Accelerating Antimicrobial Peptide Discovery with Latent Structure

Danqing Wang, Zeyu Wen, Fei Ye, Lei Li, Hao Zhou

danqingwang@ucsb.edu

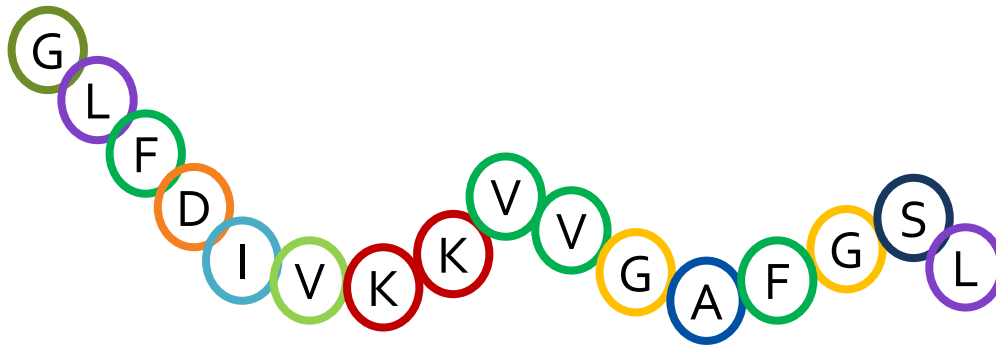UC SANTA BARBARA   ByteDance   清華大學 Tsinghua University

9/22/23

Antimicrobial Peptide (AMP)
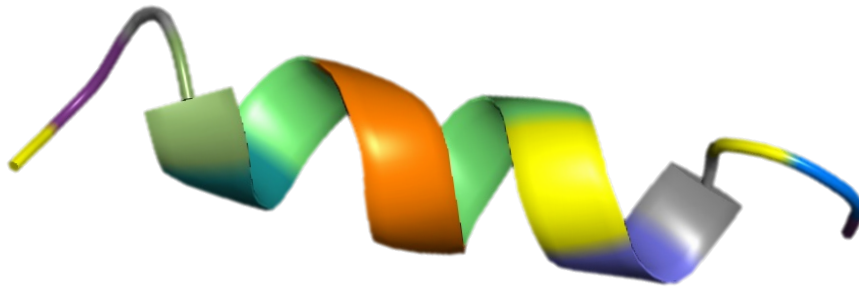
Short Amino Acid Sequence
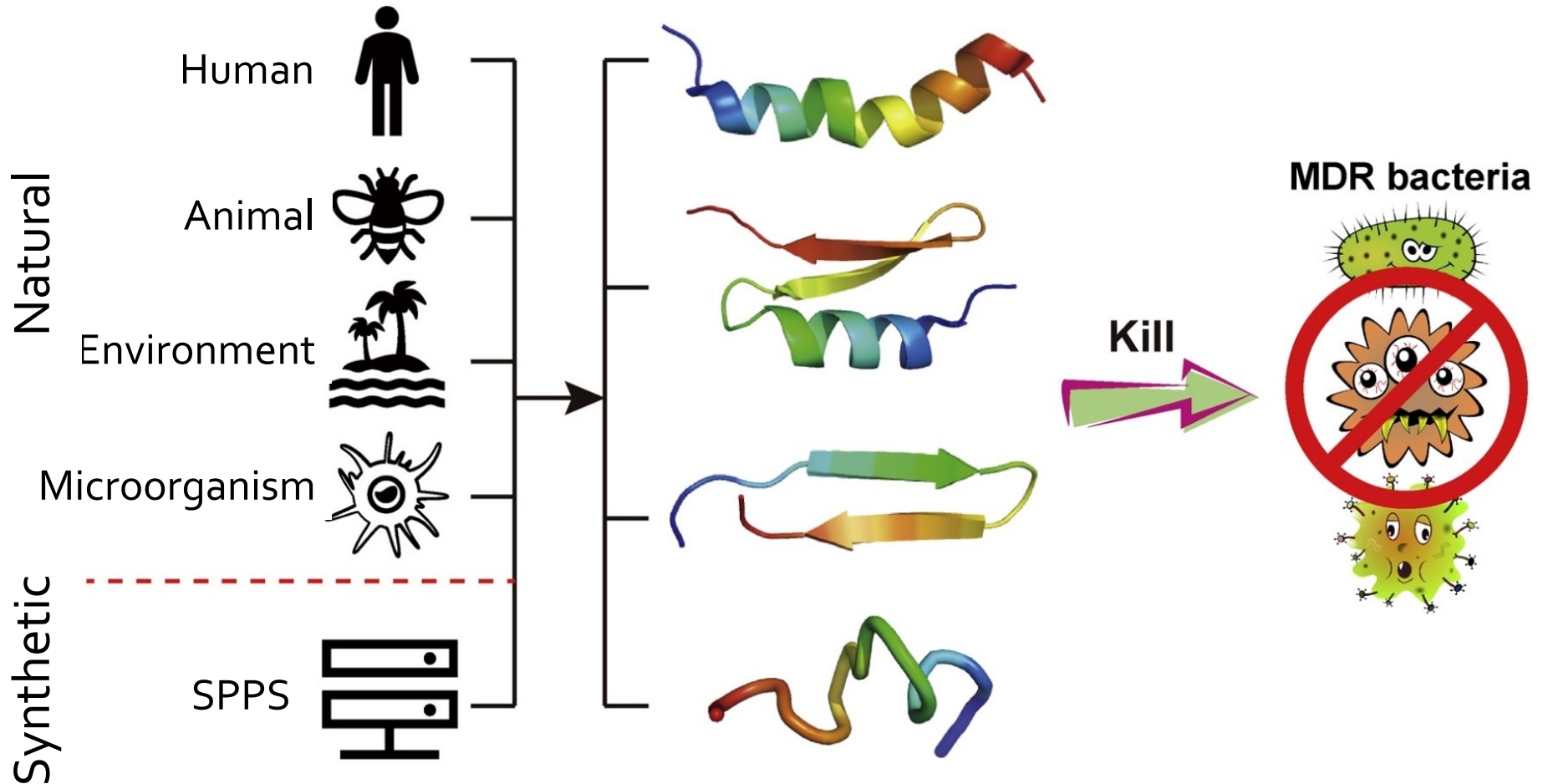

