increased GLP-1R ECD affinity of semaglutide mutant 524, a comprehensive structural biophysical analysis [50] was conducted for the 10000 structural models built by Modeller [44] with a structural electrostatic perspective.

Mutant ID	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 bridging network analysis [50] of the 4452th structural model (supplementary file **4452.pdb**) of semaglutide mutant 524, the first in supplementary file **rewriteEverything.out**. In this table, the residue naming scheme is the same as Table 1, the row with yellow background represents the interfacial salt bridge also identified in PDB entry 4ZGM [29, 52].

As described above, the role of charge-charge interactions is as important as 4.76% (2/42, supplementary file **4zgm.ic**) in the stabilization of the semaglutide backbone in complex with the extracellular domain (ECD) of GLP-1R (PDB ID: 4ZGM [29]), including only a few interfacial salt bridges (Table 1) and hydrogen bonds (Table 2). Therefore, here, a structural electrostatic perspective was used for the analysis of the 10000 structural models of semaglutide mutant 524. Specifically, four interfacial salt bridges were identified in the 4452th structural model (supplementary file **4452.pdb**) of

semaglutide mutant 524, as listed in Table 4. Among the four, three new interfacial salt bridges (Figures 3, 4 and 5) were found to be formed due to the missense mutations introduced into semaglutide's backbone, especially the missense mutation L23B_R, whereas only one salt bridge (Table 4) was found to be formed and the same as the salt bridge (Table 1) identified in PDB entry 4ZGM [29, 52].

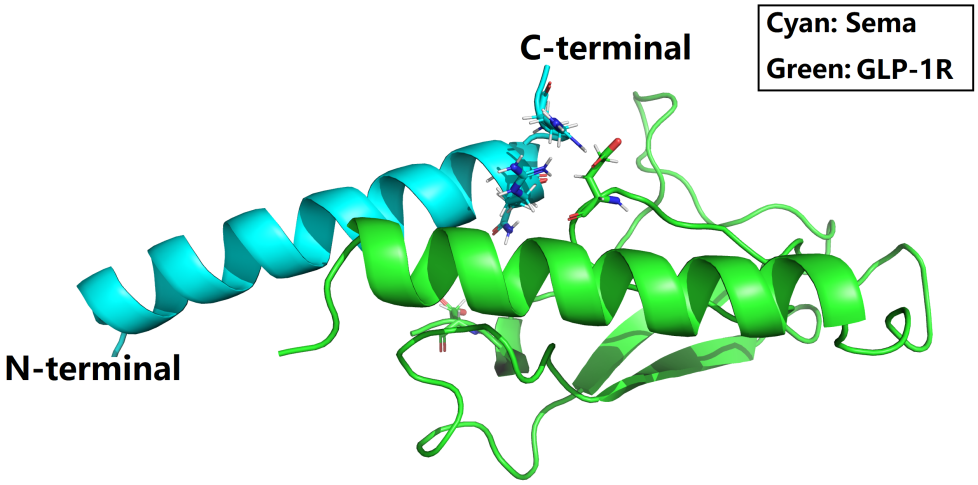

Fig 3. An overview of the complex structure of GLP-1R and semaglutide mutant 524 (Figure 2). This figure is prepared with PyMol [61] with supplementary file **4452.pdb**, and the details of the interfacial salt bridges are included in Table 4.

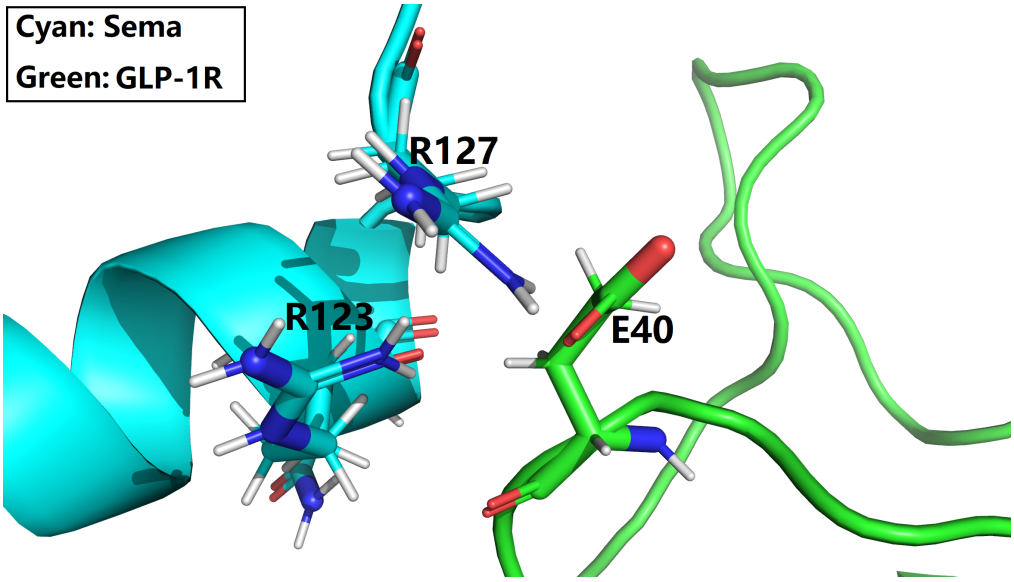

Fig 4. An zoomed-in view of the binding interface of GLP-1R and semaglutide mutant 524 (Figure 2). This figure is prepared with PyMol [61] with supplementary file **4452.pdb**, and the details of the interfacial salt bridges are included in Table 4.

