Appendix

A. Details of MMCD

A.1. Diffusion for Peptide Sequence The sequence diffusion process is to define transition probabilities between different discrete states at each step within the Markov chain, ensuring its convergence to the stationary distribution (Austin et al. 2021). In the forward process, residue types are encoded using one-hot encoding with 20 types, and the diffusion noise is represented by transition matrices $(Q^1, Q^2, ..., Q^T)$, where $[Q^t]_{ij}$ signifies the corresponding transition matrix for residue type i to type j at timestep t. Here, we employ marginal transitions (Austin et al. 2021; Ho, Jain, and Abbeel 2020) parametrized by:

$$Q^t = \alpha^t I + (1 - \alpha^t)\mathbf{1}_i m_S \tag{1}$$

where α^t transitions from 1 to 0; the transition matrix from S^0 to S^t is represented as $\overline{Q}^t = Q^1 Q^2 \dots Q^t$, and the noisy statue S^t could be defined as $q(S^t|S^0) = S^0\overline{Q}^t$; m_S is distinct from the uniform transition and represents the marginal distribution (Vignac et al. 2023) for residue type i in our peptide dataset, i.e., $\forall i, \lim_{T\to\infty} \overline{Q}^T \mathbf{1}_i = m_S$. The probability $q(S^{t-1}|S^t, S^0)$ can be computed with Bayes' rule (Vignac et al. 2023):

$$q(S^{t-1} \mid S^t, S^0) \propto S^t (Q^t)' \odot S^0 \overline{Q}^{t-1} \tag{2}$$

where $\odot$ denotes a pointwise product and Q' is the transpose of Q.

In the reverse process, we implement the $\mathcal{F}_s$ with the transformer encoder (i.e., 8 transformer blocks) and an MLP (i.e., 6 fully connected layers) in the network $\hat{p}_\theta$. For the residue representation h_i, we initialize it using residue types (i.e., the one-hot features with 20 classes) and the genetic information (e.g., the features of BLOSUM62 matrix (Styczynski et al. 2008)).

A.2. Diffusion for Peptide Structure In the structure diffusion, atom coordinates of peptide structures are represented as continuous variables in 3D space. Following (Ho, Jain, and Abbeel 2020), we introduce the Gaussian noise into the forward process ($q(\cdot|\cdot)$), defined as follows:

$$q(c_i^t|c_i^{t-1}) = \mathcal{N}(c_i^t|\sqrt{1-\beta^t}\cdot c_i^{t-1}, \beta^t I) \tag{3}$$

$$q(c_i^t|c_i^0) = \mathcal{N}(c_i^t|\sqrt{\overline{\alpha}^t}\cdot c_i^0, (1-\overline{\alpha}^t) I) \tag{4}$$

where c_i^t refers to coordinates for the i-th residue of peptide structure at timestep t, and β^t is the noise rate of structure diffusion at timestep t, which varies between 0 and 1. Formally, $\alpha^t = 1 - \beta^t$, $\overline{\alpha}^t = \prod_{\tau=1}^{t}(1-\beta^\tau)$.

In the reverse process, the EGNN in the network $\mathcal{F}_c$ is implemented as follows. First, we construct a 3D graph for each peptide, where residues serve as nodes and edges are established based on the Euclidean distance between two residues (C_α backbone atom) with a threshold set at less than 5Å. Then, we employ equivariant graph convolutional layers (EGCL) (Satorras, Hoogeboom, and Welling 2021) to learn and update the graph, as follows:

$$h^{l+1}, r^{l+1} = \mathrm{EGCL}[h^l, r^l, \mathcal{E}] \tag{5}$$

where $l = 4$ denotes the number of layers; $h = \{h_1, ..., h_N\}$ represents the set of node embeddings, which are initialized with the physicochemical data obtained from the AAIndex database (Kawashima et al. 2008) and N denotes the number of nodes; $r = \{r_1, ..., r_N\}$ is the set of residue coordinates; $\mathcal{E} = (e_{ij})$, e_{ij} denotes the feature (e.g., the distance-dependent statistical potential (Simons et al. 1999) from AAIndex features) of the edge between node i and j. Finally, we apply an attention pooling layer over the updated embeddings of all nodes to produce a summary embedding for the entire 3D graph. The MLP in the network $\mathcal{F}_c$ is implemented by the 6 fully connected layers.

