

---

# HelixFlow, SE(3)-equivariant Full-atom Design of Peptides With Flow-matching Models

---

Xuezhi Xie<sup>1,2</sup> Pedro A Valiente<sup>1</sup> Jisun Kim<sup>1</sup> Jin Sub Lee<sup>1,3</sup> Philip M Kim<sup>1,2,3</sup>

<sup>1</sup>Donnelly Centre for Cellular and Biomolecular Research, University of Toronto

<sup>2</sup>Department of Computer Science, University of Toronto.

<sup>3</sup>Department of Molecular Genetics, University of Toronto.

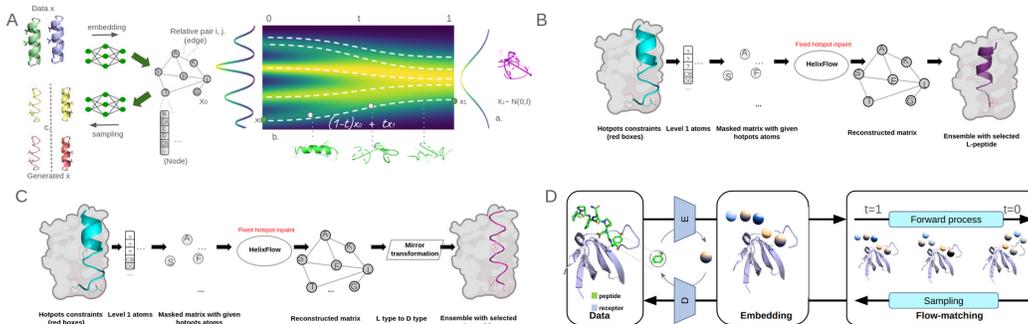
## Abstract

Peptides are becoming a major therapeutic modality. Although peptides are integral to biological processes, designing therapeutic peptides *de novo* remains a challenging prospect. In this paper, we exploit the rich biological inductive bias of amino acids and introduce HelixFlow, a flow-matching model to design full-atom peptide structures. We incorporate a hotspot-specific sequence-conditioned SE(3)-equivariant flow matching module for full-atom helical structure generation and a novel pocket-flow module to generate the binding peptides given target receptors. HelixFlow presents substantial new architectural features over the previous HelixDiff family of models including an equivariant all heavy atom representation, a transformer-based model for flow prediction, and flexible-length generation. By one-shot generation without assembling and direct coordinate generation, HelixFlow could become a more powerful tool for realistic peptide design and open a door for more concise conditional generations on the atom level. As a proof of concept, we designed an acetylated D-peptide of Insulin-like peptide 5 (INSL5) that selectively activates the relaxin family peptide receptor 4 (RXFP4). Our designed D-INSL5 peptide shows good biological activity, comparable AKT phosphorylation levels and high resistance to protease degradation, underscoring the successful integration of deep learning and structure-based modeling and simulation for target-specific peptide design.

## 1 Introduction

Computational peptide design has emerged as a pivotal field in biomedicine, driven by peptides' crucial roles in cellular processes and their therapeutic potential as unique pharmaceutical agents. Peptides are involved in nearly 40% of protein-protein interactions essential for cellular functions [8]. Renowned for their high specificity and low toxicity compared to small molecules, peptides have demonstrated commercial viability, as evidenced by several top-selling GLP-1 agonist drugs [12]. Among peptide structures,  $\alpha$ -helices are the most common secondary configuration in proteins, crucial for protein stability, DNA-binding, and membrane traversal [8]. Despite their prominence, native helical peptides pose challenges as drug candidates because they are susceptible to proteolysis and exhibit fragile conformational stability without a protein scaffold [1]. To address these limitations, peptides composed of D-amino acids present a promising alternative. These D peptides can engage in specific heterochiral interactions with natural L-protein targets, offering improved stability and reduced immunogenicity due to their mirror-image topologies. [5, 9, 10].

Previous D-peptide computational design methods like structural similarity search and HelixDiff have their strengths—providing stable D-analogs and enabling *de novo* design, respectively [4, 14]. Utilizing these techniques, we have developed D-peptide analogs capable of activating the GLP-1,



**Figure 1: Conceptual overview of HelixFlow, hotspot-specific modules and pocketFlow modules.** **A)** Peptide Embedding Utilizing 3D Graphs and Overview of HelixFlow Architecture. This approach encodes full-atom peptides as 3D graphs, with edge features represented by relative positions  $r_{ij}$ . Node features encompass amino acid types, atom types, spatial coordinates, and masking information. The process commences by sampling from a Gaussian prior that jointly models sequence and structure, followed by iterative refinement using a neural Ordinary Differential Equation (ODE). At each iteration, the SE(3) model predictions guide the flow direction, with the SE(3) model being adapted from the Equiformer architecture. **B)** The hotspot-specific conditional module flowchart to generate peptides. The module is given both the peptide length and the sequence information for the target hotspots as shown. Then the hotspot-specific inpainting module generates the rest of the regions as shown in the inpainting matrix. This model directly generates L helices since the training data were L-peptides. **C)** Hotspot-specific inpainting module flowchart to generate D-peptides. The flowchart is similar to B) but with an extra step called the mirror transformation, which transforms the generated L-helices into D-peptides **D)** Pocket-specific inpainting module flowchart to generate binding peptides. The features are transformed into irreducible representations using spherical harmonic functions. Given the receptor only, the module would generate the binding peptides through a similar inpainting process.

PTH, and GLP-2 receptors [4, 13, 11, 14]. However, the structural similarity search is limited by the D-PDB database, and represents only a small fraction of possible stable helices. Conversely, HelixDiff suffers from slow inference speeds and is constrained by fixed-length design.

Flow matching models have recently shown their ability to provide results comparable to diffusion models at a lower computational expense, and they offer an innovative solution to mitigate stochastic variability and expedite inference in the generative modeling of images and molecules. In protein design, several research groups have updated their state-of-the-art models using flow matching, like transitioning from FrameDiff[16] to FrameFlow[15]. Leveraging these advancements, our work represents one of the first forays into peptide design utilizing flow matching.

