Computers in Biology and Medicine 199 (2025) 111286

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Target-aware latent diffusion model for design of apoptosis-inducing anticancer peptides

Tiara Natasha Binte Sayuti [a], Kakuly Mittal [a], Tan Lai Heng [b], Jagath C. Rajapakse [a,*]

[a] College of Computing and Data Science, Nanyang Technological University, 50 Nanyang Ave, 639798, Singapore
[b] Collaborative Initiative, Interdisciplinary Graduate Programme, Nanyang Technological University, 61 Nanyang Dr, 637335, Singapore

ARTICLE INFO

Keywords:
Anticancer peptides
Diffusion models
Generative AI
Graph neural networks
Protein language model

ABSTRACT

Therapeutic peptides offer high specificity and low toxicity, but rational design remains difficult due to their vast chemical diversity and conformational flexibility. In oncology, B-cell lymphoma-extra large (BCL-xL), a key anti-apoptotic protein, is overexpressed in various cancers, making it an important target for peptide-based therapeutics. Existing generative models for anticancer peptide design often neglect detailed receptor structural context, limiting biological plausibility. We present a target-aware latent diffusion framework (T-LDM) that conditions sequence denoising on receptor pocket context encoded by a graph-based structural model. Conditioning is fused into the U-Net denoiser via lightweight modulation and cross attention, enabling the Latent Diffusion Model to align sampling with the local interaction topology of the binding groove while preserving protein language model priors. We evaluate distributional calibration with Jensen–Shannon divergence across key physicochemical properties, sequence quality with perplexity, motif overlap with BLEU, and structural compatibility via protein–peptide docking. A controlled ablation was performed that varies the scope of structural conditioning (none, global, pocket) and the fusion into the denoiser, letting us assess the marginal effect of pocket-aware guidance and fusion on generation quality. Overall, the study indicates that explicit pocket conditioning yields peptides that are syntactically fluent, physicochemically realistic, and more compatible with the intended BCL-xL target, while maintaining diversity and novelty. Source code is available at https://github.com/tiaranatashasayuti/T-LDM.

1. Introduction

Therapeutic peptides offer high specificity and low toxicity compared with small molecules or biologics. Anticancer peptides (ACPs), typically 5–50 amino acids long, induce cancer cell death through mechanisms such as apoptosis, membrane disruption, and inhibition of angiogenesis and proliferation [1–3]. Certain ACPs, such as AC-P19M, selectively trigger apoptosis in cancer cells while sparing normal tissue [2].

Among key therapeutic targets, the anti-apoptotic protein B-cell lymphoma-extra large (BCL-xL) plays a central role in tumor survival by binding and neutralizing pro-apoptotic factors of the BCL-2 family. Its overexpression, observed in multiple malignancies including lung and hematologic cancers, promotes chemoresistance and survival under stress [4].

However, existing peptide design approaches remain limited by their reliance on either sequence statistics or global physicochemical priors, without explicit integration of receptor structural context. Consequently, generating peptides with realistic BCL-xL-binding conformations remains a key challenge. In this study, we investigate whether incorporating receptor-derived structural constraints into a generative diffusion framework can yield peptides that are both syntactically coherent and structurally compatible with the BH3 groove of BCL-xL.

We focus primarily on BCL-xL due to its well-characterized structure, abundance of high-resolution complexes in the Apoptosis-Inducing Anticancer Peptide Database (ApInAPDB) [5], and its established role in chemoresistance. We further test model generalization across the BCL-2 family to assess whether conditioning on one receptor enables transferable binding patterns, establishing BCL-xL as both a biologically meaningful case study and a benchmark for cross-family selectivity.

Recent advances in deep generative modeling have markedly improved rational peptide design and addressed the challenges posed by the vast search space of valid sequences and conformations. However, most methods remain purely sequence-based or use simplified structural descriptors that overlook local interaction topology. To bridge this gap, we integrate diffusion-based generation with graph-based structural encoding, enabling direct three-dimensional receptor guidance during peptide generation.

* Corresponding author.
E-mail address: asjagath@ntu.edu.sg (J.C. Rajapakse).

https://doi.org/10.1016/j.compbiomed.2025.111286
Received 19 August 2025; Received in revised form 30 October 2025; Accepted 7 November 2025
Available online 18 November 2025

Latent diffusion models (LDMs) [6] have recently shown strong potential for modeling complex biomolecular distributions through iterative noise refinement. To provide biologically meaningful conditioning signals for such generative modeling, we integrate representations from both sequence and structural domains. For peptide sequences, we use ESM-2, a protein language model (PLM) trained on large datasets such as UniRef50 [7]. PLMs are used to embed amino-acid sequences into compact, information-rich representations that capture evolutionary and physicochemical dependencies across residues, offering transferable priors particularly valuable when experimentally verified data are scarce. We further fine-tune ESM-2 on BioLiP2 [8] and ApInAPDB [5] via masked language modeling (MLM) [9] to adapt it to peptide–receptor interaction contexts.

To encode structural information, we employ a graph attention network (GAT) [10]. GNNs such as GAT captures spatial and topographical information of molecular graphs. Compared with convolutional or message-passing GNNs, the attention mechanism allows the model to emphasize functionally important residues, which is critical for accurately representing binding pockets. The GAT operates with an 8 Å cutoff, balancing local chemical specificity with sufficient spatial coverage of the binding groove. By coupling PLM-derived sequence embeddings with GAT-based structural embeddings, the LDM can learn coordinated biochemical and spatial regularities, improving the realism and interpretability of generated peptide–receptor complexes.

These representations are fused in a Target-Aware Latent Diffusion Model (T-LDM) via a Modulated Prompt Network (MPN) that injects pocket embeddings as lightweight scale–shift prompts into the cross attention block of the denoising U-Net model. This soft modulation allows structural cues to influence denoising without overwriting pretrained sequence priors, achieving stable and parameter-efficient conditioning.

Generated peptides are reverse-folded with AlphaFold3 [11] and docked with HADDOCK3 [12] to assess structural fidelity and receptor compatibility. Cross-target docking on BCL-2 and MCL-1 further evaluates selectivity and transferability within the BCL-2 family. By coupling learned sequence priors with explicit structural cues, T-LDM produces peptides that are syntactically fluent and structurally compatible with receptor interfaces, offering a principled framework for receptor-guided peptide design in oncology and beyond.

To summarize, the major contributions of this study are as follows:

- **Target-aware latent diffusion model.** The model integrates receptor pocket embeddings via modulated prompts to guide peptide denoising under structural constraint in the latent diffusion model.
- **Graph-based pocket encoder.** A GAT trained at 8 Å predicts binding residues and provides compact, chemically informed pocket embeddings to use as constraints.
- **Prompt-based structural modulation.** The MPN injects pocket context as per-layer scale–shift prompts, preserving sequence priors while enabling controllable structural conditioning.
- **Structure-based validation.** Generated peptides are folded and docked to BCL-xL, and cross-evaluated on BCL-2 and MCL-1 to test selectivity and generalization.

2. Related work and motivation

Computational peptide therapeutics have advanced through multiple stages of development, each addressing different aspects of the design challenge. These can be broadly grouped into early experimental and heuristic frameworks, structure-based modeling, machine learning for sequence generation, and recent diffusion-based approaches that integrate structural awareness.

2.1. Early experimental and heuristic design

Early peptide discovery was driven by experimental library selections such as phage display, which enabled iterative panning of combinatorial peptide libraries against immobilized or in vivo targets [13–15]. Despite their effectiveness, phage display systems were limited by compositional bias and amplification artifacts, motivating refinements such as pattern enrichment and negative selection to improve hit fidelity [16,17].

In parallel, heuristic and rule-based computational frameworks emerged. These relied on handcrafted physicochemical descriptors and sequence motifs to guide de novo design. Tools such as AMP-Designer employed position-specific scoring and sequence-moment descriptors for antimicrobial peptide generation [18]. The D-descriptor and 3D-QSAR methods identified key descriptor sets linked to organism-specific selectivity and substrate specificity, successfully applied to SARS-CoV protease substrates and MHC epitopes [19]. While these approaches enabled interpretable and rapid in silico prioritization, their predictive power was constrained by dataset bias and limited transferability.

2.2. Structure-based modeling

The introduction of structural bioinformatics enabled structure-guided peptide optimization. Homology modeling and molecular docking helped identify peptide–receptor interactions that could improve stability or binding affinity. Methods such as HomoSAR combined comparative modeling with quantitative structure–activity relationships (QSAR) to enable cross-class optimization [20]. However, the effectiveness of these methods depends strongly on the availability of high-quality crystal or Nuclear Magnetic Resonance (NMR) structures, which restricts their generalization across diverse targets.

2.3. Machine learning for sequence design

The expansion of large-scale sequence repositories and structural datasets facilitated the adoption of machine learning in peptide design. Early models used motif mining, physicochemical descriptors, and classifiers such as support vector machines and random forests [21,22]. These methods achieved moderate predictive accuracy but struggled to capture higher-order dependencies and spatial context [23,24].

Deep learning introduced end-to-end feature extraction through convolutional and recurrent networks [25]. These architectures improved peptide sequence learning but lacked explicit structural awareness [26]. Generative models such as variational autoencoders (VAE) and Wasserstein autoencoders (WAE) learned smooth latent manifolds for peptide generation [27–29], yet often suffered from posterior collapse and oversimplified sequence patterns [30]. Autoregressive architectures including transformers improved sequence fluency but exhibited exposure bias and limited long-range coherence [31,32].

2.4. Diffusion and structure-conditioned generative models

Recent progress in latent diffusion modeling has provided a new direction for learning complex biomolecular distributions through iterative denoising [33,34]. Diffusion-based approaches achieve high generative coherence and controllability compared with autoregressive networks [35]. In peptide design, Diff-AMP integrated kinetic diffusion with attention mechanisms and reinforcement learning for controllable antimicrobial peptide generation [36]. SQ-DiffuPep incorporated both sequence and physicochemical property information for multimodal conditioning, improving controllability and diversity [37]. HMAMP further extended diffusion-based design to multiobjective optimization, enabling balanced control over conflicting physicochemical properties [38]. These studies illustrate the growing power of diffusion models in peptide generation, although they focus primarily on antimicrobial applications rather than apoptosis-inducing anticancer peptides.

Despite the success of generative modeling, incorporating explicit three-dimensional receptor context remains rare. Structural conditioning can enhance binding specificity, stability, and biological realism [39–41]. Graph neural networks such as Graph Convolutional Networks (GCN) and GAT provide natural representations for molecular topology and have been applied successfully in structure-informed modeling [10,42–44]. However, most existing peptide generators neglect detailed spatial constraints and rely instead on coarse descriptors or residue-level distance maps.

Multimodal frameworks such as HelixDiff, ProteinMPNN, and our earlier diffusion-based model explored integration of physicochemical properties and sequence. HelixDiff employs score-based denoising to generate all-atom helical peptides with sub-angstrom precision [45]. ProteinMPNN applies message passing to recover sequences from backbone structures with improved accuracy [46]. Our previous diffusion model conditioned generation on global physicochemical features rather than receptor structure [47]. For consistency and ease of reference, we refer to this prior diffusion model as the Physicochemical properties Conditioned–Latent Diffusion Model, abbreviated as PC-LDM, although this was not its original name. Together, these frameworks represent the current state of sequence–structure generative modeling.

2.5. Motivation

Building upon these foundations, we introduce the Target-Aware Latent Diffusion Model for the design of apoptosis-inducing anticancer peptides targeting BCL-xL. The model integrates receptor pocket embeddings derived from graph neural networks directly into the diffusion denoiser to unify structural guidance, physicochemical plausibility, and controllable sequence generation. This approach bridges the gap between sequence-based and structure-conditioned design and is benchmarked against representative generative baselines including VAE, WAE, HelixDiff, ProteinMPNN, and PC-LDM.

3. Methods

3.1. Datasets and preparation

We employed two complementary datasets to construct and train the generative framework: BioLiP2, which supplies structurally diverse peptide-protein complexes for pretraining, and ApInAPDB, which furnishes experimentally validated apoptosis-inducing peptides targeting BCL-xL for downstream adaptation.

BioLiP2 was used for pretraining of the protein language model, conditioned decoder and the target-aware GAT, thereby providing broad coverage of structural motifs across diverse peptide–protein complexes. In contrast, the ApInAPDB dataset served as a target-specific fine-tuning set, aligning the pretrained model towards anticancer peptides that specifically bind to the BH3 groove of BCL-xL. This domain-adaptation strategy enabled the model to inherit generalizable priors from BioLiP2 while specializing on the target of interest.