Alanine (A)
Cysteine (C)
...
Valine (V)

20 types

Secondary Structure
(α-helices, β-sheets)

alpha-helix (H)
beta bridge (B)
...
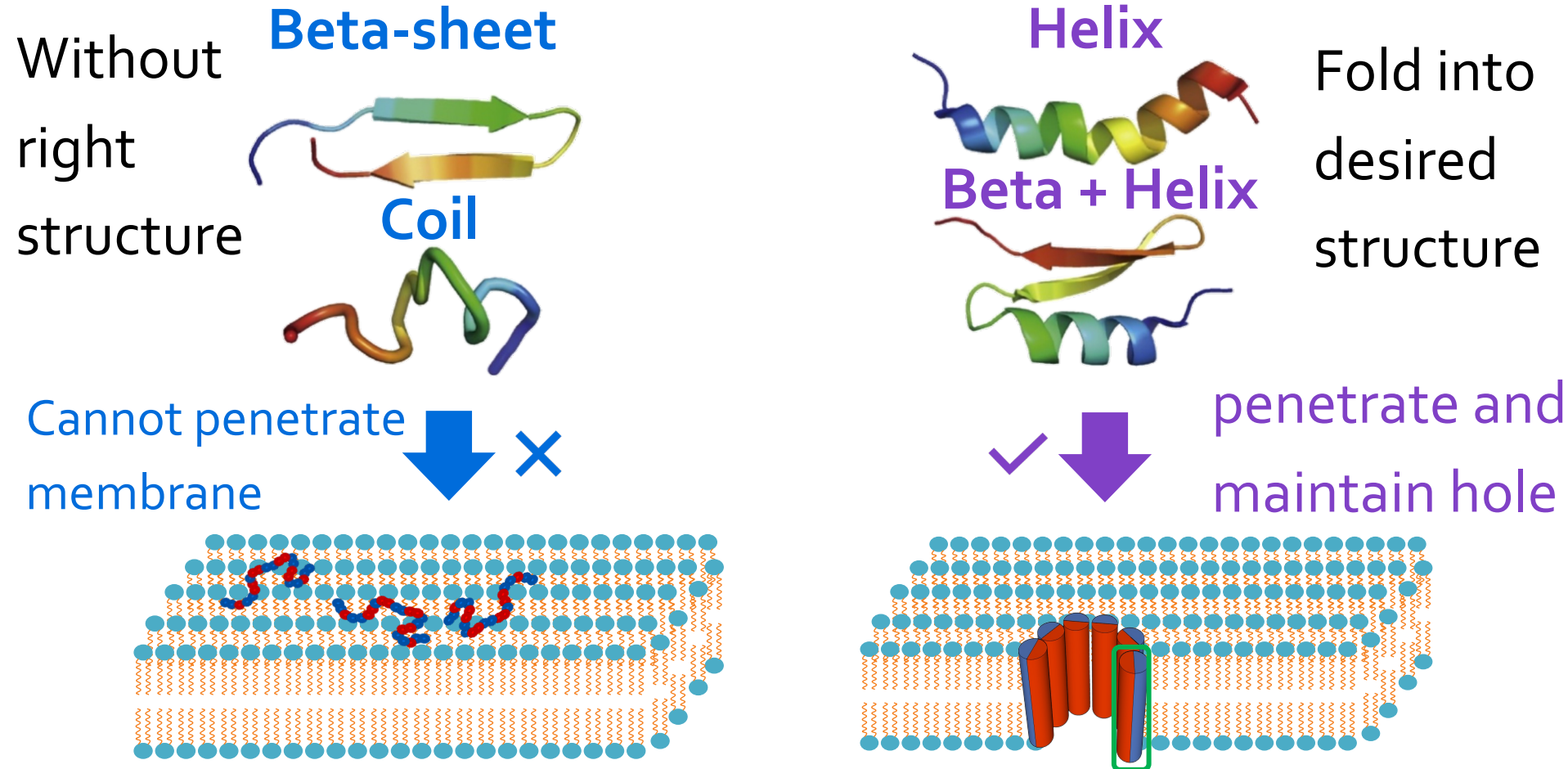other/loop (L)

