

# 基于多模态对比扩散的抗菌肽生成模型

祁志霄, 廖俊\*

中国药科大学理学院, 江苏 南京

收稿日期: 2026年5月24日; 录用日期: 2026年6月17日; 发布日期: 2026年6月23日

## 摘要

目的: 构建一种融合序列、结构与表面信息的抗菌肽生成模型, 以提高候选肽的新颖性、稳定性和功能相关性。方法: 整合公开抗菌肽数据库资源, 经序列标准化、去冗余、结构补全和质量控制后构建20104条抗菌肽样本, 并配对构建等量非抗菌肽样本。提出基于多模态对比扩散的抗菌肽生成模型SSCD (Sequence-Structure-Surface Contrastive Diffusion), 将20维氨基酸独热编码、替换矩阵与疏水性特征组成序列表征, 将主链坐标、AAindex理化描述符和残基层面表面几何拓扑特征组成结构表征; 序列分支采用Transformer进行离散扩散去噪, 结构分支采用等变图神经网络进行连续坐标去噪, 并引入跨模态对比学习和能量引导采样。结果: SSCD在Similarity、Instability、Antimicrobial和Docking score上分别达到24.93、40.05、0.87和1740, 整体优于LSTM-RNN、AMPGAN、HydrAMP和DiffAB等模型。结论: SSCD能够在统一框架下联合建模抗菌肽序列语义、三维骨架和表面特征, 提升结构合理且功能相关的抗菌肽候选物的生成, 为抗菌肽的筛选与后续研究提供了一种可行的计算策略。

## 关键词

抗菌肽, 扩散模型, 多模态学习, 对比学习

# An Antimicrobial Peptide Generation Model Based on Multimodal Contrastive Diffusion

Zhixiao Qi, Jun Liao\*

School of Science, China Pharmaceutical University, Nanjing Jiangsu

Received: May 24, 2026; accepted: June 17, 2026; published: June 23, 2026

## Abstract

**Objective:** To develop an antimicrobial peptide generation model based on multimodal contrastive

\*通讯作者。

diffusion by integrating peptide sequence, backbone structure and residue-level surface descriptors, thereby improving the novelty, stability and functional relevance of generated candidates. **Methods:** Public antimicrobial peptide resources were integrated and subjected to sequence normalization, redundancy removal, structural completion and quality control. A dataset containing 20,104 antimicrobial peptide samples was constructed, together with an equal number of non-antimicrobial peptide samples. A multimodal contrastive diffusion model termed SSSCD (Sequence-Structure-Surface Contrastive Diffusion) was proposed. Sequence representations were constructed using amino acid one-hot encodings, substitution matrices and hydrophobicity-related features, whereas structural representations incorporated backbone coordinates, AAindex physicochemical descriptors and residue-level surface geometric-topological features. A Transformer-based discrete diffusion branch was employed for sequence denoising, while an equivariant graph neural network was used for continuous coordinate denoising. Cross-modal contrastive learning and energy-guided sampling were further introduced to improve sequence-structure consistency and local geometric rationality. **Results:** SSSCD achieved Similarity, Instability, Antimicrobial and Docking scores of 24.93, 40.05, 0.87 and 1740, respectively, outperforming LSTM-RNN, AMPGAN, HydrAMP and DiffAB. **Conclusion:** SSSCD effectively integrates sequence semantics, backbone geometry and surface-level descriptors within a unified contrastive diffusion framework. The model improves the generation of structurally plausible and functionally relevant antimicrobial peptide candidates, providing a feasible computational strategy for antimicrobial peptide generation and downstream screening.

## Keywords

Antimicrobial Peptide, Diffusion Model, Multimodal Learning, Contrastive Learning

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

耐药菌感染已成为全球公共卫生和药物研发领域的重要挑战。抗菌肽具有分子量较小、结构可设计性强、作用机制多样和不易完全依赖单一靶点等特点,被认为是传统抗生素的重要补充方向之一[1]。CAMP 和 DRAMP 等数据库的持续更新,也为抗菌肽序列整理、功能注释和计算设计提供了重要数据基础[2][3]。然而,天然抗菌肽数量有限,传统筛选依赖实验验证,存在周期长、成本高和候选空间覆盖不足等问题。近年来,深度生成模型已被用于抗菌肽候选序列设计,并逐渐推动抗菌肽发现由经验筛选向数据驱动设计转变[4]。在此基础上,扩散模型进一步扩展了复杂分布建模能力,为抗菌肽从头设计和结构约束优化提供了新的技术路线[5]。