From the BioLiP2 repository, we extracted only peptide–protein co-crystal structures and applied a multi-stage filtering pipeline. Complexes were retained only if the crystallographic resolution was 5.0 Å or better (i.e., numerically $\leq$ 5.0 Å), ensuring that only high-quality structures with reliable atomic detail were included, while lower-resolution entries (>5.0 Å) were excluded. On top of that, only peptide contained at least five residues are retained. Receptor and ligand chains were normalized to remove ambiguous amino acids, failed DSSP assignments, or incomplete chains. Terminal residues were trimmed to reduce noise from poorly resolved ends, and duplicate entries were removed by enforcing one receptor–ligand pair per unique PDB ID. Binding-site residues were annotated as residues within an 8 Å heavy-atom cutoff between the ligand and receptor. These residues are labeled as binding while all remaining residues were labeled as non-binding.

After preprocessing, the number of complexes reduced from 93,389 to 8040. To prevent both homologous leakage and pocket memorization, we applied two additional filtering steps. First, we excluded all complexes containing BCL-xL, BCL-2, or MCL-1 receptors to ensure the structural conditioning network would not have prior exposure to the target binding pockets used in evaluation. Second, we constructed a peptide-identity graph at 75% global identity among the remaining complexes, extracted its connected components, and assigned entire components to the train/validation/test splits in a 70:15:15 proportion. The resulting partitions contain 5631/1209/1200 complexes drawn from 2900/1209/1200 clusters, respectively. This dual filtering strategy caps maximum peptide identity across partitions at 75% while preventing pocket memorization of the evaluation targets.

The ApInAPDB database contains 1470 unique anticancer peptide sequences that binds to various binding target such as BCL-xL, BCL-2 and MCL-1. We selected all entries annotated with BCL-xL as a binding target which resulted to 620 peptides. As structural templates are unavailable, each peptide was first aligned to the BH3 reference ligand in the BCL-xL structure (PDB ID: 4QVE) [48] using PyMOL and then docked into the canonical groove with HADDOCK3 [12,49].

The BID-bound 4QVE structure was selected as the native reference because it represents the canonical BH3 engagement of human BCL-xL with high-resolution geometry and well-defined side-chain orientations. In this complex, the BID BH3 helix binds within the classical $\alpha2$–$\alpha5$ groove and induces an expanded adjacent pocket of about 400 $Å^2$ area and 578 $Å^3$ volume, the largest among known BCL-xL–BH3 complexes. This expansion, governed by the open orientation of Glu96 and Tyr195 and the outward displacement of Phe105, defines a physiologically relevant conformation for BH3-mediated recognition and thus serves as the most representative binding geometry for apoptotic peptides.

All peptides were first structurally aligned to the reference BH3 ligand from the human BCL-xL–BID complex (PDB ID: 4QVE) using PyMOL to ensure consistent orientation within the canonical BH3 groove. Docking was then carried out in HADDOCK3 under a flexible refinement protocol guided by 5 Å ambiguous interaction restraints (AIRs) defined from interfacial proximity. Both center-of-mass and surface contact restraints were applied to maintain realistic receptor–ligand geometry during sampling. The lowest-energy structure from the final refinement stage was retained for downstream analysis. Binding-site residues were annotated, designating as pocket residues all receptor atoms within 8 Å of any peptide atom.

Quality control removed complexes lacking consistent chain contacts, leaving 609 docked BCL-xL–ACP assemblies. These were clustered at 75% sequence identity to match the BioLiP2 protocol. The resulting graph was dominated by a single biologically coherent component containing 559 peptides, with the remaining 50 distributed across 17 minor clusters. To maintain the same leakage guarantee, each component was preserved intact, yielding 559/25/25 complexes for the train/validation/test splits. Although imbalanced, this distribution reflects the real composition of validated BH3-like peptides.

3.2. Extraction of structural constraints

3.2.1. Residue-level graph construction from protein–peptide complexes

Each peptide–protein complex was converted into a residue-level graph $G = (X, E, Y)$. For every residue i, the feature vector $x_i \in \mathbb{R}^{d_F}$ aggregates (i) a 20-way amino acid identity encoded as a categorical embedding, (ii) the DSSP secondary-structure code, (iii) seven physicochemical descriptors including hydrophobicity, relative solvent accessibility and hydrogen-bond indices, and (iv) eight backbone torsion and bond-angle descriptors expressed in radians. We parameterize local geometry with eight backbone descriptors: four dihedral angles which are ω (peptide bond), ϕ (N–C_α), ψ (C_α–C), and dihedral$_O$ (carbonyl orientation), and four bond angles (N–C_α–C, C_α–C–N, C–N–C_α, and C_α–C–O). Together these 17 features capture the torsional and

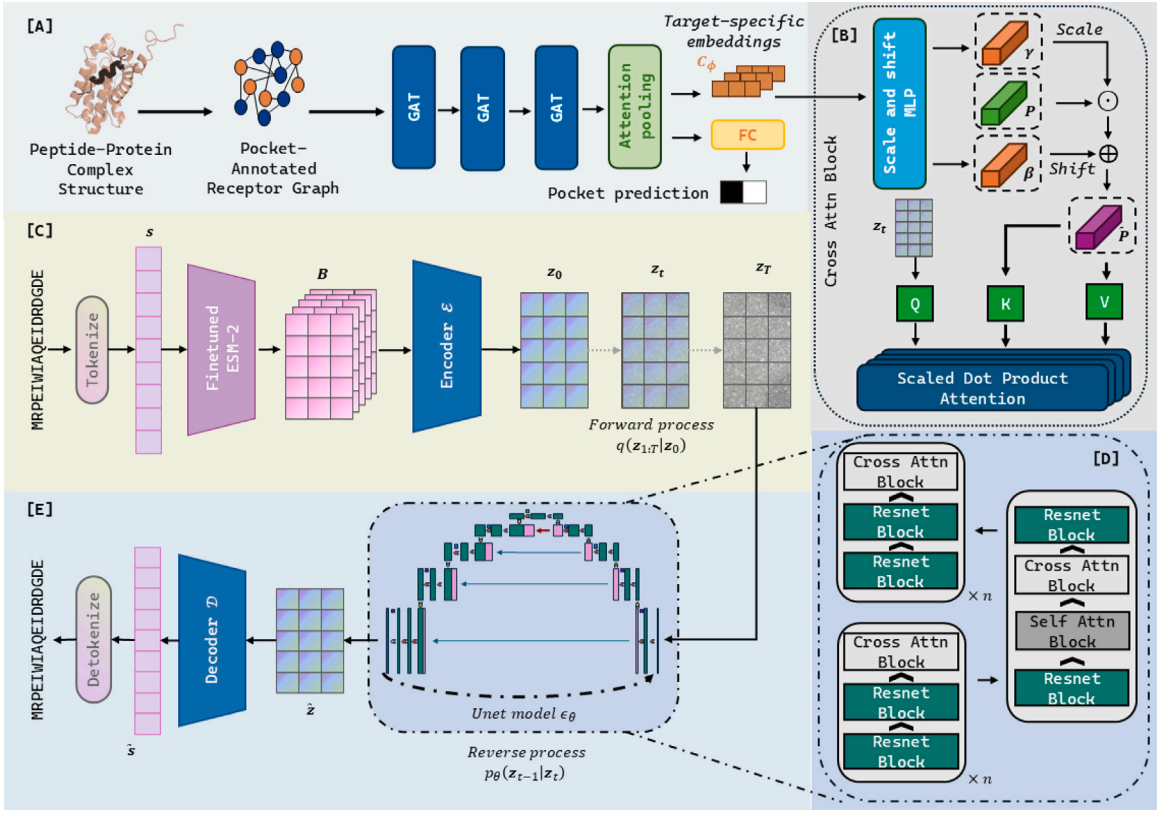

Fig. 1. Overview of the target-aware peptide generation framework. (A) Peptide–protein complexes are converted into pocket-annotated receptor graphs for node embedding and pocket prediction. (B) Target specific-embeddings are obtained via Attention readout pooling and injected into each cross attention block of the U-Net through an affine-modulated scale–shift prompt network. (C) Sequence embeddings from fine-tuned ESM-2 are compressed into a latent space and perturbed by forward diffusion. (D) Expanded view of the U-Net with cross-attention blocks integrating structural constraints. (E) Reverse diffusion integrates structural constraints and reconstructs latent vectors that are decoded into peptide sequences.

bending degrees of freedom that govern backbone conformation in a translation- and rotation-invariant manner [50].

After concatenation and trimming of chain termini, the resulting node matrix $X \in \mathbb{R}^{N\times d_F}$ where $d_F = 17$ provides deterministic, rotation-invariant inputs. The corresponding label vector $Y = \{y_i\}_{i=1}^{N}$ marks residues within 8 Å of any ligand heavy atom as binding ($y_i = 1$) and all others as non-binding ($y_i = 0$).

The edge set E represents the set of edges of the residue-level graph with $M = |E|$ representing the total number of edges. The edge set E is encoded as a matrix $E \in \mathbb{N}^{2\times M}$, where each column (v_i, v_j) denotes an undirected edge between residues i and j, defined by a 8 Å heavy-atom distance threshold [44,51], excluding chain termini.

3.2.2. GAT architecture

As shown in Fig. 1(A), we used a 3-layer GAT [10] that operates on 8 Å residue contact graphs uses attention readout with a sigmoid classifier for prediction of binding residues. The ℓ-th GAT layer updates node feature embedding $x_i^{(\ell)} \in \mathbb{R}^{d_{F'}}$ from $x_i^{(\ell-1)} \in \mathbb{R}^{d_F}$ through a learnable linear transformation via a weight matrix $W \in \mathbb{R}^{d_{F'}\times d_F}$, followed by neighbor attention aggregation.

Let the ℓ-th graph attention (GAT) layer be denoted by $\mathcal{G}^{(\ell)}$, parameterized by weights $W^{(\ell)}$. Starting from $X^{(0)} = X$, the three GAT blocks update node embeddings via multi-head masked attention, batch normalization $\mathcal{B}_\ell$, ReLU and dropout:

$$\begin{aligned} X^{(1)} &= \sigma_{\mathrm{ReLU}}\big(\mathcal{B}_1\big(\mathcal{G}^{(1)}(X^{(0)}, E, W^{(1)})\big)\big), \\ X^{(2)} &= \sigma_{\mathrm{ReLU}}\big(\mathcal{B}_2\big(\mathcal{G}^{(2)}(X^{(1)}, E, W^{(2)})\big)\big), \\ X^{(3)} &= \sigma_{\mathrm{ReLU}}\big(\mathcal{G}^{(3)}(X^{(2)}, E, W^{(3)})\big). \end{aligned} \tag{1}$$

For node i and neighbor $j \in \mathcal{N}(i)$, the attention coefficients for head k are computed as:

$$\begin{aligned} e_{ij}^{(\ell,k)} &= \sigma_{\mathrm{ReLu}}\Big(a_k^{(\ell)\top}\big[\,W_k^{(\ell)}x_i^{(\ell-1)} \,\Big\|\, W_k^{(\ell)}x_j^{(\ell-1)}\,\big]\Big), \\ \alpha_{ij}^{(\ell,k)} &= \frac{\exp\big(e_{ij}^{(\ell,k)}\big)}{\sum_{m\in\mathcal{N}(i)}\exp\big(e_{im}^{(\ell,k)}\big)}, \\ h_{i,k}^{(\ell)} &= \sum_{j\in\mathcal{N}(i)} \alpha_{ij}^{(\ell,k)}\, W_k^{(\ell)} x_j^{(\ell-1)}, \\ x_i^{(\ell)} &= \|_{k=1}^{K_\ell} h_{i,k}^{(\ell)}. \end{aligned} \tag{2}$$

Here $W_k^{(\ell)} \in \mathbb{R}^{d_h\times d_{\ell-1}}$ and $a_k^{(\ell)} \in \mathbb{R}^{2d_h}$ while $\|$ denotes vector concatenation, and $a^{(\ell)} \in \mathbb{R}^{2d_{F'}}$ is a learnable coefficient vector. For $\ell = \{1,2\}$, $K_\ell = 8$ heads and $d_h = d_{F'}$, producing a concatenated output dimension $d_\ell = d_h \cdot K_\ell$ except when at $\ell = 3$, $K_3 = 1$ to yield $X^{(3)} \in \mathbb{R}^{N\times d_h}$.

To extract the conditioning vector, we apply an attention pooling [52] to the final node embeddings $X^{(3)}$:

$$C_\phi = \sum_{n=1}^{N_i} \mathrm{softmax}(h_{\mathrm{gate}}(X^{(3)})) \tag{3}$$

with h_{gate} denoting the MLP that the attention pooling takes in. C_ϕ is broadcasted across the batch and concatenated with $X^{(3)}$ at each node position. This fused representation is passed through two fully connected output layer $\mathcal{F}_1$ and $\mathcal{F}_2$:

$$\begin{aligned} \tilde{X} &= \sigma_{\mathrm{ReLU}}\left(\mathcal{F}_1\left([X^{(3)} \,\|\, C_\phi]\right)\right) \\ L_\phi &= \mathcal{F}_2(\tilde{X}) \\ \hat{Y} &= \sigma(L_\phi) \end{aligned} \tag{4}$$

The projected attention pooling output C_ϕ is the graph-level embedding and is later used as the conditioning vector for the diffusion model.