We introduce HelixFlow, a SE(3)-equivariant flow matching model for generating flexible-length, all-atom helical peptide structures. It features an effective inpainting mechanism for one-shot  $\alpha$ -helical D peptide design that aligns targets with desired hotspot configurations. We further investigate the application of our model, referred to as the PocketFlow module, in designing peptide binders. HelixFlow outperforms HelixDiff in both unconditional generations and precise atomic-level conditional generation. As a proof of concept, we designed an acetylated D-peptide, D-INSL5Flow, targeting the relaxin family peptide receptor 4 (RXFP4) by matching critical hotspots in Insulin-like peptide 5 (INSL5), a hormone involved in metabolism and appetite regulation. While D-INSL5Flow exhibited lower potency (EC50 of 1.97  $\mu\text{M}$ ) compared to L-INSL5 (EC50 of 0.29  $\mu\text{M}$ ) in stimulating RXFP4-expressing cells, it showed comparable AKT phosphorylation levels and high resistance to protease degradation. These results suggest that HelixFlow could become a powerful tool for developing novel bioactive peptides with desired properties in early drug discovery.

## 2 Methods

We developed a flow-matching model named HelixFlow that can generate flexible, all-atom helical peptide structures of varying lengths. These peptides are represented using graph-like structures that incorporate a mixture of scalars and vectors to represent, where scalars include one-hot encoded

sequence data and atom types, and vectors are atom coordinates, as illustrated in Fig S1. To accurately capture all atomic details, we include all heavy atoms within each residue, resulting in a consistent representation of 14 atoms per residue with padding where necessary. To ensure equivariance to rotations and translations, we leverage the spherical harmonic functions to transform scalars and vectors into irreducible representations. We leverage the Equiformer architecture proposed by Liao et al. [6] as our flow prediction network, as depicted in Fig. 1A.

To be useful, we integrated a conditional hotspot-specific inpainting generation module, tailored specifically to the receptor of interest, as shown in Fig. 1B and C. Hotspot residues, defined as those crucial for target recognition, binding, and receptor activation, were first identified. Then, we constrained the structural generation of novel peptides to these hotspots, ensuring functional designs. The hotspot residue information provides contextual cues for the module to reconstruct the remaining data. This conditional generation process, focused on filling in the masked graph neural network matrix, resulted in more targeted and precise outputs. Consequently, we were able to generate a variety of realistic conformational rotamers based on hotspot residues and amino acid types, achieving superior conformation matching, particularly in D-peptide design settings.

Beyond the hotspot inpainting module, we are also exploring the use of our network for peptide binder design using the PocketFlow module (Fig. 1D). To achieve this, we collected peptide-receptor complex data from the Protein Data Bank (PDB). To properly represent the target chain, we included only key target residues which are defined as those within a 5Å distance of the binding peptide on the receptor. We then applied a similar inpainting process to generate peptide binders. However, unlike the hotspot-specific module that requires the known hotspots, this approach uses only the receptor as input to produce a complementary binding peptide.

Similar to other generative models like diffusion, flow matching learns to match the flow that transforms a given prior distribution  $p_0$  towards the data distribution  $p_1$  through a learned ODE that pushes the prior forward to data distribution by traveling from  $t_0$  to  $t_1$ . The fundamental idea is that given  $p_t$  as the probability path, the vector field  $v_t$  that generates  $p_t$  could be learned efficiently by decomposing the target probability path  $p_t$  as a mixture of tractable conditional probability paths  $p_t(x|x_1)$ ,  $t \sim U([0, 1])$ . The density is conditioned on a data point  $x_1 \sim p_{data}$  which interpolates between a prior distribution  $p_0(x|x_1) = q(x)$  and an approximate  $p_1(x|x_1) = \delta(x - x_1)$ . Given a conditional vector field  $u_t(x|x_1)$  that generates the time evolution of this conditional probability path, one then regresses against the marginal vector field with a neural network. At convergence, the learned vector  $v_\theta(x, t)$  is a neural ODE that evolves the prior distribution  $q(x)$  to the data distribution  $p_{data}(x)$ .

$$v_\theta(x, t) := E_{x_1 \sim p_t(x_1|x)}[u_t(x|x_1)] \quad (1)$$

More specific to our model, we used the Gaussian distribution as our prior, and we defined our conditional probability path as below:

$$p(x|x_1) = (1 - t)x_0 + tx_1, x_0 \sim q(x_0) \quad (2)$$

The loss is then computed by comparing the predicted vector field  $v_\theta(x, t)$  with the true conditional vector field  $u_t(x|x_1)$ , combining both sequence loss and structural loss. This loss is minimized using a neural network, which learns the vector field that transforms the prior distribution to the data distribution. The implementation also incorporates conditional generation by masking specific regions, such as the non-hotspot residues on the helical peptides, to condition the flow model on these substructures. This approach enables accurate inpainting of peptide sequences and atomic positions.

### 3 Results

#### Unconditional sampling analysis.

We used HelixFlow to unconditionally generate 2000 synthetic helical structures ranging from 14 to 20 amino acids. The generated data exhibited a similar range of physical structural features to the training data (Fig S3A-D). We also compared the results with HelixDiff. Notably, HelixFlow

has shown better sequence recovery and Rosetta scores compared with HelixDiff, indicating a better performance for  $\alpha$ -helix structure generation (Fig. S2).

### Hotspot-specific peptide generation analysis

We evaluated the performance of our hotspot-specific inpainting module on test sets, generating L-type or D-type helices that match the desired hotspot residues (Table 1). The evaluation involved calculating the RMSD between the target and generated helix structures via partial alignment of the hotspot and matched residue atoms. Impressively, 73% of the L-type 3-hotspot test cases showed RMSDs of less than 1 Å for the target hotspots. Additionally, 16.9% had RMSDs between 1 and 1.5 Å (Table 1). The D-type test is to use our model for designing D-peptides that mimic bioactive L-peptides. We applied a mirror conversion step to transform helices into generated D-peptides. Given a set of hotspots in a known L-peptide, our model conditionally generated novel D-helix peptide structures using the hotspot inpainting mechanism and mirror conversion. In 67.7% of the D-type 3-hotspot and 47.9% of the D-type 4-hotspot test cases, the RMSDs of matching residues were below 1.5 Å. HelixFlow outperforms HelixDiff in both D-type and L-type test sets.

