

Supplementary Materials for
**Combinatorial Discovery of Antibacterials via a Feature-Fusion Based
Machine Learning Workflow**

Cong Wang *et al.*

Corresponding author: Peng Zhang, and Jian Ji. Email: zhangp7@zju.edu.cn; jijian@zju.edu.cn.

Methods

Materials. Dimethyl sulfoxide (DMSO), potassium chloride, methanol, and dichloromethane (DCM) were purchased from Sinopharm Chemical Reagent Co., Ltd and used without further purification. 2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid (Hepes) was purchased from Macklin. Glucose was purchased from RHAWN. Ciprofloxacin (CF) and bacitracin (BT) were purchased from Aladdin. Benzalkonium chloride (BC) was purchased from Leyan. Triton X-100 was bought from Sigma-Aldrich. Other information of the reagents for Ugi synthesis was listed in Table S1.

Components selection for the preliminary library. A uniform manifold approximation and projection algorithm (UMAP) was used to reduce high-dimensional representations of the overall library to two-dimensional representations. As shown in Figure S1A, points of the same color represent the distribution of products containing a certain component (such as an acid) in the entire chemical space. Representative acids, amines, aldehydes, and isonitriles were artificially chosen as evenly as possible. As shown in Figure S2B, 360 combinations (yellow points) were finally selected and they distributed uniformly in the map.

Library synthesis. Stock solutions of the reagents were prepared at 0.4 M in methanol. Firstly, amines and aldehydes for preliminary screening were mixed in a 96-well PCR plate. After 15 minutes, according acids were added into the blend. Isonitriles were then added into the mixture and the plates were sealed by sealing films to reduce solution evaporation. The reaction was complete in 24 hours at room temperature for further test.

Model training. All compounds were described by PD and FD using rdkit based on previously published research. Feature engineering of PD was performed by following steps: 1) Features with 0 variance were removed; 2) Features with standard deviation higher than 0.95 and less than -0.95 were randomly removed from one of them. LSCNN imposed constraints on the representations derived from PD and FD to learn complementary features via cross-modal interaction. The loss function in the training process is given in equation (1):

$$Loss = MSE(f_{PD}(x), y) + MSE(f_{FD}(x), y) + \lambda L_{LSC} \#(1)$$

where $f_{PD}(\cdot)$ and $f_{FD}(\cdot)$ are uni-feature models, respectively. In our work, we use Euclidean distance (ED) and contrastive loss (CL) as L_{LSC} which are defined in equation (2) and equation (3):

$$L_{i,j}(ED) = \sqrt{\sum_{k=1}^d (z_{i,k} - z_{j,k})^2} \#(2)$$

$$L_{i,j}(CL) = -\log \frac{\exp(\text{sim}(z_i, z_j)/\tau)}{\sum_{k=1}^{2N} 1\{k \neq i\} \exp(\text{sim}(z_i, z_j)/\tau)} \#(3)$$

where z_i and z_j are the features extracted from PD and FD, respectively. d is the dimension of the learned features, $z_{i,k}$ is the k^{th} dimension of z_i , $\text{sim}(\cdot)$ measures the cosine similarity between the two vectors, N is the batch size, and τ is the temperature parameter. The model was optimized using RMSProp optimizer. Hyperparameters were optimized using a grid search and set as follows: $d=512$, learning rate=0.00001, batch size=16, $\lambda=1$, patience=15. The source code is available at <https://github.com/wuyuhui-zju/LSCNN>.

Minimum inhibitory concentration (MIC) assay. Methicillin-resistant *Staphylococcus aureus* (MRSA) was incubated in Tryptic Soy Broth (TSB) at 37°C overnight. Then the bacteria were diluted and calibrated to a concentration at 10^8 Colony Forming Unit (CFU)·mL⁻¹. Crude solutions from the preliminary library were diluted in DMSO with concentrations at 1 mM. H4-6, CF, BT and benzalkonium chloride (BC) solutions were prepared in DMSO at concentrations ranging from 15.36 mM to 0.12 mM. 195 μ L of MRSA suspension (diluted to 10^5 CFU·mL⁻¹) was added into a 96-well plate and then 5 μ L test solution was mixed into each well. Sterile TSB solution was used as the control and DMSO presented no antibacterial effect at such concentration. The plate was incubated at 37°C for 20 hours before the OD values at 595nm were collected by a microplate reader. MIC concentrations were determined at which the solutions were completely clear.

Minimum bactericidal concentration (MBC) assay. The MBC was defined as the lowest concentration to kill 99.9% bacteria after incubation. 10 μL solution from MIC assay was pipetted onto Tryptic Soy Agar (TSA) plates. After incubation at 37°C for 20 hours, MBC values were recorded as no bacteria colony formed on the plate.

Synthesis and purification of H1-6 (H4 as an example).

500 μL 4-(4-methylpent-3-en-1-yl)cyclohex-3-ene-1-carbaldehyde at 0.4 M in methanol was added into a 4 mL vial containing 500 μL equivalent 1-(3-aminopropyl)pyrrolidine. After 15 minutes of stirring, 500 μL 4-biphenylacetic acid and 1-isocyanobutane were added into the mixture successively. The vial was capped and kept stirring for 24 hours for reaction completion. The solvent methanol was evaporated under vacuum by rotary evaporator and the crude product was redissolved in DCM before being applied to a flash chromatography instrument (Biotage Selekt-2SV). The column chromatography eluting solution was set as a mixture of methylene chloride, methanol and formic acid (95:4:1 in volume%). The purified product was a yellowish-brown solid (47.8 mg, 40.0% yield). Nuclear magnetic resonance spectra were recorded on a Bruker AVANCE III 400 (^1H NMR at 400 MHz). Mass spectra were recorded via quadrupole-time of flight (QTOF) mass spectrometry (Agilent 6545).