3.2.3. Generation of structural constraint features

To derive structure-based constraints for peptide generation, the GAT was trained to classify binding-pocket residues in peptide–protein complexes. The network was first pretrained on BioLiP2 residue-level graphs and subsequently fine-tuned on ApInAPDB complexes, with pocket labels assigned to receptor residues whose heavy atoms lay within 8 Å of any peptide atom [44]. This two-stage procedure enabled the network to learn general structural representations from diverse complexes before specializing on the BCL-xL-specific interface.

Hyperparameter tuning and model selection were conducted using the Weights & Biases platform, which automated the exploration of learning rate, batch size, number of attention heads, and focal-loss parameters. The optimal configuration was chosen based on validation balanced accuracy. The final model employed the Adam optimizer (learning rate 5×10^{-4}, $\beta_1 = 0.9$, $\beta_2 = 0.999$) with focal loss [53] to address class imbalance ($\alpha = 1.0$, $\gamma = 2.5$). Pretraining was performed for 80 epochs with batch size 64 and a step-based learning-rate decay (factor 0.5 every 15 epochs). Fine-tuning proceeded for 70 epochs using an adaptive scheduler with patience of 5 epochs and the same decay factor. These tuning strategies ensured convergence stability and robust generalization across unseen complexes.

The architecture applied dropout of 0.3, batch normalization, and ReLU in the first two GAT layers. A linear prediction head with sigmoid activation produced residue level pocket probabilities.

From the final tuned model, node-level embeddings were aggregated via a three-step attention readout pooling to yield the structural-constraint vector $C_\phi \in \mathbb{R}^{d_{F'}}$ where $d_{F'} = 256$. As shown in Fig. 1(B), these graph-level embeddings were used to condition the peptide-generative model, guiding the diffusion process toward BCL-xL-targeting sequences with structurally coherent binding characteristics.

3.3. Generation of peptides with conditional latent diffusion model

3.3.1. Encoding using ESM-2 protein language model

To capture sequence-level biochemical representations, we employed a transformer-based ESM-2 PLM, as shown in Fig. 1(C). This PLM was pretrained on UniRef50 and adapted via a two-stage fine-tuning strategy by using the MLM objective [9]. We use ESM-2 to extract the sequence embeddings based on prior evidence from our earlier work [54], where ESM-2 yielded stronger downstream performance than ProtBERT for ACP prediction when used as a fixed feature extractor. This motivates its use in the present framework.

Given a tokenized sequence $s = \{s_l\}_{l=1}^{L}$ of length L, with masked positions $M_s \subseteq \{1, 2, \ldots, L\}$, the MLM objective aims to minimize the average negative log-likelihood of the true residues at masked positions, conditioned on the masked tokenized sequence $s \setminus M_s$:

$$\mathcal{L}_{\text{MLM}} = -\frac{1}{|M_s|} \sum_{l \in M_s} \log p(s_l \mid s \setminus M_s). \tag{5}$$

In the first stage of finetuning, peptide sequences from BioLiP2 were tokenized and padded to $L = 64$, with a fixed 15% masking rate, excluding special tokens. Training used Adam optimizer with a learning rate of 2×10^{-5}, batch size 8, for 5 epochs with linear warm-up. To preserve the general knowledge captured during pretraining while allowing domain-specific adaptation, all parameters were frozen except the final transformer block and language modeling output head [55]. This approach mitigates overfitting, particularly important in the small-data regime common in peptide therapeutics [47]. In the second stage, the BioLiP2-adapted model was further fine-tuned on ACP sequences targeting BCL-xL by using identical settings, refining sensitivity to anticancer-related residue patterns while retaining structural priors.

To enable conditional peptide generation from pretrained ESM-2 representations, we introduce encoder $\mathcal{E}$ as depicted in Fig. 1(C). Given a tokenized peptide sequence s, the ESM-2 model produces hidden representations $B \in \mathbb{R}^{d_{\text{ESM}} \times L}$, where $d_{\text{ESM}} = 1280$. The embeddings B are compressed into a lower-dimensional latent representation $z_0 = \mathcal{E}(B)$, where $z_0 \in \mathbb{R}^{d \times L}$ and $d = 128$.

3.3.2. Forward diffusion process

Given the latent vector z from encoder $\mathcal{E}$, the forward process q incrementally introduces Gaussian noise into latent embedding z_t with t representing the timestep: $z_0, z_1, \ldots, z_T$ over total timesteps T depicted in Fig. 1(C). This process, eventually turns input embedding into a Gaussian distribution at $z_T \sim \mathcal{N}(0, I)$ and is modeled as a Markov chain:

$$q(z_{1:T} \mid z_0) := \prod_{t=1}^{T} q(z_t \mid z_{t-1}) \tag{6}$$

A variance schedule β_t that denotes how much Gaussian noise is added at a timestep t over a range $\{\beta_1, \beta_2, \ldots \beta_T\}$ is introduced. Taking into consideration β_t, each transition between z_{t-1} to z_t in the forward process is defined as:

$$q(z_t \mid z_{t-1}) = \mathcal{N}(z_t; \sqrt{1-\beta_t}\, z_{t-1}, \beta_t I) \tag{7}$$

Defining $\alpha_t := \prod_{\tau=1}^{t} 1-\beta_\tau$, the marginal distribution at any timestep has the closed form:

$$q(z_t \mid z_0) = \mathcal{N}\left(z_t; \sqrt{\alpha_t} z_0, \left(1-\alpha_t\right) I\right) \tag{8}$$

and can be reparameterized as:

$$z_t(z_0, \epsilon) = \sqrt{\alpha_t} z_0 + \sqrt{1-\alpha_t}\epsilon \quad \text{for } \epsilon \sim \mathcal{N}(0, I) \tag{9}$$

This formulation enables direct sampling of z_t given any timestep t with respect to z_0 without iteratively applying (7).

3.3.3. Reverse denoising with cross-attention U-Net

To reconstruct z_0 from z_T, the reverse process p is applied through T timesteps, gradually removing noise by using a parameterized denoising model as portrayed in Fig. 1(E). This reverse process is initially defined as:

$$p(z_{0:T}) := p(z_T) \prod_{t=1}^{T} p(z_{t-1} \mid z_t) \tag{10}$$

where each step of the reverse process $p(z_{t-1} \mid z_t)$ is the conditional probability of Gaussian noise that is removed incrementally given z_t. Thus, this can be approximated as follows.

$$p(z_{t-1} \mid z_t, \theta) = \mathcal{N}(z_{t-1}; \mu(z_t, t, \theta), \Sigma(z_t, t)) \tag{11}$$

With $\Sigma_{(z_t,t)} = \beta_t I$ [56], the mean μ is learnt by a neural network. By considering the marginal distribution, The objective is to minimize the variational lower bound on the negative log-likelihood during reconstruction of z_T, which can be simplified into a cost function $\mathcal{L}$:

$$\mathcal{L}(\theta) = \mathbb{E}_{z_0,\epsilon,t}\left[\left\|\epsilon - \mu\left(\sqrt{\alpha_t} z_0 + \sqrt{1-\alpha}\epsilon, \theta, t\right)\right\|^2\right] \tag{12}$$

The objective of the LDM is to minimize the difference between ϵ and ϵ_t as expressed as a Mean Square Error (MSE) loss. We parameterized $\mu(z_t, \theta, t)$ by using a U-Net model that is well-suited for capturing both coarse and fine-grained spatial dependencies [6,57]. In each U-net block, a cross-attention block is augmented between the ResNet blocks to integrate conditioning information from the protein structure as portrayed in Fig. 1(D).

At each cross-attention block [6,58], the latent state z_t is flattened and projected via a linear transformation W_Q to produce a query matrix Q:

$$Q = W_Q \cdot Flatten(z_t) \tag{13}$$

To avoid degeneracy over the cross attention mechanism, we introduce a modulation prompt network which is an affine modulation scale and shift that takes in constrain features $C_\phi \in \mathbb{R}^{d_{F'}}$, and expands the lone context into N_k adaptive tokens $T(C_\phi) \in \mathbb{R}^{N_k \times d_{F'}}$ by modulating a small, learnable prompt bank $P \in \mathbb{R}^{N_k \times d_{F'}}$.

Concretely, we produce a per-sample *scale* γ and *shift* β for each prompt dimension using a lightweight MLP:

$$\begin{aligned} &\tilde{C}_\phi = \mathrm{LN}(C_\phi), \\ &[\gamma, \beta] = \mathrm{MLP}(\tilde{C}_\phi) \in \mathbb{R}^{N_k \times 2d_{F'}}, \\ &T(C_\phi) = \gamma \odot P + \beta \ \in \mathbb{R}^{N_k \times d_{F'}}, \end{aligned} \tag{14}$$

where $\odot$ denotes elementwise multiplication with broadcasting over the batch and LN denotes layer normalization. This affine modulation lets the model *gate* (γ) and *recenter* (β) a shared set of base tokens P based on the constraint C_ϕ, yielding a rich per-sample token sets.

The affine modulated constraint features $T(C_\phi)$ are projected to form the key K and value V matrix:

$$K = W_K \cdot T(C_\phi), \quad V = W_V \cdot T(C_\phi) \tag{15}$$

Where $W^Q \in \mathbb{R}^{n_c \times (h \cdot d)}$ and $W^K, W^V \in \mathbb{R}^{d_{F'} \times (h \cdot d)}$, with h heads and per-head dimensionality d and n_c is the no of channels, which varies depending on which U-Net block the attention module is placed at. The attention mechanism is defined as [58]:

$$\mathrm{Attention}(Q, K, V) = \mathrm{softmax}\left(\frac{QK^T}{\sqrt{d_0}}\right)V \tag{16}$$

where d_0 is the attention head dimension size.

We inject constraint features via cross-attention in every down and up block, providing conditioning at all scales while preserving the U-Net's local bias. To capture long-range dependencies efficiently, we add self-attention only at the last down, mid (bottleneck), and first up blocks, the points with maximal receptive field or cross-scale hand-offs. This concentrates global coordination where it is most effective, while maintaining pocket-aware guidance throughout the hierarchy.

3.3.4. Decoding the latent representations

After diffusion sampling, the LDM produces a denoised latent vector $\tilde{z} \in \mathbb{R}^{L \times d}$. Decoder D, illustrated in Fig. 1(E), decodes $\tilde{z}$ into tokenized logits $\tilde{s} \in \mathbb{R}^L$ via $\tilde{s} = D(\tilde{z})$, compatible with the amino acid vocabulary prior to detokenization.

The decoder D consists of an input feed-forward network, a GRU [59], and an output feed-forward network. Given the reconstructed embeddings $z \in \mathbb{R}^{d \times L}$, the first feed-forward layer applies a linear transformation $W_{\text{in}} \in \mathbb{R}^{h_{\text{GRU}} \times d}$ to obtain $z' = W_{\text{in}} z$. This is passed to a GRU layer with initial hidden state $h_0 = \mathbf{0} \in \mathbb{R}^{h_{\text{GRU}}}$, which updates according to:

$$h_1 = (1 - z') \odot n_1 + z' \odot h_0, \tag{17}$$

where the update gate τ_1, reset gate r_1, and candidate state n_1 are computed as:

$$g_1 = \sigma\left(W_{ig}\, z' + b_{ig} + W_{hg}\, h_0 + b_{hg}\right), \qquad g \in \{\tau, r\} \tag{18}$$

$$n_1 = \tanh\left(W_{in} z' + b_{in} + r_1 \odot (W_{hn} h_0 + b_{hn})\right). \tag{19}$$

with $\odot$ denoting the Hadamard product and σ the sigmoid activation. Here, W_* and b_* are learnable weight matrices and bias vectors.

The GRU output is passed through a second feed-forward network $W_{\text{out}} \in \mathbb{R}^{d_{\text{out}} \times h_{\text{GRU}}}$ with batch normalization and dropout. Batch normalization and dropout are applied after each layer to ensure stability and generalization in the learned latent manifold.

3.3.5. Model training and optimization

The LDM D_θ was trained in two phases: pretraining on BioLiP2-derived peptides for 9000 timesteps, and fine-tuning on ACP sequences targeting BCL-xL for 550 timesteps. A learning rate of 8×10^{-5} was used with the Adam optimizer. To address GPU memory constraints while maintaining effective batch size, gradient accumulation was performed every 8 steps.

