Supplementary Material for "A Conditional Denoising VAE-based Framework for Antimicrobial Peptides Generation with Preserving Desirable Properties"

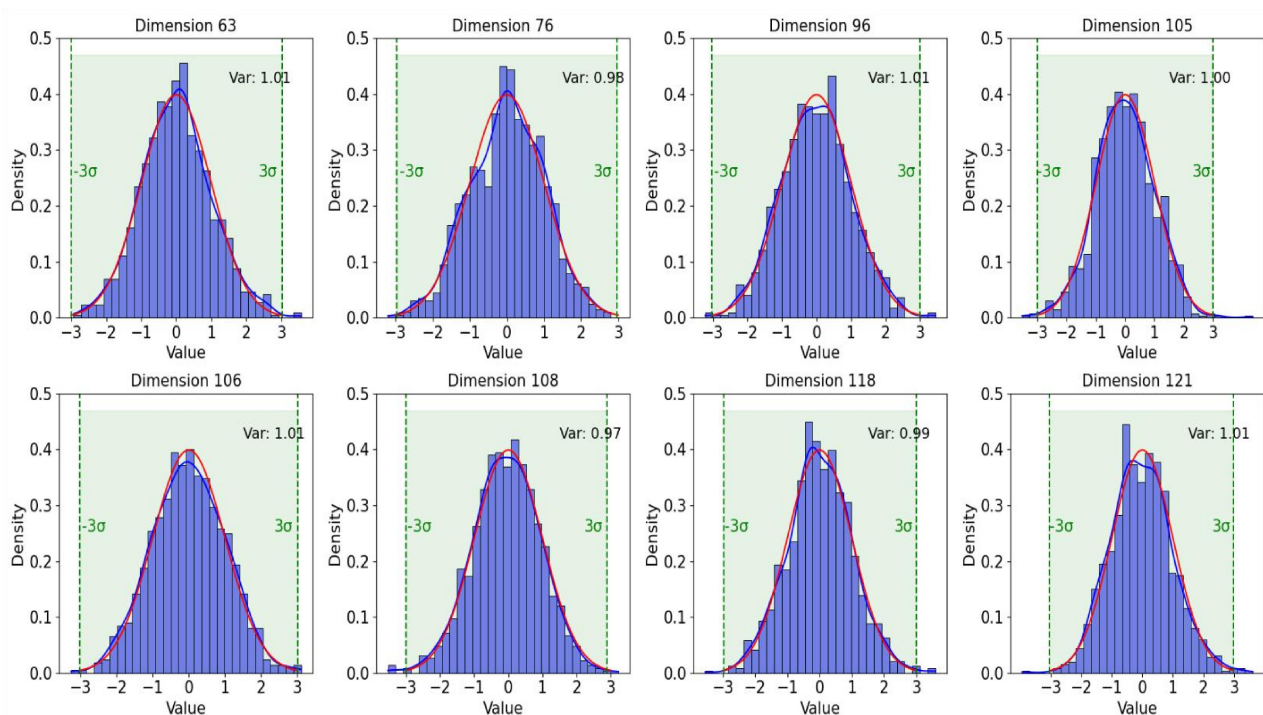## 1   Supplementary Figures

Figure S1: Comparison of latent space distribution and standard normal distribution. The red curve represents the standard normal distribution, while the blue curve illustrates the normal distribution for the corresponding dimensions, which are randomly sampled from eight out of the 128 dimensions. The light green area indicates the 3σ region of the learned distribution.
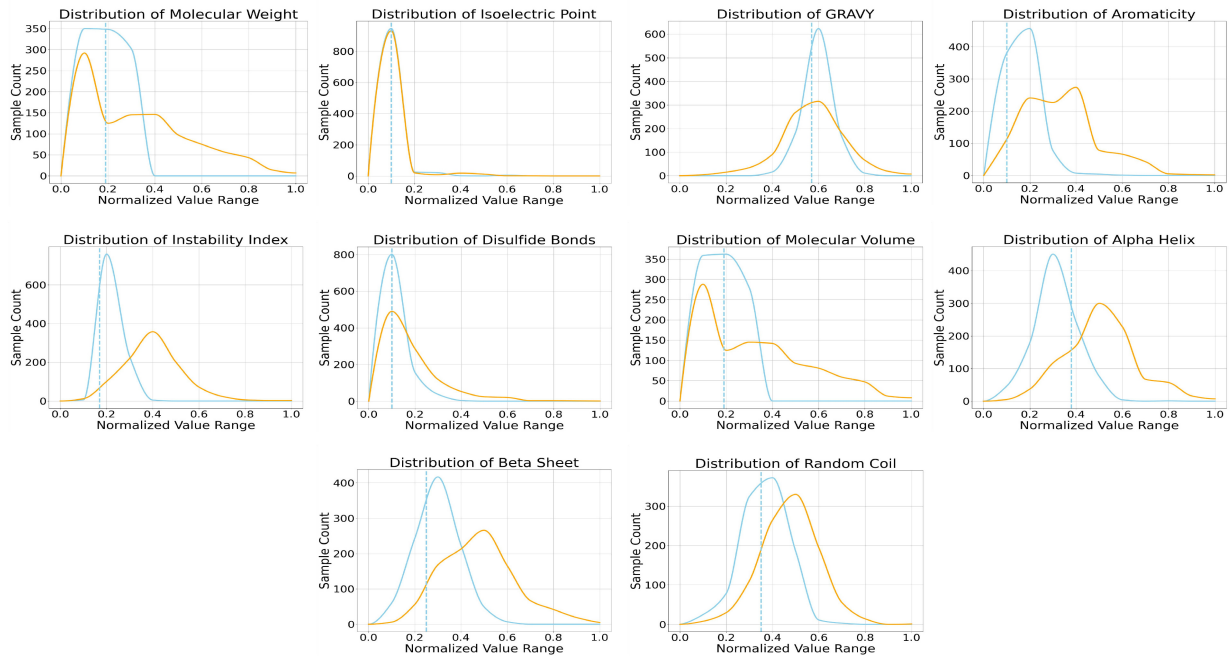
Figure S2: Results of comparison between conditional generation and unconditional generation. The analyzed parameters include molecular weight, isoelectric point, hydrophobicity, aromatic amino acid ratio, instability index, disulfide bonds, molecular