8 types

AMPs are promising treatments towards bacteria

# Structures are important to function!

Without right structure

Beta-sheet

Coil

Helix

Beta + Helix

Fold into desired structure

Cannot penetrate membrane ✗
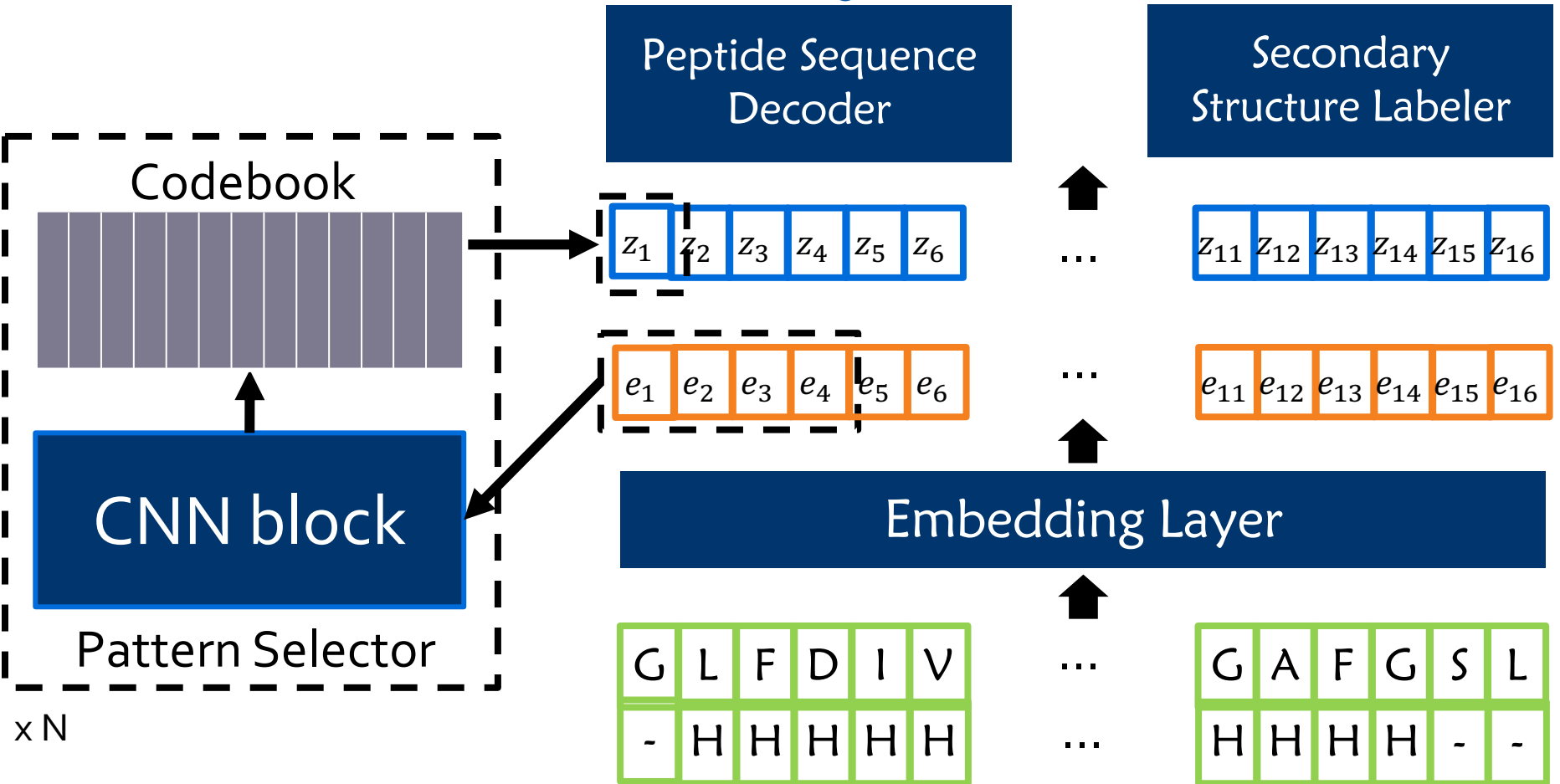
penetrate and maintain hole ✓