From Figures 3, 4, it is quite clear that there is a cluster of three charged residues, i.e., Arg123 (L23B_R in Figure 2, B_Leu_32 in 4ZGM [29, 52]) of semaglutide's backbone, Arg127 (Arg36 in Figure 1, B_ARG_36 in 4ZGM [29, 52]) of semaglutide's backbone and Glu40 of GLP-1R ECD (A_GLU_68 in 4ZGM [29, 52]), sitting at the binding interface of the semaglutide-GLP-1R complex structure (Figure 3 and 4). Here, both Arg123 and Arg127 of semaglutide's backbone (cyan cartoon, Figures 3, 4 and 5) formed interfacial salt bridges with Glu40 of GLP-1R ECD (green cartoon, Figures 3, 4 and 5), the details of which are included in Table 4 and shown in Figure 5. Specifically, it is obvious that the newly formed Arg123-Glu40 salt bridge is due to the missense mutation L23B_R (Figure 2), i.e., the substitutin of the leucine with a positively charged arginine at position 23 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 23 of semaglutide's backbone (Figure 2), because in PDB entry 4ZGM [29, 52], the distances Table 3 between the oppositely charged side chains of A_GLU_68 and B_ARG_36 are beyond the cut-off distance (4.0 Å) for salt bridge analysis as used in [50]. While in the case of 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 (Arg36 in Figure 1, B_ARG_36 in 4ZGM [29, 52]) and Glu40 of GLP-1R ECD (A_GLU_68 in 4ZGM [29, 52]), 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) is almost the cut-off distance (4.0 Å) for salt bridge analysis as used in [50].

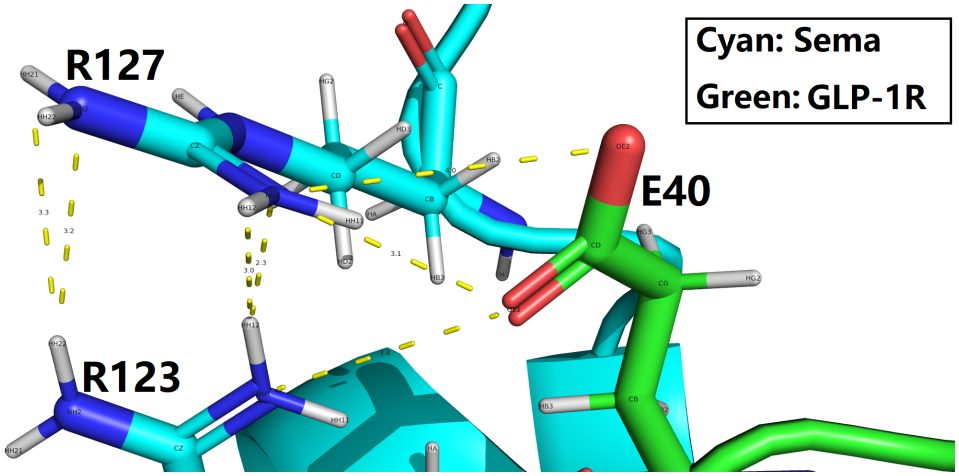

Fig 5. Salt bridges at the binding interface of GLP-1R and semaglutide mutant 524 (Figure 2). This figure is prepared with PyMol [61] with supplementary file **4452.pdb**, the details of the interfacial salt bridges are included in Table 4, and the details of other H-bond-like electrostatic interactions are included in Table 5.

Overall, from an molecular energetic point of view, the formation of these three new interfacial salt bridges (Figures 3, 4 and 5) act like three electrostatic clips to further stabilize the ligand-receptor complex structure, and contribute favorably to the binding affinity of semaglutide mutant 524 to GLP-1R ECD, with the potential of the semaglutide analogue approaching, reaching and binding faster and stronger to GLP-1R ECD, and thereby leading to improved GLP-1R activation with higher efficacy and efficiency [29]. As shown in Figure 5, nevertheless, the two positively charged side chains of Arg123 and Arg127 remain as close to each other as only a little over 3.0 Å, leading to an electrostatic repulsion between the two residues, which is energetically unfavorable. Yet, according to a set of intra-chain hydrogen bonding analysis of

supplementary file **4452.pdb**, it was found that two sets of hydrogen bond-like electrostatic interactions (Table 5) exist between the two positively charged side chains of Arg123 and Arg127, with donor-acceptor (D-A) distances at 3.23 and 3.19 Å and two $\angle ADH$ at 16.94 and 28.90˚, while a cutoff value of donor-acceptor (D-A) distance was defined as 3.0 Å [50], and a cutoff value of the angle formed by acceptor (A), donor (D) and hydrogen (H) ($\angle ADH$) was defined as 30˚. Thus, from a structural biophysical point of view, this electrostatic repulsion could be explained by the two hydrogen bond-like electrostatic interactions (Table 5) and also the presence of a negatively charged Glu40 of GLP-1R ECD, which acts like an electrostatic connector for the two positively charged Arg123 and Arg127 of semaglutide's backbone.

PDB	Acceptor (A)	Donor (D)	Hydrogen (H)	D-A (Å)	H-A (Å)	$\angle ADH$(˚)
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 bonding screening of the 4452th structural model (supplementary file **4452.pdb**) of semaglutide mutant 524. In this table, the residue naming scheme is the same as Table 1, $\angle ADH$ represents the angle formed by acceptor (A), donor (D) and hydrogen (H).

An electrostatic scaffold of salt bridges at the binding interface between GLP-1R and semaglutide analogues

In 2021 as mentioned above, a manually designed V-R exchange [43] was introduced into the backbone of semaglutide for an improve GLP-1R ECD affinity. In this article here, an automated high-throughput approach was employed for a set of systemic amino acid scanning of the backbone of semaglutide to enhance GLP-1R ECD affinity and thereafter GLP-1R activation, i.e., a set of in silico steps of structural and biophysical data generation Table 6 [47].