A.3. Diffusion Objective Through optimizing the diffusion model for peptide sequence and structure with the ELBO loss (Anand and Achim 2022), we define the objective of the diffusion process as follows:

$$\mathbb{E}[D_{KL}(q(X^{t-1}|X^t, X^0)||p_\theta(X^{t-1}|X^t))] \tag{6}$$

With the independence assumptions for peptides, the objective can be decomposed into two components, i.e., the sequence loss and the structure loss (Anand and Achim 2022):

$$\mathcal{L}_S = \mathbb{E}[D_{KL}(q(S^{t-1}|S^t, S^0)||p_\theta(S^{t-1}|S^t))] \tag{7}$$

$$\mathcal{L}_C = \mathbb{E}[D_{KL}(q(C^{t-1}|C^t, C^0)||p_\theta(C^{t-1}|C^t))] \tag{8}$$

A.4. Multi-Modal Contrastive Learning Strategy In the Inter-CL and Intra-CL loss functions, $E(\cdot,\cdot)$ is a cosine similarity function to measure mutual information, for example:

$$E(i,j) = exp(\frac{\mathcal{S}_i \cdot \mathcal{C}_j}{\tau \|\mathcal{S}_i\| \|\mathcal{C}_j\|}) \tag{9}$$

where $\mathcal{S}_i, \mathcal{C}_j$ represent the embeddings for sequences and structures of peptides, and τ is a temperature coefficient.

A.5. Model Sampling for peptide generation Based on the well-trained MMCD model, we perform the co-generation for the sequences and structures of the therapeutic peptides. First, we sample the residue types from the marginal distribution and the Cartesian coordinates of residues with four backbone atoms (N-C_α-C-O) from the Gaussian distribution. Next, we iteratively predict the noises at diffusion timestep t and denoise residue types and Cartesian coordinates of residues until $t = 0$. Furthermore, to build the all-atom 3D structure, we construct the side-chain of residues according to the local coordinates relative to the C_α position and orientation of each amino acid. And the rest of the side-chain atoms are constructed using the PeptideBulider tool (Tien et al. 2013).

B. Details of Experimental setups

B.1. Implementation Details The optimal performance of MMCD is achieved with the following hyperparameter configurations. We use a linear schedule for variances β_t with the lowest variance $\beta_1 = 1e-7$ and the highest variance $\beta_T = 2e-2$. The number of diffusion timesteps is set as 1000. The temperature coefficient τ for contrastive learning is denoted as 0.1. Moreover, we employ the Adam optimizer with a learning rate of 5e-3 to optimize the model. The number of training epochs is set as 500. We set the

hyper-parameter α as 0.9 to balance the contributions of the diffusion process for peptide generation and multi-modal contrastive tasks. All experiments are conducted on the same machine with Intel(R) Core(TM) i9-7900X CPU @ 3.30GHz and 2 GPUs (NVIDIA GeForce 3080Ti).

B.2. Datasets Following previous studies (Thi Phan et al. 2022; Zhang et al. 2023a), we collected the peptide datasets containing two distinct types, i.e., antimicrobial/anticancer peptide. For the antimicrobial dataset, therapeutic samples were derived from manually curated antimicrobial peptide (AMP) databases, combining experimentally validated peptides from 7 known AMP databases, including APD3 (Wang, Li, and Wang 2016), CAMP (Gawde et al. 2022), DBAMP (Jhong et al. 2022), DRAMP (Shi et al. 2022), SATPdb (Singh et al. 2016), YADAMP (Piotto et al. 2012), and LAMP (Zhao et al. 2013). For the anticancer dataset, therapeutic samples were extracted from 6 databases, including CancerPPD (Tyagi et al. 2015), APD3, SATPdb, DRAMP, LAMP, and DBAMP. Notably, any duplicate sequences were removed from the peptide datasets. The non-therapeutic samples were assumed to be biologically inactive for antimicrobial and anticancer types, and they were manually obtained from the UniProt database. To enhance the diversity of non-therapeutic samples, we applied the CD-hit tool (Fu et al. 2012) to remove sequences with similarity scores of $\geq 40\%$.

In addition to the peptide sequence data, the peptide structures are also necessary for this study. For the peptides with missing structures, we predicted their structures using Rosetta-based computational tools (Chaudhury, Lyskov, and Gray 2010), to ensure comprehensive coverage of structural information for the AMP and ACP datasets.