现有抗菌肽生成方法多以序列为主要输入,能够学习氨基酸组成、局部片段和简单理化属性,但往往难以同时约束三维骨架、表面暴露环境和潜在结合界面。近年来,多视图与多模态分子表征方法已被用于药物性质预测任务,并显示出融合 1D、2D 与 3D 信息能够提高模型对复杂生物功能的建模能力[6]。然而,在抗菌肽生成任务中,如何进一步联合序列、空间构象与局部表面环境进行协同建模,仍有待深入研究。已有研究表明,生成对抗网络能够用于抗菌肽候选序列设计,但其结构层面的约束能力仍有进一步提升空间。对于抗菌肽这类兼具阳离子性、两亲性和构象柔性的分子,仅从一维序列层面生成候选物,容易造成序列可读但结构或功能证据不足的问题。扩散模型在复杂分布建模方面表现出较强能力,为分子和蛋白结构生成提供了新的方法基础。同时,等变图神经网络能够在三维空间中保持旋转和平移等变性,适合用于肽链骨架构象建模[7]。

本研究以抗菌肽为对象, 提出一种基于多模态对比扩散的抗菌肽生成模型 SSSCD (Sequence-Structure-Surface Contrastive Diffusion)。该模型在同一框架下联合建模氨基酸序列、主链三维结构和残基层面表面特征, 通过序列离散扩散、结构连续扩散、跨模态对比学习和能量引导采样, 提高生成候选肽的新颖性、稳定性、抗菌活性倾向和结构合理性。

## 2. 材料与方法

### 2.1. 数据来源与预处理

本研究从公开抗菌肽数据库和综合蛋白质数据库中收集抗菌肽相关数据。其中, APD 和 CAMP 提供了大量经整理的抗菌肽序列及功能注释信息。DRAMP 和 dbAMP 进一步整合了多来源抗菌肽数据, 并提供结构、活性及机制相关注释[3][8]。UniProt 作为综合蛋白质知识库, 则用于补充蛋白质序列及功能背景信息[9]。对原始样本进行统一预处理: 仅保留由 20 种标准氨基酸组成的序列, 删除含非法字符、缺失信息或明显异常长度的样本; 对跨库重复序列进行去冗余处理; 对缺少三维结构信息的样本, 采用 AlphaFold3 相关结构预测策略补充主链结构, 并进行基本质量控制[10]。最终构建得到 20,104 条抗菌肽样本。为支持对比学习训练, 进一步配对构建等量非抗菌肽样本, 形成正负样本均衡的数据集。

数据集理化性质统计结果见表 1。所得抗菌肽数据集序列长度主要集中于 9~25 aa, 主峰约 13 aa, 平均长度为 20.69 aa。氨基酸组成方面, K、L、R、G、A 和 I 等残基占比较高, K 和 R 的富集提示数据集具有明显阳离子性, 同时保留一定疏水性, 符合典型抗菌肽依赖静电吸附和膜相互作用发挥功能的基本特征。