Problem Definition

- Goal
  - Design peptide sequences that can effectively kill bacteria

- Challenge:
  - How to generate peptides with desired secondary structures
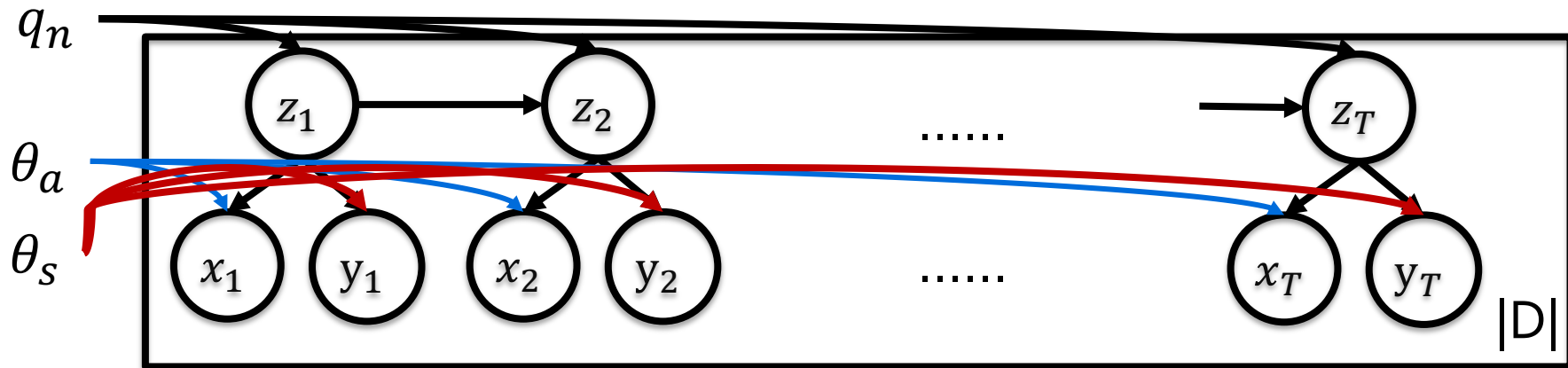 - Limited data (only 3k+ existing AMPs)

Graphical Model
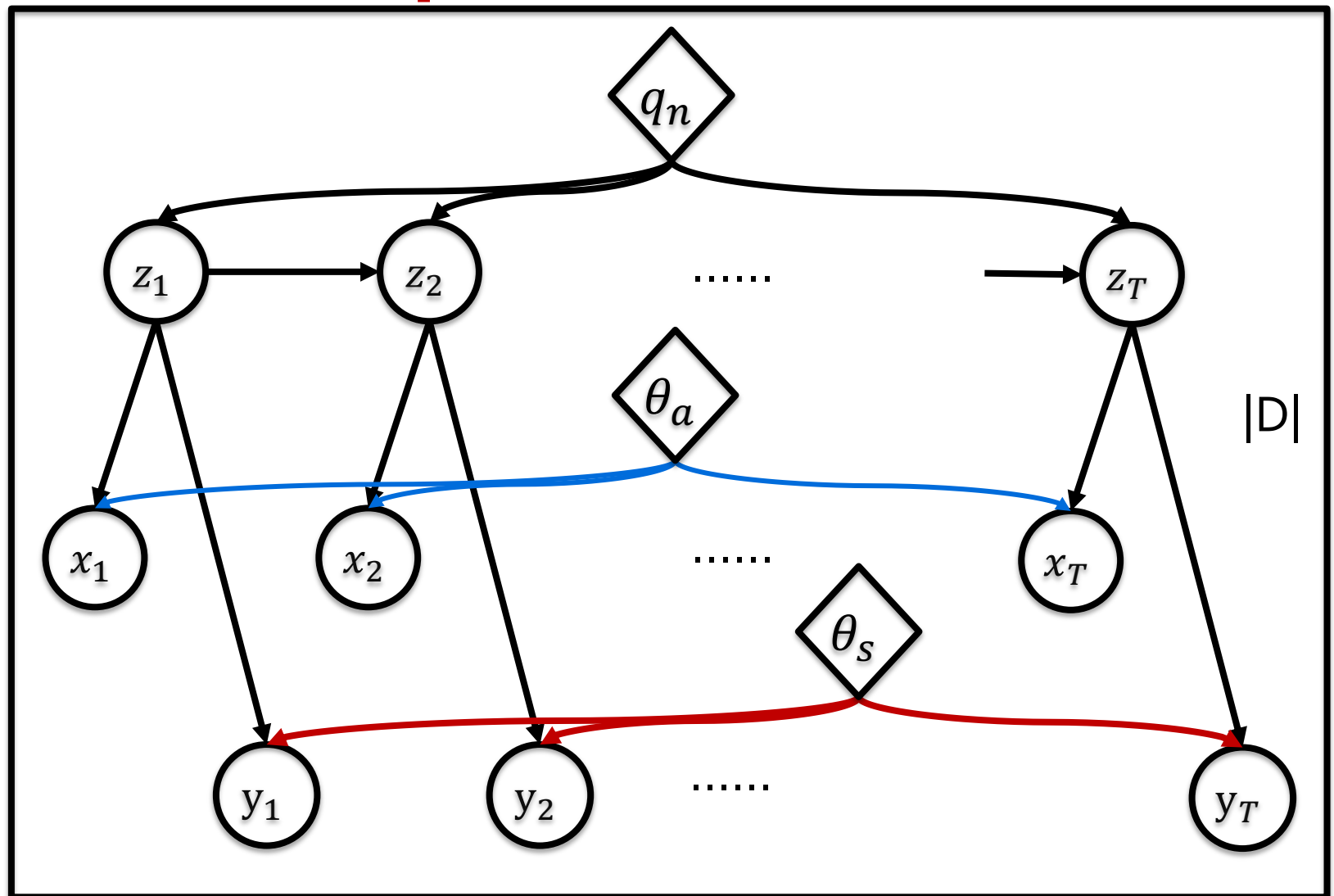
Latent Sequence-Structure AMP Model (LSSAMP)