volume, and secondary structure of the sequence (α-helix, β-sheet, and random coil). The blue curve denotes the results of conditional generation, while the orange curve denotes the results of unconditional generation. The horizontal axis corresponds to the normalized values of the respective attributes, and the vertical axis indicates the number of instances falling in each range. The dashed vertical line represents the target value for conditional generation. This figure provides a visual comparison of the number of AMPs in each interval range under the two generation modes.
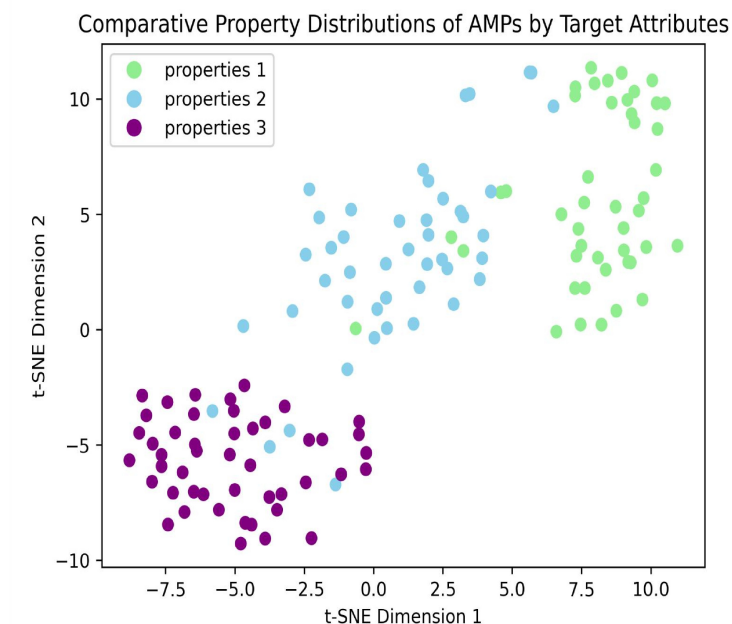
Figure S3: t-SNE Dimensionality Reduction Distribution of Ten Physicochemical Properties of AMPs. The data points are divided into three groups, represented by green, blue, and purple, corresponding to properties 1, properties 2, and properties 3, respectively.

