FEW-SHOT ACTIVE LEARNING FOR DE NOVO DUAL-TARGET PEPTIDE DESIGN WITH HIGH BIO-ACTIVITY

Tianxu Lv[1,2]*, **Bing He**[2]*, **Yuran Jia**[2], **Xiang Pan**[1]†, **& Jianhua Yao**[2]†
[1]School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi, China
[2]AI for Life Sciences Lab, Tencent, Shenzhen, China

ABSTRACT

Despite the urgent need for high bio-activity peptides in novel biomedical therapies, the *de novo* design of such peptides, especially those with dual targets, remains an unsolved challenge. Here, we introduce ORIDTP, a few-shot active learning pipeline that integrates *in silico* peptide generation with *in vitro* experimental feedback for *de novo* design of both single-target and dual-target peptides with high bio-activity. ORIDTP involves single-target or dual-target oriented peptide *de novo* generation, binding affinity maturation, and iterative reinforcement of bio-activity based on wet-laboratory feedback. Using ORIDTP, we successfully designed high bio-activity peptides targeting GLP-1R after four iterative rounds, achieving EC_{50} (half maximal effective concentration) values ranging from 35.1 pM to 8.1 pM, which outperform the natural peptide with the highest known bio-activity of 40.8 pM. Furthermore, ORIDTP successfully designed *de novo* dual-target peptides for activating GLP-1R (EC_{50} values ranging from 53.4 pM to 8.2 pM) and GCGR (EC_{50} values ranging from 0.82 nM to 0.21 nM) after four iterative rounds. The best dual-target peptide outperformed two natural peptides with the highest known bio-activity for their respective target proteins (8.2 pM versus 40.8 pM for GLP-1R, and 0.24 nM versus 1.4 nM for GCGR). ORIDTP represents a significant advancement in the rapid and effective design of dual-target peptides for therapeutic applications.

1 INTRODUCTION

Dual-target peptides, capable of simultaneously interacting with two distinct therapeutic target proteins to yield additive or synergistic effects (Liu et al. (2024)), present a promising avenue for the development of novel therapeutics for complex diseases, such as cancers (Zha et al. (2021)) and chronic conditions (Pan et al. (2021)). The ability to engineer dual-target peptide binders exhibiting high bio-activity for diverse dual-target systems opens up numerous diagnostic and therapeutic opportunities. Traditionally, the development of dual-target peptides necessitates a comprehensive understanding of the biological characteristics and binding sites of both target proteins, coupled with extensive experimental assessments (Muttenthaler et al. (2021)). This process is not only time-consuming, costly, and labor-intensive, but also requires substantial expertise from practitioners. More importantly, it poses a challenge to concurrently achieve high bio-activity against both target proteins, even when dual-target peptides are successfully developed.

Recent advancements in *de novo* design of protein binders (Cao et al. (2022); Vázquez Torres et al. (2024)), such as RFdiffusion (Watson et al. (2023)), have significantly facilitated the development of *de novo* protein-binding peptide design (Wang et al. (2023); Chen et al. (2024); Wang et al. (2024)). While existing methods have shown promising results in the design of novel peptide binders, they often overlook the improvement of peptide bio-activity, hampering the development of peptides they designed as medicines. In contrast, recent studies that integrate computational generation with experimental feedback have successfully developed promising mutated enzymes and antibodies (Jiang et al. (2024)), inspiring us to explore new strategies for the *de novo* design of peptides with high bio-activity. However, the aforementioned methods have limitations when directly applied to *de*

*Equal contribution.
†Corresponding authors.

novo design of dual-target peptide. Unlike single-target peptides that only necessitate consideration of a single optimization direction, dual-target peptides require simultaneous consideration of two different target proteins to achieve high bio-activity for both. This presents a greater challenge due to the extreme scarcity of dual-target peptide data and the complexity of optimization directions.

We reasoned that an effective iterative feedback algorithm, oriented towards dual targets, could significantly enhance model optimization and facilitate the continuous improvement of the biological activity of generated peptides for both target proteins. We developed ORIDTP, a dual-target-oriented few-shot active learning pipeline that integrates *in silico* generation with *in vitro* experiments for *de novo* design for high bio-activity peptides targeting dual targets. The effectiveness of ORIDTP in designing *de novo* high bio-activity peptide binders for both single and dual target proteins has been demonstrated through *in vitro* experiments.