LSSAMP: Training objective

q_n

θ_a

θ_s

|D|

Sequence Reconstruction   Secondary structure Prediction

$$Objective = \boxed{\sum_{i=1}^{L} \log p_{\theta_a}(x_i | z_q(x_i))} + \boxed{\sum_{i=1}^{L} \log p_{\theta_s}(y_i | z_q(x_i))}$$

$$-\boxed{\sum_{n=1}^{N}} \underline{\left\| sg[z_{e_n}(x_i)] - z_{q_n}(x_i) \right\|_2^2 + \beta \left\| [z_{e_n}(x_i)] - sg[z_{q_n}(x_i)] \right\|_2^2}$$
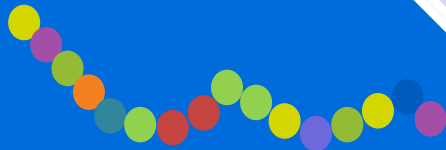
Multi-scale
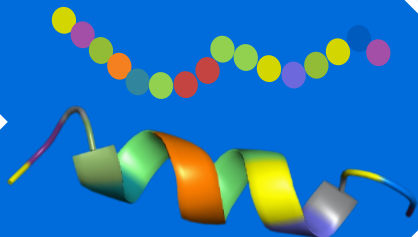codebook

Codebook loss

Peptide Data

AMP

AMP dataset (APD)

D_{AMP}: 3k

UniProt

D_r: 57k

AlphaFold

D_s: 46k

Evaluation Pipeline

Computational metric
charge, hydrophobicity, hydrophobic moment

Public classifiers
probability of AMP

Wet lab experiment
minimal inhibitory concentration (MIC)

Synthesize in wet laboratory

1. Generate candidates
2. Virtual screening using *charge, hydrophobicity, hydrophobic moment*
3. choose the top-tier in AMP classifiers
4. filter by novelty (> 5 residue)
5. check solubility based on heuristic rules
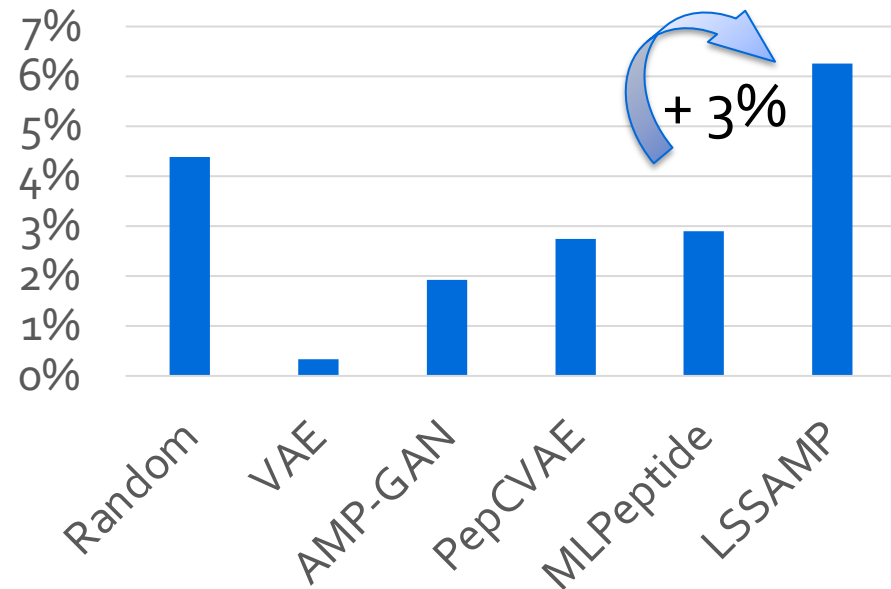
LSSAMP produces effective AMPs in wet-lab!