Colony Forming Unit counting assay. Bacteria were incubated in TSB at 37°C overnight. MRSA was diluted to a concentration of 10^8 CFU $\cdot\text{mL}^{-1}$. Then the suspension was centrifuged at 5,000 rpm for 5 minutes. TSB was removed and MRSA was resuspended in sterile phosphate buffered saline (PBS). H4-6 and BC were prepared in DMSO and mixed with MRSA suspension at 24 μM (40-fold dilution) for 6 hours. CF and BT were prepared in DMSO at 96 μM . DMSO was used as the control. 100 μL mixed solution was spread on TSA to visualize the alive bacteria. 10 μL solution was also pipetted onto the TSA plate to count the alive bacteria colony.

Bacterial killing kinetics. MRSA was incubated in TSB at 37°C overnight and diluted to 10^8 CFU $\cdot\text{mL}^{-1}$. The bacteria suspension was centrifuged and diluted in PBS to 10^5 CFU $\cdot\text{mL}^{-1}$. Then H4-6, CF, BT and BC were added to the solution at 12 μM (40-fold dilution). At the set time point (0 minute, 4 minutes, 8 minutes, 10 minutes, 20 minutes, 30 minutes, 60 minutes, 90 minutes), an aliquot of 10 μL solution was sampled and pipetted on the TSA plate to count the alive bacteria colony. DMSO was used as the control. Bacterial viability was defined as (sampled CFU counts) / (CFU counts at 0 minute).

Live/dead bacterial assay. LIVE/DEAD™ BacLight™ Bacterial Viability Kit (L7012, Invitrogen, CA, USA) was purchased from Thermo Fisher Scientific Inc. 500 μL MRSA suspensions (4×10^8 CFU $\cdot\text{mL}^{-1}$ in PBS) were treated with H4-6 and BC at 48 μM . After 5 minutes of incubation, the bacteria were centrifuged and the supernate was removed. Bacteria were resuspended in the staining solution and incubated for 15 minutes at room temperature. Finally, the staining solution was also discarded after centrifuging and PBS washing. Bacteria were observed by the confocal laser scanning microscope (Olympus FV-3000).

Cytoplasmic membrane depolarization. The procedure was applied following previous literature with slight modification^{1,2}. 3,3'-dipropylthiadicarbocyanine iodide [DiSC₃(5)] was purchased from Bidepharm. MRSA was cultured at 37°C in TSB overnight. Then the bacteria were centrifuged and washed with Hepes solution [5 mM Hepes (pH 7.4) and 20 mM glucose]. MRSA was diluted to 10^7 CFU $\cdot\text{mL}^{-1}$. The bacterial suspension was treated with 0.4 μM DiSC₃ for 1 hour. Potassium chloride was added to the suspension at a final concentration of 0.1 M. 195 μL suspension was pipetted in a 96-well plate. Its fluorescence intensity was collected by a microplate reader (excitation $\lambda = 622$ nm, emission $\lambda = 670$ nm). Once the fluorescence intensity turned stable, 5 μL of H4-6, BC and ciprofloxacin solutions were mixed into the suspension to a final concentration of 12 μM immediately. The fluorescence intensity change was recorded by the microplate reader. Hepes was used as the negative control. 2.5% Triton X-100 (final concentration) was used as the positive control.

Transmission electron microscope (TEM) characterization. The procedure was applied following previous literature with slight modification³. MRSA was incubated in TSB at 37°C overnight. The bacteria suspensions were directly treated with H4-6 at 64 μM (final concentration) for 4 hours. Then the suspensions were centrifuged and washed with PBS. PBS was then removed and 2.5% glutaraldehyde solution was applied to fix the MRSA at 4 °C overnight. The bacteria were washed by cold PBS three times before further incubation with osmium tetroxide (1%). The bacteria were rinsed by cold PBS and dehydrated by ethanol solutions (30%, 50%, 70%, 90%, and 100%). Samples were then treated with acetone followed by embedding in resin. Embedded samples were cut into ultrathin sections. Finally, the ultrathin slices were deposited on bare mesh copper grids and observed through TEM (Hitachi HT7820).

Bacteria resistance study. MIC value in the first round MIC assay was defined as MIC_0 . Then the bacteria suspension after the first round was diluted 8000-fold. The diluted suspension was deposited on another 96-well plate. H4-6 and ciprofloxacin stock solutions were added into the bacteria suspension to a final concentration of $1/2 MIC_0$. After 20 hours of incubation, MIC_1 values were recorded. In each round, the working concentration was kept at $1/2 MIC_n$ and the bacteria growth was assumed to go through 10 generations for every cycle. MIC values were collected until 100 generations.