We employ a U-Net backbone with a base channel width of 64 and channel multipliers (1, 2, 4, 8), yielding feature widths of 64, 128, 256, and 512 across the four up/down blocks. Each attention block uses h=4 heads with per-head dimensionality d = 32 (total attention width = 128). Conditioning is supplied via a modulated prompt mechanism with N_k=32 context tokens, injected through the cross-attention layers to guide denoising while preserving spatial detail.

The forward diffusion process used a cosine noise schedule over 1000 noising timesteps, while denoising was performed over 500 sampling steps using the Denoising Diffusion Implicit Models (DDIM) sampling strategy [60]. All training was conducted on a single NVIDIA A100 GPU with fixed random seed to ensure reproducibility.

This strategy ensured that the final model captured the broad physicochemical landscape of general peptide binding and the specific sequence patterns associated with therapeutic inhibition of BCL-xL.

4. Results and discussion

4.1. Comparative studies with state-of-the-art methods

4.1.1. Comparative evaluation based on generated sequences

To assess the fidelity of generated sequences, we evaluated the generated peptides using two complementary metrics: BLEU score, which quantifies n-gram overlap with reference sequences, and perplexity, which measures the likelihood of generated sequences under a fine-tuned protein language model. Higher BLEU and lower perplexity indicate better alignment with biological priors and improved linguistic fluency in the amino acid domain.

We conducted a comprehensive evaluation of six representative generative architectures spanning autoencoding, autoregressive and diffusion-based frameworks to benchmark their performance in peptide sequence design. All models were tested on identical datasets and evaluation pipelines, each contributing 2000 generated peptides conditioned on the test split, ensuring that the observed contrasts arise from architectural differences rather than disparities in sample size.

VAE and WAE [61] serve as autoencoding baselines [61]. Both encode peptide sequences into a latent manifold and reconstruct them with recurrent decoders. The VAE regularizes toward a Gaussian prior via Kullback–Leibler divergence, whereas the WAE replaces this with a Wasserstein penalty that enforces aggregate posterior matching. This often improves latent continuity but can still incur information loss or posterior collapse that limits fidelity to natural sequence statistics. To enable a controlled comparison against T-LDM, we augment both VAE and WAE with cross-attention conditioning that incorporates the pocket-specific graph-level embedding C_ϕ. Decoder token states act as queries and attend to C_ϕ after a learned projection, with residual connections and layer normalization preserving the base dynamics.

ProteinMPNN [46] is an autoregressive message-passing network that predicts residue identities sequentially along a fixed protein backbone [46]. The model propagates information through multiple rounds of node and edge updates to capture residue–residue geometry and backbone context. It achieves high sequence recovery by explicitly conditioning each residue prediction on its structural neighborhood, resulting in strong local token conformity but limited capacity for global diversity.

HelixDiff [45] is a score-based diffusion framework that generates all-atom alpha-helical peptides through a denoising process in a joint sequence–structure representation [45]. It employs a U-Net backbone

Table 1
Comparison on 2000 generated peptides (mean ± SD) across generative architectures. Best value per column appears in bold.

Model	BLEU ↑	Perplexity ↓
VAE	0.0462 ± 0.0224	79.3162 ± 60.6885
WAE	0.0508 ± 0.0372	67.4029 ± 182.6618
PC-LDM	0.0579 ± 0.0236	13.5331 ± 4.7276
HelixDiff	0.0404 ± 0.0247	13.0814 ± 7.0542
ProteinMPNN	**0.1377 ± 0.0883**	7.2946 ± 5.7495
T-LDM	0.0728 ± 0.0248	**4.8997 ± 1.3901**

to model the score function and encodes torsional, bond and side-chain angles alongside sequence tokens in a unified tensor. A hotspot-specific inpainting module constrains residue generation around known functional positions, improving local structural accuracy.

PC-LDM [47] is a conditional latent diffusion model that integrates physicochemical descriptors as external constraints [47]. Seven molecular descriptors, including hydrophobicity, charge, and molecular weight, are embedded through cross-attention to modulate the denoising network. This conditioning steers the diffusion process toward peptides that satisfy predefined physicochemical profiles while maintaining statistical proximity to reference sequences.

Table 1 summarizes sequence quality across all generators using BLEU and perplexity. Among all methods, ProteinMPNN achieved the highest BLEU, reflecting its strong ability to reproduce reference sequence fragments with high local fidelity. T-LDM ranked second, maintaining competitive sequence alignment despite prioritizing structural conditioning and generative diversity. Unlike T-LDM, ProteinMPNN lacks stochasticity as it predicts residues sequentially in a fully deterministic manner, which leads to highly similar or repeated outputs across generations.

In terms of perplexity, T-LDM attained the lowest score, indicating that its generated sequences are the most statistically probable. ProteinMPNN followed closely, reinforcing its overall strength in both local n-gram precision and global sequence realism. Interestingly, HelixDiff and PC-LDM achieved similar perplexity values, suggesting that diffusion-based approaches can also produce linguistically coherent sequences, though with reduced BLEU scores indicating less precise n-gram overlap.

Meanwhile, the traditional autoencoder models, VAE and WAE, performed worst across both metrics. Their low BLEU scores and high perplexity values reveal a tendency to generate low-quality or overly diverse outputs that deviate from realistic peptide syntax. These results reinforce the advantage of incorporating structural priors or deterministic residue selection strategies to enhance generative fidelity and biological plausibility.

To assess statistical significance, we applied Kruskal–Wallis omnibus tests to both BLEU and perplexity distributions (Supplementary Notes Table S1). These tests reject the null hypothesis of equal medians across the baseline models, indicating that the differences among model classes are highly significant. Non-parametric testing was chosen because the distributions are heavy-tailed and deviate markedly from Gaussian assumptions.

Pairwise Mann–Whitney U tests with Holm correction (Supplementary Notes Table S1) show that nearly all inter-model differences are statistically significant for both BLEU and perplexity, with the only non-significant case being the BLEU comparison between VAE and WAE. For BLEU, ProteinMPNN achieves the highest overall score and significantly outperforms all other models, while T-LDM ranks second, exceeding PC-LDM, HelixDiff, and both autoencoder baselines. For perplexity, T-LDM demonstrates the lowest values across all comparisons, indicating superior sequence fluency and consistency under the protein language model prior. Together, these results reveal a clear distinction between sequence overlap and language-model fluency.

4.1.2. Distributional calibration and physicochemical fidelity

As the baselines contribute substantially more sequences than the training set, we require a distributional metric that is insensitive to unequal sample sizes when densities are estimated with identical procedures, symmetric between the model and the reference, and remains bounded and numerically stable even when the supports only partially overlap. The Jensen–Shannon (JS) divergence meets these criteria, so we use it to quantify how well each baseline model reproduces the entire empirical distribution of key physicochemical properties.

We apply this procedure to six properties that are directly relevant to physicochemical realism and developability: net charge at pH 7, isoelectric point (pI), Grand Average of Hydropathy (GRAVY), instability index, AntiCP scores [62], and HemoPI2 score [63].

The Kernel Density Estimation (KDE) overlays in Fig. 2 illustrate the reference distributions and the model-induced shifts for each baseline. JS condenses these visual differences into a calibrated, comparable scalar per panel. In practice, close overlap of the colored curves with the gray reference, particularly in the modal regions and tails, corresponds to a small JS value, whereas broadened or shifted profiles yield larger JS. Reporting per-property JS thus provides a faithful summary of the discrepancies visible in Fig. 2 and enables fair comparison across models despite the substantial mismatch in sample counts.

In Table 2, T-LDM attains the lowest JS for net charge at pH 7, isoelectric point, and AntiCP score, and achieves the best overall mean JS, Rank 1. These gains are substantive. For example, for net charge, T-LDM reduces JS by 49% relative to the next-best baseline PC-LDM, and by 69% relative to ProteinMPNN, indicating markedly better reproduction of the reference charge distribution. For pI, T-LDM improves on the next-best baseline by 22% (vs. 0.168 for PC-LDM), suggesting that the joint handling of acidic/basic residues and sequence context is better calibrated to the training set. For AntiCP, T-LDM again achieves the smallest divergence (14% lower vs. PC-LDM), capturing both the location and dispersion of the reference propensity distribution. While other models are best on individual metrics—ProteinMPNN on GRAVY, HelixDiff on instability index, and PC-LDM on HemoPI2, T-LDM remains competitive rather than outlying on these axes, yielding the lowest mean JS across properties.

Net charge and pI are primary determinants of peptide electrostatic behavior, governing solubility, non-specific binding, and interface complementarity. T-LDM's low divergence on these axes confirms that it generates sequences within a realistic physicochemical range, faithfully capturing the reference distribution's modes without succumbing to mode collapse. Its high fidelity on AntiCP and competitive behavior on HemoPI2 further indicate preservation of therapeutic and safety-relevant profiles of the reference set.

Although VAE and WAE exhibit perfect novelty scores, their generated sequences are questionable based on previous perplexity and BLEU analyses, which revealed low sequence quality and weak alignment. Consequently, their high novelty values likely reflect randomization rather than meaningful diversity.

Finally, T-LDM achieves this distributional fidelity without sacrificing generative novelty: Novelty@75% is 0.827. ProteinMPNN shows lower novelty, while HelixDiff attains perfect novelty but with a higher mean JS than T-LDM. PC-LDM exhibits high novelty and second-best mean JS. Overall, T-LDM strikes the most effective balance, aligning with the training data's implicit physicochemical priors while producing diverse, novel sequences.

4.1.3. Comparing baseline models by docking score

We evaluate how effectively each model generates peptides that favor low-energy binding conformations by docking 100 randomly sampled peptides per model with HADDOCK3 and analyzing the resulting score distributions in Fig. 3. Lower and more negative scores indicate better predicted binding whereas large positive values typically reflect penalized, non–productively bound poses. The left plot summarizes

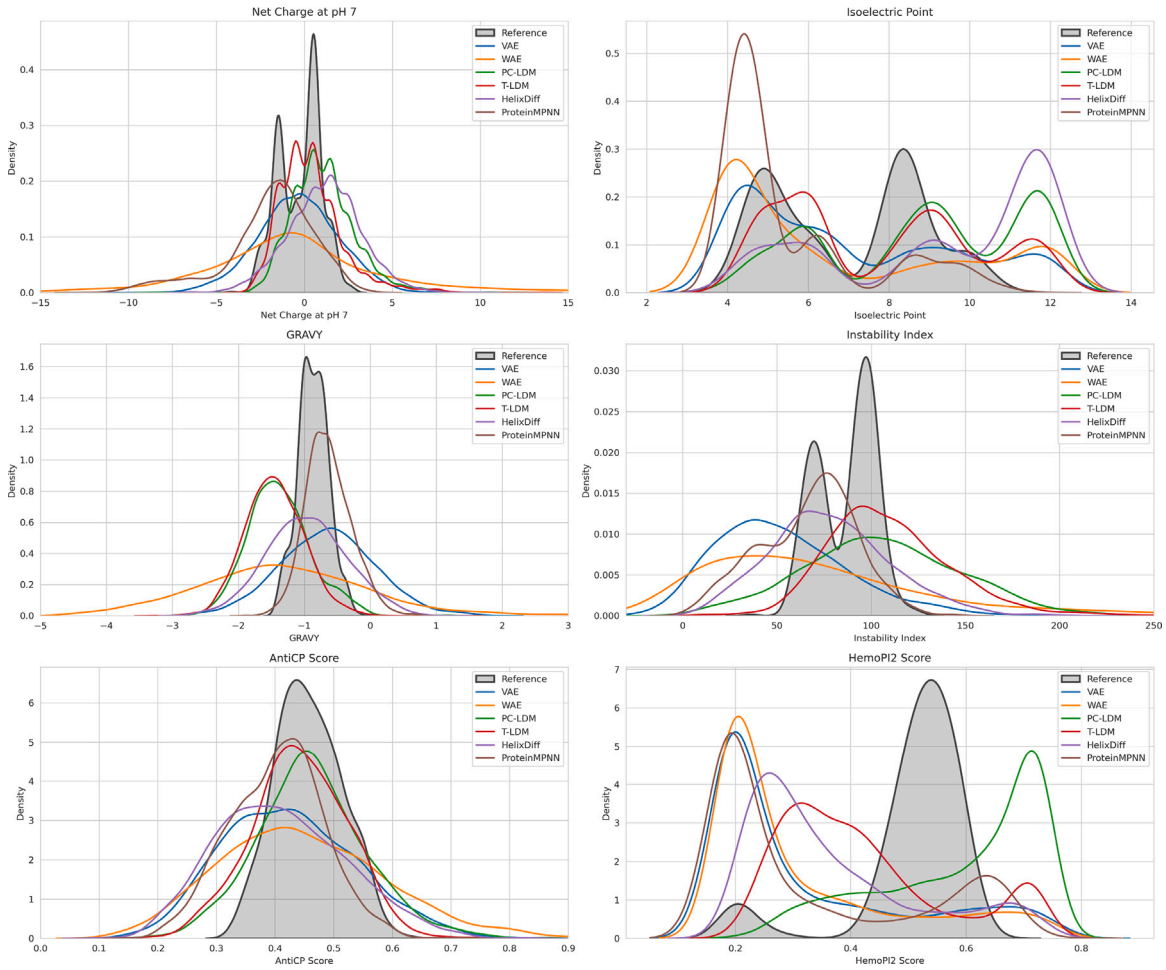

Fig. 2. Physicochemical probability across baseline models. Kernel density estimates compare the reference set (gray) to four models: PC-LDM (teal), T-LDM (orange), HelixDiff (green), and ProteinMPNN (red) for net charge (pH 7), isoelectric point, GRAVY, instability index, AntiCP score and HemoPI Score. Greater overlap with the gray curve indicates better calibration: T-LDM and PC-LDM most closely match the reference for charge and pI and track the AntiCP mode; HelixDiff exhibits broader, shifted hydropathy and higher variability; ProteinMPNN tends toward lower charge and higher pI with modest weight shifts. All models show wider molecular-weight and instability spreads than the narrow reference distribution.