**Table 1.** Physicochemical properties of the antimicrobial peptide dataset

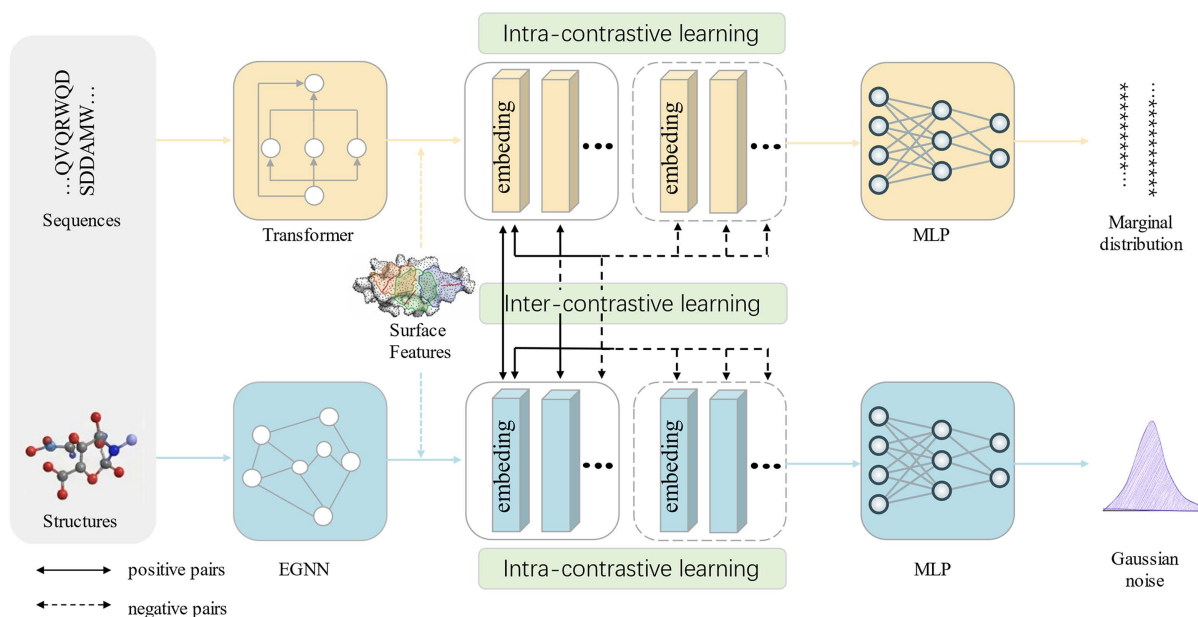
**表 1.** 抗菌肽数据集理化性质

指标	统计量	数值
碱性氨基酸比例	Mean	0.2507
酸性氨基酸比例	Mean	0.0418
极性氨基酸比例	Mean	0.1211
疏水性氨基酸比例	Mean	0.4689
平均 pI	Mean	9.4950
平均电荷	Mean	2.8596
平均疏水性	Mean	-0.2207
正电荷序列占比	Percentage	82.59%

### 2.2. SSSCD 模型总体架构

如图 1 所示, SSSCD 采用双分支扩散建模+跨模态对比对齐的总体结构。序列分支面向离散氨基酸符号, 利用 Transformer 学习序列上下文依赖并完成离散扩散去噪; Transformer 的自注意力机制适合捕获序列中不同残基位置之间的长程依赖关系[11]。结构分支面向连续三维坐标, 利用等变图神经网络 (equivariant graph neural network, EGNN) 学习肽链主链构象的去噪恢复过程; EGNN 能够在坐标更新过程中保持三维几何等变性, 适合用于分子构象建模[7]。在输入层面, 模型同时引入序列特征、结构特征和

表面特征; 在训练层面, 通过扩散损失学习真实抗菌肽分布, 通过类内对比损失区分抗菌肽与非抗菌肽, 通过跨模态匹配损失约束同一抗菌肽的序列表示和结构表示保持一致。



**Figure 1.** Overall architecture of SSSCD model

**图 1.** SSSCD 模型整体架构

### 2.3. 多模态特征构建

序列特征采用残基级表示。对于长度为  $L$  的肽序列, 每个残基由 4 类 20 维向量拼接而成, 包括氨基酸 one-hot 编码、BLOSUM62 替换得分、PAM120 替换得分和 hydrophobicity 属性向量, 形成 80 维序列嵌入; 其中疏水性刻画参考经典氨基酸亲疏水性分析思想[12]。该表示既保留氨基酸类别身份, 又引入残基替换关系、进化相似性和疏水性等功能相关先验。

结构特征以肽链主链坐标为核心。每个残基保留  $C\alpha$ 、C、N 和 O 共 4 个主链原子的三维坐标, 整条肽链可表示为  $L \times 4 \times 3$  的坐标张量。模型以  $C\alpha$  原子作为空间锚点构建残基图, 一方面根据空间距离建立局部邻接边, 另一方面显式加入相邻残基之间的顺序边, 以维持肽链拓扑连续性。节点属性采用 46 维 AAindex 及相关理化描述符, 覆盖疏水性、电荷、极性、体积、柔性、二级结构倾向和表面暴露倾向等信息。

表面特征用于补充分子局部几何和拓扑环境。当前实现并未额外构建显式表面网格, 而是基于主链坐标近似提取残基层面表面描述符, 包括平均曲率、高斯曲率、形状指数、表面法向、距质心距离、近似溶剂可及性、局部密度、局部二面角、序列相对位置、到 N 端距离、到 C 端距离和二级结构预测结果等。序列、结构和表面三类信息共同构成残基层面的多模态输入。