B.3. Baselines For the sequence generation, we compared our method against several methods:

- **LSTM-RNN** (Müller, Hiss, and Schneider 2018) was the RNN-based model to capture patterns in sequential data of peptides and generate new peptides from the learned context.
- **AMPGAN** (Oort et al. 2021) utilized the conditional GAN to generate new AMP sequences from the underlying peptide distribution, and took the target microbes, target mechanisms, and MIC50 (minimal inhibit concentration for 50%) level of AMP as the conditioning variables.
- **WAE-PSO** (Yang et al. 2022) combined a wasserstein autoencoder generative model and a particle swarm optimization forward search algorithm guided by anticancer attribute predictive model, to generate new ACPs.
- **HydrAMP** (Szymczak et al. 2023) was a VAE-based model with the MIC classifier of AMP, which learns the hidden space of meaningful peptides and enables the unconstrained generation of AMP sequences.

For the structure generation, we compared our method with:

- **APPTEST** (Timmons and Hewage 2021) employed a neural network architecture and simulated annealing methods for predicting peptide tertiary structures from their primary sequences.
- **FoldingDiff** (Wu et al. 2022) was a diffusion-based method, capturing the orientation noise of constituent residues in the diffusion process for the generation of peptide structures.
- **ProtDiff** (Trippe et al. 2023) was a diffusion-based method, leveraging an E(3) equivariant graph neural network to learn a diverse distribution over longer backbone structures.

Moreover, diffusion-based methods for sequence and structure co-design were utilized for the comparison separately in the sequence and structure generation.

- **DiffAB** (Luo et al. 2022) designed sequences, coordinates, and side-chain orientations for residues, achieving a diffusion model with equivariant rotation and translation for sequence-structure co-design.
- **SimDiff** (Zhang et al. 2023b) incorporated self-supervised learning techniques for peptide structures and constrained the mutual information of different conformers generated from the same protein under the diffusion model.

Notably, all the baseline methods were implemented by their publicly available source codes. The best or default parameters of each method were used for peptide generation.

B.4. Metrics The experimental results for the two peptide generation tasks, namely sequence/structure generation, are evaluated using the following metrics. For the sequence generation:

- **Similarity** score involves assessing the alignment score between the generated peptide sequences and the existing sequences in the corresponding peptide dataset (AMP/ACP). A lower alignment score indicates better novelty in the generated peptide sequences. This score is calculated using the PairwiseAligner from the biopython software package (Cock et al. 2009), with BLOSUM62 as the alignment scoring matrix.
- **Instability** score is a measure of peptide stability based on the amino acid composition of the generated sequences. The instability score is evaluated using the PeptideDescriptor from the modlAMP software package (Müller et al. 2017).
- **Antimicrobial** score quantifies the likelihood of AMP activity within the generated sequence. Assessment of this score has been conducted utilizing the CAMP tool (Gawde et al. 2022).
- **Anticancer** score reflects the degree of confidence in ACP activity associated with the generated sequence. Evaluation of this score has been accomplished through the utilization of the AntiCP tool (Agrawal et al. 2021).

For the structure generation:

- **Ramachandran** score assesses the energetically favorable conformations of the protein backbone by considering the allowed region. It helps gauge the overall quality and reliability of a protein structure (Hollingsworth and Karplus 2010). We extracted the allowed region scores of the generated structures using the ProCheck tool (Laskowski et al. 1996).

Methods	AMP			ACP	
	Ramachandran↑	RMSD↓	Docking↑	Ramachandran↑	RMSD↓
MMCD (w/o InterCL & IntraCL)	72.1283	2.3642	1561	72.3791	2.3681
MMCD (w/o IntraCL)	78.1142	1.9719	1681	77.4269	2.1765
MMCD (w/o InterCL)	75.2913	2.0621	1669	76.0116	2.2451
MMCD	80.4661	1.8278	1728	78.2157	2.0847

Table 1: Ablation study on the structure-level generation task.

- **RMSD** score (the root-mean-square deviation for C_α) is a measure of the structural similarity between the generated peptide structure and the original peptide structure. It quantifies the deviation between the two structures and indicates how close the generated structure is to the original one.
- **Docking** score is used to evaluate the interaction of antimicrobial peptides with bacterial membrane proteins. Since bacterial membrane proteins are the primary targets of AMP, a better docking effect implies a more promising antimicrobial property (Flórez-Castillo et al. 2020). This evaluation is performed using the ZDOCK tool (Pierce et al. 2014). Notably, due to the complex biological mechanisms of anticancer peptides, there are no special target proteins for ACP and thus the docking score is unsuitable for ACP (Kurrikoff, Aphkhazava, and Langel 2019).