Test	Hotspots	Method	rmsd (< 1Å)	rmsd (1 ~ 1.5Å)	rmsd (1.5 ~ 2Å)	rmsd (> 2Å)
L type	3	HelixFlow	73.0%	16.9%	10.1%	0%
L type	3	HelixDiff	58.0%	31.8%	5.7%	4.5%
D type	3	HelixFlow	14.6%	53.1%	29.2%	3.1%
D type	3	HelixDiff	9.4%	45.8%	41.7%	3.1%
D type	4	HelixFlow	3.1 %	44.8%	42.7%	9.4%
D type	4	HelixDiff	3.1 %	38.5%	45.8%	12.5%

Table 1: **Evaluation of the hotspot-specific conditional generation by HelixDiff and HelixFlow.** For each set of rows, the same test cases were compared between HelixDiff and HelixFlow. The lowest RMSD for each test case is calculated and summarized in the table. The hotspot columns indicate the randomly selected hotspot residues in the test cases, and the test columns show the type of the generated structures.

### Pocket-specific peptide generation analysis

We extended our helixFlow network into binder design problem called PocketFlow module. We perform the early ablation studies on two critical components to assess the performance: the structural deviation from the original binder structure and sequence similarities, as shown in table 2. We utilized the DiffAb [7] model and retrained on the same dataset as the control group. DiffAb [7] parameterized protein backbones like AlphaFold2, where the atomic positions within a residue are determined by a CA translation vector and a rotation matrix constructed from the positions of the N, CA, and C atoms. Noted, the structures generated by the PocketFlow module and DiffAb were post-optimized using OpenMM [3], a commonly used tool for energy minimization and internal structural optimization. This module is still under development, and we plan to conduct more comprehensive studies in the future.

Method	Seq identity (ARR)	RMSD
DiffAB*	7.95%	7.36
PocketFlow	16.42%	3.70

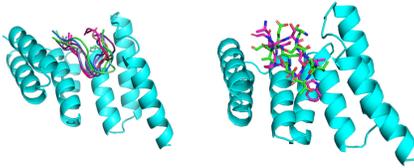


Table 2: **Evaluation of the generated peptide binders from the PocketFlow Module (left)** Each test set is evaluated by masking out the entire peptide and using two methods (DiffAb [7] and PocketFlow) to generate 96 binding peptides. The RMSD and ARR are calculated as the average value. DiffAb initially was a diffusion model for generating CDRs in antibodies targeted at antigens. We retrained the model using the same peptide-receptor complex dataset using it as a control group. *(right)* Example of Pocketflow generated sample (in green) aligned to the ground truth PDB structure (in red).

### De novo design of D-INSL5 agonists of the RXFP4 using HelixFlow

We focused on key residues of INSL5 chain B: R13, Y17, R23, and W24 to design a new D-peptide analog of INSL5 using HelixFlow. The ligand structure bound to full-length RXFP4 (PDB: 7yj4) was used as a starting point. Several retro-inverted D-helix structures were generated with HelixFlow. We evaluated the root mean square deviation (RMSD) between specific atoms in the starting helix and the generated D-helix structures. The best design, D-INSL5\_flow (RMSD = 1.44 Å), was chosen for further modeling (Fig. S4A left). We also designed D-INSL5\_mirror\_image\_search by scanning a D-PDB database with our method for converting L-peptides into stable D-peptide analogs. INSL5 was divided into two fragments, helix1 and helix2, with hotspots: R13, I16, Y17 (helix1) and S21, R23, W24 (helix2). Several D-helix structures were generated, and the best-matched peptides were joined to form D-INSL5\_mirror\_image\_search (Fig. S4A Right). Both D-INSL5\_flow and D-INSL5\_mirror\_image\_search were superimposed onto the Cryo-EM RXFP4-INSL5 structure to model RXFP4 complexes with D-peptide analogs (Fig. S4B). The RXFP4+D-peptide complexes were simulated in a POPC:PSM bilayer using 300 ns MD simulations, alongside a wild-type INSL5-RXFP4 complex for comparison. The D-peptides are retro-inverted relative to INSL5, leading to N-terminal embedding in RXFP4's transmembrane domain, unlike INSL5, which interacts with the receptor's extracellular domain. To prevent electrostatic repulsion with RXFP4's R208, the N-terminal residues of both D-peptides were acetylated. Both D-peptides stabilized near their initial positions according to their RMSD profiles, with D-INSL5\_flow showing the smallest RMSD (Fig. S5A). RMSF profiles indicated N-terminal stability of D-peptides was comparable to the C-terminal segment of INSL5 chain B (Fig. S5B). The structural superposition of representative clusters from the MD simulations revealed significant matching between D-analogs and INSL5 hotspots, with the most notable differences at R13, an exposed INSL5 hotspot (Fig. S5C-D).

#### **Experimental validation of the D-INSL5 peptide designed with HelixFlow**

To experimentally validate the designed D-peptides, the D-INSL5 peptides and L-INSL5 were chemically synthesized. We then conducted a circular dichroism analysis of the peptides in solution to determine the peptides' secondary structure. According to our findings, the two D-peptides and L-INSL5 all retain helical structures in solution (Fig. 2A). Following that, we constructed a RXFP4-expressing HEK293 cell line to test the ability of the L-INSL5 and D-INSL5-Flow design to activate the RXFP4. INSL5 binding to the RXFP4 has previously been reported to inhibit adenylyl cyclase to produce cyclic adenosine monophosphate (cAMP), thereby inhibiting protein kinase A (PKA) pathways. To assess the biological potency of L-INSL5 and the D-peptides, we used the HTRF cAMP assay with forskolin. L-INSL5 inhibited the forskolin-induced cAMP accumulation in RXFP4-expressing HEK293 cells with a half maximal effective concentration (EC50) value of 0.290  $\mu$ M. D-INSL5-flow and D-INSL5-mirror-image-search inhibited the cAMP production to a lesser extent, with an EC50 of 1.97  $\mu$ M and 3.2  $\mu$ M, respectively. Both D-INSL5 agonists stimulated the RXFP4 with efficacies ranging from 75.4 to 83.1 % compared to L-INSL5 activation (Fig. 2B).