上述表面描述符的设计主要基于抗菌肽功能与局部空间暴露环境之间的关系。已有研究表明, 抗菌肽的膜吸附、膜插入及界面识别过程不仅依赖氨基酸组成, 还与局部表面曲率、空间暴露程度及几何拓扑状态密切相关。平均曲率、高斯曲率和形状指数主要用于刻画局部表面几何形态, 可反映凸起、凹陷及平面区域等不同空间结构特征; 表面法向和局部二面角用于描述局部空间方向性与骨架弯曲状态; 近似溶剂可及性、局部密度以及距质心距离则用于反映残基暴露程度与空间分布特征; 序列相对位置及到 N/C 端距离用于补充肽链拓扑位置信息。上述特征共同构成对残基层面局部几何环境与潜在功能界面的近似描述。

## 2.4. 扩散生成与跨模态对比学习

序列分支采用离散扩散策略。前向过程中, 模型依据时间步构造类别转移矩阵, 逐步扰动原始氨基酸分布; 反向过程中, Transformer 根据带噪序列、位置嵌入和时间步嵌入预测每个位置的氨基酸概率分布, 并通过 KL 散度约束预测分布与真实序列分布的差异。

结构分支采用连续高斯扩散策略。前向过程中, 在主链坐标上逐步加入高斯噪声; 反向过程中, EGNN 根据节点特征、边特征、空间距离和时间步编码估计噪声项, 在保持旋转和平移等变性的前提下恢复合理主链结构。训练时, 结构分支以噪声预测均方误差为核心目标, 并结合键长、键角、肽键平面、Ca 距离矩阵、Ramachandran 先验和原子碰撞惩罚等辅助约束, 以提高结构物理合理性。

为增强序列与结构之间的语义一致性, SSSCD 在两条分支上分别进行自注意力池化, 获得序列级和结构级全局表示。模型设置判别投影头和匹配投影头: 判别空间用于区分抗菌肽和非抗菌肽, 匹配空间用于拉近同一抗菌肽样本的序列表示和结构表示。总体损失由序列扩散损失、结构扩散损失、几何约束损失、类内对比损失和跨模态匹配损失共同组成, 并采用动态权重平衡不同损失项。

## 2.5. 能量引导采样

在反向采样阶段, SSSCD 进一步引入能量引导机制。扩散模型首先生成当前时间步的无噪结构估计, 然后根据可微能量函数计算梯度并对坐标进行校正。能量函数主要包含键长约束、键角约束、肽键平面约束、Ramachandran 先验和原子碰撞惩罚。该机制不改变扩散主干网络, 而是在采样后半程对局部几何进行校正, 从而降低键长偏移、构象扭曲和非物理碰撞等问题。

## 2.6. 评价指标

模型性能从序列新颖性、稳定性、抗菌活性倾向和结构相互作用潜力 4 个方面进行评价。Similarity 采用 Biopython PairwiseAligner 结合 BLOSUM62 替换矩阵计算生成序列与参考抗菌肽库的平均对齐得分, 数值越低表示与已有序列差异越大[13]。Instability 用于衡量序列不稳定性倾向, 数值越低表示稳定性越好。Antimicrobial 采用外部抗菌肽预测工具输出的活性评分, 数值越高表示抗菌活性倾向越强。Docking score 采用 ZDOCK 评估生成肽与固定受体的潜在相互作用能力[14]。本研究选择 PDB 6MI7 中 LptF/LptG 跨膜核心区域作为受体, 该结构来源于大肠杆菌 LptB<sub>2</sub>FGC 复合物, 与革兰阴性菌脂多糖转运和外膜屏障形成密切相关[15]。

# 3. 结果

## 3.1. 对比实验结果

为系统评估 SSSCD 的生成效果, 本研究从序列新颖性、序列稳定性、抗菌活性倾向以及结构几何合理性四个层面选取 Similarity、Instability、Antimicrobial 和 Docking 四项指标对模型性能进行分析。其中, Similarity、Instability 与 Antimicrobial 主要反映生成序列在统计分布和功能相关性上的表现, Docking 则用评估生成的抗菌肽的活性。

将 SSSCD 与 LSTM-RNN、AMPGAN、HydrAMP 和 DiffAB 进行比较, 比较结果详见表 2。AMPGAN 代表基于生成对抗网络的抗菌肽生成思路[4]。HydrAMP 则体现了深度生成模型在高活性抗菌肽发现中的应用潜力[16]。DiffAB 代表扩散机制在蛋白质结构设计任务中的应用方向[17]。LSTM-RNN、AMPGAN 和 HydrAMP 主要生成序列, 不直接支持结构生成, 因此未计算 Docking score; DiffAB 和 SSSCD 均可进行结构相关评价。结果显示, SSSCD 在 4 项指标上均取得最优表现, 说明多模态对比扩散框架能够更有效地联合学习抗菌肽序列语义、结构空间特征和功能相关模式。