2/21 have been verified in wet-lab with high antimicrobic activity (<128)

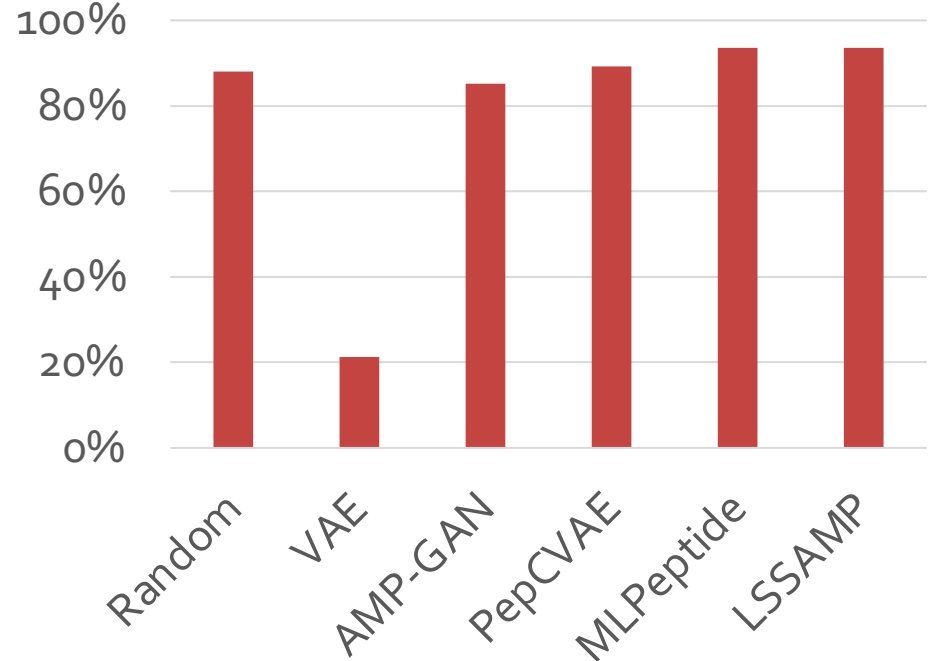| Cpd (µg/ml)	128	64	32	16	8	4	2	1	0.5	0.25	0.125	0
P1												
P2												
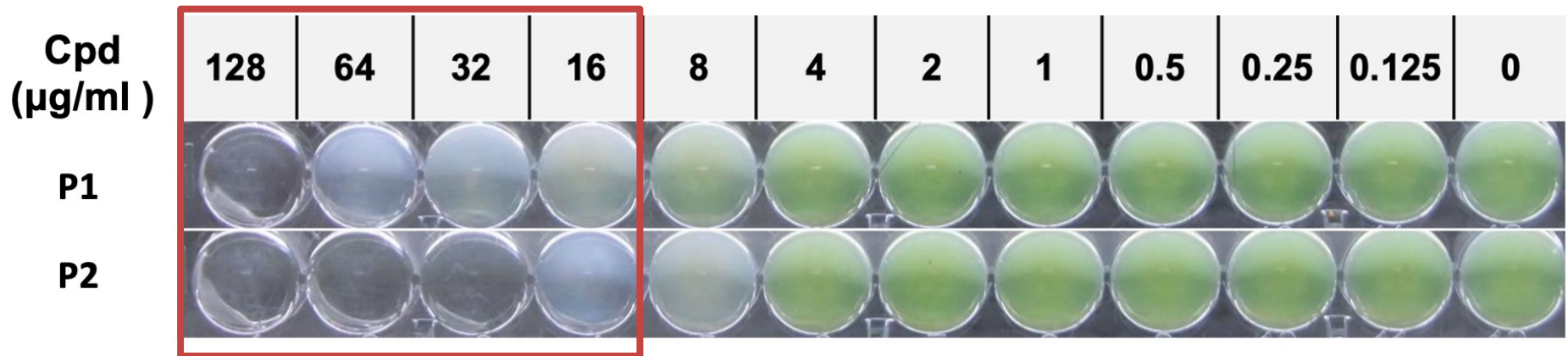
Clear liquid indicates bacteria (green) have been killed !