2 FEW-SHOT ACTIVE LEARNING FOR *de novo* PEPTIDE DESIGN

ORIDTP is a few-shot active learning pipeline that combines a target-guided diffusion model for *de novo* peptide generation, a reinforcement learning model for affinity maturation, and an affinity model for peptide candidate screening. As shown in Figure 1a, in each round of single-target peptide design, ORIDTP actively selects the most promising candidates for experimental determination. The measured results are subsequently fed back into the models to refine the single-target bio-activity landscape, facilitating the next round for peptide evolution aimed at the single protein. As shown in Figure 1b, in each round of dual-target peptide design, ORIDTP selects the candidates that show the greatest potential for both target proteins for experimental validation. The results for both targets are then fed back into the models to enhance the dual-target bio-activity landscape, guiding the next round for peptide evolution targeting dual proteins.

3 RESULTS

Iterative *de novo* design of high bio-activity single-target peptides. We first designed *de novo* peptide binders targeting a single protein. We selected glucagon-like peptide-1 receptor (GLP-1R) as our target protein due to its significant role in glucose metabolism and its therapeutic potential in the treatment of type 2 diabetes and obesity (Fujita et al. (2014)). Upon activation by agonists, GLP-1R promotes the synthesis and release of insulin, thereby lowering blood glucose in the human body. In each round, we employed ORIDTP for *de novo* peptide generation, affinity maturation, and candidate screening. ORIDTP actively selected the most promising peptides for experimental testing. Top 15 peptides were selected for synthesis and experimental validation in each round. The bio-activity was assessed by the median effect concentration (EC_{50}).

In the **first** round, five peptides exhibited sub-nanomolar biological activity with EC_{50} values ranging from 276.2 nM to 101.3 nM (Figure 2a, colored blue). The best peptide (GA-Single.R1.S5) in this round demonstrated a promising performance (EC_{50} = 101.3 nM, Figure 2b, colored blue). In the **second** round, five of selected peptides exhibited increased biological activity, with EC_{50} values ranging from 3.3 nM to 198.7 pM (Figure 2a, colored green). The best peptide in this round, GA-Single.R2.S5, exhibited an impressive EC_{50} of 198.7 pM (Figure 2b, colored green). In the **third** round, four peptides achieved further enhancement in biological activity, with EC_{50} values ranging from 80.3 pM to 44.6 pM (Figure 2a, colored pink). Notably, the best-performing peptide in this round, GA-Single.R3.S4, exhibited biological activity (EC_{50} = 44.6 pM, Figure 2b, colored pink) comparable to that of the known strongest natural peptide, GLP-1 (EC_{50} = 40.8 pM, Figure 2b, colored gray). This promising results motivated us to proceed to the next iteration, aiming to design *de novo* peptides with bio-activity surpassing that of natural peptides. Remarkably, in the **fourth** round, four peptides demonstrated increased biological activity, all outperforming the natural peptide GLP-1, with EC_{50} values ranging from 35.1 pM to 8.1 pM (Figure 2a, colored red). The best peptide, GA-Single.R4.S4, exhibited an extraordinarily high biological activity (EC_{50} = 8.1 pM, Figure 2b, colored red). The significant improvement demonstrated the effectiveness of ORIDTP in *de novo* design of high bio-activity single-target peptides by leveraging wet-lab feedback.

Iterative *de novo* design of high bio-activity dual-target peptides. After demonstrating ORIDTP's ability to design high bio-activity single-target peptides, we sought to explore its potential to design dual-target peptides. Building open the successful design of peptides targeting GLP-1R, we selected