Size (s) of the synthetic structural and biophysical data set					
Semaglutide backbone (28 Aa)			**Molecule X (100 Aa)**		
g(28,1)	$\frac{28!}{1!(27)!} \times 20^1$	560	g(100,1)	$\frac{100!}{1!(99)!} \times 20^1$	2000
g(28,2)	$\frac{28!}{2!(26)!} \times 20^2$	151200	g(100,2)	$\frac{100!}{2!(98)!} \times 20^2$	1980000
g(28,3)	$\frac{28!}{3!(25)!} \times 20^3$	26208000	g(100,3)	$\frac{100!}{3!(97)!} \times 20^3$	1293600000
g(28,4)	$\frac{28!}{4!(24)!} \times 20^4$	3276000000	g(100,4)	$\frac{100!}{4!(96)!} \times 20^4$	627396000000
g(28,5)	$\frac{28!}{5!(23)!} \times 20^5$	314496000000	g(100,5)	$\frac{100!}{5!(95)!} \times 20^5$	240920064000000

Table 6. The Size ($s = g(k,n) = \frac{k!}{n!(k-n)!} \times 20^n$) of the synthetic structural data set based on semaglutide-GLP-1R complex structure. where k represents the length of semaglutide backbone, $\underline{n}$ represents the number of missense mutations introduced into semaglutide backbone, where the value of n/k is key to ensure the overall reasonable accuracy of the synthetic structural data [47].

As a matter of fact, $\underline{n}$ in Table 6 went all the way from 1 (g(28,1) in Table 6) to 7 g(28,7) in Table 6, in comparison with the length (28) of the backbone of semaglutide, while the structural modeling (Modeller [44]) and and K_d calculations (Prodigy [45, 46]) were performed completely only for g(28,1) (as in Table 6), as from Table 6, it is quite clear that the size of the computational task is too huge even for a high performance computing facility. To this end, while a total of 564 amino acid sequences (supplementary file **rewriteEverything.out**) of semaglutide backbone were designed with the sequence template in Figure 1, the 564 site-specific missense mutations were screened from a list of g(28,7) (Table 6) semaglutide analogue canditates whose GLP-1R ECD affinity first reached as strong as 1.0×10^{-9} M, i.e., K_d as low as 1.0×10^{-9} M or lower (raw data not published).

With 564 amino acid sequences (supplementary file **rewriteEverything.out**) of semaglutide backbone in place, a new set of in silico approaches were employed for high-throughput structural modeling (Modeller [44]) and and K_d calculations (Prodigy [45, 46]), which led to the identfication 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 L23B_R, I20B_Q, V24B_R and R25B_A. As predicted by Prodigy [45, 46], the K_d of semaglutide analogue A747 to GLP-1R ECD is 1.7×10^{-07}, in comparison with the average K_d of native semaglutide to GLP-1R ECD at 3.278×10^{-06} (Figure 2). Moreover, the role of charge-charge interactions is as important as 18.18% (8/44, supplementary file **A747.ic**) for semaglutide analogue A747 to GLP-1R ECD, compared with 4.76% (2/42, supplementary file **4zgm.ic**) for native semaglutide to GLP-1R ECD, in the stabilization of the semaglutide backbone in complex with the extracellular domain (ECD) of GLP-1R (PDB ID: 4ZGM [29]).

Mutant	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 7. 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, the two rows with yellow background represents the interfacial salt bridge also identified in PDB entry 4ZGM [29, 52].

As described in Table 7, a total of 9 interfacial salt bridges (Table 7) 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 identified at the binding interface of GLP-1R ECD and native semaglutide (PDB entry 4ZGM [29, 52]). Specifically, three are a total of 7 newly formed interfacial salt bridges (Table 7) which are due to the four site-specific missense mutations L23B_R, I20B_Q, V24B_R and R25B_A. Of further interest, the 7 newly formed interfacial salt bridges constitutes a structural electrostatic scaffold (Figures 6, 7, 8, 9, 10 and 11) at the binding interface of GLP-1R ECD and semaglutide analogue A747, which consists of three sets of interfacial salt bridges, as shown by three red numbers (1, 2 and 3) within black circles in Figures 9, 10 and 11.

Mutant	Acceptor (A)	Donor (D)	Hydrogen (H)	D-A (Å)	H-A (Å)	$\angle ADH(°)$
A747	O, B_ARG_124	NH2, A_ARG_93	HH22, A_ARG_93	3.09	2.34	34.92
A747	OE2, A_GLU_100	NZ, B_LYS_117	HZ3, B_LYS_117	3.15	2.29	26.28
A747	OE1, A_GLU_99	NH1, B_ARG_124	HH12, B_ARG_124	2.75	2.08	39.63

Table 8. Interfacial side chain and main chain hydrogen bonding analysis of semaglutide analogue A747 in complex with GLP-1R ECD (supplementary file **A747.pdb**. In this table, the residue naming scheme is **Chain ID_residue name_residue number**, $\angle ADH$ represents the angle formed by acceptor (A), donor (D) and hydrogen (H).

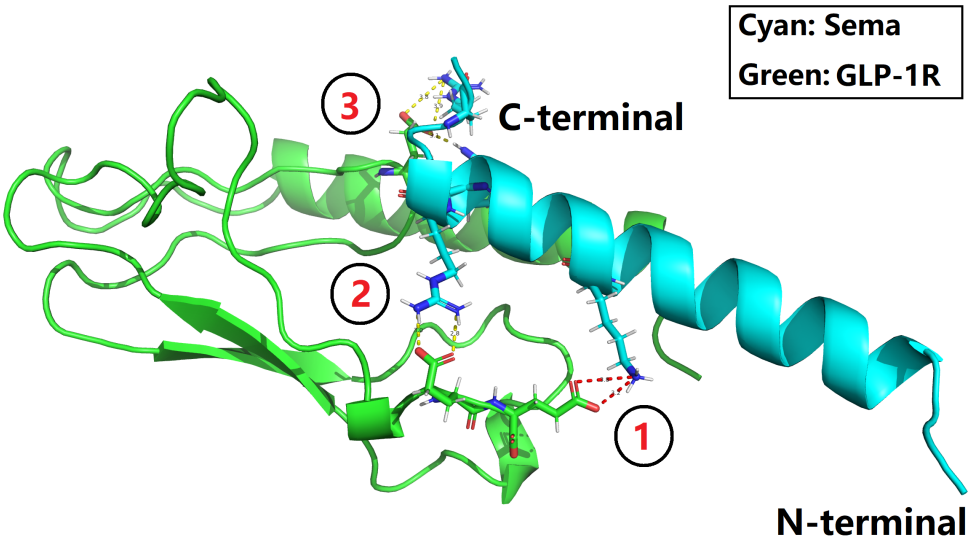

Fig 6. An overview of the complex structure of GLP-1R and semaglutide mutant A747. In this figure, the three red numbers (1, 2 and 3) within black circles represents the three sets of interfacial salt bridges (red and yellow dotted lines) of GLP-1R and semaglutide mutant A747, which constitute an electrostatic scaffold for the design of next-generation GLP-1R agonists with improved GLP-1R ECD affinity. This figure is prepared with PyMol [61] with supplementary file **A747.pdb**.

