Journal of Medicinal Chemistry

pubs.acs.org/jmc

Article

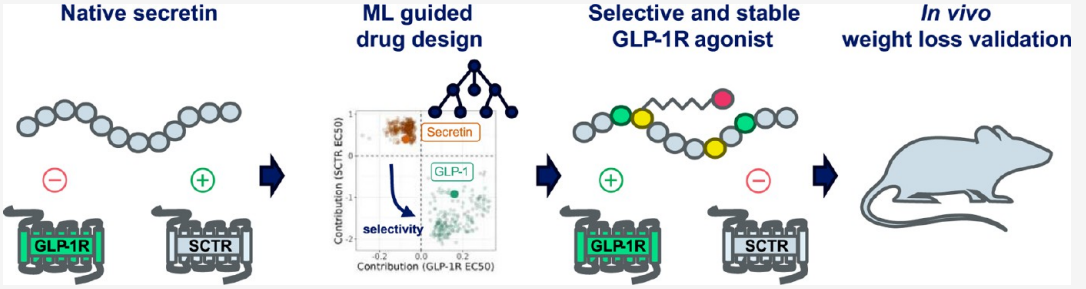
Machine-Learning-Guided Peptide Drug Discovery: Development of GLP-1 Receptor Agonists with Improved Drug Properties

Jens Christian Nielsen, Claudia Hjørringgaard, Mads Mørup Nygaard, Anita Wester, Lisbeth Elster, Trine Porsgaard, Randi Bonke Mikkelsen, Silas Rasmussen, Andreas Nygaard Madsen, Morten Schlein, Niels Vrang, Kristoffer Rigbolt, and Louise S. Dalbøge*

Cite This: *J. Med. Chem.* 2024, 67, 11814–11826

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information



ABSTRACT: Peptide-based drug discovery has surged with the development of peptide hormone-derived analogs for the treatment of diabetes and obesity. Machine learning (ML)-enabled quantitative structure−activity relationship (QSAR) approaches have shown great promise in small molecule drug discovery but have been less successful in peptide drug discovery due to limited data availability. We have developed a peptide drug discovery platform called streaMLine, enabling rigorous design, synthesis, screening, and ML-driven analysis of large peptide libraries. Using streaMLine, this study systematically explored secretin as a peptide backbone to generate potent, selective, and long-acting GLP-1R agonists with improved physicochemical properties. We synthesized and screened a total of 2688 peptides and applied ML-guided QSAR to identify multiple options for designing stable and potent GLP-1R agonists. One candidate, GUB021794, was profiled in vivo (S.C., 10 nmol/kg QD) and showed potent body weight loss in diet-induced obese mice and a half-life compatible with once-weekly dosing.

INTRODUCTION

Peptide-based therapeutics are gaining increasing attention in the pharmaceutical industry. Peptide hormones have both high receptor potency and selectivity, minimizing off-target effects and generally translating into an excellent drug safety and efficacy profile.[1,2] These features make endogenous peptides a good starting point for the development of novel peptide therapeutics. However, native unmodified peptides are rarely used as drugs because of their inherent limitations due to their very short systemic half-life and unfavorable physicochemical properties which must be circumvented to develop peptide molecules suitable for therapeutic use.[1,2]

Early drug discovery phases aim to improve the properties of candidate molecules by modifying their chemical structure. Such improvements can be achieved by rational design using an iterative and often laborious approach, where small batches of compounds are screened in multiple rounds of optimization. In contrast, when larger data sets are available, it is useful to construct mathematical models that capture the quantitative structure−activity relationship (QSAR) to guide drug design. QSAR models have been widely used in the development of small molecule therapeutics e.g., to discover novel binders[3] and elute chemical motifs necessary for binding.[4] For peptide drug discovery, however, data are often sparse, and this has limited the use of machine learning (ML) for QSAR optimization. The amount and composition of data are crucial parameters for QSAR methods,[5] which is why it is advantageous to generate data that are specifically designed for modeling purposes.

Glucagon-like peptide-1 (GLP-1) is an endogenous 30-amino acid peptide hormone produced by enteroendocrine L-cells and secreted into the hepatic portal in response to food intake. By activating GLP-1 receptors in the pancreas, native GLP-1 serves as an incretin hormone stimulating insulin release and inhibiting glucagon secretion.[6] In addition, GLP-1 is an important appetite regulator by activating central GLP-1 receptors (GLP-1R).[7] In line with this, long-acting GLP-1R

Received: February 19, 2024
Revised: June 19, 2024
Accepted: June 20, 2024
Published: July 8, 2024

Table 1. Alignment of GLP-1, Secretin, Dual-Agonist, and GLP-1R Selective Agonist GUB021794[a]

Compound	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Secretin	H	S	D	G	T	F	T	S	E	L	S	R	L	R	E	G	A	R	L	Q	R	L	L	Q	G	L	V			
GLP-1	H	A	E	G	T	F	T	S	D	V	S	S	Y	L	E	G	Q	A	A	K	E	F	I	A	W	L	V	K	G	R
Dual agonist	H	Aib	E	G	T	F	T	S	D	L	S	R	L	R	E	G	A	A	L	Q	R	F	L	Q	G	L	V			
GUB021794	H	Aib	E	G	T	F	T	S	D	V	S	Y	L	K*	E	E	A	Aib	L	Q	R	F	L	E	H	L	V			

Secretin residue GLP-1 residue Secretin and GLP-1 residue Other

[a]Origin of substitutions are highlighted in colors. * denotes attachment of half-life extender: C20DA-gGlu-2xOEG.

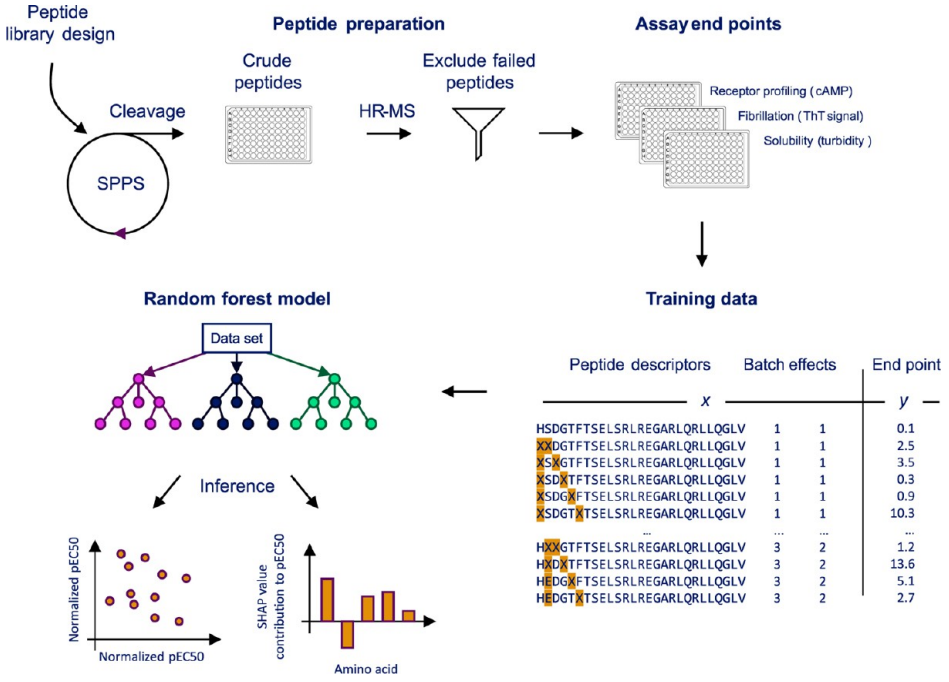


Figure 1. Overview of the data generation and data analysis workflow of the streaMLine platform. Initially a systematic library of peptides is designed, typically on the order of hundreds to thousands of peptides. The crude library of peptides are prepared using solid-phase-peptide synthesis (SPPS) and cleaved from the resin. Failed peptide samples are identified by high-resolution mass spectrometry and excluded from the analysis. A panel of high-throughput assays for determining receptor potency and physicochemical properties at different pH levels are measured. For each assay end point, a random forest model is trained and used for inferring key amino acids substitutions that determine peptide properties.

agonists such as semaglutide have been shown to be important tools in the treatment of diabetes and obesity.[8,9] However, GLP-1 is known to self-assemble into amyloid fibrils and the intrinsic physical instability of GLP-1 poses a significant challenge in synthesis and formulation.[10,11]

Recent drug development efforts have taken advantage of the varying degrees of sequence homology of GLP-1, glucagon, and glucose-dependent insulinotropic polypeptide (GIP) to engineer unimolecular dual or triple agonists targeting receptors of GLP-1, glucagon, and GIP. This approach has proven to be a successful concept for treatment intervention in diabetes and obesity.[12−17]

Alternatively, one could further envision to exploit sequence homology to obtain beneficial peptide properties not only in terms of receptor pharmacology, but also from a synthesis and/or formulation perspective.[18]

Secretin is a 27-amino acid peptide hormone that together with GLP-1 belongs to the glucagon superfamily of structurally related peptide hormones all targeting family B G-protein coupled receptors (GPCRs).[19] This family of peptides is linear peptides comprising 25 or more residues allowing them to span two key receptor domains. The C-terminal region of the peptides binds to the extracellular domain of their respective receptor whereafter the N-terminal region interacts with the core domain of the receptor enabling receptor activation.[20] The sequence identity of secretin and GLP-1 is shown in Table 1.

Contrary to other peptides of the glucagon family, such as GLP-1, secretin is not reported to aggregate.[21] Thus, secretin could serve as the starting backbone with improved physicochemical properties compared to GLP-1.

The main physiological role of secretin is to regulate water homeostasis and bicarbonate secretion from the exocrine pancreas and inhibit gastric acid secretion by activating the secretin receptor (SCTR),[19] i.e., endogenous activities we intended not to activate.