Table 2
Distributional calibration of generated peptides against the training distribution via Jensen–Shannon (JS) divergence of kernel density estimates for each property.

Model	Novelty@75% (↑)	JS Net Charge(↓)	JS pI(↓)	JS GRAVY(↓)	JS Instability(↓)	JS AntiCP(↓)	JS HemoPI2(↓)	JS Mean(↓)	JS Rank(↓)
WAE	**1.000**	0.273	0.281	0.290	0.286	0.125	0.357	0.269	6
VAE	**1.000**	0.208	0.224	0.184	0.297	0.110	0.345	0.228	5
ProteinMPNN	0.740	0.256	0.311	**0.094**	0.183	0.094	0.337	0.212	4
HelixDiff	**1.000**	0.168	0.223	0.149	**0.166**	0.123	0.351	0.196	3
PC-LDM	0.975	0.153	0.168	0.272	0.211	0.043	**0.281**	0.188	2
T-LDM	0.827	**0.078**	**0.132**	0.297	0.182	**0.037**	0.336	**0.177**	**1**

Table 3
Positive rate and median HADDOCK3 scores for Random 100 and Top 50 selections per generator.

Model	Pos. rate %	Median (Random 100)	Median (Top 50)
VAE	27.0	−36.9	−60.9
WAE	16.0	−39.3	−61.0
PC-LDM	16.0	−54.2	−95.1
HelixDiff	17.0	−57.7	−75.3
ProteinMPNN	23.0	−82.6	−101.5
T-LDM	**12.0**	**−86.8**	**−123.7**

each model's Random-100 cohort with violins, revealing central tendency and tail behavior, while the right plot focuses on the Top-50 per model to compare the high-affinity regime directly. A complementary numeric summary is provided in Table 3, including the fraction of positive scores, an indicator of non-binders. Based on the Top-50 Peptide closest median value, Fig. 4 showcases six representative docked poses per baseline model,

In both views of Fig. 5, T-LDM generates more peptides in the favorable negative-score range and shows a visibly shorter positive tail than the other models. In the Top-50 boxplots it shows a compact inter-quartile range and a strongly negative median, while avoiding the extreme outliers present for several baselines. ProteinMPNN is broadly comparable in the top-ranked area as its boxplot largely overlaps T-LDM whereas HelixDiff exhibits broader dispersion and a heavier positive tail. The autoencoder baselines VAE and WAE and the PC-LDM variant skew toward less attractive energies with wider spreads.

Non-parametric tests corroborate these impressions (Supplementary Notes Table S2). The Kruskal–Wallis omnibus is highly significant (H =

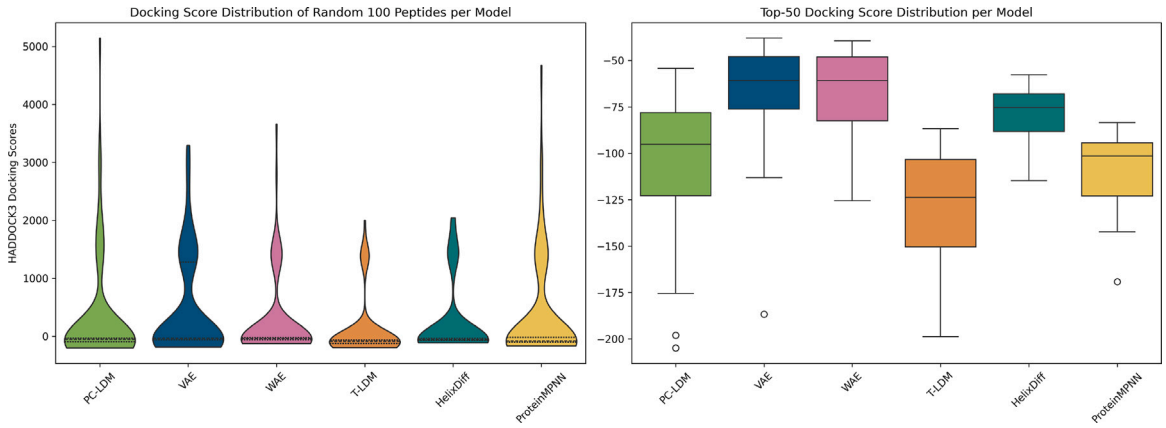

Fig. 3. HADDOCK3 docking-score distributions for each peptide generator. Violin plots summarize the full Random 100 cohort, while paired boxplots highlight the Top-N selections. Lower (more negative) scores correspond to tighter predicted binding.

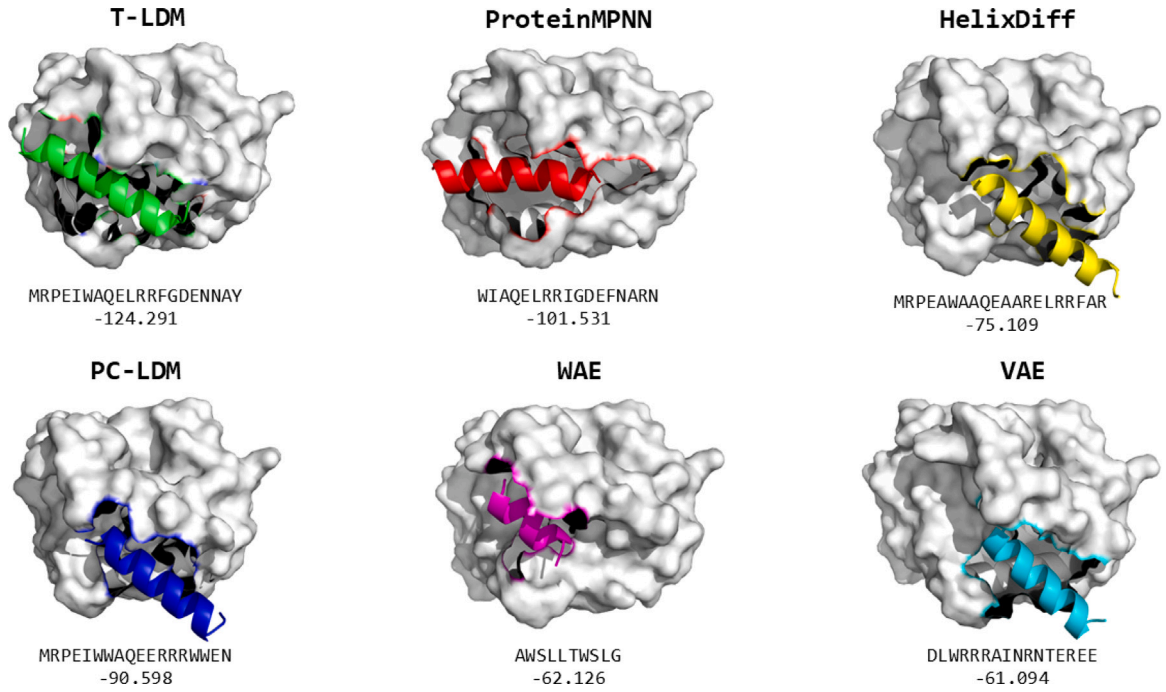

Fig. 4. For each baseline model, the peptide with a docking score nearest to the median of its top-50 ranked candidates is visualized in complex with the BCL-xL receptor surface. Below each structure, the corresponding peptide sequence and docking score are shown. T-LDM produces conformations that insert stably into the canonical binding groove, while the other baselines exhibit variable alignment and binding depth.

91.273, $p = 3.630\times10^{-18}$). Holm-adjusted Mann–Whitney contrasts show that T-LDM differs strongly from HelixDiff ($p = 2.243 \times 10^{-8}$) and from the autoencoders (VAE, 2.417×10^{-13}; WAE, 1.107×10^{-11}), and it also improves over the PC-LDM variant (4.518×10^{-4}). The comparison with ProteinMPNN is not significant after adjustment ($p = 8.243 \times 10^{-2}$), consistent with the visual overlap among their top scores. Overall, the distributional evidence and inference indicate that T-LDM delivers reliably low docking energies with controlled tails, outperforming HelixDiff and the autoencoder baselines, and performing on par with ProteinMPNN.

4.2. Ablation studies

4.2.1. GNN backbone and spatial cutoff comparison

We conducted a controlled ablation on the validation set across three GNN variants, namely GCN, GAT and EGNN, and across two spatial cutoffs at 6 Å and 8 Å. All models used mean pooling, and we evaluated metrics suited to class imbalance, including balanced accuracy, MCC, AUPRC, Brier score, and precision and recall for each class. Brier score, and per-class precision/recall. Full summary statistics (mean ± SD over 10 seeds) appear in Table 4, and the associated one-way ANOVA and Holm-adjusted post hoc tests are reported in Supplementary Notes Tables S3–S6.

At the 6 Å cutoff, GAT already dominates the comparison. It delivers balanced accuracy 0.9815 ± 0.0024, MCC 0.9591 ± 0.0079, and a Brier score of 0.0146 ± 0.0016, indicating both strong discrimination and well-calibrated probabilities. Ranking performance is nearly perfect (AUPRC 0.9991 ± 0.0003), and per-class precision/recall improve over the other backbones in tandem. ANOVA confirms a significant backbone effect for every metric (Supplementary Notes Table S3), while Holm-adjusted pairwise tests attribute those gains primarily to GAT outperforming both GCN and EGNN. GCN in turn retains a consistent but smaller edge over EGNN (Supplementary Notes Table S4). The resulting hierarchy is GAT > GCN > EGNN which holds across the imbalance-robust endpoints as well as class-wise precision and recall.

Table 4
Validation performance (mean ± SD over 10 seeds) across architectures and spatial cutoffs.

Model	Dist.	Balanced Acc. ↑	MCC ↑	Precision ↑		Recall ↑		AUPRC ↑	Brier ↓
				Class 0	Class 1	Class 0	Class 1		
GCN	6 Å	0.9340 ± 0.0023	0.8498 ± 0.0051	0.8594 ± 0.0102	0.9724 ± 0.0046	0.9513 ± 0.0087	0.9168 ± 0.0077	0.9901 ± 0.0008	0.0566 ± 0.0026
EGNN	6 Å	0.8248 ± 0.0621	0.6319 ± 0.1268	0.6994 ± 0.1163	0.9156 ± 0.0257	0.8670 ± 0.0369	0.7826 ± 0.1230	0.9356 ± 0.0271	0.2123 ± 0.0792
GAT	6 Å	0.9815 ± 0.0024	0.9591 ± 0.0079	0.9651 ± 0.0132	0.9903 ± 0.0025	0.9820 ± 0.0049	0.9809 ± 0.0075	0.9991 ± 0.0003	0.0146 ± 0.0016
GCN	8 Å	0.9734 ± 0.0013	0.9257 ± 0.0059	0.9120 ± 0.0126	0.9931 ± 0.0021	0.9817 ± 0.0058	0.9652 ± 0.0056	0.9988 ± 0.0001	0.0193 ± 0.0009
EGNN	8 Å	0.9506 ± 0.0237	0.8905 ± 0.0462	0.9025 ± 0.0517	0.9778 ± 0.0162	0.9397 ± 0.0449	0.9615 ± 0.0237	0.9852 ± 0.0242	0.0411 ± 0.0174
GAT	8 Å	**0.9993 ± 0.0008**	**0.9985 ± 0.0016**	**0.9988 ± 0.0022**	**0.9996 ± 0.0005**	**0.9990 ± 0.0014**	**0.9996 ± 0.0008**	**1.0000 ± 0.0000**	**0.0012 ± 0.0004**

Expanding the spatial neighborhood to 8 Å amplifies GAT's lead. The 8 Å configuration achieves balanced accuracy of 0.9993 ± 0.0008, MCC of 0.9985 ± 0.0016, and an AUPRC indistinguishable from unity, with a Brier score reduced to 0.0012 ± 0.0004, indicating consistently precise and well-calibrated predictions. Class-specific precision and recall remain uniformly high (≥0.9988), underscoring excellent specificity and sensitivity. ANOVA again flags significant backbone effects across all metrics (Supplementary Notes Table S5). Holm-adjusted contrasts show that GAT significantly outperforms both GCN and EGNN for every headline metric, with the single exception of AUPRC when comparing GAT and EGNN (Supplementary Notes Table S6). Differences between GCN and EGNN tighten relative to the 6 Å setting, reflecting the latter's improvements in calibration and balanced accuracy.