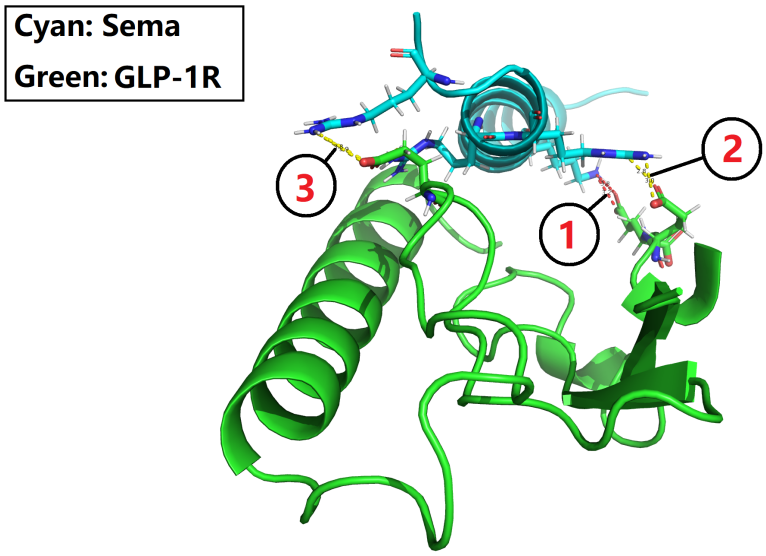

Fig 7. An sideview of the complex structure of GLP-1R and semaglutide mutant A747. In this figure, the three red numbers (1, 2 and 3) within black circles represents the three sets of interfacial salt bridges (red and yellow dotted lines) of GLP-1R and semaglutide mutant A747, which constitute an electrostatic scaffold for the design of next-generation GLP-1R agonists with improved GLP-1R ECD affinity. This figure is prepared with PyMol [61] with supplementary file **A747.pdb**.

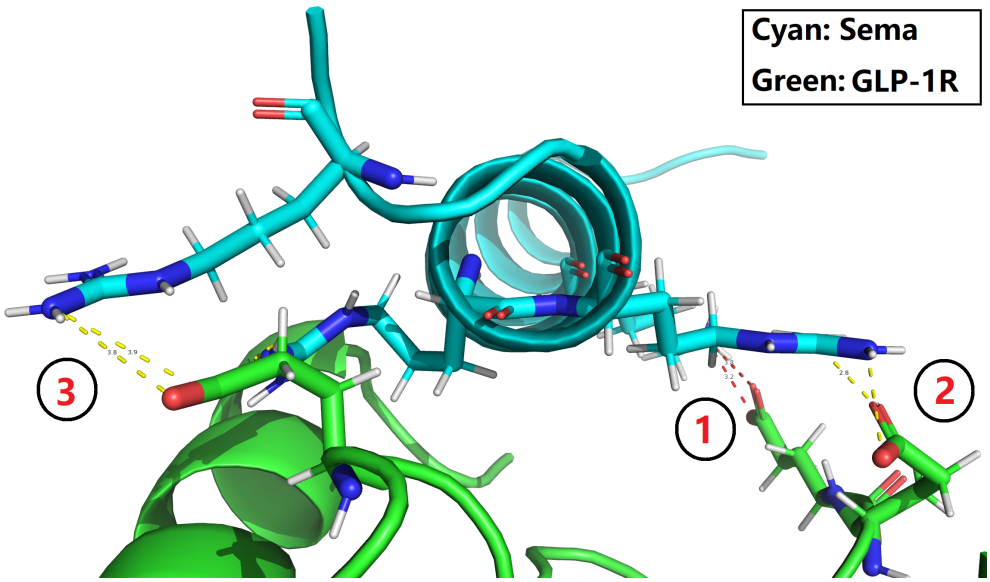

Fig 8. A zoomed-in sideview of the complex structure of GLP-1R and semaglutide mutant A747. In this figure, the three red numbers (1, 2 and 3) within black circles represents the three sets of interfacial salt bridges (red and yellow dotted lines) of GLP-1R and semaglutide mutant A747, which constitute an electrostatic scaffold for the design of next-generation GLP-1R agonists with improved GLP-1R ECD affinity. This figure is prepared with PyMol [61] with supplementary file **A747.pdb**, the details of the salt bridges are included in Table 7.