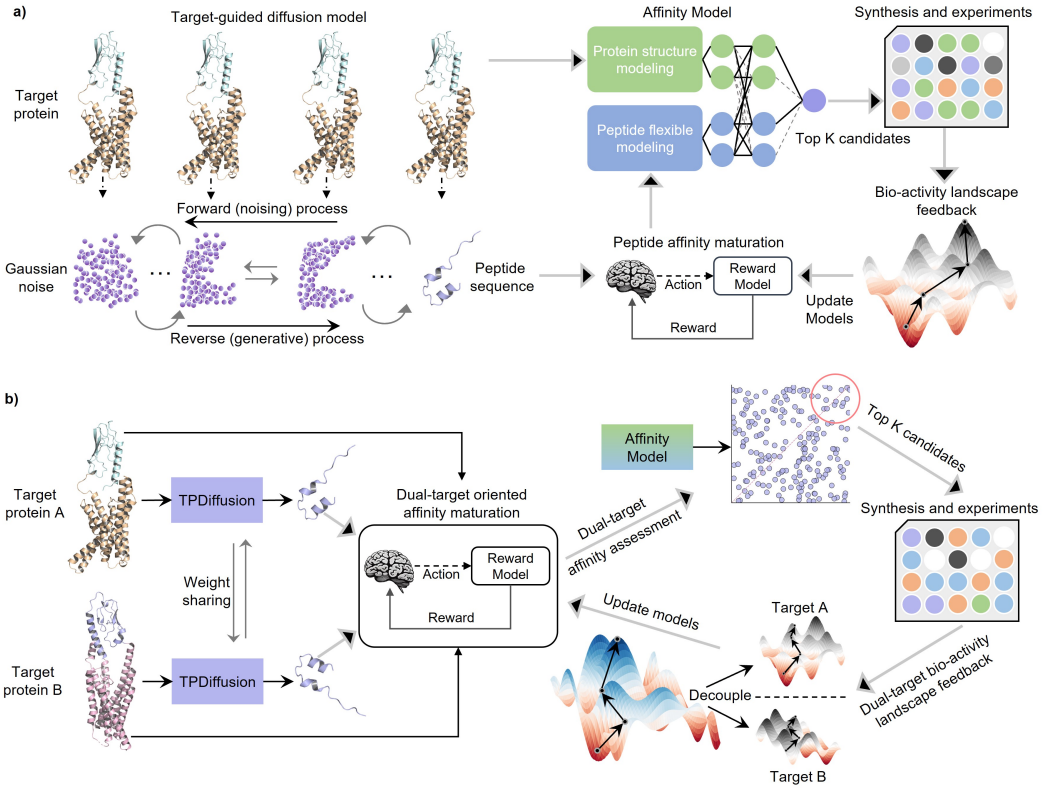

Figure 1: ORIDTP is a few-shot active learning pipeline that integrates *in silico* generation with *in vitro* experiments for *de novo* design of single-target peptides and dual-target peptides. **a)** ORIDTP consists of a target-guided diffusion model (TPDiffusion) for *de novo* peptide generation, a reinforcement learning model for peptide affinity maturation, and an affinity model for peptide screening. Experimental results of the selected peptides are fed back into the models to learn the bio-activity landscape, leading to the next round of iteration. **b)** To effectively design dual-target peptides, ORIDTP employs a dual-target oriented peptide generation, reinforcement, and screening.

the glucagon receptor (GCGR) as another target (Zhang et al. (2018)). The simultaneous modulation of both GLP-1R and GCGR could yield synergistic effects, enhancing therapeutic outcomes for conditions such as type 2 diabetes and obesity (Campbell et al. (2023)). In each round, we employed ORIDTP for dual-target-oriented *de novo* peptide generation, affinity maturation, and candidate screening. ORIDTP actively selected the most promising peptides in simultaneous activating GCGR and GLP-1R. Top 15 peptides were selected for synthesis and experimental validation in each round. The bio-activity targeting GCGR and GLP-1R was assessed by EC_{50}.

In the **first** round, three peptides demonstrated effective simultaneous activation of both GCGR and GLP-1R. The EC_{50} values for activating GCGR ranged from 29.9 nM to 6.5 nM (Figure 3a left, colored blue), while the EC_{50} values for GLP-1R ranged from 276.5 pM to 102.0 pM (Figure 3a right, colored blue). Despite the demonstrated bio-activity of these peptides against both GLP-1R and GCGR, their activities were lower than those of the known strongest natural peptides for each target. Specifically, Glucagon, which targets GCGR, has an EC_{50} of 1.4 nM (Figure 3b left), while GLP-1, which targets GLP-1R, has an EC_{50} of 40.8 pM (Figure 3b right). This indicates that designing *de novo* dual-target peptides capable of simultaneously binding both proteins while maintaining high bio-activity is challenging. In the **second** round, five peptides exhibited increased biological activity. The EC_{50} values for activating GCGR ranged from 4.6 nM to 1.5 nM (Figure 3a left, colored green), while the EC_{50} values for GLP-1R ranged from 355.3 pM to 60.7 pM (Figure 3a right, colored green). Notably, several dual-target peptides in this round demonstrated comparable efficacy to their single-target natural peptides. For instance, GA-Dual.R2.S1 showed comparable activity to Glucagon in activating GCGR (1.5 nM verse 1.4 nM), while GA-Dual R2.S5 exhibited comparable