Comparing cutoffs within each backbone reveals that enlarging the spatial radius consistently benefits all models. Both GCN and EGNN gain in balanced accuracy and MCC while reducing Brier scores, suggesting that the 8 Å context captures salient residue–residue interactions that were missed at 6 Å. GAT, however, enjoys the largest lift as its calibration and ranking metrics converge to practical optima at the broader cutoff.

Overall, the evidence points to two conclusions. First, GAT is the most effective backbone in this experiment, outperforming GCN and EGNN on every imbalance-aware metric at both spatial cutoffs, with differences that remain statistically robust after multiple-comparison control in nearly all cases. Second, the 8 Å cutoff further boosts performance across architectures and exhibits especially strong synergy with GAT. Consequently, we adopt the 8 Å GAT configuration for all downstream analyses.

4.2.2. Readout function ablation for 8 Å GAT

With 8 Å GAT established as the strongest backbone/cutoff, we ablate the graph readout under identical training, threshold–tuning, and evaluation. We compare global mean, attention-based readout [52]. Validation results (mean ± SD over 10 seeds) are summarized in Table 5 whereas omnibus and pairwise statistics appear in Supplementary Notes Table S7–S8.

All readouts achieve performance close to the upper limit, but attention delivers the strongest results across the board, covering balanced accuracy, precision for class 0 and class 1, recall for class 0 and class 1, AUPRC, and the lowest Brier score (Table 5). Mean pooling is close behind, while Set2Set is slightly weaker on discrimination and calibration. As expected at these performance level, AUPRC is indistinguishable across all readouts.

ANOVA detects a significant effect of pooling on balanced accuracy, MCC, and Brier, with other metrics not reaching the 0.05 threshold (Supplementary Notes Table S7). Holm-adjusted pairwise tests localize reliable differences primarily to Mean vs. Attention: balanced accuracy (p=4.487×10^{-2}) and MCC (p=4.484×10^{-2}) are significant (Supplementary Notes Table S8). Other pairwise contrasts, including those for Brier, do not survive Holm correction given the near-ceiling performance.

At this operating point, attention provides the best discrimination and the lowest Brier, with tight variances, whereas mean and Set2Set offer no measurable advantage.

4.2.3. Conditioning strategies and fusion mechanisms evaluation

We next ablate the denoising U-Net in our LDM to isolate the effect of the conditioning fusion block. Specifically, we compare FiLM-based modulation [64] versus cross-attention during denoising under three conditioning scopes: (i) unconstrained (no structural conditioning), (ii) global protein embeddings, and (iii) pocket-specific embeddings. The global protein embedding represents the entire structural context of the protein whereas pocket-specific embedding focuses solely on the spatial neighborhood of the binding pocket site. Each setting generates a validation cohort of 5000 peptides, enabling a controlled comparison of fusion mechanisms and conditioning scope on validation performance.

Pocket (Cross-Attention) delivers the strongest validation performance on both metrics in Table 6. BLEU improves by roughly 17% relative to Pocket (FiLM) and by about 4% over the Global (FiLM). Perplexity falls by approximately 27% versus Pocket (FiLM) and by 35% to 45% relative to the global baselines and the unconstrained LDM, indicating closer adherence to biological priors encoded by the language model.

The Kruskal–Wallis tests in Supplementary Notes Table S9 decisively reject equal medians for both BLEU and perplexity, motivating pairwise evaluation. Holm-adjusted Mann–Whitney U outcomes in Supplementary Notes Table S9 confirm that Pocket (Cross-Attention) significantly outperforms every comparator on both endpoints. Within the global setting, FiLM exceeds cross-attention on BLEU while cross-attention delivers lower perplexity, signaling that fusion choice has the largest impact when pocket-aware embeddings are available.

Together these observations show that pocket-aware conditioning supplies the most informative structural context, and cross-attention is the fusion mechanism that exploits it best. Pocket (Cross-Attention) therefore remains the preferred configuration for downstream sequence generation.

4.2.4. Docking performance across BCL-xL conformers and other BCL-2 family receptors

To examine target specificity and conformational robustness, we re-docked the same one hundred peptides that previously bound to the native BCL-xL structure (4QVE). The test panel comprised three additional BCL-xL conformers (1R2D, 3ZLR, 4QVF) and two other BCL-2 family receptors (BCL-2 and MCL-1).

The 4QVF [48] structure (BIM BH3) maintains the canonical helical binding seen for BH3 peptides, while BID in 4QVE uniquely unwinds at the C-terminus. Relative to BID, BIM is associated with a closed orientation of Glu96 and Tyr195 (versus the open orientation observed with BID), reflecting local packing changes in the groove. Including 4QVF tests whether generated peptides retain favorable binding when the groove contracts. The 1R2D [65] conformer captures an alternative BH3-groove state of BCL-xL with subtly shifted loop packing and side-chain orientations around the P2–P4 region. Although not complexed with a peptide, it samples a compact binding-competent groove similar to BH3-bound forms, thereby testing tolerance to modest conformational rearrangements within the binding cleft. In contrast, 3ZLR [66] is a small-molecule complex in which a benzothiazole-hydrazone ligand displaces Phe105 and remodels the P2 pocket, creating a markedly different local geometry from any peptide-bound form. This structure tests robustness against non-peptidic groove conformations.

Table 5
Validation performance (mean ± SD over 10 seeds) for GAT at 8Åacross pooling/readout functions.

Pooling	Balanced Acc. ↑	MCC ↑	Precision ↑		Recall ↑		AUPRC ↑	Brier ↓
			Class 0	Class 1	Class 0	Class 1		
Mean	0.9993 ± 0.0008	0.9985 ± 0.0016	0.9988 ± 0.0022	0.9996 ± 0.0005	0.9990 ± 0.0014	0.9996 ± 0.0008	1.0000 ± 0.0000	0.0012 ± 0.0004
Set2Set	0.9985 ± 0.0018	0.9967 ± 0.0041	0.9970 ± 0.0041	0.9993 ± 0.0008	0.9982 ± 0.0022	0.9989 ± 0.0015	1.0000 ± 0.0000	0.0015 ± 0.0008
Attention	**0.9999 ± 0.0003**	**0.9999 ± 0.0004**	**1.0000 ± 0.0000**	**0.9999 ± 0.0002**	**0.9998 ± 0.0006**	**1.0000 ± 0.0000**	**1.0000 ± 0.0000**	**0.0006 ± 0.0008**

Table 6
Comparison of conditioning strategies and fusion mechanisms. Metrics are reported as mean ±SD over 5,000 generated peptides.

Condition	BLEU ↑	Perplexity ↓
Unconstrained	0.176 ± 0.078	10.774 ± 6.828
Global (FiLM)	0.221 ± 0.095	9.142 ± 7.446
Global (Cross-Attn)	0.173 ± 0.0970	9.183 ± 4.323
Pocket (FiLM)	0.198 ± 0.115	8.209 ± 6.830
Pocket (Cross-Attn)	**0.231 ± 0.087**	**5.957 ± 2.514**

Table 7
Docking success rates for 100 random attempts per receptor.

	BCL-xL				Other BCL-2 family	
Receptor	4QVE	1R2D	3ZLR	4QVF	BCL-2	MCL-1
Success rate (%)	100	98	98	97	79	18

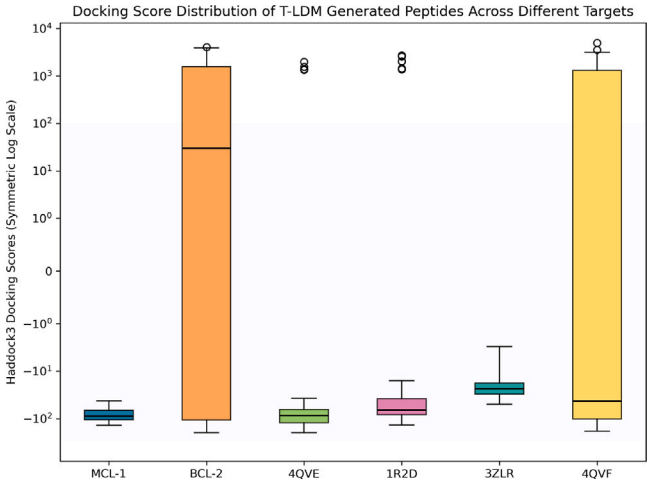

Fig. 5. Docking score distributions (lower is better) for successful runs across BCL-xL conformers and other BCL-2 family receptors. Counts of successful docking are annotated. The main panel uses a symlog scale with a linear inset near zero. Binding is strongest and most consistent for 4QVE, slightly weaker for 1R2D, degraded for 3ZLR, and unstable for 4QVF. Cross-family transfer is limited, with weak binding to BCL2 and sparse but favorable cases for MCL1.

Together, these conformers span the structural spectrum from peptide-induced compact and expanded grooves to ligand-induced rearrangements, while BCL-2 [67] and MCL-1 [68] provide cross-family comparisons for selectivity. For each receptor we measured both the probability of achieving a valid docking pose and the conditional score distribution, where lower scores indicate stronger binding. Fig. 5 visualizes these distributions, and Table 7 summarizes success rates.

The native 4QVE structure achieved complete success with the most favorable and compact score distribution. The 1R2D conformer remained highly tractable with only mild weakening, indicating limited sensitivity to moderate structural deviations. The 3ZLR pocket yielded uniformly weaker scores, consistent with a less compatible groove geometry. The 4QVF structure produced wide and heavy-tailed distributions with occasional extreme penalties, suggesting unstable or partially collapsed docking modes.

Table 8
Docking success rates (%, 100 attempts per receptor).

	MCL-1 Cond. Peptides		BCL-2 Cond. Peptides	
Receptor	MCL-1	BCL-xL	BCL-2	BCL-xL
Success rate (%)	100	85	100	82

Cross-family transfer was limited. BCL-2 accepted 79 of 100 peptides but produced broadly unfavorable scores with a pronounced right skew, while MCL-1 succeeded for only 18 peptides. Among those few successes, the scores were strong, yet the low success rate indicates strong selection bias toward easy cases.

Nonparametric tests found in Supplementary Notes Table S10 confirm these trends. The Kruskal–Wallis omnibus test detected strong differences across receptors ($p = 8.690 \times 10^{-18}$). Dunn contrasts indicate that the native 4QVE binds significantly better than 3ZLR, 4QVF, and BCL-2, and marginally better than 1R2D. The 1R2D conformer remains stronger than 3ZLR and comparable to 4QVF, while BCL2 performs significantly worse than both native and 1R2D. MCL-1 differs from several receptors, outperforming 3ZLR, 4QVF, and BCL-2 but not distinguishable from 4QVE or 1R2D. However, MCL-1 produced only eighteen successful dockings, its conditional statistics should be interpreted with caution.

Considering both success rates and score distributions, the generated peptides bind most consistently and favorably to the primary BCL-xL, retain strong affinity for moderate conformational variants such as 1R2D, and lose stability for ligand-free or open-pocket states like 3ZLR and 4QVF. Cross-family generalization is weak, with BCL-2 showing poor compatibility and MCL-1 favorable only in a few selective cases. Overall, the results indicate that the target-aware latent diffusion model produces sequences optimally tuned for the primary BCL-xL target while maintaining limited but interpretable robustness to structural variation within the family.

4.2.5. *Conditioned peptide docking selectivity*

To evaluate whether the conditioning mechanism transfers meaningful structural preferences across related receptors, we analyzed the docking performance of peptides generated under distinct conditioning targets. Each conditioned peptide set of 100 sequences per target was docked to both its native receptor and cross-family receptors using HADDOCK3, and the resulting score distributions were compared on a symmetric log scale as shown in Fig. 6.