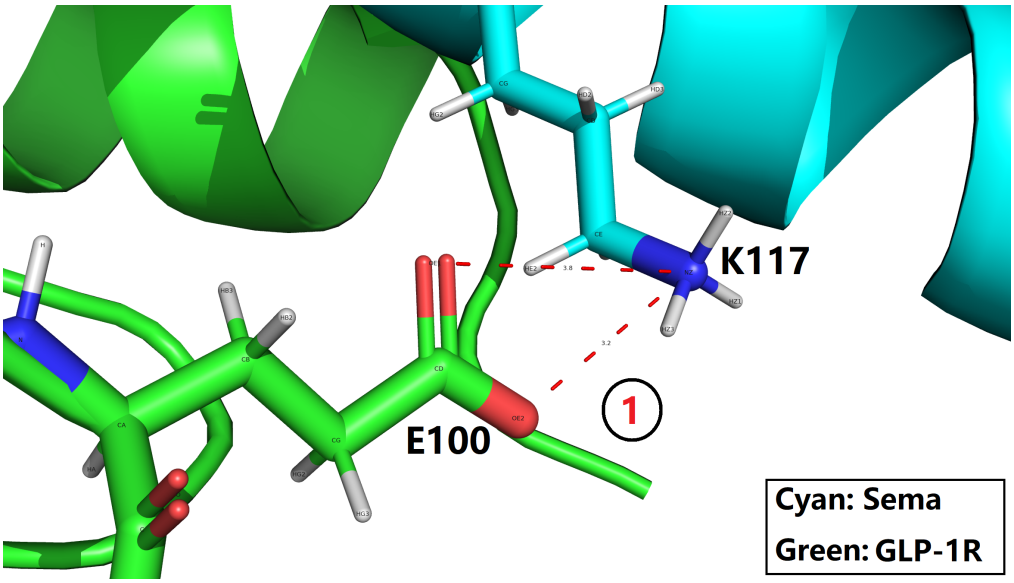

Fig 9. The first set of interfacial salt bridges (red dotted lines) of GLP-1R and semaglutide mutant A747, the details of which are included in Table 7. This figure is prepared with PyMol [61] with supplementary file **A747.pdb**.

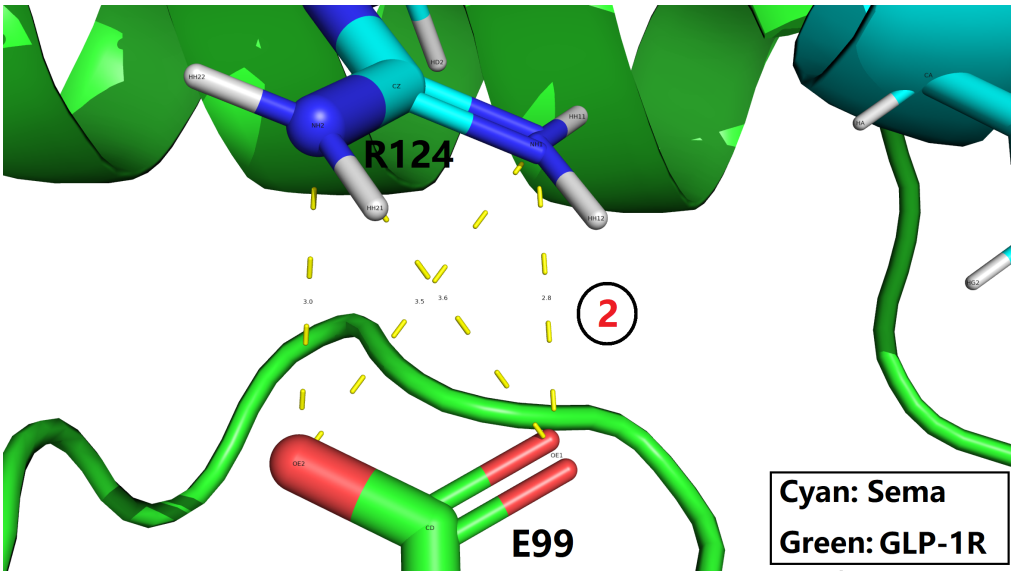

Fig 10. The second set of interfacial salt bridges (yellow dotted lines) of GLP-1R and semaglutide mutant A747, the details of which are included in Table 7. This figure is prepared with PyMol [61] with supplementary file **A747.pdb**.

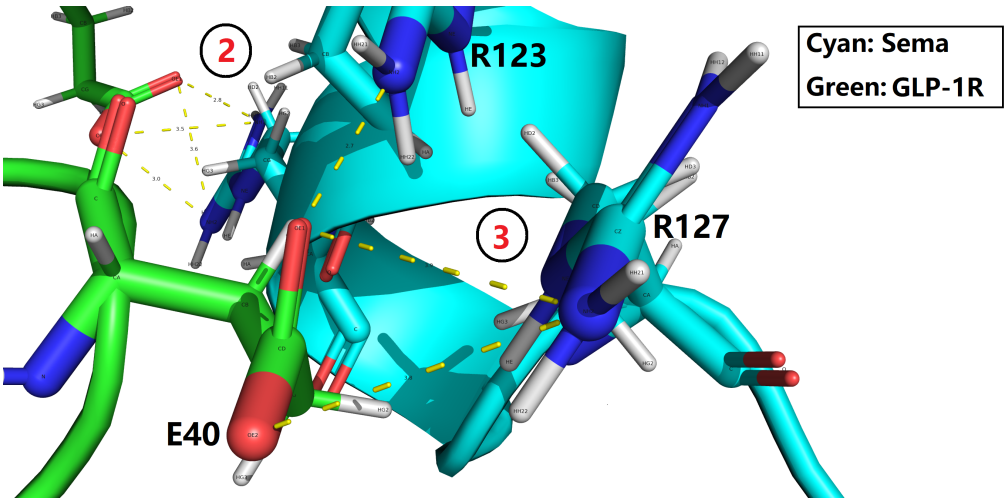

Fig 11. The third set of interfacial salt bridges (yellow dotted lines) of GLP-1R and semaglutide mutant A747, the details of which are included in Table 7. This figure is prepared with PyMol [61] with supplementary file **A747.pdb**.

Conclusion and Discussion

This study presents a systematic and computationally-driven framework for the design of semaglutide analogues, emphasizing structural and electrostatic optimization to enhance GLP-1R activation. In addition to 564 semaglutide analogues, this article identified an electrostatic scaffold of four intermolecular salt bridges (Figures 6, 7, 8, 9, 10 and 11) at the semaglutide-GLP-1R binding interface. These salt bridges stabilize the ligand-receptor complex and improve binding affinity and activation, providing a robust template for the development of next-generation GLP-1R agonists.