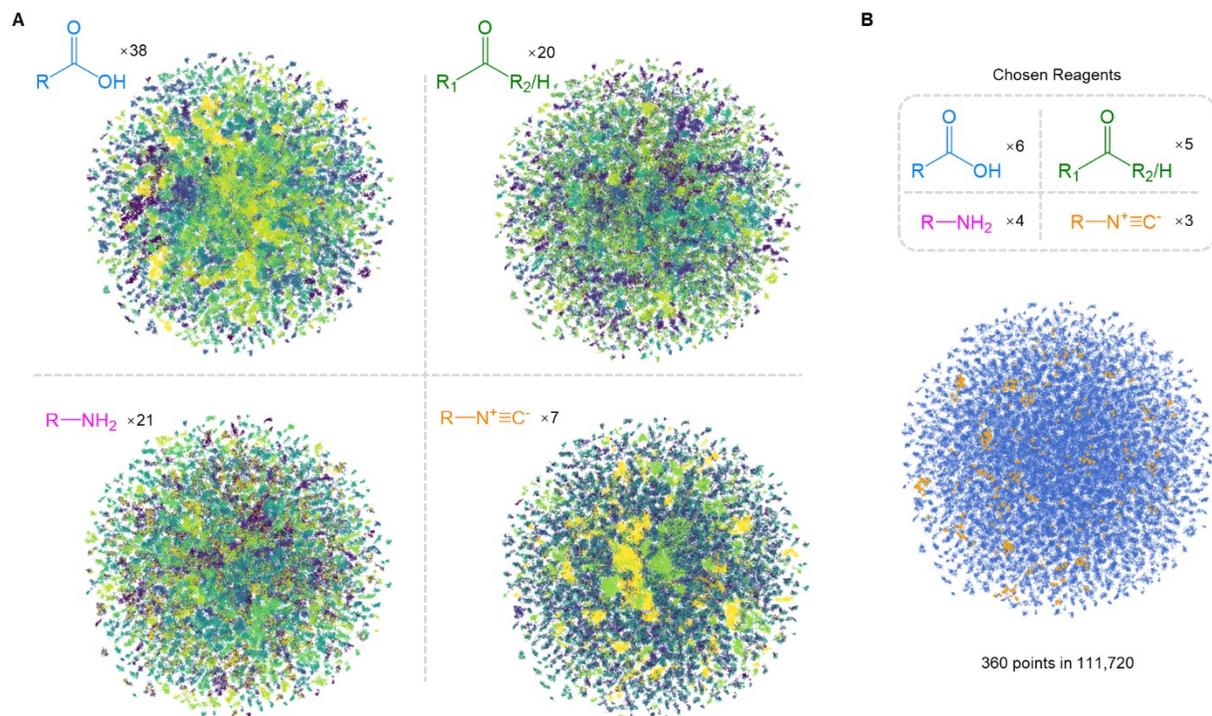


Figure S1. The preliminary library construction based on the chemical diversity. A) The obtained chemical space using UMAP dimensional reduction on physiochemical descriptors. Each point represents an Ugi product and each color represents a type of reagent. Generally, the distance between two points presents the similarity of the two Ugi products' properties. (B) The total number of chosen reagents and the corresponding products in the chemical space. The chosen products are labeled as yellow dots, and others are labeled as blue dots.

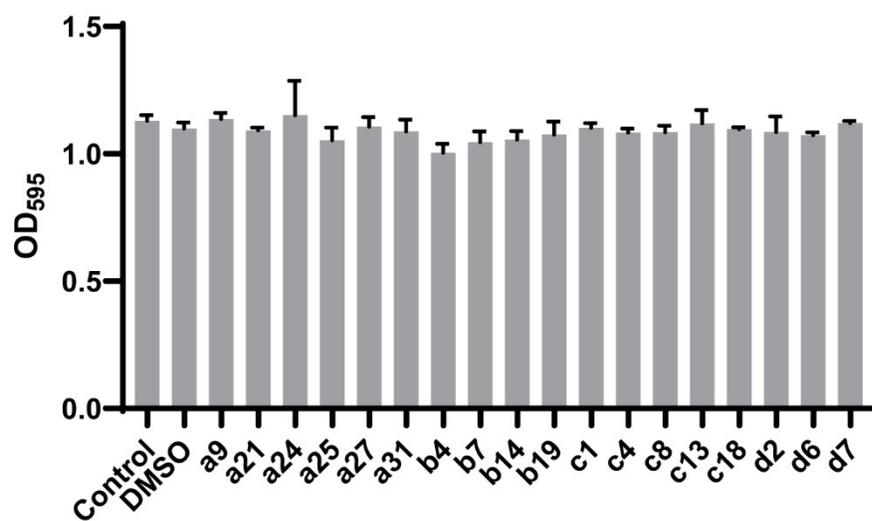


Figure S2. Antibacterial activity evaluation of the reagent for the preliminary library.

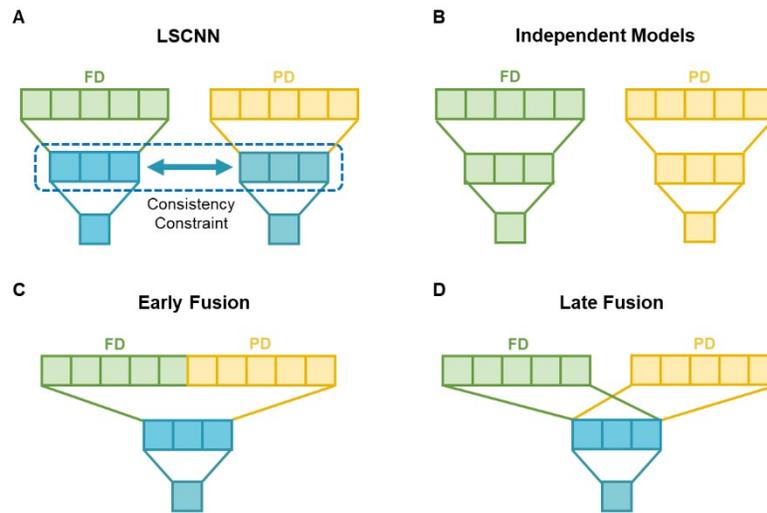


Figure S3. Different feature fusion architectures. A) Latent-Space Constraint Neural Network (LSCNN), which applied the consistency constraint to certain middle layers of the two models. B) Models trained independently using PD and FD. C) Early fusion. PD and FD were directly concatenated and input into the neural network. D) Late fusion. PD and FD were concatenated at the hidden layer.