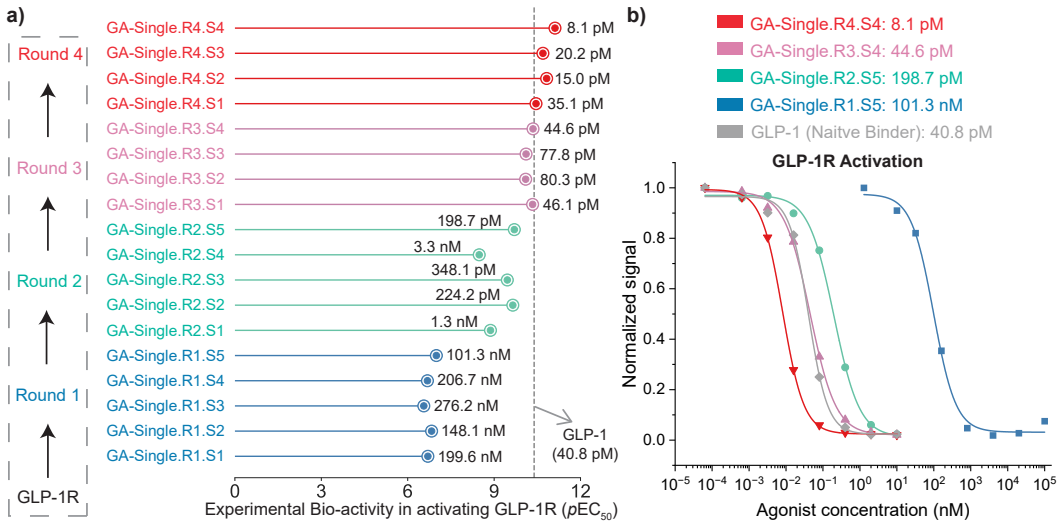

Figure 2: ***De novo* iterative design and characterization of high bio-activity peptides in activating GLP-1R using ORIDTP**. **a**, In four rounds of iterative design, each round generates peptides that exhibit higher biological activity against GLP-1R than those from the previous round. The *de novo* designed peptides with the highest bio-activity in rounds 1, 2, 3, and 4 have EC_{50} values of 101.3 nM, 198.7 pM, 44.6 pM, and 8.1 pM, respectively. In the fourth round, four peptides demonstrated activities ranging from 35.1 pM to 8.1 pM, all of which outperform the known strongest natural peptide, GLP-1 (EC_{50}: 40.8 pM). **b**, Activity profile the first evolved peptide GA-Single.R1.S5 (EC_{50}: 101.3 nM, colored blue), the second evolved peptide GA-Single.R2.S5 (EC_{50}: 198.7 pM, colored green), the third evolved peptide GA-Single.R3.S4 (EC_{50}: 44.6 pM, colored pink), the fourth evolved peptide GA-Single.R4.S4 (EC_{50}: 8.1 pM, colored red), and the natural peptide GLP-1 (EC_{50}: 40.8 pM, colored gray) over HEK293 cells stably expressing GLP-1R and CRE-luciferase.