We then investigated whether D-INSL5 analogs can induce the downstream signaling pathway activation through binding to RXFP4. The signal transduction pathways activated by INSL5 at RXFP4 include the activation of MAPK/ERK pathway, as well as AKT signaling pathway [2]. We determined whether stimulating the RXFP4 with the D-INSL5 analogs would induce ERK and AKT phosphorylation. Both D-INSL5 analogs triggered comparable p-ERK and p-AKT level to L-INSL5 in RXFP4-expressing HEK293 cells (Fig. 2C-E). Finally, we examined the protease resistance of both D-INSL5 analogs. In terms of therapeutic applications, D-peptides have the advantage of longer half-life in serum due to protease-resistance. L-INSL5 was almost completely degraded within 2 h, while over 85% of both D-INSL5 analogs can remain after 2.5 h incubation with proteinase K (Fig. 2F). Consequently, the D-INSL5 agonists exhibited high protease resistance, thereby expecting a longer half-life and higher therapeutic potency.

## **4 Conclusion**

In this work, we developed HelixFlow, a flow-matching SE(3)-equivariant model, capable of generating all-atom  $\alpha$ -helix peptide structures and enabling conditional peptide design with matching hotspots. The effectiveness of direct conditional design is a crucial feature of HelixFlow, leveraging ordinary differential equations for smoother data-to-noise transformations and more precise synthesis through a reverse flow-matching process. This study highlights the capability of flow-matching generative models, such as HelixFlow, as peptide design tools, demonstrating their capacity to outperform classic diffusion-based approaches. Our findings illustrate the benefits of using a flow-matching

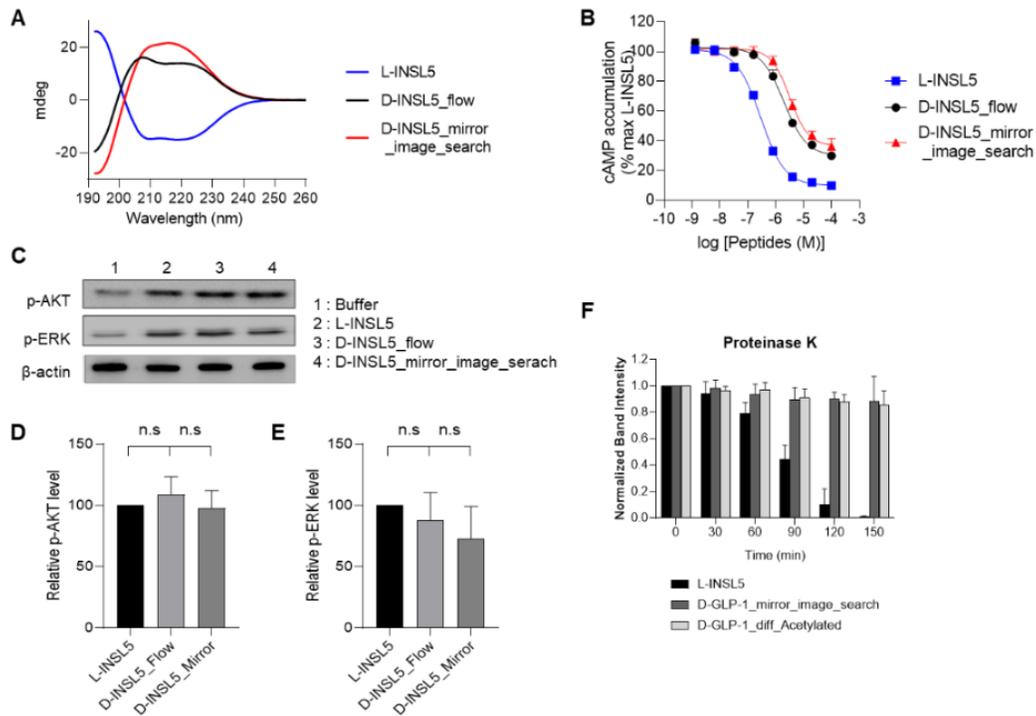


Figure 2: **Experimental validation of D-INSL5 agonists designed with HelixFlow.** A) Circular dichroism measurements of the D-INSL5 designs and L-INSL5 in solution. B) Activity profile of L-INSL5 and D-INSL5 peptides in HEK293 cells stably expressing RXFP4. C) Western blots showing the p-ERK and p-AKT levels induced by the D-INSL5 analogs and the native L-INSL5 at 20  $\mu$ M concentration. D-E) Quantification of the p-ERK levels (D) and p-AKT levels (E) induced by D-INSL5 agonists relative to L-INSL5 treated cells after 4 h of incubation. n.s., not significant. F) Quantification of the remaining D-INSL5 analogs and L-INSL5 after Proteinase K treatment in 30 min intervals. Intensities of peptide bands were normalized to the intensity of the untreated peptide ( $T_0$ ).

SE(3)-equivariant model, including enhanced sample generation quality, a robust conditional generation pipeline, and greater stability in constructing D-peptide analogs. Future work would involve creating an algorithm capable of designing new peptides that can interact with proteins in a controlled manner, enhancing their bioactivity.

## References

- [1] A. M. Ali, J. Atmaj, N. Van Oosterwijk, M. R. Groves, and A. Dömling. Stapled peptides inhibitors: a new window for target drug discovery. *Computational and structural biotechnology journal*, 17:263–281, 2019.
- [2] S. Y. Ang, D. S. Hutchinson, N. Patil, B. A. Evans, R. A. Bathgate, M. L. Halls, M. A. Hossain, R. J. Summers, and M. Kocan. Signal transduction pathways activated by insulin-like peptide 5 at the relaxin family peptide rxfp4 receptor. *British journal of pharmacology*, 174(10): 1077–1089, 2017.
- [3] P. Eastman, J. Swails, J. D. Chodera, R. T. McGibbon, Y. Zhao, K. A. Beauchamp, L.-P. Wang, A. C. Simmonett, M. P. Harrigan, C. D. Stern, et al. Openmm 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS computational biology*, 13(7):e1005659, 2017.
- [4] M. Garton, S. Nim, T. A. Stone, K. E. Wang, C. M. Deber, and P. M. Kim. Method to generate highly stable d-amino acid analogs of bioactive helical peptides using a mirror image of the entire pdb. *Proceedings of the National Academy of Sciences*, 115(7):1505–1510, 2018.