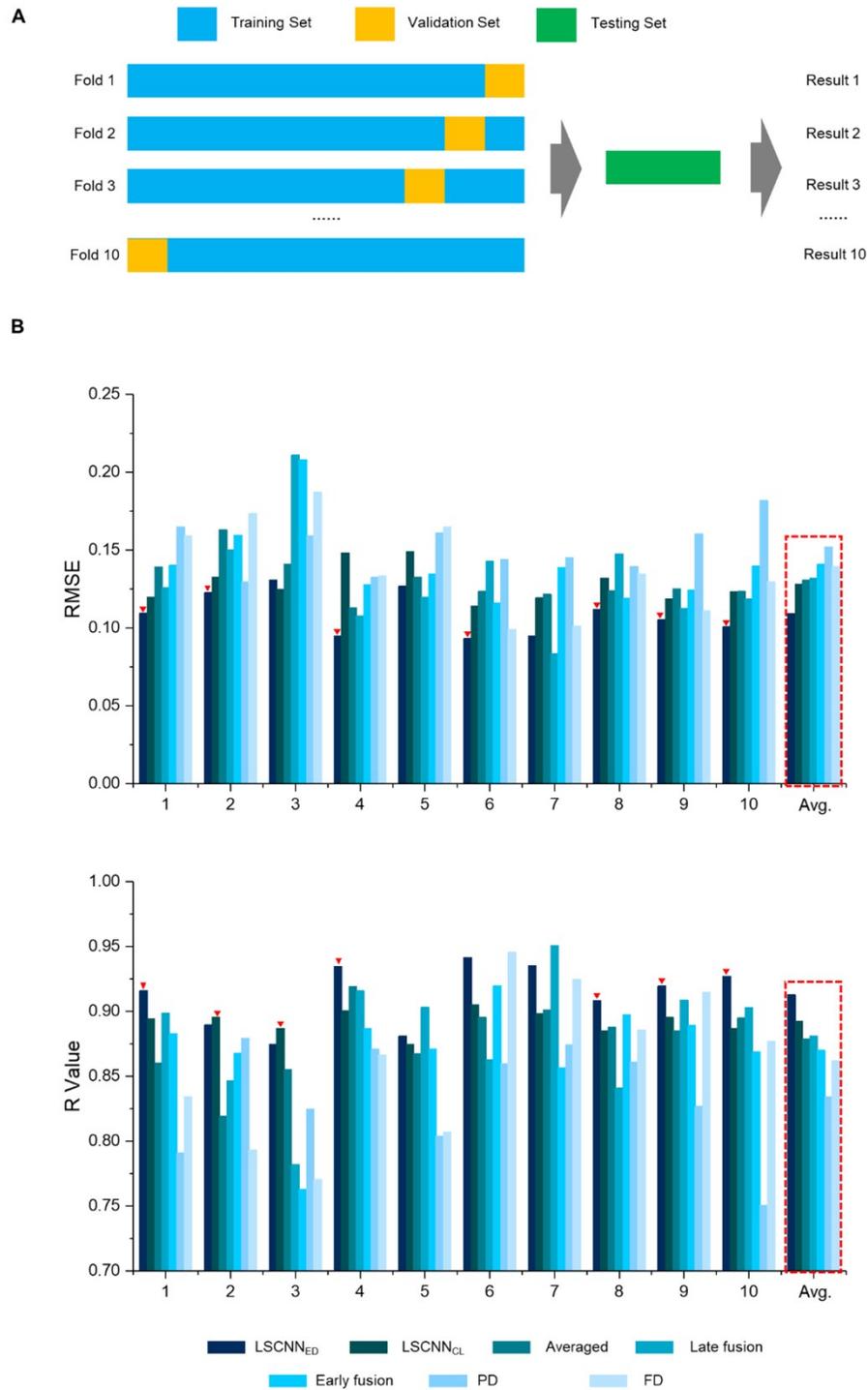


Figure S4. The results of the 10-fold cross-validation on our in-house dataset. A) The process of the 10-fold cross-validation. 1/6 of the dataset was set as the fixed leave-out testing set. The remaining data was set as training and validation sets in turn at the ratio of 9:1. 10 results were obtained in total for each model. B) The R and RMSE of OD values of different models on the testing set in each fold. The red triangles pointed out those LSCNN models which achieved the best index in the test of that fold. Averaged denoted averaging the output of MLPs trained with PD and FD respectively.

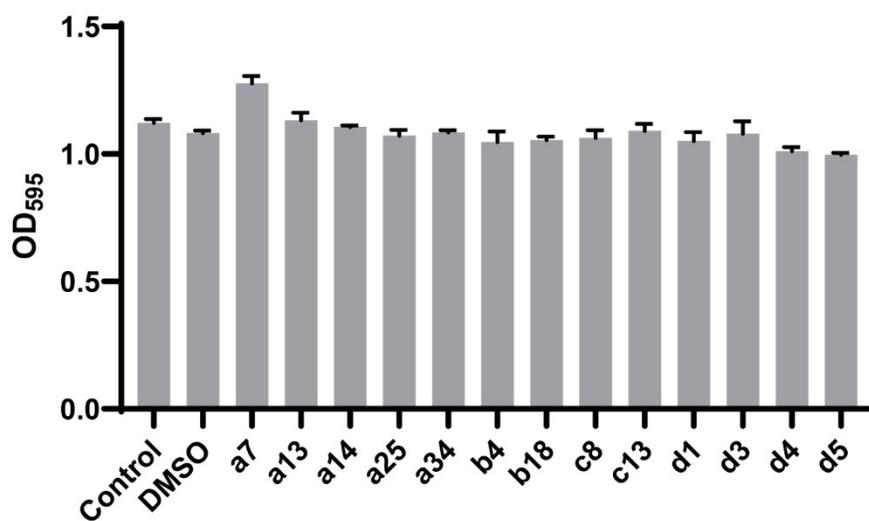


Figure S5. Antibacterial results of the reagents for the predicted combinations.