Hence, we aimed to leverage the more favorable physicochemical properties of secretin to develop a selective and physicochemical stable GLP-1R agonist based on the secretin backbone. With this aim, we exploited an innovative ML-based peptide drug discovery platform termed streaMLine and demonstrated how streaMLine effectively facilitates

https://doi.org/10.1021/acs.jmedchem.4c00417
J. Med. Chem. 2024, 67, 11814−11826

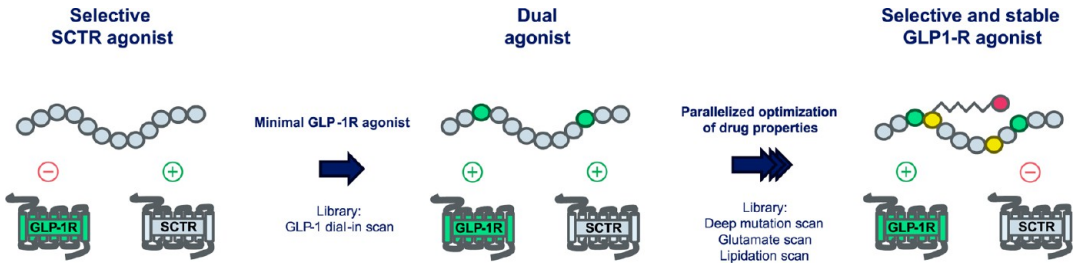


Figure 2. Schematic representation of the parallelized development process for generating a selective and stable GLP-1R agonist based on the secretin backbone. The two-step process for converting secretin into a preclinical drug candidate. First, a minimal GLP-1R agonist was developed by introducing only necessary GLP-1 residues to provide activation of GLP-1R. Second, a parallelized workflow was initiated where a deep mutational scan, a glutamate scan, and a lipidation scan provided a blueprint for generating various soluble, physically stable, and half-life extended GLP-1R agonist.

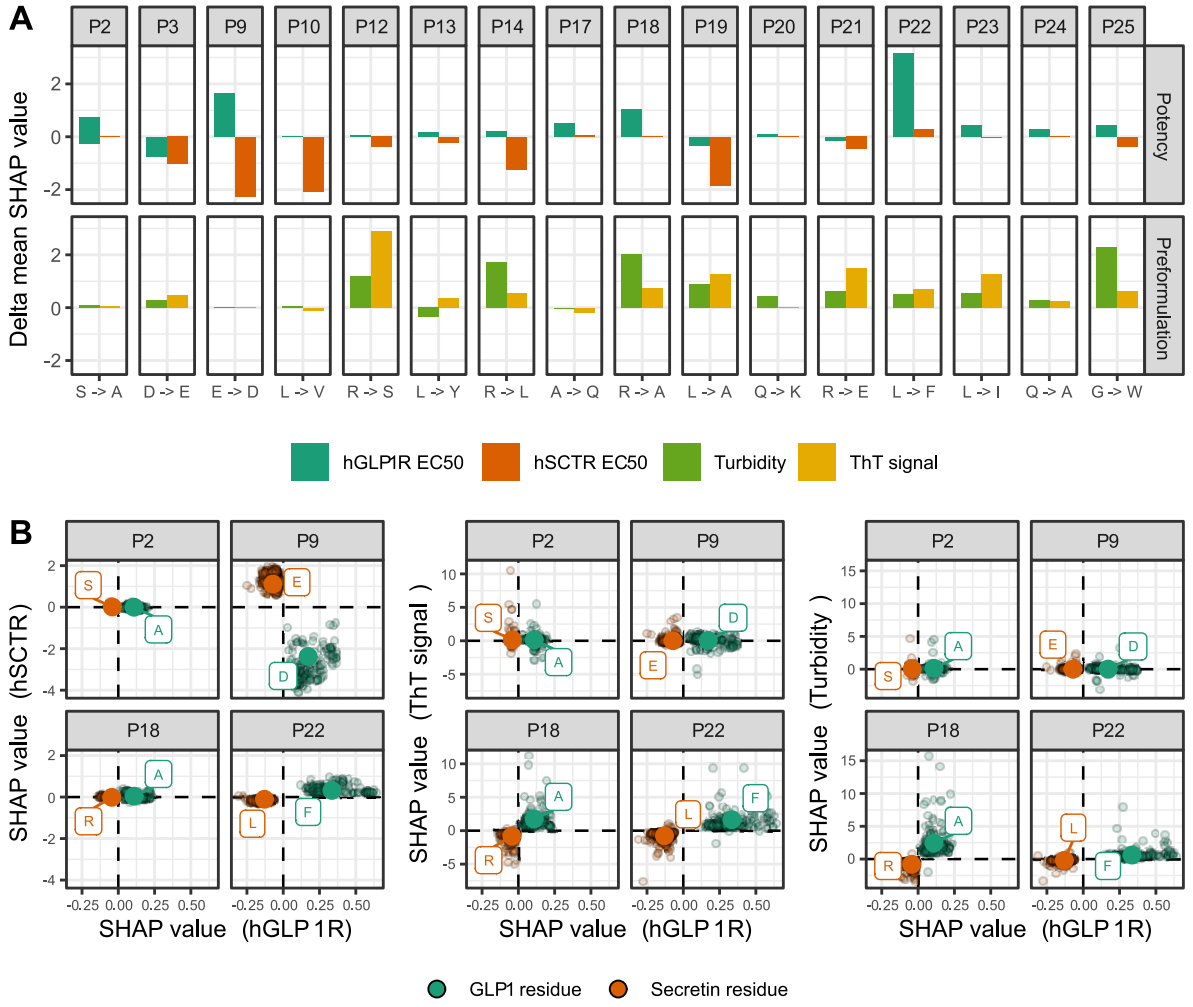


Figure 3. Overview of substitution-effects from a GLP-1 dial-in scan in the secretin backbone. (A) Effect of introducing GLP-1 residues into the secretin backbone. For each assay end point, a random forest model was trained on 768 peptides and used to compute SHAP values determining the level of contribution of each amino acid substitution. Delta mean SHAP values denote the contribution of substituting the secretin residue with the corresponding GLP-1 residue. (B) Detailed overview of SHAP values for selected positions, where substitutions were introduced to obtain analogs with dual GLP-1R and SCTR potency. Small points denote SHAP values per individual peptide and large points denote mean SHAP value.

accelerated development of novel peptide-based therapeutics by rigorous design and ML-driven analysis of large peptide libraries.

RESULTS

The streaMLine Platform. The streamline platform is a drug development tool where peptide libraries are designed, synthesized, and screened to provide large data sets suitable for machine learning (ML)-enabled quantitative structure−activity

https://doi.org/10.1021/acs.jmedchem.4c00417
J. Med. Chem. 2024, 67, 11814−11826

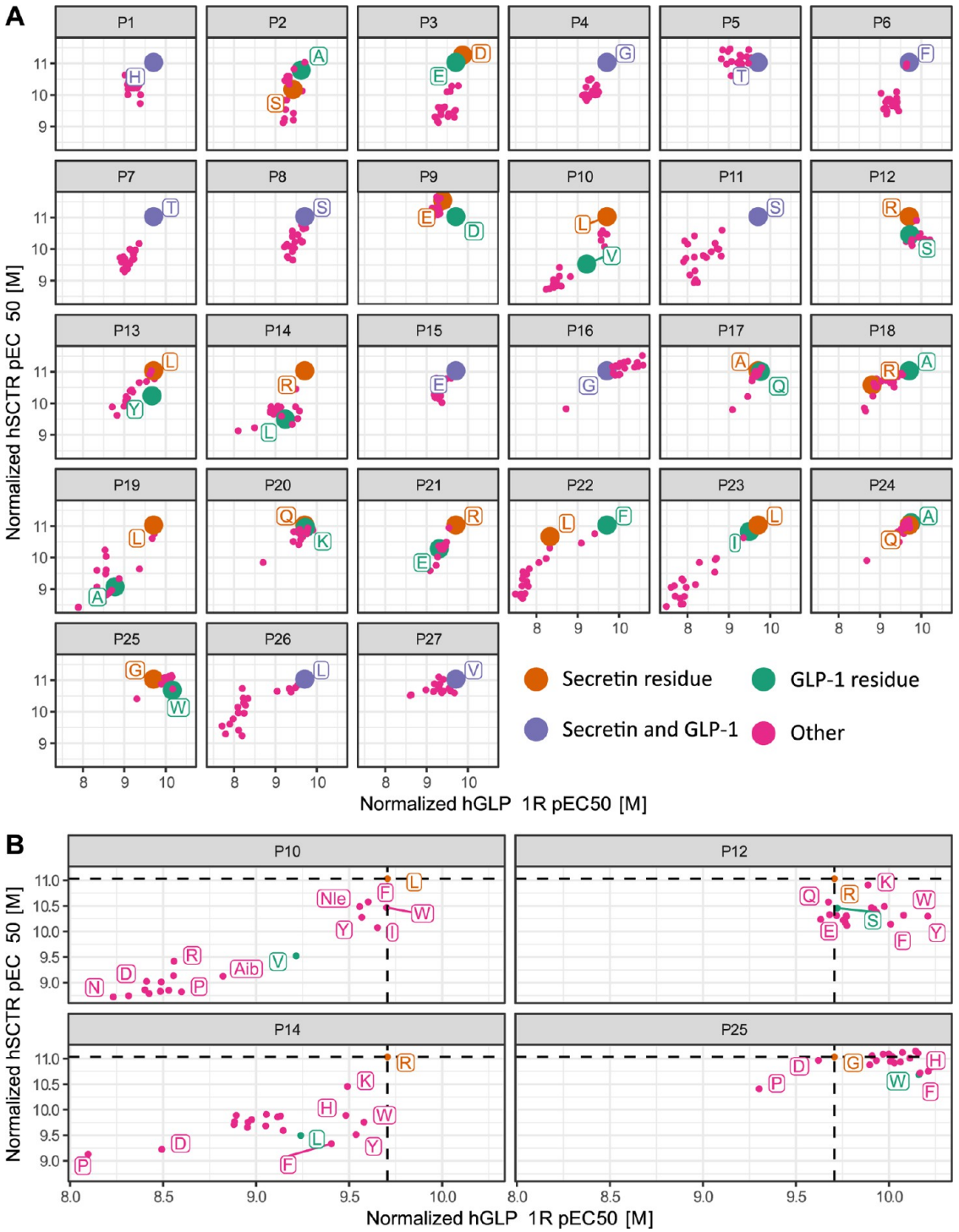


Figure 4. Overview of substitution-effects from a deep mutational scan (DMS) in a secretin derived GLP-1R and SCTR dual agonist. (A) Effect of single mutations in all positions. For GLP-1R and SCTR potency, random forest models were trained on 1152 peptides encoded using z-scales.[23] From the models, the effect of single mutations was computed to normalize for assay batch effects. (B) Detailed overview of selected positions highlighting the effect of individual substitutions.

relationship (QSAR) approaches. An overview of the learning cycles in the streamline platform is illustrated in Figure 1.

In the streaMLine platform, peptides are synthesized using solid-phase peptide synthesis (SPPS) in a plate format. The crude peptide libraries are screened directly in functional potency assays and in preformulation assays for determination of, for example, fibrillation and solubility. Each peptide library is analyzed by using high-resolution mass spectrometry (HRMS) to determine purity. The average purity per library ranges between 30 and 50%, and peptide samples with less than 10% purity are excluded from further analysis.