**Table 2.** Comparison of generation performance among different models**表 2.** 不同模型生成性能比较

模型	Similarity	Instability	Antimicrobial	Docking
LSTM-RNN	39.61	45.08	0.85	-
AMPGAN	38.30	51.52	0.86	-
HydrAMP	31.06	59.63	0.81	-
DiffAB	28.98	43.36	0.80	1608
SSSCD	24.93	40.05	0.87	1740

### 3.2. 消融实验结果

为验证关键模块贡献, 分别移除表面特征、跨模态对比学习、能量引导采样和结构分支, 消融实验结果见表 3。结果显示, 完整 SSSCD 在 Similarity、Instability 和 Antimicrobial 上均最优。去除结构分支后下降最明显, Similarity 升至 27.31, Instability 升至 42.59, Antimicrobial 降至 0.81, 表明结构信息对学习抗菌肽序列-构象关系具有基础作用。去除表面特征、跨模态对比或能量引导后也出现不同程度退化, 说明这些模块分别在功能相关表征、模态一致性和几何校正方面发挥增益作用。

**Table 3.** Ablation experiment results of SSSCD**表 3.** SSSCD 消融实验结果

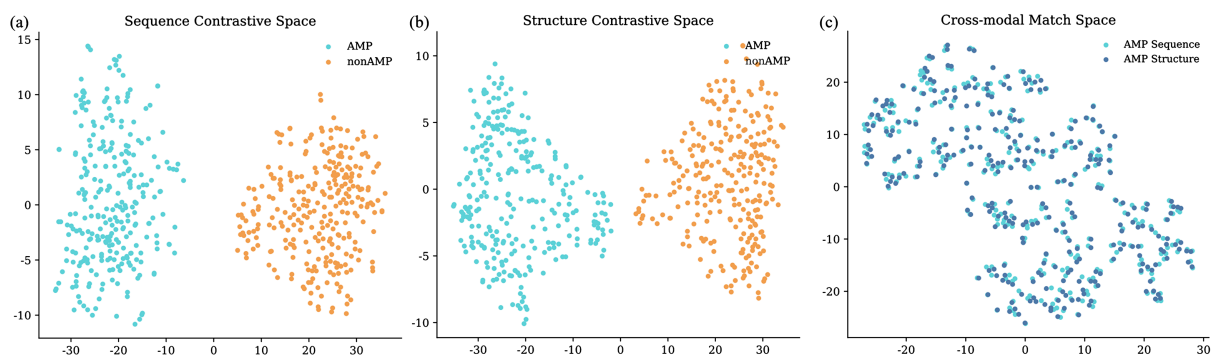
配置	Similarity	Instability	Antimicrobial
SSSCD	24.93	40.05	0.87
w/o surface	25.12	41.78	0.80
w/o contrastive	25.44	40.63	0.84
w/o energy guidance	25.28	40.47	0.83
w/o structure branch	27.31	42.59	0.81

### 3.3. 表征可视化分析

采用 t-SNE 对测试集中抗菌肽和非抗菌肽的全局表示进行降维可视化[18], 结果如图 2 所示, 其中 a、b、c 分别为序列对比空间、结构对比空间与跨模态匹配空间的可视化结果。序列对比空间中, 两类样本形成较清晰的双簇结构, 二维质心距离为 40.2342, 高维 cosine silhouette 系数为 0.8295。结构对比空间也呈现明显分离趋势, 二维质心距离为 40.1621, 高维 cosine silhouette 系数为 0.7970。跨模态匹配空间中, 抗菌肽序列表示与结构表示高度重叠, 二维质心距离仅为 0.2483, 平均余弦相似度为 0.9341。上述结果表明, SSSCD 不仅能够分别学习序列和结构空间中的类别判别特征, 还能够将同一抗菌肽的序列语义和骨架结构映射到一致的共享表示空间。

### 3.4. 生成样本理化性质

生成抗菌肽理化性质统计结果见表 4, 其各项理化性质均与参考数据集接近。生成样本的碱性氨基酸比例、酸性氨基酸比例、疏水性氨基酸比例、极性氨基酸比例、平均 pI、平均电荷、平均疏水性和正电荷序列占比整体接近参考分布, 提示模型较好地学习了抗菌肽的阳离子性、两亲性和电荷-疏水平衡关系。