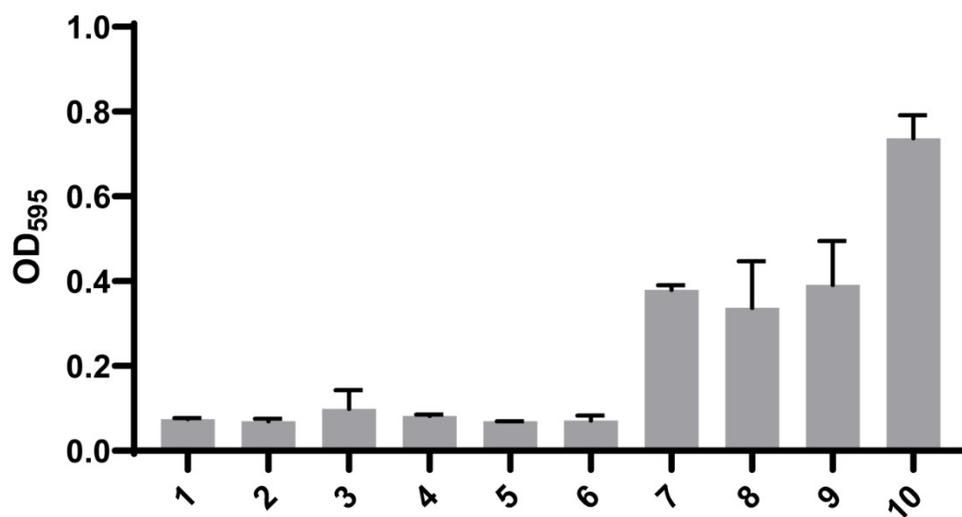
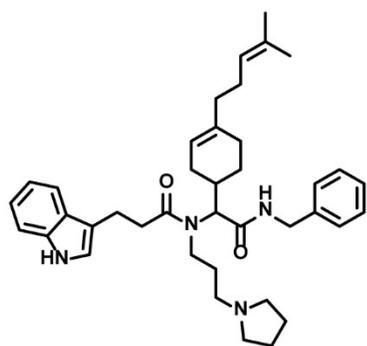
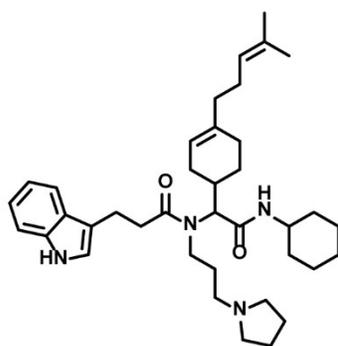


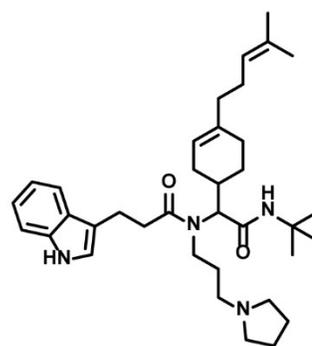
Figure S6. Antibacterial activity evaluation of the predicted combinations.



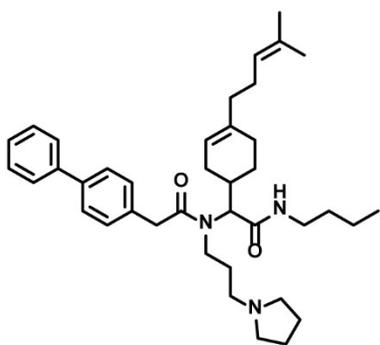
H1



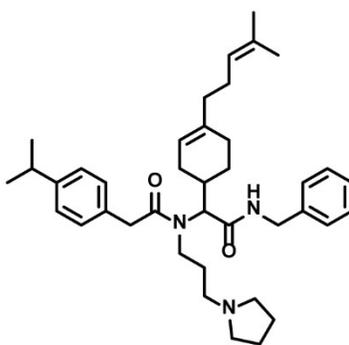
H2



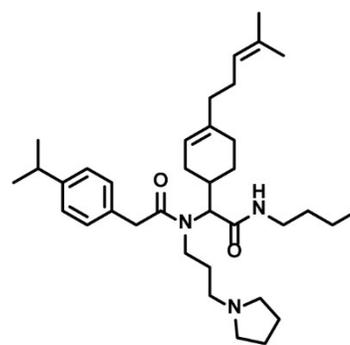
H3



H4



H5



H6

Figure S7. Predicted Ugi products structure.