C. Details of Experimental Results

C.1. Ablation study Table 1 indicates the structure-level comparisons between MMCD with its variants on both AMP and ACP datasets.

C.2. Peptide-docking analysis For therapeutic peptides, the biological activity stems from their specific binding interactions (docking) with target proteins. To assess the validity of the generated peptide structures, we conducted evaluations on the docking results with target proteins and the localized amino acid patterns of structures. The reference peptide with length 25 (sequence: FKCRRWQWRMKKLGAPSITCVRRAF) is randomly selected from the AMP dataset (antimicrobial peptides). The methods (i.e., MMCD, SimDiff, DiffAB, FoldingDiff, and ProtDiff) are employed to generate corresponding structures based on the sequence of the reference peptide. Herein, we use the lipopolysaccharide structure (PDB ID: 6MI7 (Li, Orlando, and Liao 2019)) on the outer membrane of bacteria as the target protein for molecular docking, and the docking score is evaluated with the ZDOCK tool (Pierce et al. 2014). The residues within a 5Å proximity between peptides (i.e., the reference and generated structures) and the active pocket of target protein in docking complexes are extracted using the Pymol tool (Version 1.8.4 open-source) to visualize their binding residues (Miller et al. 2021).

References

Agrawal, P.; Bhagat, D.; Mahalwal, M.; Sharma, N.; and Raghava, G. P. S. 2021. AntiCP 2.0: An Updated Model for Predicting Anticancer Peptides. *Briefings in Bioinformatics*, 22(3): bbaa153.

Anand, N.; and Achim, T. 2022. Protein Structure and Sequence Generation with Equivariant Denoising Diffusion Probabilistic Models. arxiv:2205.15019.

Austin, J.; Johnson, D. D.; Ho, J.; Tarlow, D.; and van den Berg, R. 2021. Structured Denoising Diffusion Models in Discrete State-Spaces. In *Advances in Neural Information Processing Systems*, volume 34, 17981–17993. Curran Associates, Inc.

Chaudhury, S.; Lyskov, S.; and Gray, J. J. 2010. PyRosetta: A Script-Based Interface for Implementing Molecular Modeling Algorithms Using Rosetta. *Bioinformatics*, 26(5): 689–691.

Cock, P. J. A.; Antao, T.; Chang, J. T.; Chapman, B. A.; Cox, C. J.; Dalke, A.; Friedberg, I.; Hamelryck, T.; Kauff, F.; Wilczynski, B.; and de Hoon, M. J. L. 2009. Biopython: Freely Available Python Tools for Computational Molecular Biology and Bioinformatics. *Bioinformatics*, 25(11): 1422–1423.

Flórez-Castillo, J. M.; Rondón-Villareal, P.; Ropero-Vega, J. L.; Mendoza-Espinel, S. Y.; Moreno-Amézquita, J. A.; Méndez-Jaimes, K. D.; Farfán-García, A. E.; Gómez-Rangel, S. Y.; and Gómez-Duarte, O. G. 2020. Ib-M6 Antimicrobial Peptide: Antibacterial Activity against Clinical Isolates of Escherichia Coli and Molecular Docking. *Antibiotics*, 9(2): 79.

Fu, L.; Niu, B.; Zhu, Z.; Wu, S.; and Li, W. 2012. CD-HIT: Accelerated for Clustering the next-Generation Sequencing Data. *Bioinformatics*, 28(23): 3150–3152.

Gawde, U.; Chakraborty, S.; Waghu, F. H.; Barai, R. S.; Khanderkar, A.; Indraguru, R.; Shirsat, T.; and Idicula-Thomas, S. 2022. CAMPR4: A Database of Natural and Synthetic Antimicrobial Peptides. *Nucleic Acids Res*, 51(D1): D377–D383.

Ho, J.; Jain, A.; and Abbeel, P. 2020. Denoising Diffusion Probabilistic Models. In *Advances in Neural Information Processing Systems*, volume 33, 6840–6851. Curran Associates, Inc.

Hollingsworth, S. A.; and Karplus, P. A. 2010. A Fresh Look at the Ramachandran Plot and the Occurrence of Standard Structures in Proteins. 1(3-4): 271–283.