**Figure 2.** t-SNE visualization of SSSCD representation spaces

**图 2.** SSSCD 表征空间 t-SNE 可视化

**Table 4.** Physicochemical properties of generated antimicrobial peptides

**表 4.** 生成抗菌肽理化性质

指标	数据集数值	生成数值
碱性氨基酸比例	0.2507	0.2410
酸性氨基酸比例	0.0418	0.0423
疏水性氨基酸比例	0.4689	0.4539
极性氨基酸比例	0.1211	0.1258
平均 pI	9.4950	9.6317
平均电荷	2.8596	2.8131
平均疏水性	-0.2207	-0.2247
正电荷序列占比	82.59%	83.00%

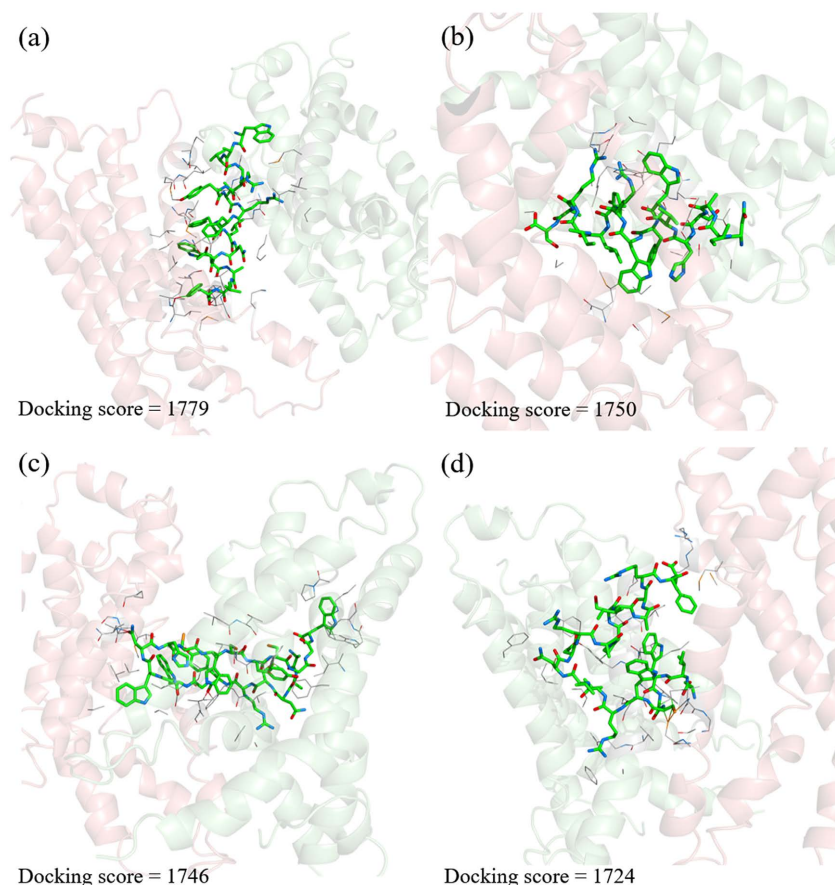
### 3.5. 代表性候选肽对接分析

随机选取 4 条生成抗菌肽进行可视化分析, 以 PDB 6MI7 的 F/G 链界面作为受体区域进行 ZDOCK 对接, 对接结果如图 3 所示。结果显示 4 条候选肽均定位于 F/G 链形成的界面沟槽附近, 并与周围残基形成连续或半连续接触。AMP\_1 主要偏向 F 链侧界面, AMP\_2 表现为较紧凑的 G 链核心区域结合, AMP\_3 呈跨链桥接式结合, AMP\_4 则更集中于 G 链局部螺旋附近。4 条候选肽的 Docking score 分别为 1779、1746、1750 和 1720, 提示其具有一定受体界面适配潜力。

## 4. 讨论

本研究结果表明, 将序列、结构和表面特征纳入统一扩散生成框架, 可有效改善抗菌肽候选物的综合质量。与传统序列生成模型相比, SSSCD 不仅降低了生成序列与已有抗菌肽库的平均相似性, 还改善了不稳定性指数和抗菌活性倾向; 与结构扩散模型相比, SSSCD 进一步通过表面特征、跨模态匹配和能量引导机制增强了生成结果的功能相关性和物理合理性。