- [5] G. Kreil. D-amino acids in animal peptides. *Annual review of biochemistry*, 66(1):337–345, 1997.
- [6] Y.-L. Liao and T. Smidt. Equiformer: Equivariant graph attention transformer for 3d atomistic graphs. *arXiv preprint arXiv:2206.11990*, 2022.
- [7] S. Luo, Y. Su, X. Peng, S. Wang, J. Peng, and J. Ma. Antigen-specific antibody design and optimization with diffusion-based generative models for protein structures. In *Advances in Neural Information Processing Systems*, 2022.
- [8] E. Petsalaki and R. B. Russell. Peptide-mediated interactions in biological systems: new discoveries and applications. *Current opinion in biotechnology*, 19(4):344–350, 2008.
- [9] A. E. Rabideau and B. L. Pentelute. A d-amino acid at the n-terminus of a protein abrogates its degradation by the n-end rule pathway. *ACS central science*, 1(8):423–430, 2015.
- [10] M. Uppalapati, D. J. Lee, K. Mandal, H. Li, L. P. Miranda, J. Lowitz, J. Kenney, J. J. Adams, D. Ault-Riché, S. B. Kent, et al. A potent d-protein antagonist of vegf-a is nonimmunogenic, metabolically stable, and longer-circulating in vivo. *ACS chemical biology*, 11(4):1058–1065, 2016.
- [11] P. A. Valiente, S. Nim, J. Kim, and P. M. Kim. Computational design of potent and selective d-peptide agonists of the glucagon-like peptide-2 receptor. *Journal of Medicinal Chemistry*, 66(15):10342–10353, 2023.
- [12] L. Wang, N. Wang, W. Zhang, X. Cheng, Z. Yan, G. Shao, X. Wang, R. Wang, and C. Fu. Therapeutic peptides: Current applications and future directions. *Signal Transduction and Targeted Therapy*, 7(1):48, 2022.
- [13] X. Xie, P. A. Valiente, and P. M. Kim. Helixgan a deep-learning methodology for conditional de novo design of  $\alpha$ -helix structures. *Bioinformatics*, 39(1):btad036, 2023.
- [14] X. Xie<sup>12</sup>, P. A. Valiente, J. Kim, and P. M. Kim<sup>123</sup>. Helixdiff: Hotspot-specific full-atom design of peptides using diffusion models.
- [15] J. Yim, A. Campbell, A. Y. Foong, M. Gastegger, J. Jiménez-Luna, S. Lewis, V. G. Satorras, B. S. Veeling, R. Barzilay, T. Jaakkola, et al. Fast protein backbone generation with se (3) flow matching. *arXiv preprint arXiv:2310.05297*, 2023.
- [16] J. Yim, B. L. Trippe, V. De Bortoli, E. Mathieu, A. Doucet, R. Barzilay, and T. Jaakkola. Se (3) diffusion model with application to protein backbone generation. *arXiv preprint arXiv:2302.02277*, 2023.

## Supplementary Methods

### Design strategy to generate the D-GLP-1 agonist

We superimposed D-INSL5\_flow and D-INSL5\_mirror\_image\_search structures onto the Cryo\_EM structure of RXFP4 bound to INSL5 to build the 3D structure of the RXFP4 in complex with both D-peptide analogs (Fig. S4B). The RXFP4+D-peptide complexes were then embedded in a POPC: PSM (1:1) bilayer before evaluating its binding mode stability using 300 ns MD simulations (Fig. S4B). We also simulated the RXFP4 bound to the wild-type INSL5 as a control. The D-peptides analog designed in the current work are retro-inverted compared to INSL5. This difference in the direction means that D-INL5 peptides have their N-terminal residue embedded in the transmembrane domain of the RXFP4, while INSL5, its chain B C-terminal residue interacts with the receptor's extracellular domain. The role of R208 in the TM core of RXFP4 to stabilize the C-terminal negative charge of INSL5 chain B has been previously reported. Then, we acetylated the N-terminal residue of both D-peptides to prevent any electrostatic repulsion of the D-peptides N-terminal residues with R208.

### Circular dichroism (CD) analysis

Peptide secondary structure determination was carried out using a Jasco J-810 spectropolarimeter. Lyophilized peptide powders were dissolved in 20 % acetic acid (v/v) for D-INSL5 peptides and 20 % acetic acid (v/v) with 0.5 mM DTT for L-INSL5. Peptide concentrations were 50  $\mu$ M for L-INSL5, D-INSL5\_Flow, and D-INSL5\_Mirror. Samples were read using a 0.1-cm cuvette path length with ten accumulations per run, 50 nm/min scanning speed. All spectra were background subtracted.

### cAMP inhibition assay

Inhibition of forskolin-induced cAMP accumulation by L-INSL5 peptide and D-INSL5 peptides, was measured by a HTRF cAMP Gs Dynamic kit (Revity) following manufacturer's protocol. Briefly, HEK293T cells expressing RXFP4 were trypsinized from subconfluent culture and seeded in a 96-well low-volume plate at a density of 2,000 cells per well. The cells were stimulated with different concentrations of peptides plus 1.5  $\mu$ M forskolin. After 2 h of incubation at 37  $^{\circ}$ C, cAMP d2 reagent,

and cAMP Eu-Cryptate antibodies were added to each well. After incubation at room temperature for 30 min in the dark, the plate was read using a Synergy 2 plate reader (BioTek) with excitation at 330 nm and emission at 620 nm and 665 nm. Data were normalized to the value of forskolin-treated sample and were used to calculate the EC<sub>50</sub> value by fitting it to a non-linear sigmoidal curve using GraphPad Prism 8.