Jhong, J.-H.; Yao, L.; Pang, Y.; Li, Z.; Chung, C.-R.; Wang, R.; Li, S.; Li, W.; Luo, M.; Ma, R.; Huang, Y.; Zhu, X.; Zhang, J.; Feng, H.; Cheng, Q.; Wang, C.; Xi, K.; Wu, L.-C.; Chang, T.-H.; Horng, J.-T.; Zhu, L.; Chiang, Y.-C.;

Wang, Z.; and Lee, T.-Y. 2022. dbAMP 2.0: Updated Resource for Antimicrobial Peptides with an Enhanced Scanning Method for Genomic and Proteomic Data. *Nucleic Acids Res*, 50(D1): D460–D470.

Kawashima, S.; Pokarowski, P.; Pokarowska, M.; Kolinski, A.; Katayama, T.; and Kanehisa, M. 2008. AAindex: Amino Acid Index Database, Progress Report 2008. *Nucleic Acids Res*, 36(Database issue): D202–205.

Kurrikoff, K.; Aphkhazava, D.; and Langel, Ü. 2019. The future of peptides in cancer treatment. *Current Opinion in Pharmacology*, 47: 27–32.

Laskowski, R. A.; Rullmann, J. A. C.; MacArthur, M. W.; Kaptein, R.; and Thornton, J. M. 1996. AQUA and PROCHECK-NMR: Programs for Checking the Quality of Protein Structures Solved by NMR. *J Biomol Nmr*, 8(4): 477–486.

Li, Y.; Orlando, B. J.; and Liao, M. 2019. Structural Basis of Lipopolysaccharide Extraction by the LptB2FGC Complex. *Nature*, 567(7749): 486–490.

Luo, S.; Su, Y.; Peng, X.; Wang, S.; Peng, J.; and Ma, J. 2022. Antigen-Specific Antibody Design and Optimization with Diffusion-Based Generative Models for Protein Structures.

Miller, E. B.; Murphy, R. B.; Sindhikara, D.; Borrelli, K. W.; Grisewood, M. J.; Ranalli, F.; Dixon, S. L.; Jerome, S.; Boyles, N. A.; Day, T.; Ghanakota, P.; Mondal, S.; Rafi, S. B.; Troast, D. M.; Abel, R.; and Friesner, R. A. 2021. Reliable and Accurate Solution to the Induced Fit Docking Problem for Protein–Ligand Binding. *J Chem Theory Comput*, 17(4): 2630–2639.

Müller, A. T.; Gabernet, G.; Hiss, J. A.; and Schneider, G. 2017. modlAMP: Python for Antimicrobial Peptides. *Bioinformatics*, 33(17): 2753–2755.

Müller, A. T.; Hiss, J. A.; and Schneider, G. 2018. Recurrent Neural Network Model for Constructive Peptide Design. *J Chem Inf Model*, 58(2): 472–479.

Oort, C. M. V.; Ferrell, J. B.; Remington, J. M.; Wshah, S.; and Li, J. 2021. AMPGAN v2: Machine Learning Guided Design of Antimicrobial Peptides.

Pierce, B. G.; Wiehe, K.; Hwang, H.; Kim, B.-H.; Vreven, T.; and Weng, Z. 2014. ZDOCK Server: Interactive Docking Prediction of Protein–Protein Complexes and Symmetric Multimers. *Bioinformatics*, 30(12): 1771–1773.

Piotto, S. P.; Sessa, L.; Concilio, S.; and Iannelli, P. 2012. YADAMP: Yet Another Database of Antimicrobial Peptides. *Int J Antimicrob Ag*, 39(4): 346–351.

Satorras, V. G.; Hoogeboom, E.; and Welling, M. 2021. E (n) equivariant graph neural networks. In *International conference on machine learning*, 9323–9332. PMLR.

Shi, G.; Kang, X.; Dong, F.; Liu, Y.; Zhu, N.; Hu, Y.; Xu, H.; Lao, X.; and Zheng, H. 2022. DRAMP 3.0: An Enhanced Comprehensive Data Repository of Antimicrobial Peptides. *Nucleic Acids Res*, 50(D1): D488–D496.

Simons, K. T.; Ruczinski, I.; Kooperberg, C.; Fox, B. A.; Bystroff, C.; and Baker, D. 1999. Improved recognition of native-like protein structures using a combination of sequence-dependent and sequence-independent features of proteins. *Proteins*, 34(1): 82–95.