The peptide libraries are designed in a highly systematic manner, where each substitution is observed multiple times in combination with other substitutions. Typically, peptide libraries consist of hundreds to thousands of peptides, which enables robust evaluation of each substitution in multiple

https://doi.org/10.1021/acs.jmedchem.4c00417
J. Med. Chem. 2024, 67, 11814−11826

chemical contexts, i.e., different backbones. The peptide sequences together with assay data are used as training data to construct random forest models[22] describing the relationship between peptide sequence and assay end point. In the training data, the systematic peptide library is encoded using amino acid descriptors (z-scales[23] or one-hot encoding), and potential laboratory batch effects are incorporated in the model to normalize, e.g., synthesis and assay plate differences. For each assay end point, a new model is trained and used for inferring the key amino acid substitutions affecting the end point. Model inference is done either by (1) correcting assay data for batch effects and amino acid similarity and thereby computing normalized assay measurements for individual peptides, (2) or by computing Shapley Additive explanation (SHAP) values.[24] SHAP values are used to explain the effect of each amino acid substitution on the end point and can thus be used to infer the key drivers in the data set. After model inference, the most promising substitutions are assessed in a new peptide library design.

Data points obtained on crude peptides are challenging to interpret individually, but in the context of the systematically designed peptide libraries, the random forest model provides accurate guidance for identifying the effect of substitutions. The performances of all models generated in this study are given in Figure S1.

Development of Dual GLP-1R-SCTR Agonists. We applied the streaMLine platform to generate selective GLP-1R agonists with suitable physicochemical parameters starting from the secretin backbone.

The native secretin peptide has no GLP-1R potency, hence to identify both desired substitutions a starting point with some GLP-1R potency was needed. We therefore aimed at generating first a dual GLP-1R-SCTR agonist that could be used as an intermediate for further optimization (Figure 2).

A peptide library was designed to evaluate the effect of introducing GLP-1 residues into the native secretin. Non-conserved residues (position 2−3, 9−10, 12−14, and 17−25), were changed into the corresponding GLP-1 residue, one at a time or in combinations (Table 1). The library (768 peptides) was screened by determining GLP-1R and SCTR potency (EC50), fibril formation (ThT assay), and solubility (turbidity), and random forest models were trained to determine the relationship between measured end points and the amino acid sequence of peptides. From these models, we computed SHAP values to determine the level of contribution of each substitution.[22] Substitutions with positive SHAP values increase the end point, while substitutions with negative SHAP values decrease the end point.

Amino acid positions 2, 9, 18, and 22 had the highest positive SHAP values for GLP-1R EC50, thus being critical for improving GLP-1R potency. Conversely, positions 3, 9, 10, 14, and 19 exhibited the most negative SHAP values for SCTR EC50, hence being critical for abolishing SCTR potency or enhancing GLP-1R selectivity (Figure 3A).

In addition to potency determination, the propensity for fibril formation and a reduction in solubility was most pronounced when mutating amino acid positions 12, 14, 18, 19, 21, 23, and 25. The introduction of GLP-1 residues at these positions could thus negatively affect the physicochemical properties of a peptide (Figure 3B).

Based on these learnings, five substitutions were introduced in the secretin backbone to achieve an agonist with dual activity on GLP-1R and SCTR (Table 1). Mutations 9D, 18A, and 22F increased GLP-1R potency. Likewise, 2A was found to increase the GLP-1R potency. The substitution of alanine for 2-aminoisobutyric acid (Aib) is well-known to prevent DPP-4 proteolytic cleavage of the GLP-1 backbone without compromising GLP-1R potency[9,25] hence 2Aib was introduced. The 3E mutation prevented the isomerization of the native aspartic acid residue in secretin without influencing receptor potency. Next, a comprehensive sequence exploration was performed on the dual GLP-1R-SCTR agonist, i.e., a deep mutational scan.

Development of Selective GLP-1R Agonists. A deep mutational scan (DMS) was designed such that all-natural amino acids (except cysteine and methionine) were introduced in all sequence positions, either as single mutations or as double mutations. The library consisted of 1152 peptides, which were screened for GLP-1R and SCTR potency. Based on these data, we trained random forest models on the relationship between all assay end points and the peptide amino acid sequence. The models were used to normalize for batch effects (synthesis and assay plate) and the resulting pEC50 values for each single mutant are shown in Figure 4.

The DMS identified several receptor-selectivity-promoting substitutions. For each position, substitution maps were obtained allowing us to navigate toward desired properties, including GLP-1R potency and/or improved receptor selectivity. Rather than identifying a single compound with desired properties through an iterative design process, the DMS generated a solution space of possible amino acid substitutions from which peptide candidates could be designed and synthesized. A selection of substitutions is described below.

We identified amino acid positions 9, 12, and 25 where substitutions could significantly improve GLP-1R selectivity by increasing GLP-1R potency and decreasing SCTR potency (Figures 4A, B and S2). At position 12, several substitutions improved GLP-1R selectivity. 12Y most effectively improved GLP-1R potency and reduced SCTR potency, whereas 12E only reduced SCTR potency. Aromatic residues 25H, 25F, 25Y, and 25W improved GLP-1R potency while also reducing SCTR potency, with 25H being the most effective. At position 9, only the native GLP-1 residue 9D improved potency and selectivity.

In addition, amino acid positions 10, 14, and 19 were identified to improve selectivity by decreasing SCTR potency with a neglectable effect on GLP-1R potency (Figures 4A, B and S2). 10I and 10V reduced SCTR potency without significantly compromising GLP-1R potency. A similar effect was seen by substituting position 14 to F, Y, or L.

Positions 16, 18, and 22 could be substituted to enhance GLP-1R selectivity. For position 16, all mutations, except P, increased GLP-1R potency (Figure S2). For positions 18 and 22, 18A, 18Aib, 18L, 18, 22F, 22W, and 22Y considerably improved GLP-1R potency while only marginally affecting SCTR potency.

Improving Solubility and Conjugation of Fatty Acid. In parallel with DMS, we systematically investigated the effect and tolerability of glutamate substitution and derivatization with half-life extenders (HLEs) in our dual GLP-1R-SCTR agonist. Glutamate substitutions can be used to modulate the isoelectric point of a peptide, thereby improving solubility at the desired formulation pH.[26] Fatty acid conjugation is a well-described technology broadly applied to extend the half-life of peptides from minutes to hours. Native secretin and GLP-1

https://doi.org/10.1021/acs.jmedchem.4c00417
J. Med. Chem. 2024, 67, 11814−11826

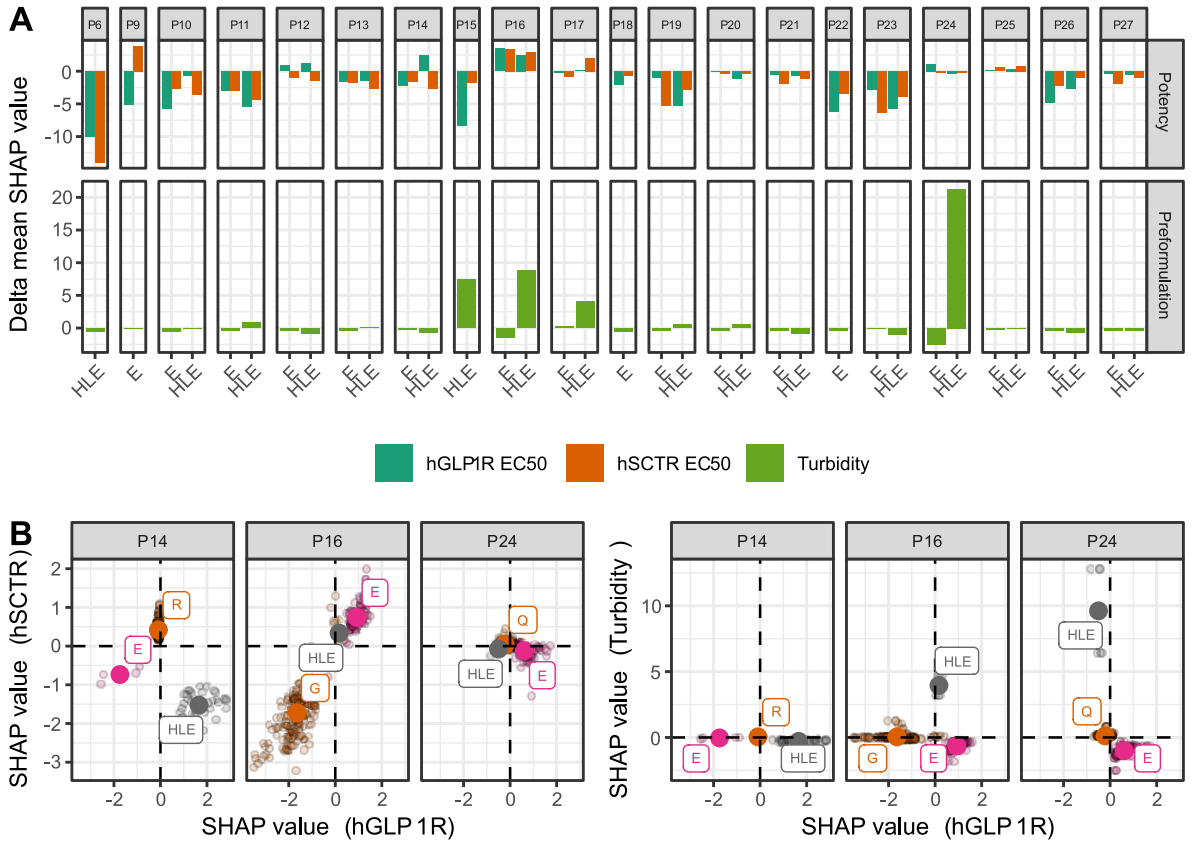


Figure 5. Overview of substitution-effects from glutamate scan and half-life extender (HLE) scan in a secretin-derived GLP-1R and SCTR dual agonist. (A) Effect of introducing HLEs or glutamate. For each assay end point, a random forest model was trained on 576 peptides and used to compute SHAP values determining the level of contribution of each amino acid substitution. Delta mean SHAP values denote the contribution of substituting the backbone residue with either a glutamate or HLE. (B) Detailed overview of SHAP values for selected positions that tolerate HLE derivatization or glutamate substitution for improving half-life and solubility, respectively. Small points denote SHAP values per individual peptide and large points denote mean SHAP value.

have reported half-lives in the range of 2−4 min.[27,28] Conjugation with fatty diacids facilitates strong binding to serum albumin, thus reducing renal clearance and enzymatic degradation.[28,29]

Fatty acid conjugation was investigated by attachment to the epsilon nitrogen of lysines via linker moieties. The combined fatty acid and linker is here referred to as the half-life extender (HLE). Six different HLEs were evaluated in each position representing different lengths of fatty diacids, octadecanedioic acid (C18DA) and eicosanedioic acid (C20DA), and varying combinations of linker moieties, L-γ-glutamyl (gGlu) and 3,8-dioxa-aminooctanoic acid (OEG).