### **Western blot**

After cell starvation (0 % FBS, 6 h), HEK293 cells expressing RXFP4 were treated 20  $\mu$ M of L or D-INSL5 peptides for 4 h for ERK and AKT pathway activation. Cells were lysed with lysis buffer (10 mM Tris-HCl pH 7.4, 1% SDS, 100 mM NaCl, 1 mM EDTA, 1 $\times$  protease inhibitor mixture (Sigma)) for 30 min at 4  $^{\circ}$ C. Protein samples were separated on a SDS-PAGE gel and transferred to PVDF membranes. Transferred samples were immunoblotted with primary antibodies, followed by incubation with horseradish peroxidase-conjugated secondary antibodies (Cell Signaling) and detected using enhanced chemiluminescence (Invitrogen). For quantification, band intensities were quantified using ImageJ software<sup>38</sup> and normalized to the  $\beta$ -actin loading control values. Relative band intensity was presented as a ratio compared to the value of the L-INSL5.

### **Protease stability assays**

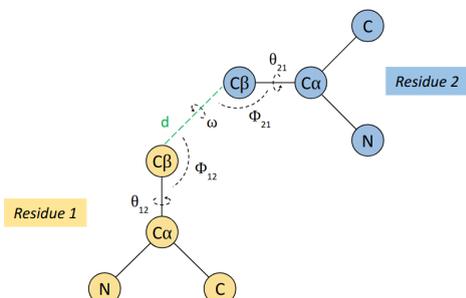
For protease stability assay, 50  $\mu$ M of peptide were diluted in 80  $\mu$ l of reaction buffer (10 mM Tris-HCl, 10 mM NaCl, pH 7.4, 5  $\mu$ M CaCl<sub>2</sub>), and 12  $\mu$ l were removed for the un-treated T0 sample. Proteinase K (Bioshop) was then added to a final 100  $\mu$ g/ml concentration. Samples were incubated at 37  $^{\circ}$ C, and 30  $\mu$ l was removed after each time point, and protease activity was blocked by the addition of 10 mM PMSF. Protease-inactivated samples were frozen at -20  $^{\circ}$ C until further use. Frozen samples were analyzed by SDS-PAGE. Gels were stained using Coomassie Brilliant Blue dye. The densitometry of bands was determined using ImageJ software. All samples were normalized to their respective untreated sample (T0).

### **Statistical analysis**

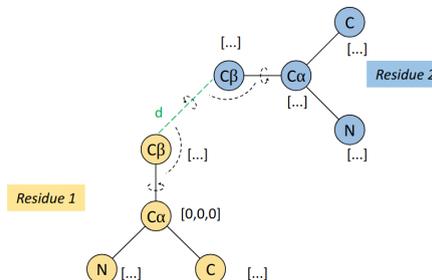
Statistical significance was analyzed by a two-tailed unpaired Student's t-test using MS Excel. A *P* value of less than 0.05 was considered statistically significant.

## Supplementary Figures.

A. Invariant



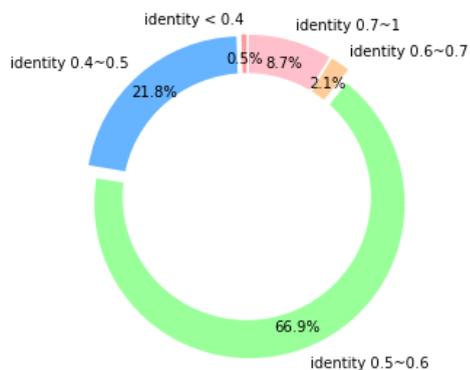
B. Equivariant



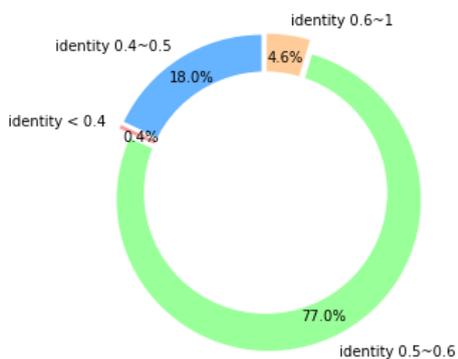
**Figure S1. Graphical scheme of the features encoded in HelixDiff (invariant) and HelixFlow (equivariant) to generate novel helices.**

Helices are encoded as the combined vectors of the primary sequence information using one-hot encoding and structural information A) Invariant Features. The structural information is represented as phi ( $\phi$ ), psi ( $\psi$ ), omega ( $\omega$ ), and bond angles, while the side chains are defined using the five chi ( $\chi$ ) angles. B). Equivariant features. The encoding process involves categorizing different tensor types (scalars and vectors) based on their behavior under rotation, using spherical harmonics that transform into irreducible representations. Features can be mixed with scalars and vectors. Scalars, also called L0 vectors, are referred to the invariant features like atom types, and residues types. Vectors refer to the atom coordinates in our model, which are L1 vectors. The L0 and L1 vectors represent angular frequency under rotation.

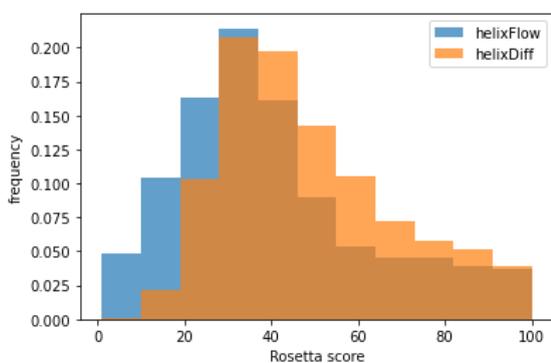
A.



B.



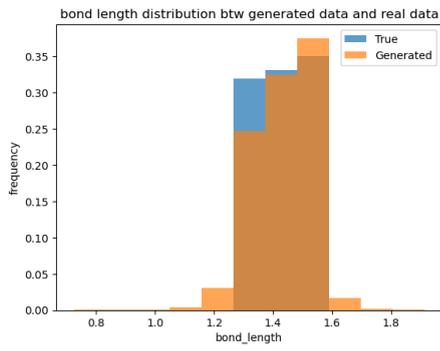
C.