2 Supplementary Tables

Table S1: Performance Comparison among Different Generation Models

Methods	HEMO <0.5	TOXI <0.5
LSTM	0.831	0.630
AMP-GAN	0.393	0.914
Pep-GAN	**0.945**	0.945
WAE	0.338	0.881
AMPEMO	0.443	0.842
MoFormer	0.881	0.936
Our model	0.941	**0.976**

Table S2: Distribution of AMPs Under Three Different Conditions

property ID	Molecular Weight	Isoelectric Point	GRAVY	Aromaticity	Instability Index	Disulfide Bonds	Molecular Volume	Alpha Helix	Beta Sheet	Random Coil
Pro1	0.19	0.19	0.57	0.05	0.17	0.03	0.19	0.41	0.14	0.35
Pro2	0.30	0.20	0.70	0.10	0.30	0.11	0.34	0.38	0.19	0.23
Pro3	0.36	0.15	0.39	0.14	0.27	0.23	0.35	0.29	0.20	0.13

Table S3: Top 10 Generated AMPs Preserving Most Similar Properties to That of Nisin

Sequence ID	Sequence	Molecular Weight	Isoelectric Point	GRAVY	Aromaticity	Instability Index	Disulfide Bonds	Molecular Volume	Alpha Helix	Beta Sheet	Random Coil
Nisin	**ISILEKAIIMLNPIMEKLFVTELVMKTEYSIHTN**	**3963.77**	**5.56**	**0.54**	**0.06**	**17.85**	**1490**	**4987.6**	**0.41**	**0.14**	**0.35**
Seq1	GIEKIVQIAKASAYAIATIAKTAGISIYEASDDIG	3553.02	4.78	0.51	0.06	17.18	2980	4397.7	0.31	0.17	0.29
Seq2	AYAIIVKGGAAGDYAGYKAIVDA	2257.54	6.01	0.61	0.13	9.27	4470	2780.9	0.35	0.17	0.30
Seq3	IQIAATVAFGGVGGYAKIIGEIEKISE	2735.14	5.01	0.69	0.07	42.71	1490	3419.9	0.37	0.22	0.26
Seq4	AGAYYAKAEAIIVVAIGAHETG	2175.44	5.40	0.72	0.09	6.62	2980	2682.5	0.32	0.14	0.41
Seq5	DGAAESSIGGLDGVTKCGIATAEIIISYAKKYIYA	3549.99	4.78	0.37	0.08	36.00	4470	4332.3	0.31	0.23	0.26
Seq6	EIYAIAAAASFIKHGADIGGI	2088.36	5.32	0.89	0.09	14.70	1490	2589.2	0.33	0.19	0.33
Seq7	FEGAIGYIHKIAGAIEIAIKDEI	2471.84	4.83	0.61	0.08	42.62	1490	3101.8	0.39	0.13	0.30
Seq8	GGAAAVGEAIYAIVRIAEAYIGKTIGS	2621.98	6.14	0.81	0.07	7.60	2980	3245.3	0.33	0.22	0.33
Seq9	KVYAEIYIAIAGAAGAGVKDIKDYGSVFEV	3118.81	4.78	0.55	0.13	6.36	4470	3863.3	0.40	0.16	0.27
Seq10	AIGVIGAIAFAIEYKAEYVYREAEVIK	2987.45	4.95	0.66	0.15	10.56	4470	3740.2	0.44	0.07	0.37