A library of 576 peptides was designed where each HLE at each position was examined in backbones comprising 0, 1, or 2 glutamate mutations. All positions were screened, except a few positions in the pharmacophore essential for receptor activation. For the glutamate substitution screening, positions 1−5, 7, and 8 were excluded, with position 15 already being a glutamate. For the HLE screening, positions 1−5, 7−9, 18, and 22 were excluded. The library was screened for GLP-1R and SCTR potency and turbidity. For each end point, we computed SHAP values to determine the contribution of each mutation relative to the backbone residue (Figure 5).

Evaluating the effect of glutamate mutations on GLP-1R potency revealed several amino acid positions where glutamate was tolerated, i.e., positions 12, 16, 17, 19, 20, 21, 24, 25, and 27. Position 12, 16, and 24 were found to have a positive impact on GLP-1R potency compared to the backbone residue. Among these positions, we found the largest reduction in turbidity from introducing 16E and 24E, indicating improved solubility (Figure 5B).

For the HLEs, we found no major difference in potency across the different fatty acids and/or linker combinations and therefore analyzed the different HLEs as a single substitution. Evaluating the effect of attaching HLEs on the GLP-1R potency revealed several positions where HLEs were tolerated, i.e., positions 10, 12, 14, 16, 17, 20, 21, 24, 25, and 27. For positions 12, 14, and 16, the attachment of HLEs was found to have positive effects on GLP-1R potency compared to the backbone residue, with positions 12 and 14 also inducing GLP-1R selectivity. This selectivity effect of positions 12 and 14 was consistent with the observations from the DMS, where GLP-1R selectivity could also be improved by mutating these positions (Figure 4). Importantly, no effect on turbidity was observed when HLEs were conjugated to positions 12 (Figure 5A) and 14 (Figure 5A, B).

Fine-Tuning Potency, Selectivity, and Physicochemical Properties. Previous sections described our parallel peptide development process. The conjugation of HLEs could dramatically alter the properties of a peptide.[26] We therefore set out to investigate if an HLE was compatible with positions and substitutions found to be selectivity-inducing in the deep

https://doi.org/10.1021/acs.jmedchem.4c00417
J. Med. Chem. 2024, 67, 11814−11826

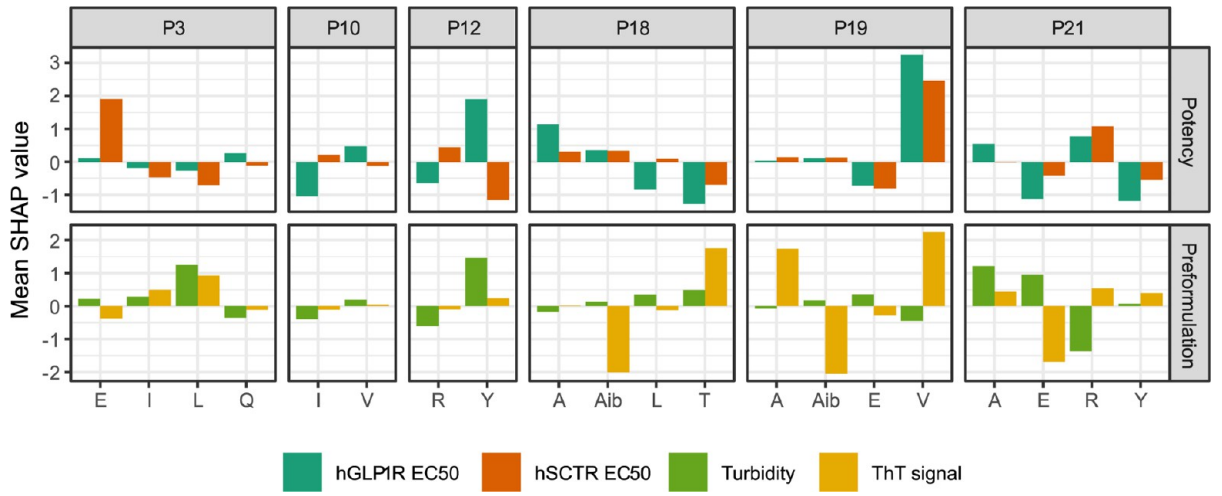


Figure 6. Overview of substitution-effects from selected substitutions in secretin derived selective GLP-1R agonist. Effect of combining selected substitutions identified in the deep mutational scan, HLE scan, and glutamate scan. For each assay end point a random forest model was trained on 192 peptides and used to determine SHAP values determining the level of contribution of each amino acid substitution. Mean SHAP values denote the contribution of each substitution relative to the data set mean.

Table 2. Profiling of Optimized Secretin-Derived GLP-1R Agonist

compound	hGLP1R EC50 (nM)	hSCTR EC50 (nM)	selectivity ratio[a]	solubility pH 7.0 and 8.0 (mg/mL)	fibrillation pH 7.0 and 8.0	chemical stability pH 7.0 and 8.0 (% degradation)	rat half-life, i.v. (h)
secretin	2300	0.0023	10^{-6}	NA	no	8.6/9	NA
GLP-1	0.002	800	400,000	NA	no	3.8/8.1	NA
GUB021794	0.018	190	10,556	>10	no	1.4/0.55	22

[a]Selectivity ratio was calculated as hSCTR EC50 divided by hGLP1R EC50.

mutational scan. Furthermore, this library would enable us to rank the mutations relative to each other and according to desired end points to identify an optimal combination of substitutions providing a selective GLP-1R agonist with desired physicochemical properties. We designed a library of 192 compounds where selected substitutions were evaluated at positions 3, 10, 12, 18, 19, and 21 in the dual agonist backbone (Table 1) being lipidated at position 14 with the HLE: C18DA-gGlu-2xOEG. The selected substitutions all induce selectivity in differing degrees, while some also have different physicochemical properties, such as opposite charges, hydrophobicity, helix-inducing effect, etc. In addition to receptor potency, we also assessed the solubility and the propensity for forming amyloid fibrils using a Thioflavin T (ThT) assay. We trained random forest models for each end point and calculated SHAP values to determine the contribution of each mutation to the relevant end points in a lipidated backbone (Figure 6).

In terms of selectivity, we found that 10V and 12Y provided the greatest positive effect by increasing the GLP-1R potency and decreasing the SCTR potency. For 12Y this also led to increased turbidity, meaning that this mutation should be accompanied by mutations that improve solubility.

Interestingly, three mutations were identified to diminish the propensity for fibril formation, 18Aib, 19Aib, and 21E, where the two former did not affect potency on the two receptors. We further used SHAP values to determine the interaction effect of these mutations that prevent fibril formation and found that the reduction in fibril formation was not additive when combining 18Aib and 19Aib (Figure S3). Thus, only one of the mutations is necessary to provide a reduction in fibrillation propensity.

Design and Characterization of Final Candidate. Based on all of the substitution options identified, we designed GUB021794. The peptide sequence is shown in Table 1. We aimed for a peptide candidate with minimal (<10 fold) GLP-1R potency loss compared to native GLP-1 but with a 10,000-fold selectivity ratio (GLP-1R over SCTR). The peptide candidate should be soluble, chemically- and physically stable at neutral pH and have a pharmacokinetic profile compatible with once-weekly dosing in humans. The *in vitro* characterization and pharmacokinetic properties of GUB021794 are summarized in Table 2. GUB021794 is potent and selective with a GLP-1R potency of 0.018 nM and an hSCTR potency of 190 nM.

For GUB021794, we modified the dual-acting backbone at positions 10, 12, 14, 16, 18, 24, and 25. The pronounced GLP-1R potency and high selectivity were achieved by combining substitutions identified in the DMS to have a positive effect on those end points. From the DMS (Figure 4), we identified positions where substitutions could induce selectivity by decreasing SCTR potency, while also improving GLP-1R potency or not significantly impacting GLP-1R potency. For this design, we decided to pursue 12Y and 25H as both were found to improve GLP-1R potency while also decreasing SCTR potency. 10V was introduced to further decrease SCTR potency without significantly reducing GLP-1R potency. In the glutamate scan (Figure 5), we identified two positions where glutamate had a positive impact on GLP-1R potency while also decreasing the turbidity of the peptides. Hence, glutamates were introduced in positions 16 and 24 to achieve a solubility of at least 10 mg/mL. The HLE scan (Figure 5) identified two positions, 12 and 14, where the attachment of a HLE-induced selectivity without compromising solubility. For the design of