Peptides conditioned on MCL-1 exhibited strong binding to their intended receptor, achieving the most negative docking energies and a 100% success rate as reported in Table 8. When the same peptides were redirected to BCL-xL, their median affinity weakened modestly with an 85% docking success rate, indicating partial structural compatibility between the two grooves. In a similar manner, BCL-2 conditioned peptides bound effectively to BCL-2 with 100% success but lost binding efficiency when redirected to BCL-xL with 82% success. These results show that the conditioning network captures target specific features that are partly transferable to homologous BCL-2 family members but remain strongest at the original conditioning receptor.

Statistical testing presented in Supplementary Notes Table S11 confirmed significant differences among receptor and condition groups. The Kruskal–Wallis test yielded $H = 74.700$ and $p = 4.202 \times 10^{-16}$. Dunn post hoc comparisons with Holm correction showed clear contrasts

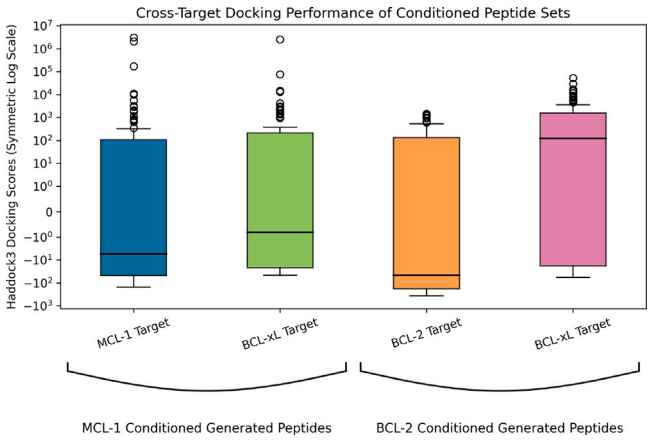

Fig. 6. Cross-target docking performance of conditioned peptide sets on a symmetric log scale. Each box shows the distribution of HADDOCK3 scores for random 100 peptides conditioned on the indicated source target and evaluated on the target shown on the x-axis. Lower scores indicate stronger predicted binding.

between cognate and non-cognate pairs such as MCL-1 versus BCL-xL for MCL-1 conditioned peptides with $p = 1.060 \times 10^{-2}$ and BCL-2 versus BCL-xL for BCL-2 conditioned peptides with $p = 2.489 \times 10^{-15}$.

Taken together, these results demonstrate that the target aware conditioning framework preserves receptor specific preferences while retaining moderate flexibility toward related anti-apoptotic proteins. The observed selectivity patterns confirm that structural conditioning effectively guides peptide generation toward sequences optimized for the intended binding pocket, consistent with the model objective.

5. Conclusion

We presented a receptor-aware generative framework, T-LDM, that fuses graph-encoded pocket context with latent diffusion to design apoptosis-inducing peptides. By extracting residue-level constraints with an 8 Å GAT and injecting them as modulated prompts into a U-Net denoiser, T-LDM aligns generative sampling with the local interaction topology of the BCL-xL groove while preserving sequence priors from a protein language model. This design achieves three complementary outcomes. First, the GAT structural encoder is accurate and well-calibrated, enabling reliable conditioning constraint embeddings. Second, pocket-centric cross-attention during denoising materially improves sequence quality and linguistic fluency relative to global or unconstrained variants (significant gains in BLEU and perplexity under non-parametric tests with Holm correction). Third, among contemporary baselines, T-LDM attains the lowest perplexity and the best overall JS divergence across physicochemical properties (best-in-class for net charge, pI, and AntiCP; lowest mean JS), while preserving novelty relative to the training corpus.

Functionally, docking studies show that T-LDM concentrates probability mass on low-energy BCL-xL binders and limits the fraction of penalized, non-productive poses, outperforming autoencoder baselines and complementing the strong median energies observed for ProteinMPNN. Cross-family evaluations further suggest that while transfer to BCL-2/MCL-1 is limited, BCL-xL conformational variants can be accommodated with moderate sensitivity, supporting the use of pocket-specific conditioning when structural heterogeneity is expected.

This work establishes T-LDM as a principled approach for peptide design that is simultaneously syntax-faithful, physicochemically realistic, and structurally compatible with a therapeutic receptor. Beyond ACPs, the modular design of our framework which spans across ESM-based language modeling and structural GNN conditioning, can be adapted to other therapeutic contexts, such as antimicrobial, antiviral, or cell-penetrating peptides. Furthermore, the interpretability and robustness of the model open the door for integration with wet-lab screening pipelines, facilitating a loop between in silico generation and experimental validation.

In summary, this work establishes a target-aware, data-efficient framework for rational peptide design, capable of producing candidates with high linguistic fidelity and biochemical realism, accelerating the discovery of next-generation peptide therapeutics, development timelines, and expanding the frontiers of precision medicine.

CRediT authorship contribution statement

Tiara Natasha Binte Sayuti: Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Kakuly Mittal:** Writing – review & editing, Software, Methodology, Data curation. **Tan Lai Heng:** Writing – review & editing, Methodology. **Jagath C. Rajapakse:** Writing – review & editing, Methodology, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was supported by AcRF Tier-2 grant MOE-T2EP20224-0004 of Ministry of Education, Singapore.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.compbiomed.2025.111286.

References

[1] L.V. Bauso, V. La Fauci, S. Munaò, et al., Biological activity of natural and synthetic peptides as anticancer agents, Int. J. Mol. Sci. 25 (13) (2024) http://dx.doi.org/10.3390/ijms25137264, URL https://www.mdpi.com/1422-0067/25/13/7264.

[2] M.K. Shin, B.-Y. Jang, K.-B. Bu, et al., De novo design of AC-P19M, a novel anticancer peptide with apoptotic effects on lung cancer cells and anti-angiogenic activity, Int. J. Mol. Sci. 23 (24) (2022) http://dx.doi.org/10.3390/ijms232415594, URL https://www.mdpi.com/1422-0067/23/24/15594.

[3] M. Librizzi, C. Martino, M. Mauro, et al., Natural anticancer peptides from marine animal species: Evidence from in vitro cell model systems, Cancers 16 (1) (2024) http://dx.doi.org/10.3390/cancers16010036, URL https://www.mdpi.com/2072-6694/16/1/36.

[4] D. Park, A.T. Magis, R. Li, et al., Novel small-molecule inhibitors of bcl-XL to treat lung cancer, Cancer Res. 73 (17) (2013) 5485–5496, http://dx.doi.org/10.1158/0008-5472.CAN-12-2272.

[5] N. Faraji, A. Doustmohammadi, S.S. Arab, et al., ApInAPDB: a database of apoptosis-inducing anticancer peptides, Sci. Rep. 12 (2022) 21341, http://dx.doi.org/10.1038/s41598-022-25530-6.

[6] R. Rombach, A. Blattmann, D. Lorenz, et al., High-resolution image synthesis with latent diffusion models, in: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, IEEE Computer Society, Los Alamitos, CA, USA, 2022, pp. 10674–10685, http://dx.doi.org/10.1109/CVPR52688.2022.01042, URL https://doi.ieeecomputersociety.org/10.1109/CVPR52688.2022.01042.

[7] Z. Lin, H. Akin, R. Rao, et al., Evolutionary-scale prediction of atomic-level protein structure with a language model, Science 379 (6637) (2023) 1123–1130, http://dx.doi.org/10.1126/science.ade2574.

[8] C. Zhang, X. Zhang, L. Freddolino, et al., BioLiP2: an updated structure database for biologically relevant ligand–protein interactions, Nucleic Acids Res. 52 (D1) (2023) D404–D412, http://dx.doi.org/10.1093/nar/gkad630.

[9] J. Devlin, M. Chang, K. Lee, et al., BERT: pre-training of deep bidirectional transformers for language understanding, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT Minneapolis, MN, USA, Association for Computational Linguistics, 2019, pp. 4171–4186, http://dx.doi.org/10.18653/V1/N19-1423.