LSSAMP outperforms baselines in biological attributes and comparable in AMP classifiers
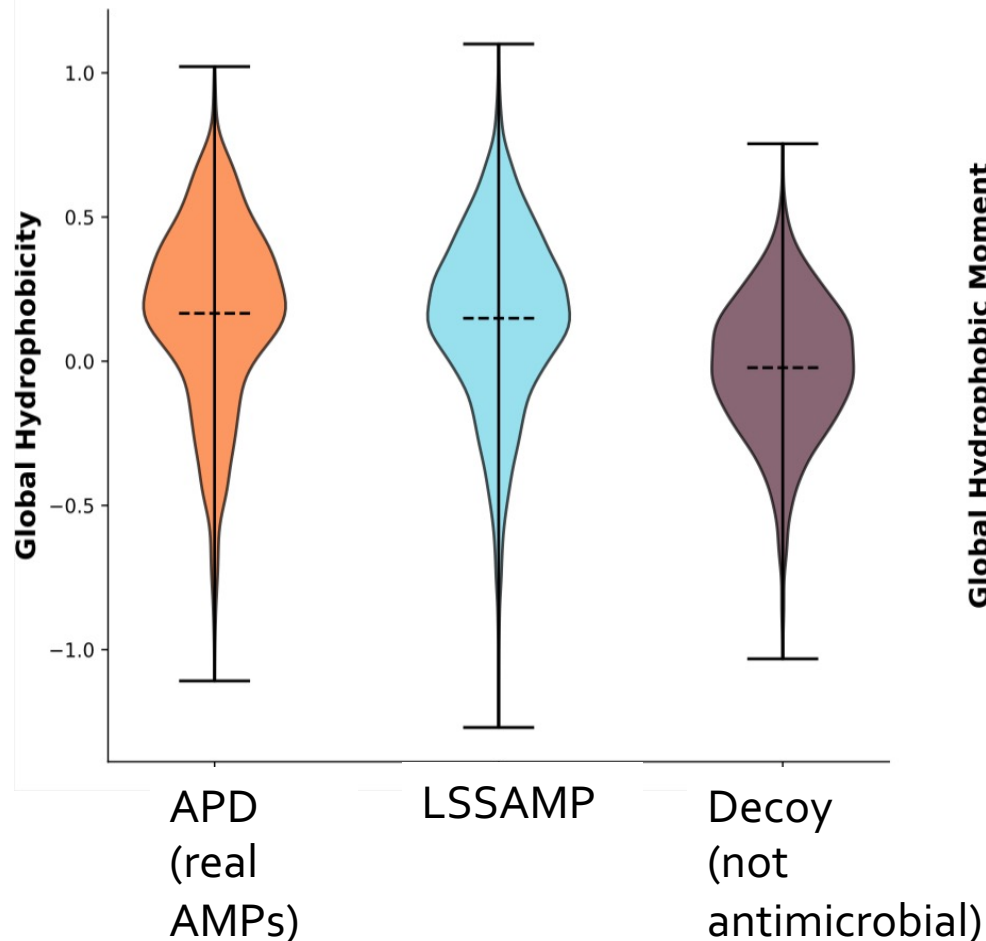
Proportion Satisfying Attribute Condition

Bar chart. Y-axis: 0% to 7%. Categories: Random ≈4.4%, VAE ≈0.3%, AMP-GAN ≈1.9%, PepCVAE ≈2.8%, MLPeptide ≈2.9%, LSSAMP ≈6.2%. Annotation: + 3%

Avg Score of AMP Classifiers

Bar chart. Y-axis: 0% to 100%. Categories: Random ≈88%, VAE ≈21%, AMP-GAN ≈85%, PepCVAE ≈89%, MLPeptide ≈93%, LSSAMP ≈93%
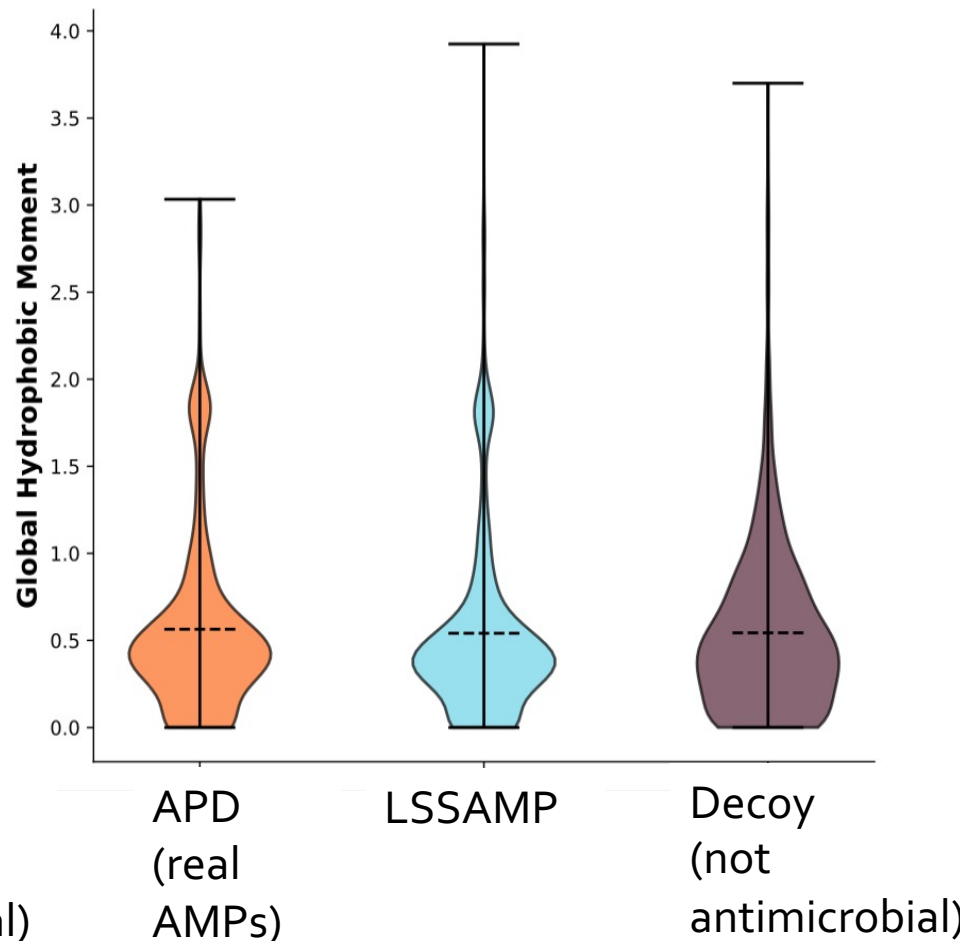
13