https://doi.org/10.1021/acs.jmedchem.4c00417
J. Med. Chem. 2024, 67, 11814−11826

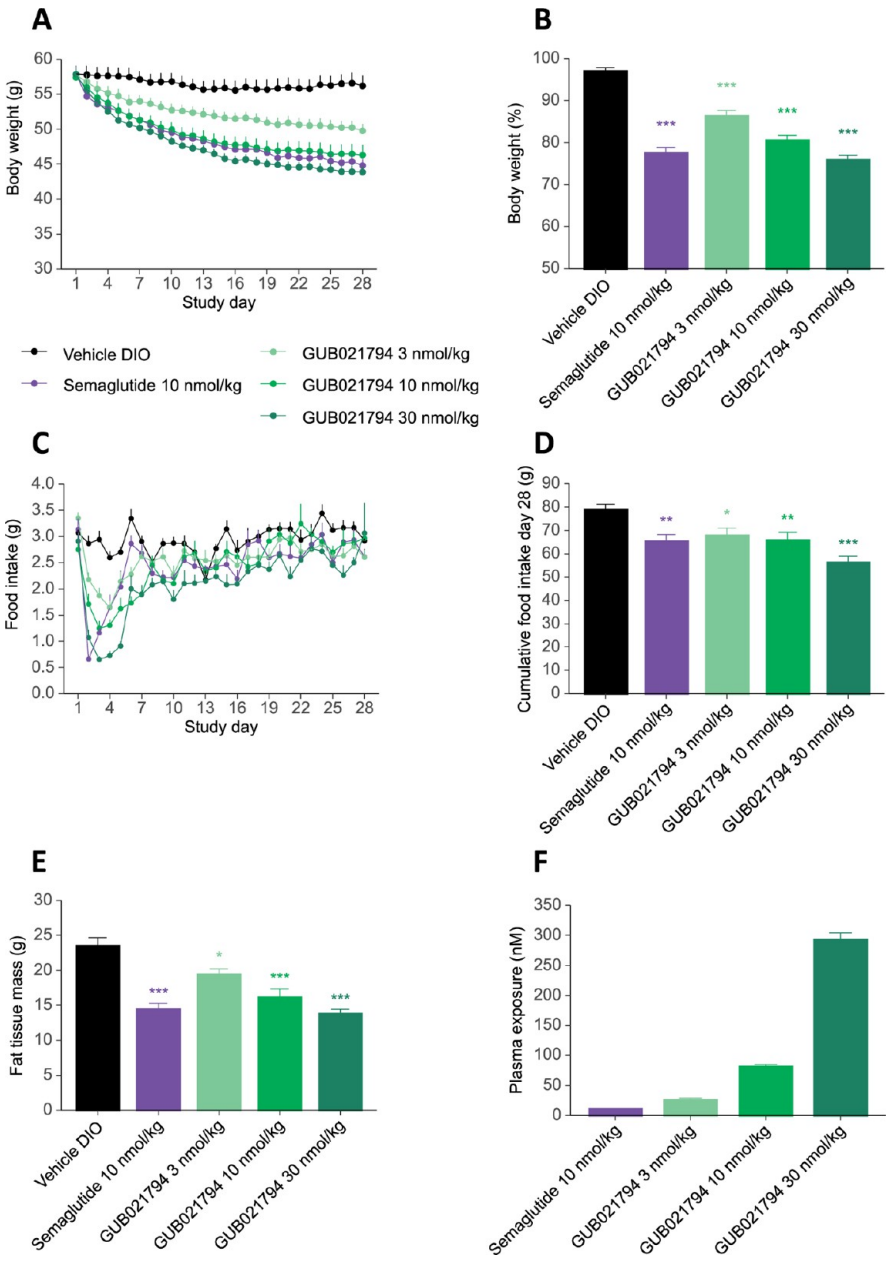


Figure 7. GUB021794 dose dependently reduces body weight and adipose tissue mass in DIO mice. DIO mice were treated subcutaneously once daily (QD) with vehicle, semaglutide (10 nmol/kg), and ascending doses of GUB021794 (3, 10, or 30 nmol/kg) for 28 days. (A) Body weight. (B) Body weight change (% of day 1) at the end of the study. (C) Daily food intake. (D) Cumulative food intake day 28. (E) Fat mass day 28. (F) Plasma exposure day 29, (samples taken 24 h after dosing). Values expressed as mean of $n = 10$ + SEM. Dunnett's test one-factor linear model. *: $P < 0.05$, **: $P < 0.01$, ***: $P < 0.001$ compared to Vehicle.

GUB021794, the HLE was attached to position 14 to improve GLP-1R potency and selectivity, while maintaining high solubility. Finally, we identified two positions where Aib was found to decrease the propensity for forming amyloid fibrils (Figure 6). In GUB021794, 18Aib was introduced to preserve the nonfibrillating properties at neutral pH.

While improving solubility and potency, the substitution of positions 24 and 25 additionally removes a sequence motif, QG, prone to cause chemical instability due to the potential deamidation of glutamine positioned next to glycine. Additionally, the isomerization of aspartic acid was addressed by substituting aspartic acid in position 3 for glutamic acid and by introducing the sterically hindered valine in position 10 next to 9D. Therefore, we obtain a candidate with high chemical stability at neutral pH.

Based on the sequence homology of peptides in the glucagon superfamily of structurally related peptide hormones, GUB021794 was tested for cross-reactivity toward other receptors in the family. GUB021794 showed no activation of the GIPR, GLP-2R, or the GCGR when tested at concentrations up to 3000 nM (Table S2).

The albumin binding affinity of fatty diacids correlates with the length of the fatty diacids; C18- and C20 diacids have the highest affinity.[29] A C20DA-gGlu-2xOEG has previously been described to provide a once-weekly dosing regimen in humans[12] while 2Aib is known to protect against DPP-4 proteolytic cleavage.[9,25] Thus, to achieve a long-acting drug

https://doi.org/10.1021/acs.jmedchem.4c00417
J. Med. Chem. 2024, 67, 11814−11826

candidate compatible with once-weekly dosing, we attached C20DA-gGlu-2xOEG to position 14, while 2Aib was introduced to prevent DPP-4 proteolytic cleavage.

GUB021794 demonstrated a prolonged pharmacokinetic half-life of 22 h after intravenous dosing in rats. Such a pharmacokinetic profile has previously been shown to translate to a once-weekly pharmacokinetic profile in humans.[29,30]

Metabolic Effects of GUB021794 in DIO Mice. GUB021794 was characterized in a mouse model of diet-induced obesity (DIO). DIO mice were treated subcutaneously once daily with vehicle, semaglutide (10 nmol/kg), or GUB021794 (3, 10, or 30 nmol/kg) for 28 days. 28-day treatment with GUB021794 promoted a dose-dependent reduction in body weight, on par with an equimolar dose (10 nmol/kg) of semaglutide (Figure 7A, B). Food intake (Figure 7C) was dose-dependently reduced throughout the study with maximal effect on treatments days 2−3 (Figure 7D). The marked reduction in body weight and food intake was reflected by a substantial reduction in whole-body adipose tissue mass (Figure 7E). Plasma exposure in DIO mice (Figure 7F) was dose-dependent and supported the extended half-life of GUB021794.

Overall, the preclinical data demonstrate that GUB021794 is a potent long acting GLP-1R agonist promoting robust body weight loss in DIO mice comparable to clinically approved GLP-1R agonists, such as semaglutide.[31,32]

DISCUSSION

In this study, we demonstrated the capability of the streaMLine platform to accelerate the drug discovery process of peptide pharmaceuticals. We systematically explored secretin as a peptide backbone to generate potent, selective, and long-acting GLP-1R agonists without the intrinsic physical instability previously reported for GLP-1.[10,11] A total of 2688 peptides were synthesized, screened, and analyzed using an ML-guided approach to map out a solution space for generating GLP-1R agonists derived from the secretin backbone. Based on this solution space, we designed and characterized a novel GLP-1R agonist, GUB021794, that showed potent body weight loss in DIO mice, a pharmacokinetic profile compatible with once-weekly dosing in humans, and beneficial physicochemical properties.

The key features employed in the streaMLine platform are (1) systematic peptide library design where each mutation is observed multiple times in multiple different backbones, (2) analysis of large libraries consisting of 100−1000 s of crude peptides, and (3) ML-guided peptide design. The screening of peptide crudes leads to faster data generation, as the laborious peptide purification following SPPS can be omitted. This leads to faster generation of lead molecules as well as potential identification of better compounds as a larger chemical space can be screened using the same amount of resources. Despite the contribution of synthesis byproducts in the individual assay measurements of the peptide crudes, we observe that these contributions have a negligible effect on the overall data interpretation when analyzed as part of a larger systematic peptide library design.

Following the crude screening campaign of this study, we subsequently purified and profiled GUB021794, and verified the conclusions obtained from the crude libraries. Aside from verifying the conclusions obtained during the crude screenings, GUB021794 also demonstrated that the thorough screening approach enabled the generation of a very potent therapeutic peptide. Despite being derived from the alternative starting point secretin, GUB021794 showed a weight-lowering effect in DIO mice on par with semaglutide, a clinically approved GLP-1 receptor agonists.[31,32]

In addition to verifying our screening approach by in-depth characterization of GUB021794, several of the key amino acid substitutions identified in this study have been reported previously. For example, the amino acid positions being important for GLP-1R potency (Figure 2A), agreed with previous studies reporting the high importance of phenylalanine in position 22[33] and loss in GLP-1R potency when introducing 2S[34] and 9E into a GLP-1 context.[25] Furthermore, we found position 19 to be important for SCTR potency in agreement with previous SAR efforts on the secretin backbone.[35] The physicochemical screening enabled the identification and validation of mutations associated with fibril formation of GLP-1R agonists. We identified three substitutions preventing fibril formation, 18Aib, 19Aib, and 21E which could be responsible for the formation of amyloid fibrils of GLP-1R agonists previously reported.[10,11]

Previous studies using QSAR for peptide optimization have been limited to case studies on peptide families with large data sets publicly available such as antimicrobial peptides (AMPs),[36] Major histocompatibility complex (MHC)[37] and antitumor activity of peptides.[38] Other related studies have extracted information about QSAR from smaller publicly available data sets. Deng et al. collected a data set of 141 unique angiotensin-converting enzyme (ACE) inhibitory dipeptides, and fitted a regression model to IC50 values with model predictions being validated by synthesis and in vitro testing of five novel ACE inhibitory dipeptides.[39] Although not evaluating the physicochemical properties of the developed peptides, the study by Deng et al. highlights the great potential of peptide QSAR, even based on a limited data set. Our study demonstrates the importance of co-optimizing potency and physicochemical properties to avoid developing potent but unstable peptides.

In addition, we have demonstrated how the platform allows multiparameter screening in a parallelized workflow, thus enabling timeline-efficient identification of learnings to obtain peptides with drug properties early in a project. This accelerated approach further enables fast evaluation of the *in vitro*/*in vivo* correlation. In case, the peptide design needs re-evaluation, and the already generated data sets provide alternative solutions to be evaluated without significant timeline delay.

While the aim of the present work was to dial out the secretin receptor potency and leverage the more favorable physicochemical properties of secretin, the streaMLine platform is highly applicable for the development of other peptide-based therapies such as polyagonists targeting multiple receptors, since it enables fine-tuning potency on multiple receptors and allows for exploration of a large chemical solution space. Another application of streaMLine could be the development and potency optimization of nonendogenous peptides identified *de novo* by display technologies including mRNA display,[40] or in the development of peptides containing noncanonical amino acids.[41] Recent advances in artificial intelligence (AI) with pretrained deep learning models such as AlphaFold[42] and protein language models (PLMs)[43,44] have enabled improved prediction of protein and peptide properties. For example, de novo design of protein binders has been drastically improved recently[45] making peptide-based therapies

https://doi.org/10.1021/acs.jmedchem.4c00417
J. Med. Chem. 2024, 67, 11814−11826

approachable for previously undruggable targets. Applying QSAR methods as demonstrated in this study could aid in quickly advancing *de novo* peptide hits to the clinical stage.

CONCLUSIONS

In summary, we have developed an ML-based drug discovery platform to systematically identify design solutions for optimizing peptide-based drugs. As a validation, we exploited the secretin peptide backbone as a scaffold to determine a solution space for designing novel GLP-1R agonists with a selective receptor profile and improved physicochemical properties. This unprecedented QSAR approach enables accelerated development of peptide drug candidates on multiple parameters and demonstrates a novel approach that can be applied for future developments of peptide drug candidates.