activity to GLP-1 in activating GLP-1R (60.7 pM verse 40.8 pM, Figure 3b right). In the **third** round, five of selected peptides achieved further enhancements in biological activity. The EC_{50} values for activating GCGR ranged from 1.05 nM to 0.28 nM (Figure 3a left, colored pink), while the EC_{50} values for GLP-1R ranged from 228.4 pM to 44.0 pM (Figure 3a right, colored pink). Among these, the best-performing peptide, GA-Dual.R3.S3, not only demonstrated comparable efficacy to GLP-1 in activating GLP-1R (44.0 pM verse 40.8 pM, Figure 3b right), but also exhibited higher bio-activity in activating GCGR compared to Glucagon (0.32 nM verse 1.4 nM, Figure 3b left). These promising results motivated us to pursue the next iteration to design a dual-target peptide with higher bio-activity than both GLP-1 and Glucagon. Remarkably, in the **fourth** round, three peptides, GA-Dual.R4.S2, GA-Dual.R4.S3, and GA-Dual.R4.S4, achieved this challenging goal (Figure 3a, colored by red). In particular, GA-Dual.R4.S4 demonstrated an EC_{50} of 8.2 pM for targeting GLP-1R (Figure 3b right, colored red), which is a 5.0-fold increase in activity than GLP-1. Meanwhile, it exhibited an EC_{50} of 0.24 nM for targeting GCGR (Figure 3b left, colored red), representing a 5.8-fold increase in activity compared to Glucagon. These findings highlight the potential of GA-Dual.R4.S2, GA-Dual.R4.S3, and GA-Dual.R4.S4 as prime candidates for dual-target therapeutic applications, as both effectively combines the desirable properties of both natural peptides while enhancing their individual activities. The above observations not only demonstrate the effectiveness of ORIDTP in designing *de novo* dual-target peptides with high biological activity, but also highlight ORIDTP's ability to continuously improve the bio-activity of designed peptides through wet-laboratory feedback and iterative refinement.

4 Discussion

Here, we have developed ORIDTP, an innovative few-shot active learning pipeline that combines computational peptide design and wet-laboratory feedback to design high bio-activity peptides for single targets or dual targets. ORIDTP is composed of a target-guided diffusion model for *de*