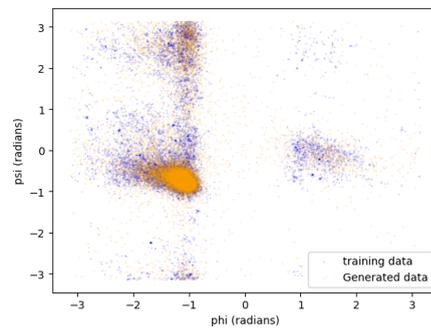
**Figure S2 Sequence identity of the helical structures at fixed length 14 amino acids generated using HelixDiff and HelixFlow, as well as Rosetta score.**

**A)** HelixDiff. **B)** HelixFlow. Helices with 14 amino acids were randomly generated and compared to the nearest natural helices from the training data to calculate sequence identity for two algorithms. HelixDiff was only applied to fixed-length helices. **C)** Rosetta distribution regarding fixed length 14 generations by HelixFlow and HelixDiff

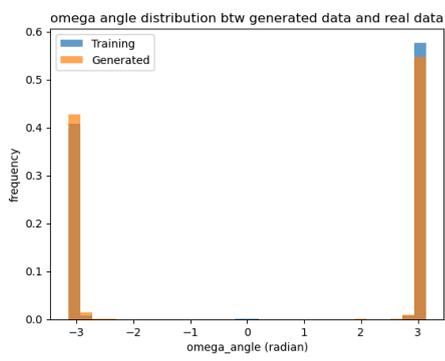
A.



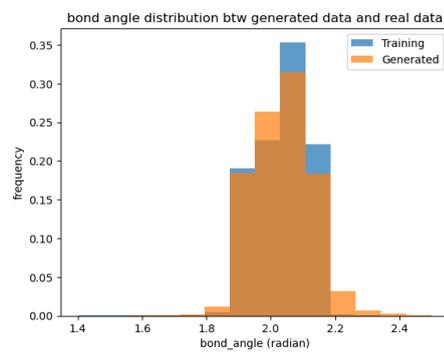
B.



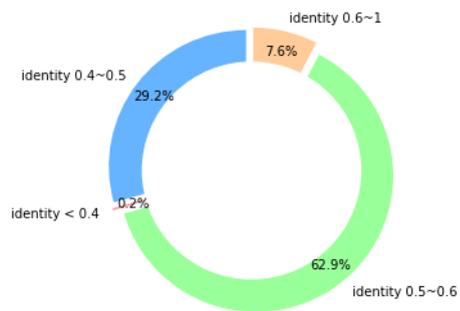
C.



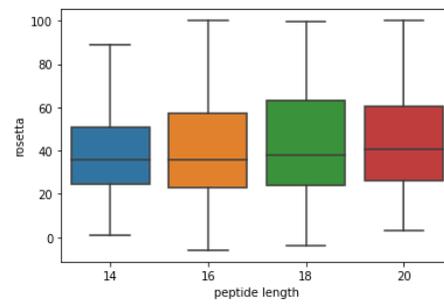
D.



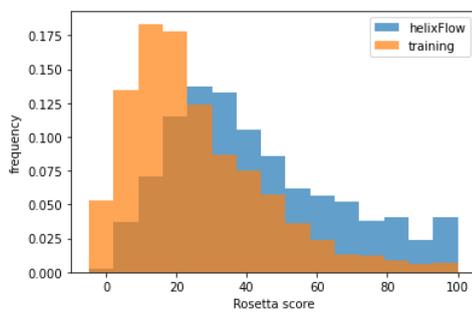
E.



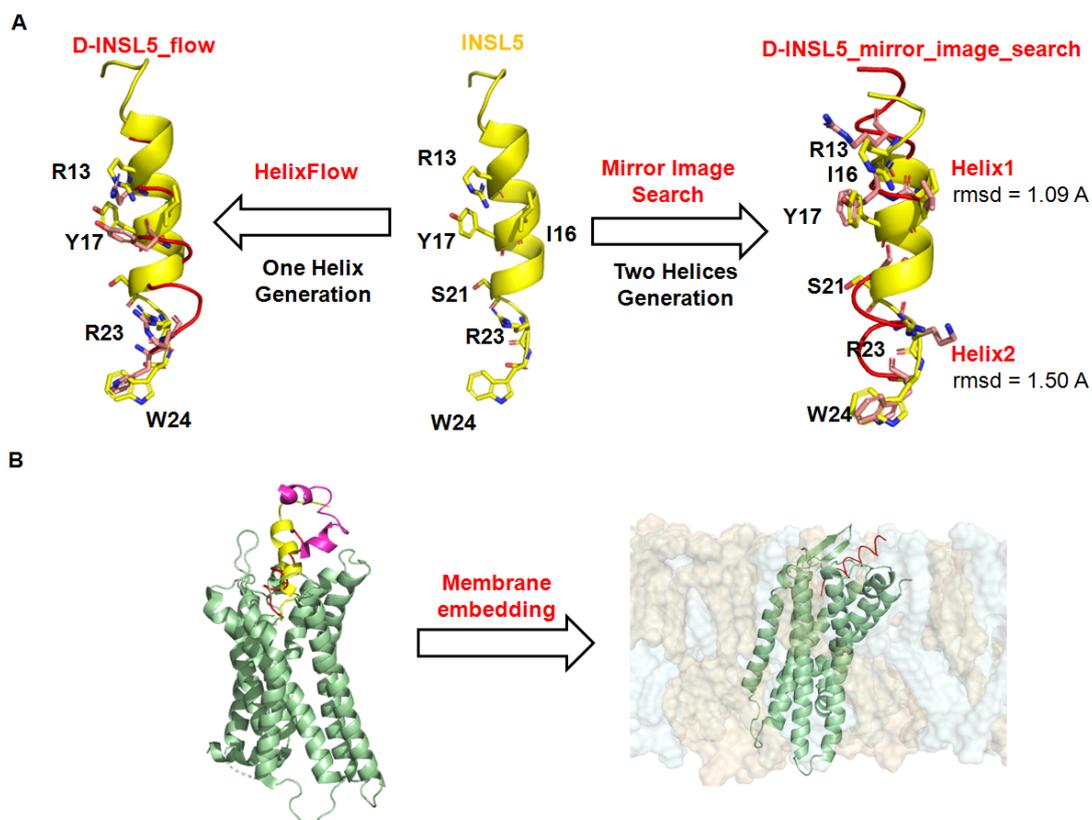
F.