EXPERIMENTAL SECTION

General Procedure for Peptide Library Synthesis. Peptide libraries were synthesized using standard Fmoc-based solid phase methods on a SyroII fully automated parallel peptide synthesizer (MultiSynTech GmbH, Germany) equipped with a heating block. Tentagel S RAM resin with a loading of 0.23−0.25 mmol/g (Rapp polymer GmbH, Germany) was used. Fmoc deprotection was performed in two stages by treating the resin with 40% piperidine/DMF (0.2 M HOBt (1-hydroxybenzotriazole)) for 3 min at 45 °C followed by 20% piperidine/DMF (0.1 M HOBt) for 7 min at 75 °C. Asp, Cys, and His residues were Fmoc-deprotected at room temperature; i.e. 40% piperidine/DMF (0.2 M HOBt) for 3 min followed by 20% piperidine/DMF (0.1 M HOBt) for 15 min. Coupling was performed using 0.5 M Fmoc-amino acid in 0.5 M oxyma in DMF (except His which was dissolved in NMP) added in 5-fold excess to the resin and activated by 5-fold DIC (*N*,*N*′-diisopropyl-carbodiimide). Coupling conditions were single or double couplings for 15 min at 75 °C, except for His and Cys residues, which were double coupled for 15 min at 50 °C. Also, amino acids coupled after Aib were double coupled. The resin was washed 5× with DMF after Fmoc deprotection and 3× after couplings.

For peptides lipidated at the lysine residue, Boc-protected amino acid was incorporated as the N-terminal residue, and the lipidation position was incorporated as lysine (Mtt). The Mtt group (4-methyltrityl) was removed by treating the resin 3 times with 75% HFIP (1,1,1,3,3,3-hexafluoro-propan-2-ol), 5% TIPS in DCM (dichloromethane) for 10 min. The resin was washed with 10% DIPEA in DCM, followed by a 3× DMF wash. The linker residue(s) were coupled to the epsilon-amino group of the deprotected lysine using double- or triple-standard coupling conditions prior to coupling the mono-*t*-butyl-protected fatty diacid. The fatty acid was double coupled using 2 equiv building blocks.

For N-terminal lipidated peptides, linker residues and mono-*t*-butyl protected fatty diacid were coupled to the N-terminus using the conditions described above.

After synthesis, the resin was washed with DCM and dried, and the polypeptide was cleaved from the resin by a 45 min treatment with TFA (trifluoroacetic acid)/TES (triethylsilane)/DODT(2,2′-(ethylenedioxy)diethanethiol)/water (93/2.5/2.5/2.0) at 40 °C followed by precipitation and wash using cold diethyl ether.

Crude Peptide Libraries. The peptides were characterized by HR-MS (Waters, Denmark) and quantified by LC-CAD (ThermoFisher Scientific, Denmark). Finally, the crude peptide libraries were lyophilized.

Purified GUB021974. The crude peptide was dissolved in acetonitrile/water and purified by reverse phase HPLC using a Waters preparative HPLC with C18 column (Reprosil Gold 200 Å, 5 μm, 30 mm × 250 mm), preparative pumps (waters 2545), UV/vis detector (Waters 2489) and a Waters fraction collector III. The mobile phase was run with a gradient of buffer A (0.1% TFA in H_2O) and buffer B (0.1% TFA in acetonitrile at a flow rate of 20 mL/min at room temperature). Relevant fractions were analyzed, pooled, and lyophilized. Finally, the peptide was freeze-dried using a Telstar benchtop freeze-drieder. Peptide purity and mass were determined by analytical RP-HPLC-MS on an ACQUITY UPLC Peptide CSH C18 column (Waters, ACQUITY UPLC Peptide CSH, C18, 130 Å, 1.7 μm, 2.1 mm × 100 mm) using a Waters Acquity HPLC System equipped with a 3100 Mass Detector. Analysis was performed by gradient elution with buffer A (0.3% TFA in H_2O) and buffer B (0.3% TFA in acetonitrile) at a temperature of 40 °C (gradients used were 40−60%B over 14 min). GUB021794: purity = 95%; calc. mass: 3862.3871; m/z = 1288.2 (m + 3/3).

In Vitro Functional Assay: General Procedure for the Determination of hGLP-1R Potency. A CHO-K1 cell line stably overexpressing the hGLP-1R was obtained from Euroscreen (FAST-0145L), expanded, aliquoted, and frozen. An aliquot was thawed and plated in DPBS with 0.05% casein and 0.5 mM IBMX (2000 cells/well) in a 384-well format. The cells were then immediately stimulated for 30 min at room temperature with graded doses of test compound using human GLP-1(7−36) (synthesized at Gubra) as a positive control. cAMP accumulation was measured using a Cisbio assay for Gs-coupled receptors (cat. no. 62AM4PEC), where the assay reagents were added as per the manufacturer's instructions and time-resolved fluorescence energy transfer recorded after one hour on a CLARIOstar (BMG Labtech) plate reader.

In Vitro Functional Assay: General Procedure for Determination of hSCTR Potency. Frozen division arrested CHO-K1 cells stably overexpressing the hSCTR was obtained from PerkinElmer (ES-712-AF). Alternatively, a CHO-K1 cell line stably overexpressing the hSCTR was obtained from Euroscreen (FAST-0161L), expanded, aliquoted, and frozen. An aliquot was thawed and plated in DPBS with 0.05% casein and 0.5 mM IBMX as 2000 cells/well in a 384-well format. The cells were then immediately stimulated for 30 min at room temperature with graded doses of test compound using human secretin (synthesized at Gubra) as a positive control. cAMP accumulation was measured using a Cisbio assay for Gs-coupled receptors (cat. no. 62AM4PEC), where the assay reagents were added as per the manufacturer's instructions and time-resolved fluorescence energy transfer recorded after 1 h on a CLARIOstar (BMG Labtech) plate reader.

Biophysics: General Procedure for Determination of Solubility and Fibril Formation. Peptides were dissolved in buffers (50 mM sodium phosphate at pH 7.0) to 267 μM and incubated for 1−2 h at room temperature. The samples were then divided into two replicates of 80 μL in a black 384-well plate (μ-clear, Greiner Bio-One) and mixed with Thioflavin T (ThT) to a final concentration of 4 μM. The plate was centrifuged for 2 min at 2000 rpm to remove air bubbles, sealed, and placed in a plate reader (CLARIOstar, BMG). First, the turbidity of the samples was measured as the absorbance at 600 nm as an indication of precipitated material. Second, the plate reader temperature was set to 40 °C and the fluorescence was measured every 10 min for 72 h by exciting the ThT at 450 nm and measuring the emission at 480 nm. Samples were stressed by shaking the plate at 700 rpm (linear) for five min before every measurement, and fibril formation was determined as the average emission for each sample.

Data Modeling. A random forest model[22] was constructed for each in vitro endpoint. Features consisted of the peptide sequences as well as the specific plate number for synthesis and assay, respectively. For the deep mutational scan, peptide sequences were encoded using z-scales,[23] and for the remaining data sets, peptide sequences were one-hot encoded (binary classification of the presence or absence of each amino acid in each position). The models were implemented in the statistical programming language R using the randomForest package version 4.6.14, and SHAP values were computed based on the random forest models using the treeSHAP package version 0.1.1.

Ethics. All animal experiments were approved by The Danish Animal Experiments Inspectorate (License No. 2017-15-0201-01378). Animal experiments were conducted in agreement with internal Gubra bioethical guidelines that are fully compliant with

https://doi.org/10.1021/acs.jmedchem.4c00417
J. Med. Chem. 2024, 67, 11814−11826

internationally accepted principles for the care and use of laboratory animals.

Pharmacokinetics in Rats. Male Spraque Dawley rats were obtained from Janvier (Janvier Laboratories, France) at 6 weeks of age. The animals were single-housed on the day of arrival. After 1 week of acclimatization, the animals (n = 3) were codosed with two test peptides; one compound was dosed intravenously and the other was dosed subcutaneously. In a parallel group (n = 3), the same two test peptides were dosed with opposite routes of administration. This allowed for profiling both the intravenous and subcutaneous administration routes and obtaining key pharmacokinetics parameters. Following dose administration, blood samples (90 μL) were collected in K_2EDTA tubes at time points: 0.17, 1, 3, 6, 24, 30, 48, 72, and 96 h post-dosing. Plasma was obtained by centrifugation of the blood samples (3000 × g, 10 min 4 °C), and the plasma samples were frozen at −70 °C until analysis.

Pharmacokinetics parameters were calculated by noncompartmental analysis (NCA) of the individual plasma concentration−time profile from the animals. NCA was performed in the R-package NonCompart, using the log-linear trapezoidal method for the estimation of AUC and AUMC. $AUC_{0-\infty}$ was extrapolated <20%.

LC-MSMS Analysis. Plasma concentrations were measured by LC-MSMS using electrospray ionization and multiple reaction monitoring. Calibration standards and quality control (QC) samples were prepared in a species-match matrix. In brief, 15 μL calibration standards, QC, and study samples were extracted by protein precipitation using 60 μL of methanol followed by the addition of 45 μL of Milli-Q water. Samples were shaken (800 rpm) at room temperature for 5 min before centrifugation (2570 × g, 40 min, 4 °C), and the supernatant was transferred to a LoBind PCR plate. Samples were analyzed on a Thermo Triscend UHPLC system coupled to a Sciex API 6500+ mass spectrometer. Samples were subject to online SPE cleanup on an HLB column (1 × 50 mm, Waters) before being loaded onto an Aeris Peptide XB C18 column (3.6 μm, 100 Å, 2 × 50 mm, Phenomenex, Torrance, USA) analytical column. The mobile phases consisted of acetonitrile and Milli-Q H_2O both containing 0.1% V/V formic acid.

Pharmacodynamics: Effect on Weight-Loss. Male C57BL/6JRj mice were obtained from Janvier (Janvier Laboratories, France), and group-housed in a controlled environment (12/12 h light-dark cycle, lights off at 3 PM, 22 °C ± 2 °C, humidity 50% ± 10%). Each animal was identified by an implantable subcutaneous microchip (PetID Microchip, E-vet, Haderslev, Denmark). Mice arrived at 5 weeks of age and were offered tap water and regular chow (Altromin 1324, Brogaarden, Denmark) or diet-induced obesity (DIO) was induced by offering ad libitum access to a 60% high-fat diet (5.15 kcal/g; 60 kcal% fat, 20 kcal% carbohydrates, 20% kcal% protein; D12492, Ssnif, Soest, Germany) for 46 weeks before study start. The mice were kept on a diet throughout the study. Mice were randomized according to body weight in week −1. The mice were SC-dosed once daily with Vehicle (50 mM phosphate, 3.5% mannitol) Lean (n = 10), Vehicle DIO (n = 10), positive control semaglutide (Ozempic, Novo Nordisk A/S, Denmark) 10 nmol/kg (n = 10) or the GLP-1R agonists; GUB021794, at 3, 10, or 30 nmol/kg (n = 10 mice/group) for 28 days. Body weight and food intake were measured daily from days −3 to 28. Whole-body lean/fat tissue composition was analyzed by noninvasive EchoMRI scanning (EchoMRI-900, Houston, TX, USA) performed on study day 28. Terminal exposure was measured using LC-MSMS on samples obtained 24 h after the last dosing.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jmedchem.4c00417.