结构分支在消融实验中贡献最大, 说明抗菌肽生成不能只关注氨基酸排列, 还需要考虑空间构象与功能界面。抗菌肽常通过正电荷区域与细菌膜表面发生静电吸附, 并依赖疏水片段参与膜扰动, 因此序列特征、骨架构象和表面暴露环境之间存在天然耦合关系。SSSCD 通过 EGNN 建模主链坐标, 并将 AAindex 理化信息和残基层面表面描述符整合到结构图中, 有助于模型学习这种耦合规律。



**Figure 3.** Docking conformations of representative generated antimicrobial peptides with PDB 6MI7  
**图 3.** 代表性生成抗菌肽与 PDB 6MI7 受体对接构象

跨模态对比学习是模型的另一关键设计。t-SNE 和余弦相似度结果显示, 抗菌肽序列表示与结构表示能够在共享空间中保持高度一致, 说明模型并非简单拼接多源特征, 而是在样本级别建立了序列语义与空间构象之间的对应关系。这种表示对齐有助于提高生成序列与生成结构之间的一致性, 也为后续候选肽聚类、相似性搜索和功能解释提供了基础。

需要指出的是, 本研究仍存在一定局限。首先, 生成样本的抗菌活性评价主要依赖计算预测模型和分子对接结果, 而这些指标本质上属于间接功能估计, 尚不能替代真实生物实验验证。当前常用抗菌肽预测器主要基于已有数据库训练, 其判别能力容易受到训练集分布偏倚影响, 对于具有全新序列模式或非常规作用机制的候选肽, 其预测结果可能存在不确定性。同时, Docking score 更多反映的是局部结构匹配与界面几何相容性, 其本身并不等同于真实结合自由能, 也难以完整描述膜穿透、构象重排、溶剂效应以及动态膜扰动等复杂生物过程。

此外, 抗菌肽的真实活性不仅取决于静态结构, 还与细菌膜组成、离子环境、蛋白酶降解、聚集行为及宿主细胞相容性等多种因素相关。因此, 即使生成肽在计算指标上表现较优, 仍可能在体外抑菌实验、溶血实验或细胞毒性实验中出现活性不足或毒性偏高等问题。这表明计算生成模型与真实生物体系之间仍存在一定“模拟 - 实验鸿沟”。

另一方面, 本研究中的结构信息部分来源于 AlphaFold3 等预测结构, 而预测构象本身可能与真实溶液环境中的动态构象存在偏差, 进一步影响后续对接与功能评估结果。当前 SSSCD 更关注抗菌肽生成质量与模态一致性, 尚未系统引入毒性、稳定性、半衰期、免疫原性、合成可行性及多目标优化约束。未来

可进一步结合分子动力学模拟、高通量湿实验筛选以及主动学习策略, 以提高生成候选肽的真实可转化性与生物可靠性。

## 5. 结论

本研究提出了一种基于多模态对比扩散的抗菌肽生成模型 SSSCD。该模型联合利用氨基酸序列、主链三维结构和残基层面表面特征, 通过 Transformer 离散扩散分支、EGNN 连续结构扩散分支、跨模态对比学习和能量引导采样, 实现了抗菌肽候选序列与结构的协同生成。实验结果显示, SSSCD 在序列新颖性、稳定性、抗菌活性倾向和受体对接评分方面均优于对比模型; 消融实验证实结构分支、表面特征、对比学习和能量引导均对模型性能有积极贡献。总体而言, SSSCD 为抗菌肽智能设计提供了一种可扩展的计算框架, 可用于抗菌肽候选物快速生成、初筛和后续实验验证。