***N*-(2-(benzylamino)-1-(4-(4-methylpent-3-en-1-yl)cyclohex-3-en-1-yl)-2-oxoethyl)-3-(1*H*-indol-3-yl)-*N*-(3-(pyrrolidin-1-yl)propyl)propanamide (H1):**

Yellow solid. ¹H NMR (400 MHz, CDCl₃): δ 8.38 – 7.75 (m, 2H), 7.68 – 7.27 (m, 5H), 7.23 – 6.83 (m, 5H), 5.35 (d, *J* = 40.9 Hz, 1H), 5.14 – 4.99 (m, 1H), 4.68 – 4.38 (m, 1H), 3.68 (s, 4H), 3.11 (ddd, *J* = 7.9, 6.9, 0.9 Hz, 3H), 2.79 – 1.85 (m, 13H), 1.77 – 1.52 (m, 7H), 1.26 (s, 5H), 0.86 (dt, *J* = 17.7, 7.6 Hz, 5H). QTOF ESI MS: *m/z* calculated for C₃₉H₅₂N₄O₂ [M+H]⁺: 609.4163; Found 609.4163.

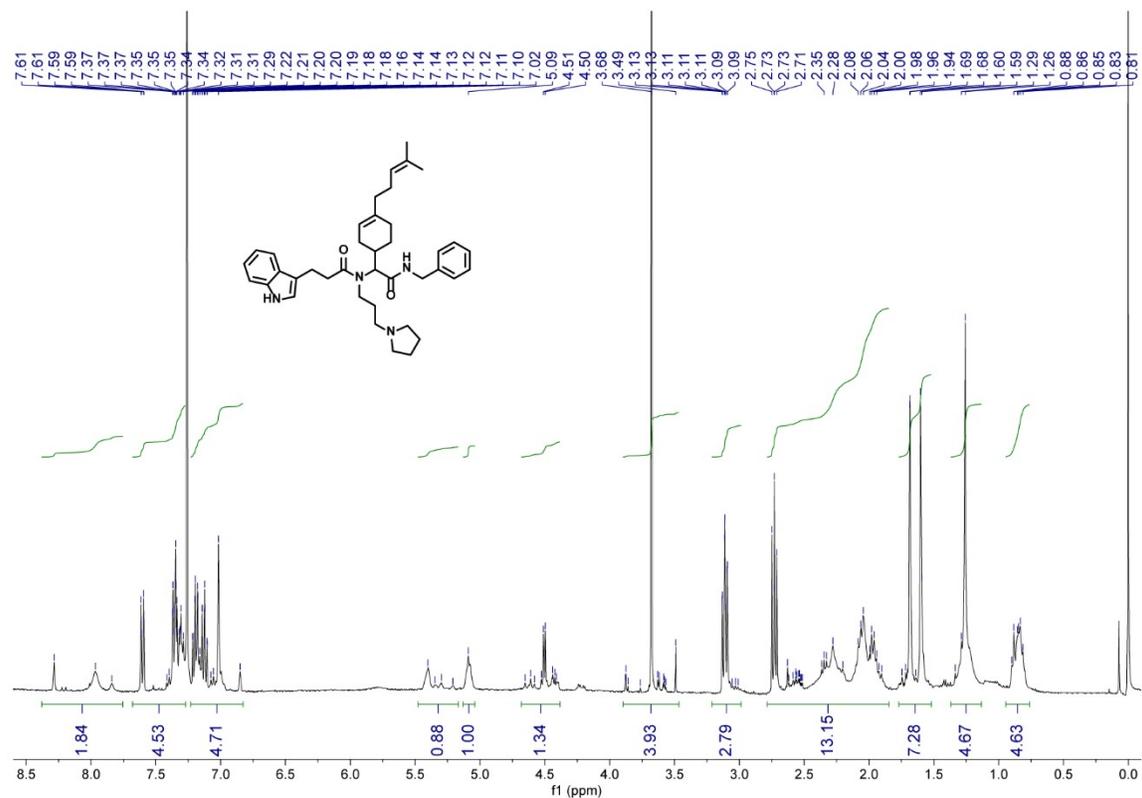


Figure S8. ¹H NMR spectrum for H1.