Model performance, detailed positional data overview, interaction effects, characterization of GUB021794, and receptor data (PDF)

Molecular formular string sheet (CSV)

■ AUTHOR INFORMATION

Corresponding Author

Louise S. Dalbøge − *Gubra, Hørsholm 2970, Denmark;* orcid.org/0000-0001-7773-6313; Email: lsd@gubra.dk

Authors

Jens Christian Nielsen − *Gubra, Hørsholm 2970, Denmark;* orcid.org/0000-0002-6522-9987

Claudia Hjørringgaard − *Gubra, Hørsholm 2970, Denmark*

Mads Mørup Nygaard − *Gubra, Hørsholm 2970, Denmark*

Anita Wester − *Gubra, Hørsholm 2970, Denmark*

Lisbeth Elster − *Gubra, Hørsholm 2970, Denmark*

Trine Porsgaard − *Gubra, Hørsholm 2970, Denmark*

Randi Bonke Mikkelsen − *Gubra, Hørsholm 2970, Denmark;* orcid.org/0000-0002-4848-5743

Silas Rasmussen − *Gubra, Hørsholm 2970, Denmark*

Andreas Nygaard Madsen − *Gubra, Hørsholm 2970, Denmark*

Morten Schlein − *Gubra, Hørsholm 2970, Denmark*

Niels Vrang − *Gubra, Hørsholm 2970, Denmark*

Kristoffer Rigbolt − *Gubra, Hørsholm 2970, Denmark*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jmedchem.4c00417

Author Contributions

J.C.N., N.V., K.R., and L.S.D. conceptualized the study. C.H., A.W., L.E., R.B.M., M.S., carried out the in vitro experiments. T.P., S.R., A.N.M. caried out the in vivo experiments. J.C.N., C.H., M.M.N., A.W. analyzed the data. J.C.N., C.H., and L.S.D. wrote the paper. All authors reviewed and approved the final paper. J.C.N. and C.H. contributed equally to this work.

Notes

The authors declare the following competing financial interest(s): All authors are employed at Gubra.

†Joint first author.

■ ACKNOWLEDGMENTS

We thank Henrik Björk Hansen for his help in revising the manuscript.

■ REFERENCES

(1) Fosgerau, K.; Hoffmann, T. Peptide Therapeutics: Current Status and Future Directions. *Drug Discovery Today* **2015**, *20* (1), 122−128, DOI: 10.1016/j.drudis.2014.10.003.

(2) Wang, L.; Wang, N.; Zhang, W.; Cheng, X.; Yan, Z.; Shao, G.; Wang, X.; Wang, R.; Fu, C. Therapeutic Peptides: Current Applications and Future Directions. *Signal Transduct. Target. Ther.* **2022**, *7*, 48.

(3) Schüß, C.; Vu, O.; Mishra, N. M.; Tough, I. R.; Du, Y.; Stichel, J.; Cox, H. M.; Weaver, C. D.; Meiler, J.; Emmitte, K. A.; Beck-Sickinger, A. G. Structure−Activity Relationship Study of the High-Affinity Neuropeptide Y $_4$ Receptor Positive Allosteric Modulator VU0506013. *J. Med. Chem.* **2023**, *66* (13), 8745−8766.

(4) Verma, J.; Khedkar, V.; Coutinho, E. 3D-QSAR in Drug Design - A Review. *Curr. Top Med. Chem.* **2010**, *10* (1), 95−115.

(5) Vamathevan, J.; Clark, D.; Czodrowski, P.; Dunham, I.; Ferran, E.; Lee, G.; Li, B.; Madabhushi, A.; Shah, P.; Spitzer, M.; Zhao, S. Applications of Machine Learning in Drug Discovery and Development. *Nat. Rev. Drug Discov* **2019**, *18* (6), 463−477.

(6) Drucker, D. J. Mechanisms of Action and Therapeutic Application of Glucagon-like Peptide-1. *Cell Metab.* **2018**, 740−756, DOI: 10.1016/j.cmet.2018.03.001.