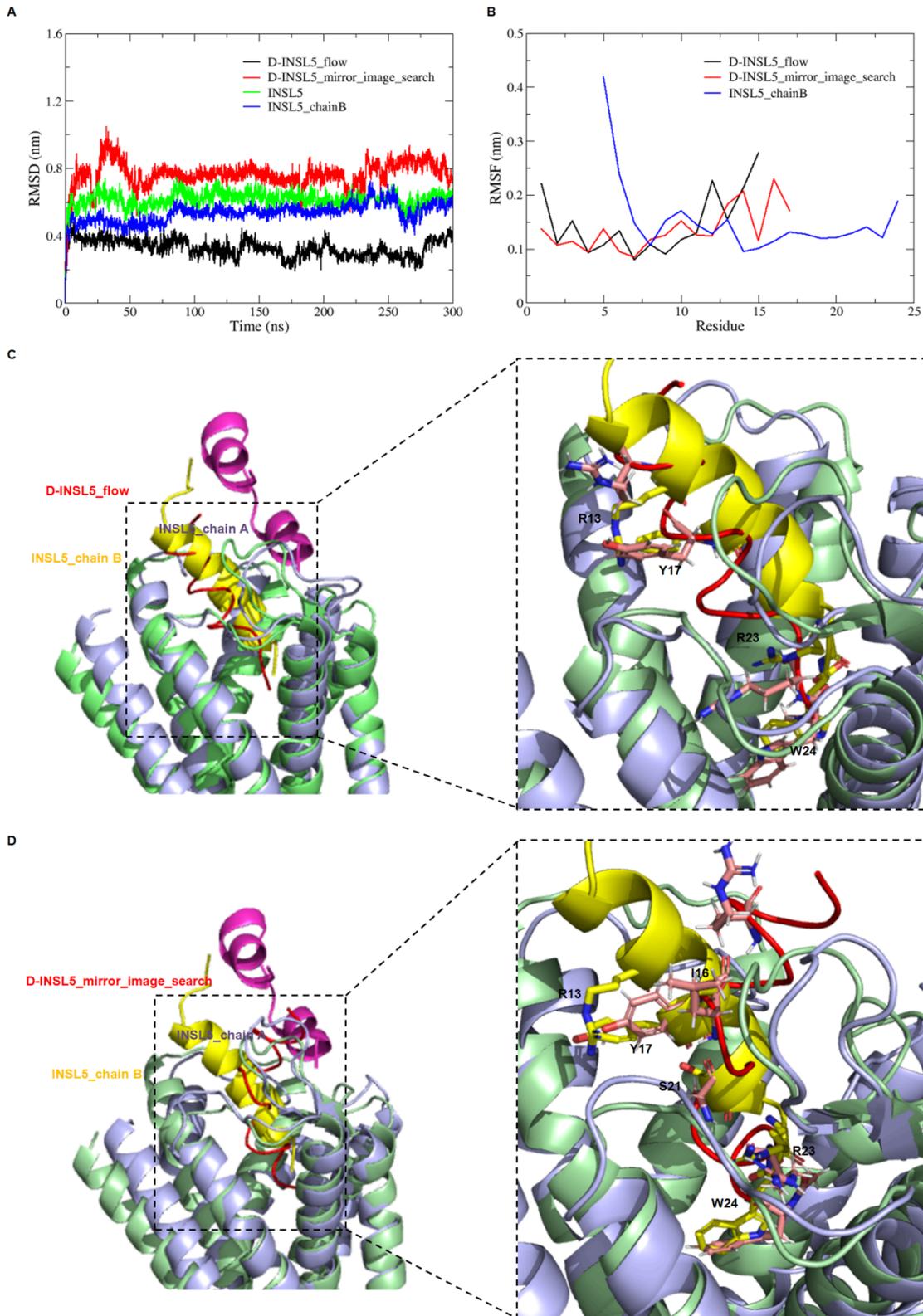
G.



**Figure S3. Assessment of the novel helices generated with HelixFlow at random lengths.** A-D) Structural features regarding bond length, psi, phi, omega, and bond angles between 2k generated samples and training data. E) Sequence identities compared with training data. F) Rosetta score distribution by generated peptide length using HelixFlow. G) Comparing the rosetta scores distribution of the training data and the ones generated HelixFlow.



**Figure S4. Design of the D-INSL5 analogs.** **A)** Strategy for design two novel D-INSL5 analogs using HelixFlow (left) and the mirror image search of a D-PDB database (right). For the mirror image search design, the INSL5 structure was divided into two overlapping peptides named helix1, and helix2. Helix1 extends from E10 to A20, while helix2 runs from I16 to W24. Critical hotspots for the INSL5 function chosen for the design were colored as licorice. **B)** Structural superposition of D-INSL5\_flow analog (red) over the INSL5 chain B (yellow) structure bound to the RXFP4 (light green) (pdb code: 7yju). In the right, it is shown the RXFP4 coupled to D-INSL5\_flow embedded in a POPC:PSM (1:1) membrane. The orange (POPC) and gray (PSM) surface represent the lipids structure.



**Figure S5. Design and modeling the 3D structure of D-GLP-1\_diff analogs bound to the GLP-1 receptor.** **A)** Root mean square deviation (RMSD) of the heavy atoms of INSL5 and the D-INSL5 analogs bound to the RXFP4. **B)** Root mean square fluctuation (RMSF) per residue of the heavy atoms of INSL5 and the D-INSL5 analogs bound to the RXFP4. **C)** Structural superposition of the most representative cluster extracted from the MD simulation of RXFP4 bound to D-INSL5\_flow over the most representative structure extracted from the MD

simulation of the INSL5 in complex with RXFP4. **D)** Structural superposition of the most representative cluster extracted from the MD simulation of RXFP4 bound to D-INSL5\_mirror\_image\_search over the most representative structure extracted from the MD simulation of the INSL5 in complex with RXFP4. We showed as licorice the hotspots and matching residues (red) in the L (yellow) and D-peptide (red), respectively. In the hotspots and matching residues, the nitrogen and oxygen atoms were colored in blue and red, respectively.