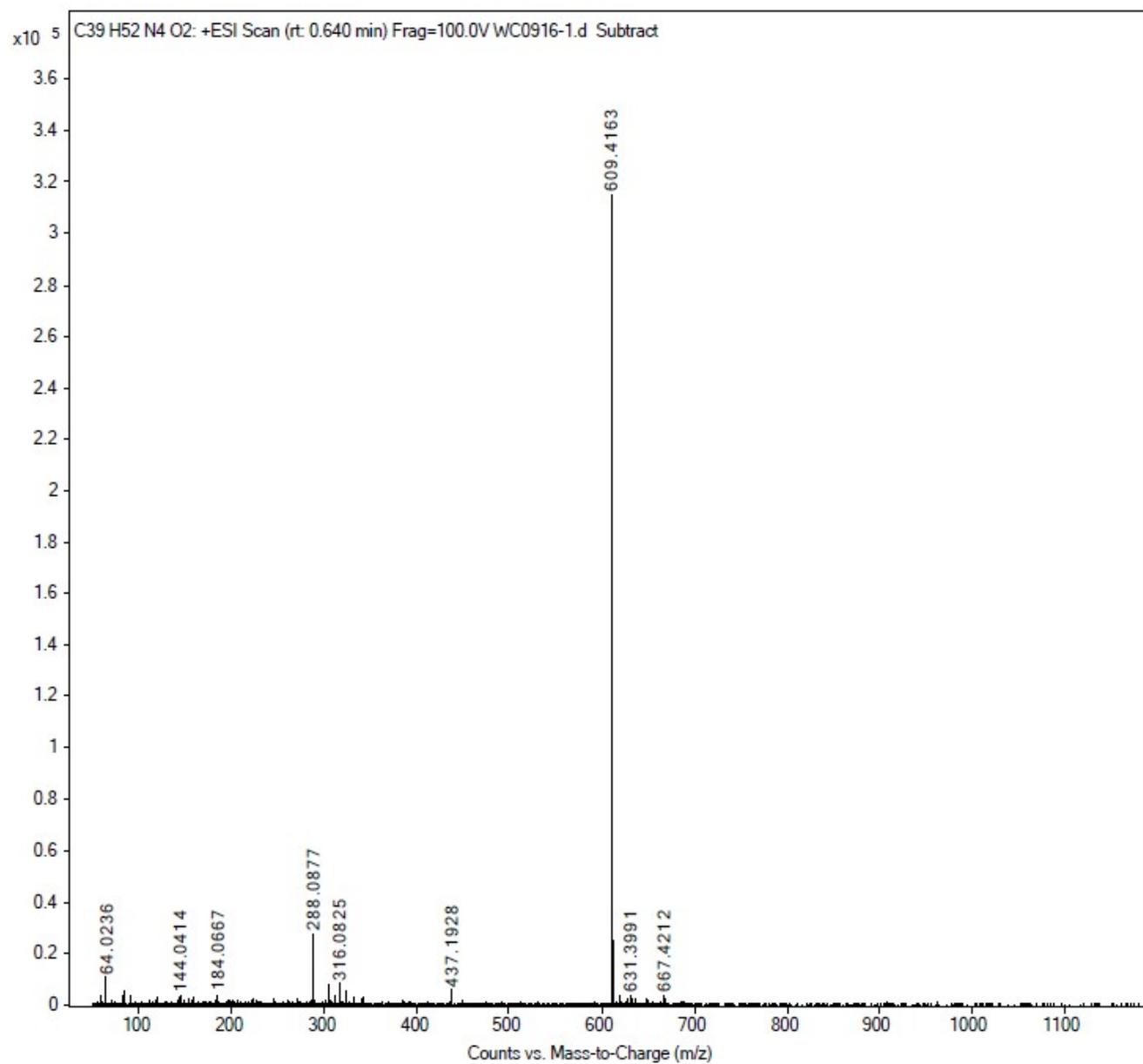


Figure S9. Mass spectrum for H1.

***N*-2-(cyclohexylamino)-1-(4-(4-methylpent-3-en-1-yl)cyclohex-3-en-1-yl)-2-oxoethyl)-3-(1*H*-indol-3-yl)-*N*-(3-(pyrrolidin-1-yl)propyl)propanamide (H2):**

Reddish brown solid. ¹H NMR (400 MHz, CDCl₃): δ 8.47 (d, *J* = 87.7 Hz, 2H), 7.49 (dd, *J* = 105.7, 7.7 Hz, 2H), 7.12 (p, *J* = 7.4 Hz, 3H), 5.56 – 5.19 (m, 1H), 5.07 (s, 1H), 3.81 – 2.48 (m, 13H), 2.39 – 1.42 (m, 25H), 1.40 – 0.60 (m, 9H). QTOF ESI MS: *m/z* calculated for C₃₈H₅₆N₄O₂ [M+H]⁺: 601.4476; Found 601.4477.

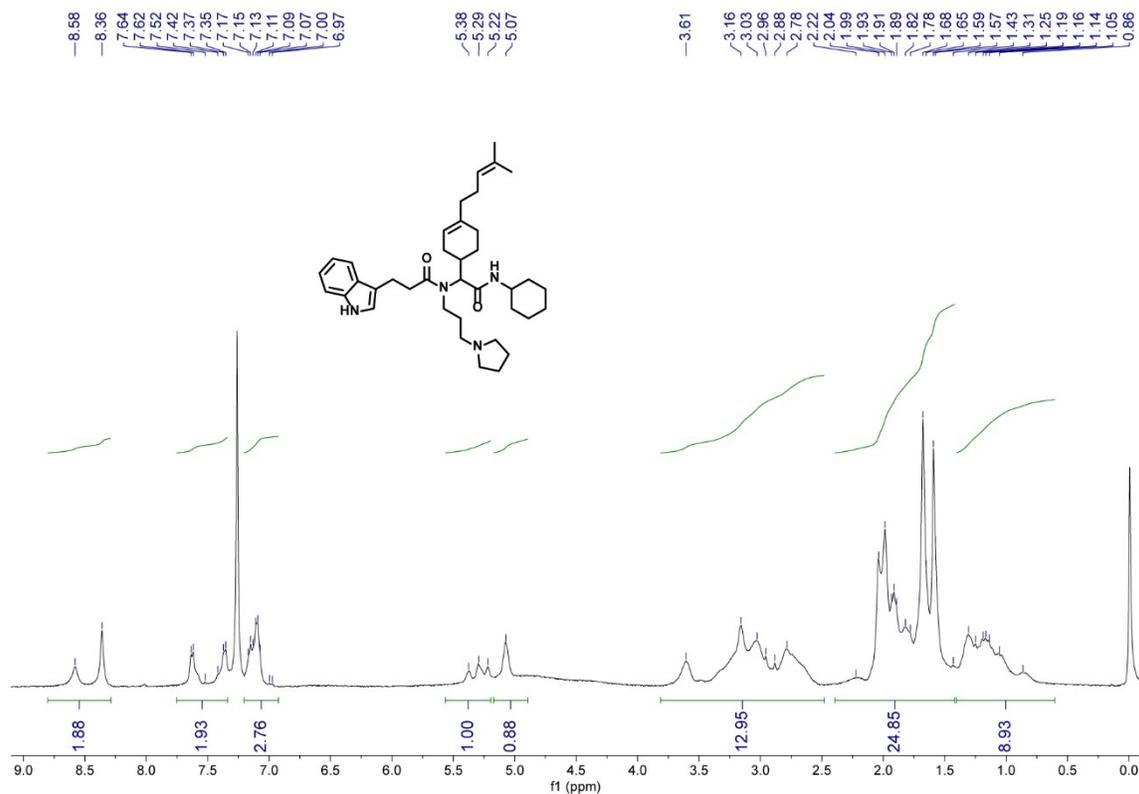


Figure S10. ¹H NMR spectrum for H2.