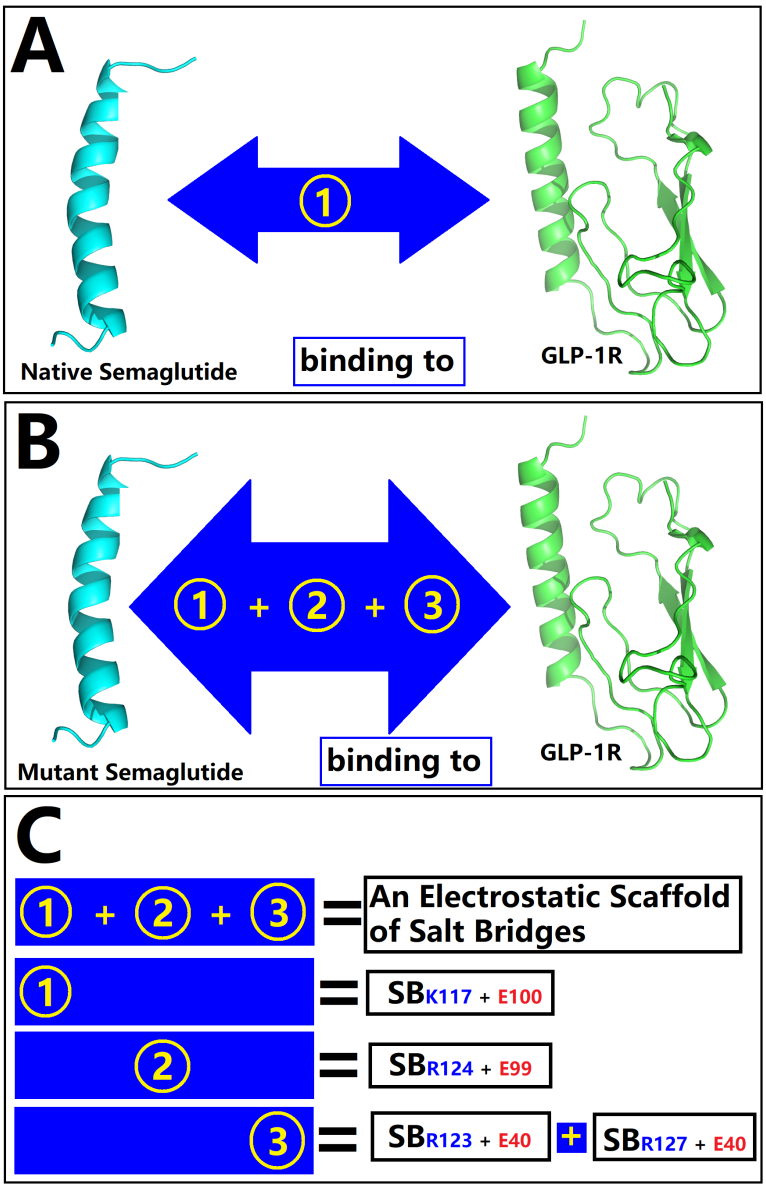

Fig 12. An electrostatic scaffold of three sets of salt bridges (Figures 9, 10 and 11) at the binding interface of GLP-1R ECD and semaglutide analogues with faster (K_{on}) and stronger (K_d) receptor binding and slower receptor dissociation K_{off} [29].

This study not only advances GLP-1R agonist design by introducing a robust electrostatic scaffold but also establishes a generalizable framework for peptide

engineering. By balancing enhanced receptor affinity with safety considerations (e.g., immunogenicity, DPP4 digestion), the work highlights opportunities to refine therapeutic outcomes for metabolic disorders. Future directions include the experimental validation of the 564 analogues, exploration of multifunctional multi-target peptide-based therapeutics with improved efficacy and clinical utility.

Supporting information

S1 File. **4zgm.ic** A summary of the residue-residue pairs at the interface of semaglutide and GLP-1R ECD [29, 45, 46, 52].

S2 File. **rewriteEverything.out** A summary of the 564 semaglutide analogues with improved GLP-1R ECD affinity to enhance GLP-1R activation.

S3 File. **4452.pdb** A homology structural built by Modeller [44] for the complex structure of GLP-1R ECD and semaglutide analogues with three site-specific missense mutations L23B_R, I20B_Q and V24B_R.

S4 File. **A747.pdb** A homology structural built by Modeller [44] for the complex structure of GLP-1R ECD and semaglutide analogues with four site-specific missense mutations L23B_R, I20B_Q, V24B_R and R25B_A.

S5 File. **A747.ic** A summary of the residue-residue pairs at the interface of semaglutide mutant A747 and GLP-1R ECD [29, 45, 46, 52].

Acknowledgments

The author is grateful to the communities of structural biology, biophysics, medicinal and computational chemistry and algorithm design, for the continued accumulation of knowledge and data for drug discovery & design, and for the continued development of tools (hardware, software and algorithm) for drug discovery & design.

References

1. Sims EK, Carr ALJ, Oram RA, DiMeglio LA, Evans-Molina C. 100 years of insulin: celebrating the past, present and future of diabetes therapy. Nature Medicine. 2021;27(7):1154–1164. doi:10.1038/s41591-021-01418-2.

2. Vecchio I, Tornali C, Bragazzi NL, Martini M. The Discovery of Insulin: An Important Milestone in the History of Medicine. Frontiers in Endocrinology. 2018;9. doi:10.3389/fendo.2018.00613.

3. Hirsch IB. Insulin Analogues. New England Journal of Medicine. 2005;352(2):174–183.

4. Slieker LJ, Brooke GS, Chance RE, Fan L, Hoffmann JA, Howey DC, et al. Insulin and IGF-I Analogs: Novel Approaches to Improved Insulin Pharmacokinetics. 1994; p. 25–32.

5. Bilotta F. Short acting insulin analogues in intensive care unit patients. World Journal of Diabetes. 2014;5(3):230.

6. Miikkulainen K, Caruso A, Mast O, Zhang R, Borisenko O. Systematic literature review of use of blood glucose monitoring in phase III clinical studies of insulin analogs. BMC Endocrine Disorders. 2016;16(1).

7. Scapin G, Dandey VP, Zhang Z, Strickland C, Potter CS, Carragher B. Insulin Receptor ectodomain in complex with one insulin molecule; 2018.

8. Li W. Delving Deep into the Structural Aspects of the BPro28-BLys29 Exchange in Insulin Lispro: A Structural Biophysical Lesson. 2020;.

9. Kruger DF, Novak LM. Role of ultrafast-acting insulin analogues in the management of diabetes. Journal of the American Association of Nurse Practitioners. 2019;31(9):537–548.