[10] P. Veličković, G. Cucurull, A. Casanova, et al., Graph attention networks, in: International Conference on Learning Representations, 2018.
[11] J. Abramson, J. Adler, J. Dunger, et al., Accurate structure prediction of biomolecular interactions with AlphaFold 3, Nature (2024) http://dx.doi.org/10.1038/s41586-024-07487-w, URL https://www.nature.com/articles/s41586-024-07487-w_reference.pdf.
[12] M. Giulini, V. Reys, J.M.C. Teixeira, et al., HADDOCK3: A modular and versatile platform for integrative modeling of biomolecular complexes, J. Chem. Inf. Model. 65 (13) (2025) 7315–7324, http://dx.doi.org/10.1021/acs.jcim.5c00969.
[13] B.K. Kay, L. Castagnoli, Mapping protein-protein interactions with phage-displayed combinatorial peptide libraries, Curr. Protoc. Pharmacol. 17 (2003) http://dx.doi.org/10.1002/0471143030.CB1704S17, URL http://www.ncbi.nlm.nih.gov/pubmed/18228422.
[14] M.G. Kolonin, J. Sun, K.A. Do, C.I. Vidal, Y. Ji, K.A. Baggerly, R. Pasqualini, W. Arap, Synchronous selection of homing peptides for multiple tissues by in vivo phage display, FASEB J. 20 (2006) 979–981, http://dx.doi.org/10.1096/FJ.05-5186FJE, URL https://faseb.onlinelibrary.wiley.com/doi/10.1096/fj.05-5186fje.
[15] L.A. Landon, S.L. Deutscher, S.L. Deutscher, Combinatorial discovery of tumor targeting peptides using phage display, J. Cell. Biochem. 90 (2003) 509–517, http://dx.doi.org/10.1002/JCB.10634, URL https://onlinelibrary.wiley.com/doi/full/10.1002/jcb.10634.
[16] T. Miki, K. Namii, K. Seko, S. Kakehi, G. Moro, H. Mihara, Pattern enrichment analysis for phage selection of stapled peptide ligands, Chem. Sci. 13 (2022) 12634–12642, http://dx.doi.org/10.1039/d2sc04058a, URL https://pubs.rsc.org/en/content/articlepdf/2022/sc/d2sc04058a.
[17] J.X. Huang, S.L. Bishop-Hurley, M.E. Cooper, Development of anti-infectives using phage display: Biological agents against bacteria, viruses, and parasites, Antimicrob. Agents. Chemother. 56 (2012) 4569–4582, http://dx.doi.org/10.1128/AAC.00567-12, URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3421897.
[18] D. Juretić, D. Vukičević, N. Ilić, N. Antcheva, A. Tossi, Computational design of highly selective antimicrobial peptides, J. Chem. Inf. Model. 49 (2009) 2873–2882, http://dx.doi.org/10.1021/CI900327A, URL https://bib.irb.hr/datoteka/437670.2009_JChemInfo.pdf.
[19] J.B. Bhonsle, D. Venugopal, D.P. Huddler, A.J. Magill, R.P. Hicks, Application of 3D-QSAR for identification of descriptors defining bioactivity of antimicrobial peptides, J. Med. Chem. 50 (2007) 6545–6553, http://dx.doi.org/10.1021/JM070884Y, URL https://europepmc.org/article/MED/18062663.
[20] M.R. Borkar, R.R.S. Pissurlenkar, E.C. Coutinho, HomoSAR: bridging comparative protein modeling with quantitative structural activity relationship to design new peptides., J. Comput. Chem. 34 (2013) 2635–2646, http://dx.doi.org/10.1002/JCC.23436, URL https://dblp.uni-trier.de/db/journals/jcc/jcc34.html#BorkarPC13.
[21] S. Basith, B. Manavalan, T. Hwan Shin, et al., Machine intelligence in peptide therapeutics: A next-generation tool for rapid disease screening, Med. Res. Rev. 40 (4) (2020) 1276–1314, http://dx.doi.org/10.1002/med.21658.
[22] G. Agüero-Chapin, D. Galpert-Cañizares, D. Domínguez-Pérez, et al., Emerging computational approaches for antimicrobial peptide discovery, Antibiotics 11 (7) (2022) http://dx.doi.org/10.3390/antibiotics11070936, URL https://www.mdpi.com/2079-6382/11/7/936.
[23] M.D.C. Aguilera-Puga, F. Plisson, Structure-aware machine learning strategies for antimicrobial peptide discovery, Sci. Rep. 14 (1) (2024) 11995, http://dx.doi.org/10.1038/s41598-024-62419-y.
[24] S. Chen, T. Lin, R. Basu, et al., Design of target specific peptide inhibitors using generative deep learning and molecular dynamics simulations, Nat. Commun. 15 (1) (2024) http://dx.doi.org/10.1038/s41467-024-45766-2.
[25] L. Lai, Y. Liu, B. Song, et al., Deep generative models for therapeutic peptide discovery: A comprehensive review, ACM Comput. Surv. 57 (6) (2025) http://dx.doi.org/10.1145/3714455.
[26] M. Defresne, S. Barbe, T. Schiex, Protein Design with Deep Learning, Int. J. Mol. Sci. 22 (21) (2021) http://dx.doi.org/10.3390/ijms222111741, URL https://pmc.ncbi.nlm.nih.gov/articles/PMC8584038/.
[27] S.N. Dean, J.A.E. Alvarez, D. Zabetakis, et al., PepVAE: Variational autoencoder framework for antimicrobial peptide generation and activity prediction, Front. Microbiol. 12 - 2021 (2021) http://dx.doi.org/10.3389/fmicb.2021.725727, URL https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2021.725727.
[28] J. Trinquier, G. Uguzzoni, A. Pagnani, et al., Efficient generative modeling of protein sequences using simple autoregressive models, Nat. Commun. 12 (1) (2021) http://dx.doi.org/10.1038/s41467-021-25756-4.
[29] L. Yang, G. Yang, Z. Bing, et al., Accelerating the discovery of anticancer peptides targeting lung and breast cancers with the wasserstein autoencoder model and PSO algorithm, Brief. Bioinform. 23 (5) (2022) bbac320, http://dx.doi.org/10.1093/bib/bbac320.
[30] D.W. Bin Dai, The usual suspects? Reassessing blame for VAE posterior collapse, in: Proceedings of the 37th International Conference on Machine Learning, ICML '20, JMLR.org, 2020, pp. 2313–2322.
[31] Z. Liu, J. Li, M. Zhu, Alleviating exposure bias for neural machine translation via contextual augmentation and self distillation, IEEE/ACM Trans. Audio Speech Lang. Process. 31 (2023) 2079–2089, http://dx.doi.org/10.1109/TASLP.2023.3277245.
[32] A. Katharopoulos, A. Vyas, N. Pappas, et al., Transformers are RNNs: fast autoregressive transformers with linear attention, in: Proceedings of the 37th International Conference on Machine Learning, JMLR.org, 2020, pp. 5156–5165.
[33] A. Alakhdar, B. Poczos, N. Washburn, Diffusion models in de novo drug design, J. Chem. Inf. Model. 64 (19) (2024) 7238–7256, http://dx.doi.org/10.1021/acs.jcim.4c01107.
[34] C. Shi, C. Wang, J. Lu, et al., Protein sequence and structure co-design with equivariant translation, in: The Eleventh International Conference on Learning Representations, 2023, URL https://openreview.net/forum?id=pRCMXcfdihq.
[35] K.E. Wu, K.K. Yang, R. van den Berg, et al., Protein structure generation via folding diffusion, Nat. Commun. 15 (1059) (2024) http://dx.doi.org/10.1038/s41467-024-45051-2.
[36] R. Wang, T. Wang, L. Zhuo, J. Wei, X. Fu, Q. Zou, Diff-AMP: tailored designed antimicrobial peptide framework with all-in-one generation, identification, prediction and optimization, Brief. Bioinform. 25 (2) (2024) bbae078, http://dx.doi.org/10.1093/bib/bbae078.
[37] L. Tan, L. Wang, X. Ren, Q. Zou, X. Yao, X. Zeng, X. Fu, SQ-DiffuPep: A multimodal information-guided quantitative latent diffusion model for antimicrobial peptide discovery, Inf. Fusion 121 (2025) 103119, http://dx.doi.org/10.1016/j.inffus.2025.103119.
[38] L. Wang, Y. Liu, X. Fu, X. Ye, J. Shi, G.G. Yen, Q. Zou, HMAMP: Designing highly potent antimicrobial peptides using a hypervolume-driven multiobjective deep generative model, J. Med. Chem. 68 (8) (2025) 8346–8360, http://dx.doi.org/10.1021/acs.jmedchem.4c03073.
[39] F. Imrie, T.E. Hadfield, A.R. Bradley, et al., Deep generative design with 3D pharmacophoric constraints, Chem. Sci. 12 (2021) 14577–14589, http://dx.doi.org/10.1039/D1SC02436A, URL http://dx.doi.org/10.1039/D1SC02436A.
[40] W. Zhung, H. Kim, W. Kim, 3D molecular generative framework for interaction-guided drug design, Nat. Commun. 15 (1) (2024) http://dx.doi.org/10.1038/s41467-024-47011-2.
[41] C. Wang, H.H. Ong, S. Chiba, et al., GLDM: hit molecule generation with constrained graph latent diffusion model, Brief. Bioinform. 25 (3) (2024) http://dx.doi.org/10.1093/bib/bbae142.
[42] F. Scarselli, M. Gori, A.C. Tsoi, et al., The graph neural network model, IEEE Trans. Neural Netw. 20 (1) (2009) 61–80, http://dx.doi.org/10.1109/TNN.2008.2005605.
[43] M. Xu, W. Huang, M. Xu, et al., 3D conformational generative models for biological structures using graph information-embedded relative coordinates, Molecules 28 (1) (2023) http://dx.doi.org/10.3390/molecules28010321, URL https://www.mdpi.com/1420-3049/28/1/321.
[44] K. Jha, S. Saha, H. Singh, Prediction of protein–protein interaction using graph neural networks, Sci. Rep. 12 (1) (2022) 8360, http://dx.doi.org/10.1038/s41598-022-12201-9.
[45] X. Xie, P.A. Valiente, J. Kim, P.M. Kim, HelixDiff, a score-based diffusion model for generating all-atom α-helical structures, ACS Central Sci. 10 (5) (2024) 1001–1011, http://dx.doi.org/10.1021/acscentsci.3c01488.
[46] J. Dauparas, I. Anishchenko, N. Bennett, H. Bai, R.J. Ragotte, L.F. Milles, B.I.M. Wicky, A. Courbet, R.J. de Haas, N. Bethel, P.J.Y. Leung, T.F. Huddy, S. Pellock, D. Tischer, F. Chan, B. Koepnick, H. Nguyen, A. Kang, B. Sankaran, A.K. Bera, N.P. King, D. Baker, Robust deep learning–based protein sequence design using proteinmpnn, Science 378 (6615) (2022) 49–56, http://dx.doi.org/10.1126/science.add2187.
[47] T.N.B. Sayuti, W. Conghao, J.C. Rajapakse, Generating apoptosis-inducing anticancer peptides targeting BCL-xL using latent diffusion models on small datasets, in: ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, 2025, pp. 1–5, http://dx.doi.org/10.1109/ICASSP49660.2025.10888225.
[48] S. Rajan, M. Choi, K. Baek, H.S. Yoon, Bh3 induced conformational changes in Bcl-Xl revealed by crystal structure and comparative analysis, Proteins: Struct. Funct. Bioinform. 83 (7) (2015) 1262–1272, http://dx.doi.org/10.1002/prot.24816.
[49] J.M.C. Teixeira, R.V. Honorato, M. Giulini, et al., Haddocking/haddock3: v3.0.0-beta.5 (version 3.0.0-beta.5) [computer software], 2024, http://dx.doi.org/10.5281/zenodo.10527751.
[50] K. Yi, B. Zhou, Y. Shen, et al., Graph denoising diffusion for inverse protein folding, in: Proceedings of the 37th International Conference on Neural Information Processing Systems, NIPS '23, Curran Associates Inc., Red Hook, NY, USA, 2023.
[51] M. Sobieraj, P. Setny, Entropy-based distance cutoff for protein internal contact networks, Proteins 89 (2021) 1333–1339, http://dx.doi.org/10.1002/PROT.26154, URL https://onlinelibrary.wiley.com/doi/10.1002/prot.26154.
[52] O. Vinyals, S. Bengio, M. Kudlur, Order matters: Sequence to sequence for sets, in: International Conference on Learning Representations, 2016, arXiv preprint arXiv:1511.06391.
[53] T.-Y. Lin, P. Goyal, R. Girshick, et al., Focal Loss for Dense Object Detection, IEEE Trans. Pattern Anal. Mach. Intell. 42 (02) (2020) 318–327, http://dx.doi.org/10.1109/TPAMI.2018.2858826, URL https://doi.ieeecomputersociety.org/10.1109/TPAMI.2018.2858826.

[54] T.N. Binte Sayuti, S. Cheng, S. Ajith, A.H. Bin Abdul Samad, J.C. Rajapakse, Enhancing protein language model with feature integration for anticancer peptide prediction, in: 2024 IEEE EMBS International Conference on Biomedical and Health Informatics, BHI, 2024, pp. 1–8, http://dx.doi.org/10.1109/BHI62660.2024.10913878.

[55] A. Merchant, E. Rahimtoroghi, E. Pavlick, et al., What happens to BERT embeddings during fine-tuning? in: Proceedings of the Third BlackboxNLP Workshop on Analyzing and Interpreting Neural Networks for NLP, Association for Computational Linguistics, 2020, pp. 33–44, http://dx.doi.org/10.18653/v1/2020.blackboxnlp-1.4, URL https://aclanthology.org/2020.blackboxnlp-1.4/.

[56] J. Ho, A. Jain, P. Abbeel, Denoising diffusion probabilistic models, in: Proceedings of the 34th International Conference on Neural Information Processing Systems, NIPS '20, Curran Associates Inc., Red Hook, NY, USA, 2020, p. 12.

[57] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional networks for biomedical image segmentation, in: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Springer International Publishing, Munich, Germany, 2015, pp. 234–241.

[58] A. Vaswani, N. Shazeer, N. Parmar, et al., Attention is all you need, in: Advances in Neural Information Processing Systems, vol. 30, Curran Associates, Inc., 2017, URL https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.

[59] K. Cho, B. van Merriënboer, C. Gulcehre, et al., Learning phrase representations using RNN encoder–decoder for statistical machine translation, in: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP, Association for Computational Linguistics, Doha, Qatar, 2014, pp. 1724–1734, http://dx.doi.org/10.3115/v1/D14-1179, URL https://aclanthology.org/D14-1179/.

[60] J. Song, C. Meng, S. Ermon, Denoising diffusion implicit models, in: 9th International Conference on Learning Representations, Austria, 2021, URL https://openreview.net/forum?id=St1giarCHLP.

[61] P. Das, T. Sercu, K. Wadhawan, I. Padhi, S. Gehrmann, F. Cipcigan, V. Chenthamarakshan, H. Strobelt, C. dos Santos, P.-Y. Chen, Y.Y. Yang, J.P.K. Tan, J. Hedrick, J. Crain, A. Mojsilovic, Accelerated antimicrobial discovery via deep generative models and molecular dynamics simulations, Nat. Biomed. Eng. 5 (6) (2021) 613–623, http://dx.doi.org/10.1038/s41551-021-00689-x.

[62] P. Agrawal, D. Bhagat, M. Mahalwal, et al., AntiCP 2.0: an updated model for predicting anticancer peptides, Brief. Bioinform. 22 (3) (2020) bbaa153, http://dx.doi.org/10.1093/bib/bbaa153.

[63] A.S. Rathore, N. Kumar, S. Choudhury, N.K. Mehta, G.P.S. Raghava, Prediction of hemolytic peptides and their hemolytic concentration, Commun. Biol. 8 (1) (2025) 176, http://dx.doi.org/10.1038/s42003-025-07615-w.

[64] E. Perez, F. Strub, H. de Vries, V. Dumoulin, A.C. Courville, FiLM: Visual reasoning with a general conditioning layer, in: AAAI, 2018.

[65] M.K. Manion, J.W. O'Neill, C.D. Giedt, K.M. Kim, K.Y. Zhang, D.M. Hockenbery, Bcl-XL mutations suppress cellular sensitivity to antimycin A*, J. Biol. Chem. 279 (3) (2004) 2159–2165, http://dx.doi.org/10.1074/jbc.M306021200, URL https://www.sciencedirect.com/science/article/pii/S0021925820687798.

[66] G. Lessene, P.E. Czabotar, B.E. Sleebs, K. Zobel, K.N. Lowes, J.M. Adams, J.B. Baell, P.M. Colman, K. Deshayes, W.J. Fairbrother, J.A. Flygare, P. Gibbons, W.J.A. Kersten, S. Kulasegaram, R.M. Moss, J.P. Parisot, B.J. Smith, I.P. Street, H. Yang, D.C.S. Huang, K.G. Watson, Structure-guided design of a selective BCL-XL inhibitor, Nat. Chem. Biol. 9 (6) (2013) 390–397, http://dx.doi.org/10.1038/nchembio.1246.

[67] B. Ku, C. Liang, J.U. Jung, B.-H. Oh, Evidence that inhibition of BAX activation by BCL-2 involves its tight and preferential interaction with the BH3 domain of BAX, Cell Res. 21 (4) (2011) 627–641, http://dx.doi.org/10.1038/cr.2010.149.

[68] E. Fire, S.V. Gullá, R.A. Grant, A.E. Keating, Mcl-1–bim complexes accommodate surprising point mutations via minor structural changes, Prot. Sci. 19 (3) (2010) 507–519, http://dx.doi.org/10.1002/pro.329.