LSSAMP generations share similar distribution as Real AMPs

Hydrophobicity
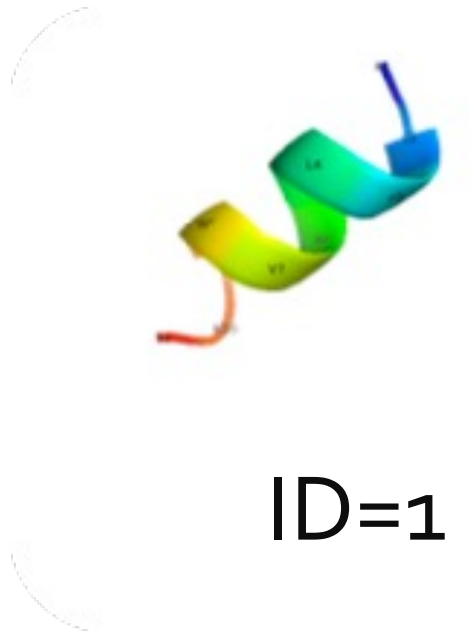
Hydrophobic moment

Peptides found by LSSAMP

ID	Sequence	Secondary Structure	c	H	uh	D
1	FLPLVRVWAKLI	--HHHHHHHHHH	2.0	0.4708	0.7227	6
2	FLSVIKGVWAASLPKQFCAVTAKC	--HHHHHHHHHHHHHHHHHH HHT--	3.0	0.3338	0.6599	8

ID=1

ID=2

Sum up

- Antimicrobial activities highly depend on secondary structure

- LSSAMP: generate desired secondary structure
  - fine-grained control on position
  - multi-scale features

- Data augmentation is important in AMP generation

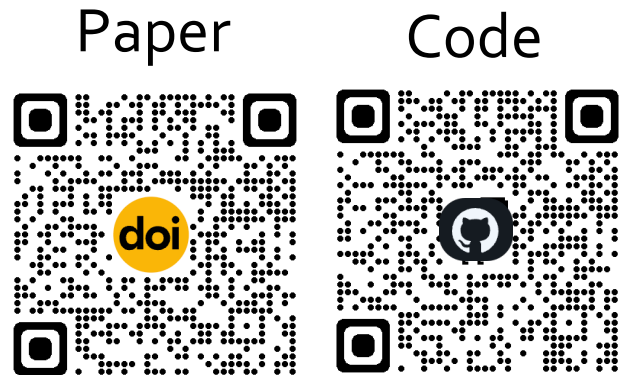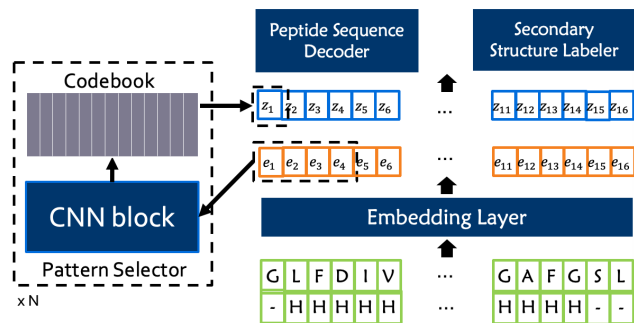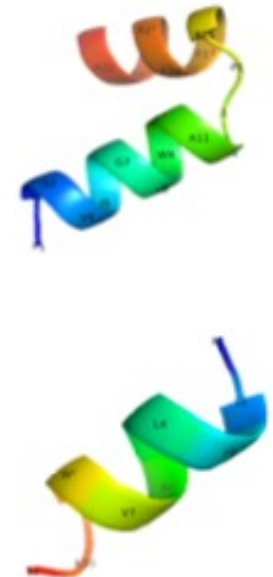- Computational metric and wet-lab verify that LSSAMP is effective

Take Away

- Incorporate structure information into sequence generation

- Evaluate model generation based on biological mechanism

- Data augmentation via Alphafold2

Accelerating Antimicrobial Peptide Discovery with Latent Structure (LSSAMP)

Danqing Wang

(danqingwang@ucsb.edu)

Paper

Code

DOI

Code: LSSAMP