10. Adams GG, Meal A, Morgan PS, Alzahrani QE, Zobel H, Lithgo R, et al. Characterisation of insulin analogues therapeutically available to patients. PLOS ONE. 2018;13(3):e0195010.

11. Sharma AK, Taneja G, Kumar A, Sahu M, Sharma G, Kumar A, et al. Insulin analogs: Glimpse on contemporary facts and future prospective. Life Sciences. 2019;219:90–99.

12. Caparrotta TM, Evans M. PEGylated insulin Lispro, (LY2605541)-a new basal insulin analogue. Diabetes, Obesity and Metabolism. 2013;16(5):388–395.

13. Francis D, Chacko AM, Anoop A, Nadimuthu S, Venugopal V. In: Evolution of biosynthetic human insulin and its analogues for diabetes management. Elsevier; 2024. p. 191–256.

14. Rosenstock J, Juneja R, Beals JM, Moyers JS, Ilag L, McCrimmon RJ. The Basis for Weekly Insulin Therapy: Evolving Evidence With Insulin Icodec and Insulin Efsitora Alfa. Endocrine Reviews. 2024;45(3):379–413. doi:10.1210/endrev/bnad037.

15. Derewenda U, Derewenda Z, Dodson GG, Hubbard RE, Korber F. Molecular Structure of insulin: The insulin monomer and its assembly. British Medical Bulletin. 1989;45(1):4–18.

16. Slieker LJ, Brooke GS, DiMarchi RD, Flora DB, Green LK, Hoffmann JA, et al. 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. 1997;40(14):S54–S61.

17. Knudsen LB, Lau J. The Discovery and Development of Liraglutide and Semaglutide. Frontiers in Endocrinology. 2019;10:1–32.

18. Bucheit JD, Pamulapati LG, Carter N, Malloy K, Dixon DL, Sisson EM. Oral Semaglutide: A Review of the First Oral Glucagon-Like Peptide 1 Receptor Agonist. Diabetes Technology & Therapeutics. 2020;22(1):10–18.

19. Husain M, Birkenfeld AL, Donsmark M, Dungan K, Eliaschewitz FG, Franco DR, et al. Oral Semaglutide and Cardiovascular Outcomes in Patients with Type 2 Diabetes. New England Journal of Medicine. 2019;381(9):841–851.

20. Rodbard HW, Rosenstock J, Canani LH, Deerochanawong C, Gumprecht J, Lindberg SØ, et al. Oral Semaglutide Versus Empagliflozin in Patients With Type 2 Diabetes Uncontrolled on Metformin: The PIONEER 2 Trial. Diabetes Care. 2019;42(12):2272–2281.

21. Xiao Y, Sun L. Semaglutide in weight management. The Lancet. 2019;394(10205):1226.

22. Anderson SL, Beutel TR, Trujillo JM. Oral semaglutide in type 2 diabetes. Journal of Diabetes and its Complications. 2020;34(4):107520.

23. Røder ME. Clinical potential of treatment with semaglutide in type 2 diabetes patients. Drugs in Context. 2019;8:1–11.

24. Aroda VR, Rosenstock J, Terauchi Y, Altuntas Y, Lalic NM, Villegas ECM, et al. PIONEER 1: Randomized Clinical Trial of the Efficacy and Safety of Oral Semaglutide Monotherapy in Comparison With Placebo in Patients With Type 2 Diabetes. Diabetes Care. 2019;42(9):1724–1732.

25. Pratley R, Amod A, Hoff ST, Kadowaki T, Lingvay I, Nauck M, et al. Oral semaglutide versus subcutaneous liraglutide and placebo in type 2 diabetes (PIONEER 4): a randomised, double-blind, phase 3a trial. The Lancet. 2019;394(10192):39–50.

26. Longwell CK, Hanna S, Hartrampf N, Sperberg RAP, Huang PS, Pentelute BL, et al. Identification of N-Terminally Diversified GLP-1R Agonists Using Saturation Mutagenesis and Chemical Design. ACS Chemical Biology. 2020;16(1):58–66. doi:10.1021/acschembio.0c00722.

27. Zhang H, Wu T, Wu Y, Peng Y, Wei X, Lu T, et al. Binding sites and design strategies for small molecule GLP-1R agonists. European Journal of Medicinal Chemistry. 2024;275:116632. doi:10.1016/j.ejmech.2024.116632.

28. Zhang X, Belousoff MJ, Liang YL, Danev R, Sexton PM, Wootten D. Structure and dynamics of semaglutide- and taspoglutide-bound GLP-1R-Gs complexes. Cell Reports. 2021;36(2):109374. doi:10.1016/j.celrep.2021.109374.

29. Lau J, Bloch P, Schäffer L, Pettersson I, Spetzler J, Kofoed J, et al. Discovery of the Once-Weekly Glucagon-Like Peptide-1 (GLP-1) Analogue Semaglutide. Journal of Medicinal Chemistry. 2015;58(18):7370–7380.

30. Dang T, Yu J, Cao Z, Zhang B, Li S, Xin Y, et al. Endogenous cell membrane interactome mapping for the GLP-1 receptor in different cell types. Nature Chemical Biology. 2024;doi:10.1038/s41589-024-01714-1.

31. Wright SC, Motso A, Koutsilieri S, Beusch CM, Sabatier P, Berghella A, et al. GLP-1R signaling neighborhoods associate with the susceptibility to adverse drug reactions of incretin mimetics. Nature Communications. 2023;14(1). doi:10.1038/s41467-023-41893-4.

32. Arrick G, Sticker D, Ghazal A, Lu Y, Duncombe T, Gwynne D, et al. Cephalopod-inspired jetting devices for gastrointestinal drug delivery. Nature. 2024;doi:10.1038/s41586-024-08202-5.

33. Abramson A, Caffarel-Salvador E, Khang M, Dellal D, Silverstein D, Gao Y, et al. An ingestible self-orienting system for oral delivery of macromolecules. Science. 2019;363(6427):611–615. doi:10.1126/science.aau2277.