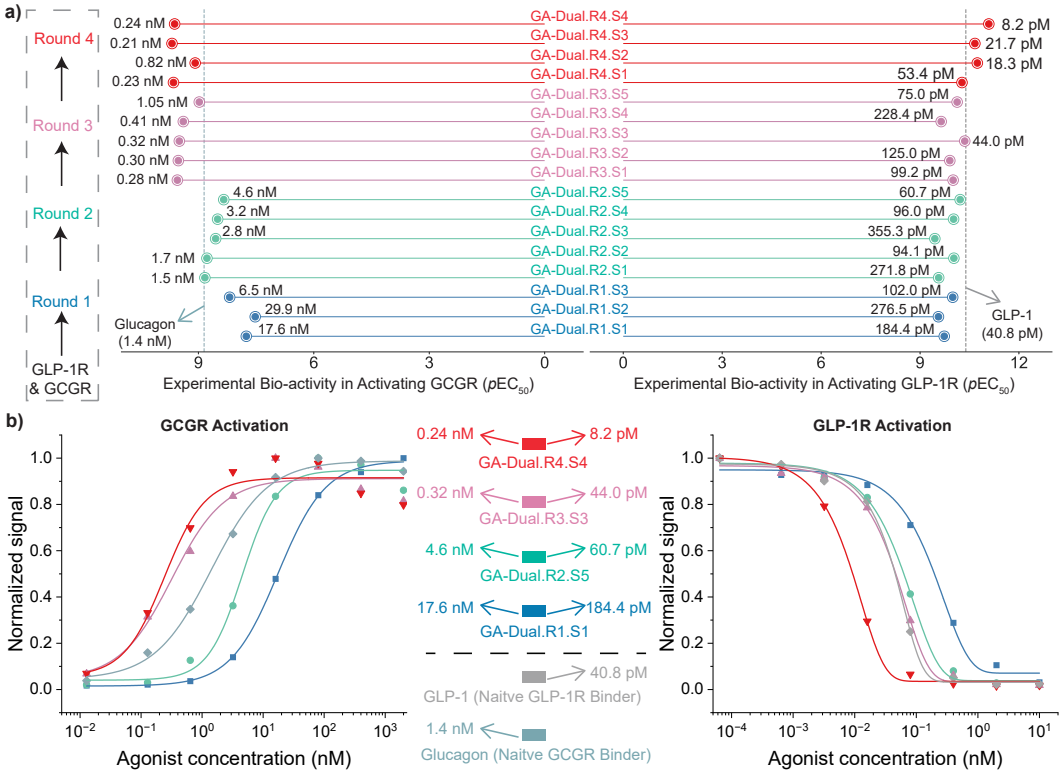

Figure 3: ***De novo* iterative design of dual-target peptides with high bio-activity in activating GLP-1R and GCGR using ORIDTP. a**, In four rounds of dual-target-oriented iterative design, each round generated peptides that exhibited increased biological activity targeting both receptors compared to the previous round. The known strongest natural peptides targeting GLP-1R and GCGR are GLP-1 (EC_{50}: 40.8 pM) and Glucagon (EC_{50}: 1.4 nM), respectively. In the third round, GA-Dual.R3.S3 demonstrated an activation effect on GLP-1R comparable to that of GLP-1 (44.0 pM verse 40.8 pM), while it activated GCGR more effectively than Glucagon (0.32 nM verse 1.4 nM). In the fourth round, three peptides, including GA-Dual.R4.S2 (EC_{50}: 0.82 nM for GCGR and 18.3 pM for GLP-1R), GA-Dual.R4.S3 (EC_{50}: 0.21 nM for GCGR and 21.7 pM for GLP-1R), and GA-Dual.R4.S4 (EC_{50}: 0.24 nM for GCGR and 8.2 pM for GLP-1R), exhibited higher bio-activities in activating GCGR and GLP-1R compared to their natural counterparts. **b**, Activity profile of the first evolved peptide GA-Dual.R1.S1 (EC_{50}: 17.6 nM for GCGR and 0.18 nM for GLP-1R, colored blue), the second evolved peptide GA-Dual.R2.S5 (EC_{50}: 4.6 nM for GCGR and 60.7 pM for GLP-1R, colored green), the third evolved peptide GA-Dual.R3.S3 (EC_{50}: 0.32 nM for GCGR and 44.0 pM for GLP-1R, colored pink), and the fourth evolved peptide GA-Dual.R4.S4 (EC_{50}: 0.24 nM for GCGR and 8.2 pM for GLP-1R, colored red) in activating GCGR (left) and GLP1R (right).

novo peptide generation, a reinforcement learning model for affinity maturation, and a geometric graph model for candidate screening. During each iterative round, ORIDTP actively selected the most promising peptides for experimental validation, the results of which were then fed back into ORIDTP for optimization, facilitating the next round of iteration. Using ORIDTP, we successfully designed high bio-activity peptides that activate GLP-1R. After three iterative rounds, ORIDTP was able to design *de novo* peptides with bio-activity comparable to that of the most potent known natural peptide, GLP-1 (44.6 pM verse 40.8 pM). Remarkably, after four iterative rounds, ORIDTP designed novel peptides that outperform the bio-activity of GLP-1, with EC_{50} ranging from 35.1 pM to 8.1 pM. This demonstrates the significant potential of ORIDTP in the rapid and efficient design of highly bioactive peptides for therapeutic applications. Moreover, after four rounds of dual-target-oriented iterations, we successfully designed high bio-activity dual-target peptides, capable of simultaneously activating GLP-1R and GCGR. In comparison to the strongest known natural peptides for each respective target, two *de novo* designed dual-target peptides demonstrated superior bio-activity. The

most promising peptide GA-Dual.R4.S4 demonstrated an EC_{50} of 8.2 pM for GLP-1R, which is 5.0-fold increase in activity than GLP-1. Meanwhile, it exhibited an EC_{50} of 0.24 nM for GCGR, representing a 5.8-fold increase in activity compared to Glucagon. We anticipate ORIDTP will facilitate the rapid design of high bio-activity single-target and dual-target peptides, accelerating the development of peptides for a wide range of functional applications.

ACKNOWLEDGMENT

This work is supported in part by the National Natural Science Foundation of China under grants W2411054, U21A20521 and 62271178, the Postgraduate Research & Practice Innovation Program of Jiangsu Province KYCX23_2524, National Foreign Expert Project of China under Grant G2023144009L, Zhejiang Provincial Natural Science Foundation of China (LR23F010002), Wuxi Health Commission Precision Medicine Project (J202106), Jiangsu Provincial Six Talent Peaks Project (YY-124), and the construction project of Shanghai Key Laboratory of Molecular Imaging (18DZ2260400).

We thank Sanyou Biopharmaceuticals Co., Ltd. for support in conducting the wet-lab experiments.

REFERENCES

Jonathan E Campbell, Timo D Müller, Brian Finan, Richard D DiMarchi, Matthias H Tschöp, and David A D'Alessio. Gipr/glp-1r dual agonist therapies for diabetes and weight loss—chemistry, physiology, and clinical applications. *Cell Metabolism*, 2023.

Longxing Cao, Brian Coventry, Inna Goreshnik, Buwei Huang, William Sheffler, Joon Sung Park, Kevin M Jude, Iva Marković, Rameshwar U Kadam, Koen HG Verschueren, et al. Design of protein-binding proteins from the target structure alone. *Nature*, 605(7910):551–560, 2022.