Singh, S.; Chaudhary, K.; Dhanda, S. K.; Bhalla, S.; Usmani, S. S.; Gautam, A.; Tuknait, A.; Agrawal, P.; Mathur, D.; and Raghava, G. P. 2016. SATPdb: A Database of Structurally Annotated Therapeutic Peptides. *Nucleic Acids Res*, 44(D1): D1119–D1126.

Styczynski, M. P.; Jensen, K. L.; Rigoutsos, I.; and Stephanopoulos, G. 2008. BLOSUM62 miscalculations improve search performance. *Nature biotechnology*, 26(3): 274–275.

Szymczak, P.; Możejko, M.; Grzegorzek, T.; Jurczak, R.; Bauer, M.; Neubauer, D.; Sikora, K.; Michalski, M.; Sroka, J.; Setny, P.; Kamysz, W.; and Szczurek, E. 2023. Discovering Highly Potent Antimicrobial Peptides with Deep Generative Model HydrAMP. *Nat Commun*, 14(1): 1453.

Thi Phan, L.; Woo Park, H.; Pitti, T.; Madhavan, T.; Jeon, Y.-J.; and Manavalan, B. 2022. MLACP 2.0: An Updated Machine Learning Tool for Anticancer Peptide Prediction. *Comput Struct Biotec*, 20: 4473–4480.

Tien, M. Z.; Sydykova, D. K.; Meyer, A. G.; and Wilke, C. O. 2013. PeptideBuilder: A Simple Python Library to Generate Model Peptides. *Peerj*, 1: e80.

Timmons, P. B.; and Hewage, C. M. 2021. APPTEST Is a Novel Protocol for the Automatic Prediction of Peptide Tertiary Structures. *Brief Bioinform*, 22(6): bbab308.

Trippe, B. L.; Yim, J.; Tischer, D.; Baker, D.; Broderick, T.; Barzilay, R.; and Jaakkola, T. 2023. Diffusion Probabilistic Modeling of Protein Backbones in 3D for the Motif-Scaffolding Problem. arxiv:2206.04119.

Tyagi, A.; Tuknait, A.; Anand, P.; Gupta, S.; Sharma, M.; Mathur, D.; Joshi, A.; Singh, S.; Gautam, A.; and Raghava, G. P. S. 2015. CancerPPD: A Database of Anticancer Peptides and Proteins. *Nucleic Acids Res*, 43(Database issue): D837–843.

Vignac, C.; Krawczuk, I.; Siraudin, A.; Wang, B.; Cevher, V.; and Frossard, P. 2023. DiGress: Discrete Denoising Diffusion for Graph Generation. arxiv:2209.14734.

Wang, G.; Li, X.; and Wang, Z. 2016. APD3: The Antimicrobial Peptide Database as a Tool for Research and Education. *Nucleic Acids Res*, 44(D1): D1087–D1093.

Wu, K. E.; Yang, K. K.; van den Berg, R.; Zou, J. Y.; Lu, A. X.; and Amini, A. P. 2022. Protein Structure Generation via Folding Diffusion. arxiv:2209.15611.

Yang, L.; Yang, G.; Bing, Z.; Tian, Y.; Huang, L.; Niu, Y.; and Yang, L. 2022. Accelerating the Discovery of Anticancer Peptides Targeting Lung and Breast Cancers with the Wasserstein Autoencoder Model and PSO Algorithm. *Brief Bioinform*, 23(5): bbac320.

Zhang, H.; Saravanan, K. M.; Wei, Y.; Jiao, Y.; Yang, Y.; Pan, Y.; Wu, X.; and Zhang, J. Z. H. 2023a. Deep Learning-Based Bioactive Therapeutic Peptide Generation and Screening. *J Chem Inf Model*, 63(3): 835–845.

Zhang, Z.; Xu, M.; Lozano, A.; Chenthamarakshan, V.; Das, P.; and Tang, J. 2023b. Pre-Training Protein Encoder via Siamese Sequence-Structure Diffusion Trajectory Prediction. arxiv:2301.12068.

Zhao, X.; Wu, H.; Lu, H.; Li, G.; and Huang, Q. 2013. LAMP: A Database Linking Antimicrobial Peptides. *PLoS One*, 8(6): e66557.