(7) Secher, A.; Jelsing, J.; Baquero, A. F.; Hecksher-sørensen, J.; Cowley, M. A.; Dalbøge, L. S.; Hansen, G.; Grove, K. L.; Pyke, C.; Raun, K.; Schäffer, L.; Tang-christensen, M.; Verma, S.; Witgen, B. M.; Vrang, N.; Knudsen, L. B. The Arcuate Nucleus Mediates GLP-1 Receptor Agonist Liraglutide-Dependent Weight Loss. *J. Clin. Invest.* **2014**, *124* (10), 4473−4488.
(8) Wilding, J. P. H.; Batterham, R. L.; Calanna, S.; Davies, M.; Van Gaal, L. F.; Lingvay, I.; McGowan, B. M.; Rosenstock, J.; Tran, M. T. D.; Wadden, T. A.; Wharton, S.; Yokote, K.; Zeuthen, N.; Kushner, R. F. Once-Weekly Semaglutide in Adults with Overweight or Obesity. *New England Journal of Medicine* **2021**, *384* (11), 989−1002.
(9) Lau, J.; Bloch, P.; Schä, L.; Pettersson, I.; Spetzler, J.; Kofoed, J.; Madsen, K.; Knudsen, L. B.; Mcguire, J.; Steensgaard, B.; Strauss, M.; Gram, D. X.; Knudsen, S. M.; Nielsen, F. S.; Thygesen, P.; Reedtz-Runge, S.; Kruse, T. Discovery of the Once-Weekly Glucagon-Like Peptide-1 (GLP-1) Analogue Semaglutide. *J. Med. Chem.* **2015**, *58*, 7370−7380, DOI: 10.1021/acs.jmedchem.5b00726.
(10) Přáda Brichtová, E.; Krupová, M.; Bouř, P.; Lindo, V.; Gomes dos Santos, A.; Jackson, S. E. Glucagon-like Peptide 1 Aggregates into Low-Molecular-Weight Oligomers off-Pathway to Fibrillation. *Biophys. J.* **2023**, *122*, 2475−2488, DOI: 10.1016/j.bpj.2023.04.027.
(11) Zapadka, K. L.; Becher, F. J.; Gomes dos Santos, A. L.; Jackson, S. E. Factors Affecting the Physical Stability (Aggregation) of Peptide Therapeutics. *Interface Focus* **2017**, *7*, No. 20170030, DOI: 10.1098/rsfs.2017.0030.
(12) Coskun, T.; Sloop, K. W.; Loghin, C.; Alsina-Fernandez, J.; Urva, S.; Bokvist, K. B.; Cui, X.; Briere, D. A.; Cabrera, O.; Roell, W. C.; Kuchibhotla, U.; Moyers, J. S.; Benson, C. T.; Gimeno, R. E.; D'Alessio, D. A.; Haupt, A. LY3298176, a Novel Dual GIP and GLP-1 Receptor Agonist for the Treatment of Type 2 Diabetes Mellitus: From Discovery to Clinical Proof of Concept. *Mol. Metab.* **2018**, *18*, 3−14.
(13) Coskun, T.; Urva, S.; Roell, W. C.; Qu, H.; Loghin, C.; Moyers, J. S.; O'Farrell, L. S.; Briere, D. A.; Sloop, K. W.; Thomas, M. K.; Pirro, V.; Wainscott, D. B.; Willard, F. S.; Abernathy, M.; Morford, L.; Du, Y.; Benson, C.; Gimeno, R. E.; Haupt, A.; Milicevic, Z. LY3437943, a Novel Triple Glucagon, GIP, and GLP-1 Receptor Agonist for Glycemic Control and Weight Loss: From Discovery to Clinical Proof of Concept. *Cell Metab* **2022**, *34*, 1234.
(14) Finan, B.; Yang, B.; Ottaway, N.; Smiley, D. L.; Ma, T.; Clemmensen, C.; Chabenne, J.; Zhang, L.; Habegger, K. M.; Fischer, K.; Campbell, J. E.; Sandoval, D.; Seeley, R. J.; Bleicher, K.; Uhles, S.; Riboulet, W.; Funk, J.; Hertel, C.; Belli, S.; Sebokova, E.; Conde-Knape, K.; Konkar, A.; Drucker, D. J.; Gelfanov, V.; Pfluger, P. T.; Müller, T. D.; Perez-Tilve, D.; DiMarchi, R. D.; Tschöp, M. H. A Rationally Designed Monomeric Peptide Triagonist Corrects Obesity and Diabetes in Rodents. *Nat. Med.* **2015**, *21* (1), 27−36.
(15) Finan, B.; Ma, T.; Ottaway, N.; Müller, T. D.; Habegger, K. M.; Heppner, K. M.; Kirchner, H.; Holland, J.; Hembree, J.; Raver, C.; Lockie, S. H.; Smiley, D. L.; Gelfanov, V.; Yang, B.; Hofmann, S.; Bruemmer, D.; Drucker, D. J.; Pfluger, P. T.; Perez-Tilve, D.; Gidda, J.; Vignati, L.; Zhang, L.; Hauptman, J. B.; Lau, M.; Brecheisen, M.; Uhles, S.; Riboulet, W.; Hainaut, E.; Sebokova, E.; Conde-Knape, K.; Konkar, A.; DiMarchi, R. D.; Tschöp, M. H. Unimolecular Dual Incretins Maximize Metabolic Benefits in Rodents, Monkeys, and Humans. *Sci. Transl Med.* **2013**, *5* (209), 1−18.
(16) Zimmermann, T.; Thomas, L.; Baader-Pagler, T.; Haebel, P.; Simon, E.; Reindl, W.; Bajrami, B.; Rist, W.; Uphues, I.; Drucker, D. J.; Klein, H.; Santhanam, R.; Hamprecht, D.; Neubauer, H.; Augustin, R. BI 456906: Discovery and Preclinical Pharmacology of a Novel GCGR/GLP-1R Dual Agonist with Robust Anti-Obesity Efficacy. *Mol. Metab* **2022**, *66*, No. 101633.
(17) Aronne, L. J.; Sattar, N.; Horn, D. B.; Bays, H. E.; Wharton, S.; Lin, W.-Y.; Ahmad, N. N.; Zhang, S.; Liao, R.; Bunck, M. C.; Jouravskaya, I.; Murphy, M. A.; Fretes, J. O.; Coronel, M. J.; Gutnisky, L. L.; Frechtel, G. D.; Gelersztein, E.; Aizenberg, D.; Maldonado, N.; Pereira, M.; Santos, Q. G.; Calil Salim, C.; Canani, L. H.; Halpern, B.; Russo, L. A.; Bodart, J.; Neto, D. V.; Augusto, G.; Leite, S.; Yang, Y.-C.; Lin, W.-Y.; Huang, C.-N.; Huang, K.-C.; Fitz-Patrick, D.; Pau, C. H.; Toth, P. D.; Freeman, G. H.; Gardner, D. F.; Wynne, A. G.; Loy, J.; Horn, D. B.; Mehra, P. K.; Layle, S.; Bergthold, J. H.; de Souza, J.; Nadar, V. K.; Albizu Angulo, G. R.; Cohen, K. R.; Smith, T. R.; Aronne, L. J.; Vaughn, M.; Alcantara-Gonzalez, A. A.; Forman, S. B.; Agaiby, J. M.; Geller, S. A.; Fraser, N. J.; Jenders, R. A.; Barbel-Johnson, K. M.; Mayfield, R. K.; Vance, C. D.; Prier, K. T.; Murray, A. V.; Lillestol, M. J.; Denham, D. S.; Park, J. Y.; Klein, E. J.; Bays, H. E.; Philis-Tsimikas, A.; Bressler, P. E.; Reed, J. C.; Aslam, S.; Rosenstock, J.; Frias, J. P.; Klaff, L. J.; Brazg, R.; Gomez-Cuellar, M.; Connery, L.; Van, J. T.; Selam, J.-L.; Kim, J.; Blake, D.; Gabriel, J.; Arora, S.; McCartney, M. J.; Solano, R. K.; Brodie, S. K.; Nardandrea, J. P. Continued Treatment With Tirzepatide for Maintenance of Weight Reduction in Adults With Obesity. *JAMA* **2024**, *331* (1), 38.
(18) Nielsen, J.; Rigbolt, K.; Bech, E.; Lundh, M.; Magotti, P.; Ballarín-Gonzáles, B.; Pedersen, S.; Vrang, N. Ham15−52 Analogues With Improved Amylin Receptor (Hamy3r) Potency. WO_2022_063925_A1, 2022. https://www.lens.org/lens/patent/089-819-804-977-189/frontpage (accessed 2023−10−10).
(19) Chey, W. Y.; Chang, T. M. Secretin, 100 Years Later. *Journal of Gastroenterology.* **2003**, *38*, 1025−1035.
(20) Hoare, S. R. J. Mechanisms of Peptide and Nonpeptide Ligand Binding to Class B G-Protein-Coupled Receptors. *Drug Discov Today* **2005**, *10* (6), 417−427.
(21) Horváth, D.; Dürvanger, Z.; Menyhárd, K. D.; Sulyok-Eiler, M.; Bencs, F.; Gyulai, G.; Horváth, P.; Taricska, N.; Perczel, A. Polymorphic Amyloid Nanostructures of Hormone Peptides Involved in Glucose Homeostasis Display Reversible Amyloid Formation. *Nat. Commun.* **2023**, *14* (1), 4621.
(22) Breiman, L. Random Forests. *Mach Learn* **2001**, *45*, 5−32.
(23) Sandberg, M.; Eriksson, L.; Jonsson, J.; Sjöström, M.; Wold, S. New Chemical Descriptors Relevant for the Design of Biologically Active Peptides. A Multivariate Characterization of 87 Amino Acids. *J. Med. Chem.* **1998**, *41* (14), 2481−2491.
(24) Lundberg, S.; Lee, S.-I. *A Unified Approach to Interpreting Model Predictions*; 2017.
(25) Manandhar, B.; Ahn, J. M. Glucagon-like Peptide-1 (GLP-1) Analogs: Recent, New Possibilities, and Therapeutic Implications. *J. Med. Chem.* **2015**, *58* (3), 1020.
(26) Ebrahimi, S. B.; Samanta, D. Engineering Protein-Based Therapeutics through Structural and Chemical Design. *Nat. Commun.* **2023**, *14* (1), 2411.
(27) Lee, S.; Lee, D. Y. Glucagon-like Peptide-1 and Glucagon-like Peptide-1 Receptor Agonists in the Treatment of Type 2 Diabetes. *Ann. Pediatr Endocrinol Metab* **2017**, *22* (1), 15−26.
(28) Laurila, S.; Rebelos, E.; Honka, M.-J.; Nuutila, P. Pleiotropic Effects of Secretin: A Potential Drug Candidate in the Treatment of Obesity? *Front Endocrinol (Lausanne)* **2021**, *12*, No. 737686.
(29) Kurtzhals, P.; Østergaard, S.; Nishimura, E.; Kjeldsen, T. Derivatization with Fatty Acids in Peptide and Protein Drug Discovery. *Nat. Rev. Drug Discovery* **2023**, *22*, 59−80, DOI: 10.1038/s41573-022-00529-w.
(30) Kruse, T.; Hansen, J. L.; Dahl, K.; Schäffer, L.; Sensfuss, U.; Poulsen, C.; Schlein, M.; Hansen, A. M. K.; Jeppesen, C. B.; Dornonville De La Cour, C.; Clausen, T. R.; Johansson, E.; Fulle, S.; Skyggebjerg, R. B.; Raun, K. Development of Cagrilintide, a Long-Acting Amylin Analogue. *J. Med. Chem.* **2021**, *64* (15), 11183−11194.
(31) Gabery, S.; Salinas, C. G.; Paulsen, S. J.; Ahnfelt-Rønne, J.; Alanentalo, T.; Baquero, A. F.; Buckley, S. T.; Farkas, E.; Fekete, C.; Frederiksen, K. S.; Helms, H. C. C.; Jeppesen, J. F.; John, L. M.; Pyke, C.; Nøhr, J.; Lu, T. T.; Polex-Wolf, J.; Prevot, V.; Raun, K.; Simonsen, L.; Sun, G.; Szilvásy-Szabó, A.; Willenbrock, H.; Secher, A.; Knudsen, L. B.; Hogendorf, W. F. J. Semaglutide Lowers Body Weight in Rodents via Distributed Neural Pathways. *JCI Insight* **2020**, *5* (6), No. e133429, DOI: 10.1172/jci.insight.133429.
(32) Knudsen, L. B.; Lau, J. The Discovery and Development of Liraglutide and Semaglutide. *Front Endocrinol (Lausanne)* **2019**, *10*, 155.
(33) Gallwitz, B.; Witt, M.; Paetzold, G.; Morys-Wortmann, C.; Zimmermann, B.; Eckart, K.; Fölsch, U. R.; Schmidt, W. E. Structure/

Activity Characterization of Glucagon-Like Peptide-1. *Eur. J. Biochem.* **1994**, *225* (3), 1151−1156.
(34) Sarrauste de Menthière, C.; Chavanieu, A.; Grassy, G.; Dalle, S.; Salazar, G.; Kervran, A.; Pfeiffer, B.; Renard, P.; Delagrange, P.; Manechez, D.; Bakes, D.; Ktorza, A.; Calas, B. Structural Requirements of the N-Terminal Region of GLP-1-[7−37]-NH2 for Receptor Interaction and CAMP Production. *Eur. J. Med. Chem.* **2004**, *39* (6), 473−480.
(35) Dong, M.; Le, A.; Te, J. A.; Pinon, D. I.; Bordner, A. J.; Miller, L. J. Importance of Each Residue within Secretin for Receptor Binding and Biological Activity. *Biochemistry* **2011**, *50* (14), 2983−2993.
(36) Waghu, F. H.; Gawde, U.; Gomatam, A.; Coutinho, E.; Idicula-Thomas, S. A QSAR Modeling Approach for Predicting Myeloid Antimicrobial Peptides with High Sequence Similarity. *Chem. Biol. Drug Des* **2020**, *96* (6), 1408−1417.
(37) Hoof, I.; Peters, B.; Sidney, J.; Pedersen, L. E.; Sette, A.; Lund, O.; Buus, S.; Nielsen, M. NetMHCpan, a Method for MHC Class I Binding Prediction beyond Humans. *Immunogenetics* **2009**, *61* (1), 1−13.
(38) Radman, A.; Gredičak, M.; Kopriva, I.; Jerić, I. Predicting Antitumor Activity of Peptides by Consensus of Regression Models Trained on a Small Data Sample. *Int. J. Mol. Sci.* **2011**, *12* (12), 8415−8430.
(39) Deng, B.; Ni, X.; Zhai, Z.; Tang, T.; Tan, C.; Yan, Y.; Deng, J.; Yin, Y. New Quantitative Structure−Activity Relationship Model for Angiotensin-Converting Enzyme Inhibitory Dipeptides Based on Integrated Descriptors. *J. Agric. Food Chem.* **2017**, *65* (44), 9774−9781.
(40) Huang, Y.; Wiedmann, M. M.; Suga, H. RNA Display Methods for the Discovery of Bioactive Macrocycles. *Chem. Rev.* **2019**, *119* (17), 10360−10391.
(41) Goto, Y.; Suga, H. The RaPID Platform for the Discovery of Pseudo-Natural Macrocyclic Peptides. *Acc. Chem. Res.* **2021**, *54* (18), 3604−3617.
(42) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583−589.
(43) Lin, Z.; Akin, H.; Rao, R.; Hie, B.; Zhu, Z.; Lu, W.; Smetanin, N.; Verkuil, R.; Kabeli, O.; Shmueli, Y.; dos Santos Costa, A.; Fazel-Zarandi, M.; Sercu, T.; Candido, S.; Rives, A. Evolutionary-Scale Prediction of Atomic-Level Protein Structure with a Language Model. *Science* **1979**, *379* (6637), 1123−1130.
(44) Rives, A.; Meier, J.; Sercu, T.; Goyal, S.; Lin, Z.; Liu, J.; Guo, D.; Ott, M.; Zitnick, C. L.; Ma, J.; Fergus, R. Biological Structure and Function Emerge from Scaling Unsupervised Learning to 250 Million Protein Sequences. *Proc. Natl. Acad. Sci. U. S. A.* **2021**, *118* (15), No. e2016239118, DOI: 10.1073/pnas.2016239118.
(45) Bennett, N. R.; Coventry, B.; Goreshnik, I.; Huang, B.; Allen, A.; Vafeados, D.; Peng, Y. P.; Dauparas, J.; Baek, M.; Stewart, L.; DiMaio, F.; De Munck, S.; Savvides, S. N.; Baker, D. Improving de Novo Protein Binder Design with Deep Learning. *Nat. Commun.* **2023**, *14* (1), 2625.

https://doi.org/10.1021/acs.jmedchem.4c00417
J. Med. Chem. 2024, 67, 11814−11826