Sijie Chen, Tong Lin, Ruchira Basu, Jeremy Ritchey, Shen Wang, Yichuan Luo, Xingcan Li, Dehua Pei, Levent Burak Kara, and Xiaolin Cheng. Design of target specific peptide inhibitors using generative deep learning and molecular dynamics simulations. *Nature Communications*, 15(1): 1611, 2024.

Hiroki Fujita, Tsukasa Morii, Hiromi Fujishima, Takehiro Sato, Tatsunori Shimizu, Mihoko Hosoba, Katsushi Tsukiyama, Takuma Narita, Takamune Takahashi, Daniel J Drucker, et al. The protective roles of glp-1r signaling in diabetic nephropathy: possible mechanism and therapeutic potential. *Kidney international*, 85(3):579–589, 2014.

Kaiyi Jiang, Zhaoqing Yan, Matteo Di Bernardo, Samantha R Sgrizzi, Lukas Villiger, Alisan Kayabolen, BJ Kim, Josephine K Carscadden, Masahiro Hiraizumi, Hiroshi Nishimasu, et al. Rapid in silico directed evolution by a protein language model with evolvepro. *Science*, pp. eadr6006, 2024.

Han Liu, Yujie Hong, Hui Chen, Xianggui Wang, Jiale Dong, Xiaoqian Li, Zihan Shi, Qian Zhao, Longyuan Zhou, JiaXin Wang, et al. Dual activation of gcgr/glp1r signaling ameliorates intestinal fibrosis via metabolic regulation of histone h3k9 lactylation in epithelial cells. *Acta Pharmaceutica Sinica B*, 2024.

Markus Muttenthaler, Glenn F King, David J Adams, and Paul F Alewood. Trends in peptide drug discovery. *Nature reviews Drug discovery*, 20(4):309–325, 2021.

Qi Pan, Shushan Lin, Yu Li, Liang Liu, Xiaoping Li, Xianglei Gao, Jiangyu Yan, Baohua Gu, Xiaofeng Chen, Wenjia Li, et al. A novel glp-1 and fgf21 dual agonist has therapeutic potential for diabetes and non-alcoholic steatohepatitis. *EBioMedicine*, 63, 2021.

Susana Vázquez Torres, Philip JY Leung, Preetham Venkatesh, Isaac D Lutz, Fabian Hink, Huu-Hien Huynh, Jessica Becker, Andy Hsien-Wei Yeh, David Juergens, Nathaniel R Bennett, et al. De novo design of high-affinity binders of bioactive helical peptides. *Nature*, 626(7998):435–442, 2024.

Fanhao Wang, Yuzhe Wang, Laiyi Feng, Changsheng Zhang, and Luhua Lai. Target-specific de novo peptide binder design with diffpepbuilder. *Journal of Chemical Information and Modeling*, 2024.

Yi Wang, Hui Tang, Lichao Huang, Lulu Pan, Lixiang Yang, Huanming Yang, Feng Mu, and Meng Yang. Self-play reinforcement learning guides protein engineering. *Nature Machine Intelligence*, 5(8):845–860, 2023.

Joseph L Watson, David Juergens, Nathaniel R Bennett, Brian L Trippe, Jason Yim, Helen E Eisenach, Woody Ahern, Andrew J Borst, Robert J Ragotte, Lukas F Milles, et al. De novo design of protein structure and function with rfdiffusion. *Nature*, 620(7976):1089–1100, 2023.

Shuai Zha, Ho-Fai Chau, Wai Yin Chau, Lai Sheung Chan, Jun Lin, Kwok Wai Lo, William Chi-Shing Cho, Yim Ling Yip, Sai Wah Tsao, Paul J Farrell, et al. Dual-targeting peptide-guided approach for precision delivery and cancer monitoring by using a safe upconversion nanoplatform. *Advanced Science*, 8(5):2002919, 2021.

Haonan Zhang, Anna Qiao, Linlin Yang, Ned Van Eps, Klaus S Frederiksen, Dehua Yang, Antao Dai, Xiaoqing Cai, Hui Zhang, Cuiying Yi, et al. Structure of the glucagon receptor in complex with a glucagon analogue. *Nature*, 553(7686):106–110, 2018.