34. Conlon JM. Evolution of the insulin molecule: insights into structure-activity and phylogenetic relationships. Peptides. 2001;22(7):1183–1193. doi:10.1016/s0196-9781(01)00423-5.

35. Ludvik B, Giorgino F, Jódar E, Frias JP, Landó LF, Brown K, et al. Once-weekly tirzepatide versus once-daily insulin degludec as add-on to metformin with or without SGLT2 inhibitors in patients with type 2 diabetes (SURPASS-3): a randomised, open-label, parallel-group, phase 3 trial. The Lancet. 2021;398(10300):583–598.

36. Urva S, Coskun T, Loghin C, Cui X, Beebe E, O'Farrell L, et al. The Novel Dual GIP and GLP -1 Receptor Agonist, Tirzepatide, Transiently Delays Gastric Emptying Similarly to Selective Long-Acting GLP -1 Receptor Agonists. Diabetes, Obesity and Metabolism. 2020;.

37. Mariam Z, Niazi SK. Glucagon-like peptide agonists: A prospective review. Endocrinol Diabetes Metab. 2024;7(1):e462.

38. Kalra S, Bhattacharya S, Kapoor N. Contemporary Classification of Glucagon-Like Peptide 1 Receptor Agonists (GLP1RAs). Diabetes Therapy. 2021;12(8):2133–2147.

39. Seetharaman R. IcoSema's leap forward: new data from COMBINE 3 paves the way. J Basic Clin Physiol Pharmacol. 2024;.

40. Westergaard L, Alifrangis L, Buckley ST, Coester HV, Klitgaard T, Kristensen NR, et al. Pharmacokinetic properties of a once-weekly fixed-ratio combination of insulin icodec and semaglutide compared with separate administration of each component in individuals with type 2 diabetes mellitus. Clin Drug Investig. 2024;44(11):849–861.

41. Lingvay I, Asong M, Desouza C, Gourdy P, Kar S, Vianna A, et al. Once-weekly insulin icodec vs once-daily insulin degludec in adults with insulin-naive type 2 diabetes: The ONWARDS 3 randomized clinical trial. JAMA. 2023;330(3):228–237.

42. Philis-Tsimikas A, Asong M, Franek E, Jia T, Rosenstock J, Stachlewska K, et al. Switching to once-weekly insulin icodec versus once-daily insulin degludec in individuals with basal insulin-treated type 2 diabetes (ONWARDS 2): a phase 3a, randomised, open label, multicentre, treat-to-target trial. Lancet Diabetes Endocrinol. 2023;11(6):414–425.

43. Li W. Strengthening Semaglutide-GLP-1R Binding Affinity via a Val27-Arg28 Exchange in the Peptide Backbone of Semaglutide: A Computational Structural Approach. Journal of Computational Biophysics and Chemistry. 2021;20(05):495–499.

44. Webb B, Sali A. Protein Structure Modeling with MODELLER. In: Methods in Molecular Biology. Springer US; 2020. p. 239–255.

45. Vangone A, Bonvin AM. Contacts-based prediction of binding affinity in protein-protein complexes. eLife. 2015;4.

46. Xue LC, Rodrigues JP, Kastritis PL, Bonvin AM, Vangone A. PRODIGY: a web server for predicting the binding affinity of protein-protein complexes. Bioinformatics. 2016; p. btw514.

47. Li W. In Silico Generation of Structural and Intermolecular Binding Affinity Data with Reasonable Accuracy: Expanding Horizons in Drug Discovery and Design. 2024;doi:10.20944/preprints202405.1739.v1.

48. Gilbert MP, Pratley RE. GLP-1 Analogs and DPP-4 Inhibitors in Type 2 Diabetes Therapy: Review of Head-to-Head Clinical Trials. Frontiers in Endocrinology. 2020;11. doi:10.3389/fendo.2020.00178.

49. Deacon CF. Dipeptidyl peptidase 4 inhibitors in the treatment of type 2 diabetes mellitus. Nature Reviews Endocrinology. 2020;16(11):642–653. doi:10.1038/s41574-020-0399-8.

50. Li W. How do SMA-linked mutations of *SMN1* lead to structural/functional deficiency of the SMA protein? PLOS ONE. 2017;12(6):e0178519.

51. Berman H, Henrick K, Nakamura H. Announcing the worldwide Protein Data Bank. Nature Structural & Molecular Biology. 2003;10(12):980–980.

52. Reedtz-Runge S. Crystal structure of Semaglutide peptide backbone in complex with the GLP-1 receptor extracellular domain; 2015.

53. Wu D, Sun J, Xu T, Wang S, Li G, Li Y, et al. Stacking and energetic contribution of aromatic islands at the binding interface of antibody proteins. Immunome Research. 2010;6(Suppl 1):S1.

54. Davis FP, Sali A. PIBASE: a comprehensive database of structurally defined protein interfaces. Bioinformatics. 2005;21(9):1901–1907.

55. Jubb HC, Pandurangan AP, Turner MA, Ochoa-Montaño B, Blundell TL, Ascher DB. Mutations at protein-protein interfaces: Small changes over big surfaces have large impacts on human health. Progress in Biophysics and Molecular Biology. 2017;128:3–13.

56. Wells JA, McClendon CL. Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. Nature. 2007;450(7172):1001–1009.

57. Antolíková E, Žáková L, Turkenburg JP, Watson CJ, Hančlová I, Šanda M, et al. Non-equivalent Role of Inter- and Intramolecular Hydrogen Bonds in the Insulin Dimer Interface. Journal of Biological Chemistry. 2011;286(42):36968–36977.

58. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Research. 2018;46(W1):W296–W303.

59. Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Research. 2014;42(W1):W252–W258.

60. Li W. Optimizing GLP-1R Agonist: A Computational Semaglutide Analogue with 112-fold Enhanced Binding Affinity to GLP-1R. 2024;doi:10.20944/preprints202406.1032.v1.

61. DeLano WL. Pymol: An open-source molecular graphics tool. CCP4 Newsletter On Protein Crystallography. 2002;40:82–92.