## 参考文献

- [1] Wang, G., Li, X. and Wang, Z. (2016) APD3: The Antimicrobial Peptide Database as a Tool for Research and Education. *Nucleic Acids Research*, **44**, D1087-D1093. <https://doi.org/10.1093/nar/gkv1278>
- [2] Waghu, F.H., Barai, R.S., Gurung, P. and Idicula-Thomas, S. (2016) CAMP<sub>3</sub>: A Database on Sequences, Structures and Signatures of Antimicrobial Peptides: Table 1. *Nucleic Acids Research*, **44**, D1094-D1097. <https://doi.org/10.1093/nar/gkv1051>
- [3] Shi, G., Kang, X., Dong, F., Liu, Y., Zhu, N., Hu, Y., et al. (2022) DRAMP 3.0: An Enhanced Comprehensive Data Repository of Antimicrobial Peptides. *Nucleic Acids Research*, **50**, D488-D496. <https://doi.org/10.1093/nar/gkab651>
- [4] Van Oort, C.M., Ferrell, J.B., Remington, J.M., Wshah, S. and Li, J. (2021) AMPGAN V2: Machine Learning-Guided Design of Antimicrobial Peptides. *Journal of Chemical Information and Modeling*, **61**, 2198-2207. <https://doi.org/10.1021/acs.jcim.0c01441>
- [5] Ho, J., Jain, A. and Abbeel, P. (2020) Denoising Diffusion Probabilistic Models. *Advances in Neural Information Processing Systems*, **33**, 6840-6851.
- [6] Li, S., Fan, J., He, H., Zhou, R. and Liao, J. (2025) MolP-PC: A Multi-View Fusion and Multi-Task Learning Framework for Drug ADMET Property Prediction. *Chinese Journal of Natural Medicines*, **23**, 1293-1300. [https://doi.org/10.1016/s1875-5364\(25\)60945-9](https://doi.org/10.1016/s1875-5364(25)60945-9)
- [7] Satorras, V.G., Hoogeboom, E. and Welling, M. (2021) E(n) Equivariant Graph Neural Networks. *Proceedings of the 38th International Conference on Machine Learning*, Virtual Event, 18-24 July 2021, 9323-9332.
- [8] Yao, L., Guan, J., Xie, P., Chung, C., Zhao, Z., Dong, D., et al. (2025) dbAMP 3.0: Updated Resource of Antimicrobial Activity and Structural Annotation of Peptides in the Post-Pandemic Era. *Nucleic Acids Research*, **53**, D364-D376. <https://doi.org/10.1093/nar/gkae1019>
- [9] Bateman, A., Martin, M., Orchard, S., Magrane, M., Ahmad, S., Alpi, E., et al. (2023) UniProt: The Universal Protein Knowledgebase in 2023. *Nucleic Acids Research*, **51**, D523-D531. <https://doi.org/10.1093/nar/gkac1052>
- [10] Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., et al. (2024) Accurate Structure Prediction of Biomolecular Interactions with AlphaFold 3. *Nature*, **630**, 493-500. <https://doi.org/10.1038/s41586-024-07487-w>
- [11] Vaswani, A., Shazeer, N., Parmar, N., et al. (2017) Attention Is All You Need. *Advances in Neural Information Processing Systems*, 5998-6008.
- [12] Kyte, J. and Doolittle, R.F. (1982) A Simple Method for Displaying the Hydropathic Character of a Protein. *Journal of Molecular Biology*, **157**, 105-132. [https://doi.org/10.1016/0022-2836\(82\)90515-0](https://doi.org/10.1016/0022-2836(82)90515-0)
- [13] Cock, P.J.A., Antao, T., Chang, J.T., Chapman, B.A., Cox, C.J., Dalke, A., et al. (2009) Biopython: Freely Available Python Tools for Computational Molecular Biology and Bioinformatics. *Bioinformatics*, **25**, 1422-1423. <https://doi.org/10.1093/bioinformatics/btp163>
- [14] Chen, R., Li, L. and Weng, Z. (2003) ZDOCK: An Initial-Stage Protein-Docking Algorithm. *Proteins: Structure, Function, and Bioinformatics*, **52**, 80-87. <https://doi.org/10.1002/prot.10389>
- [15] Li, Y., Orlando, B.J. and Liao, M. (2019) Structural Basis of Lipopolysaccharide Extraction by the LptB2FGC Complex. *Nature*, **567**, 486-490. <https://doi.org/10.1038/s41586-019-1025-6>
- [16] Szymczak, P., Możejko, M., Grzegorzec, T., Jurczak, R., Bauer, M., Neubauer, D., et al. (2023) Discovering Highly Potent Antimicrobial Peptides with Deep Generative Model HydrAMP. *Nature Communications*, **14**, Article No. 1453.

<https://doi.org/10.1038/s41467-023-36994-z>

- [17] Luo, S., Ma, J., Peng, J., Peng, X., Su, Y. and Wang, S. (2022) Antigen-Specific Antibody Design and Optimization with Diffusion-Based Generative Models for Protein Structures. *Advances in Neural Information Processing Systems*, **35**, 9754-9767. <https://doi.org/10.52202/068431-0709>
- [18] Van der Maaten, L. and Hinton, G. (2008) Visualizing Data Using t-SNE. *Journal of Machine Learning Research*, **9**, 2579-2605.