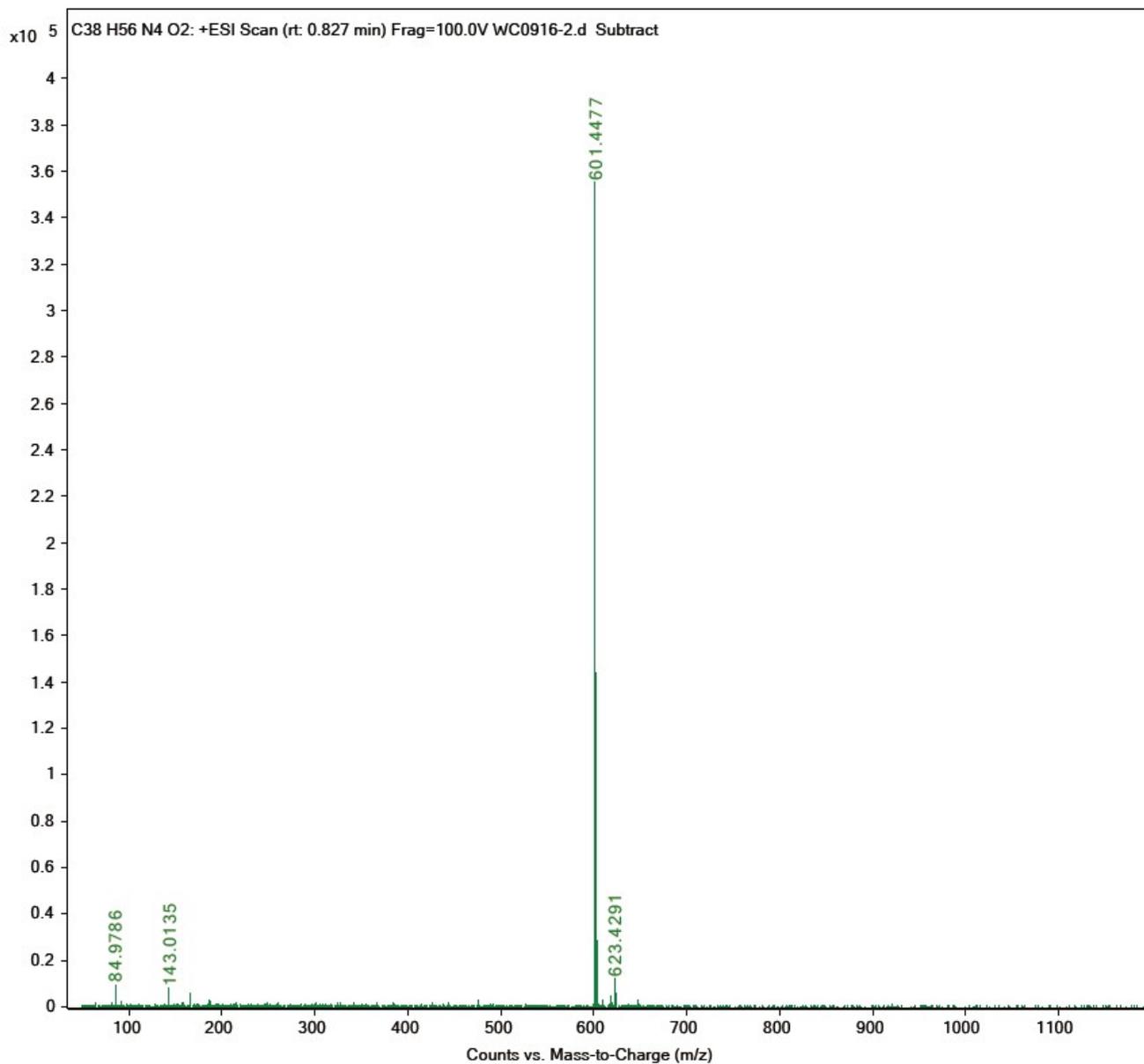


Figure S11. Mass spectrum for H2.

***N*-(2-(*tert*-butylamino)-1-(4-(4-methylpent-3-en-1-yl)cyclohex-3-en-1-yl)-2-oxoethyl)-3-(1*H*-indol-3-yl)-*N*-(3-(pyrrolidin-1-yl)propyl)propanamide (H3):**
 Bright yellow solid. ^1H NMR (400 MHz, CDCl_3): δ 8.41 (d, $J = 79.3$ Hz, 2H), 7.72 – 7.32 (m, 2H), 7.13 (dq, $J = 14.6, 7.2$ Hz, 3H), 5.43 – 5.17 (m, 1H), 5.08 (s, 1H), 3.48 – 2.47 (m, 21H), 2.13 – 1.62 (m, 13H), 1.43 – 1.12 (m, 9H), 0.92 (d, $J = 64.3$ Hz, 2H). QTOF ESI MS: m/z calculated for $\text{C}_{36}\text{H}_{54}\text{N}_4\text{O}_2$ $[\text{M}+\text{H}]^+$: 575.4320; Found 575.4323.

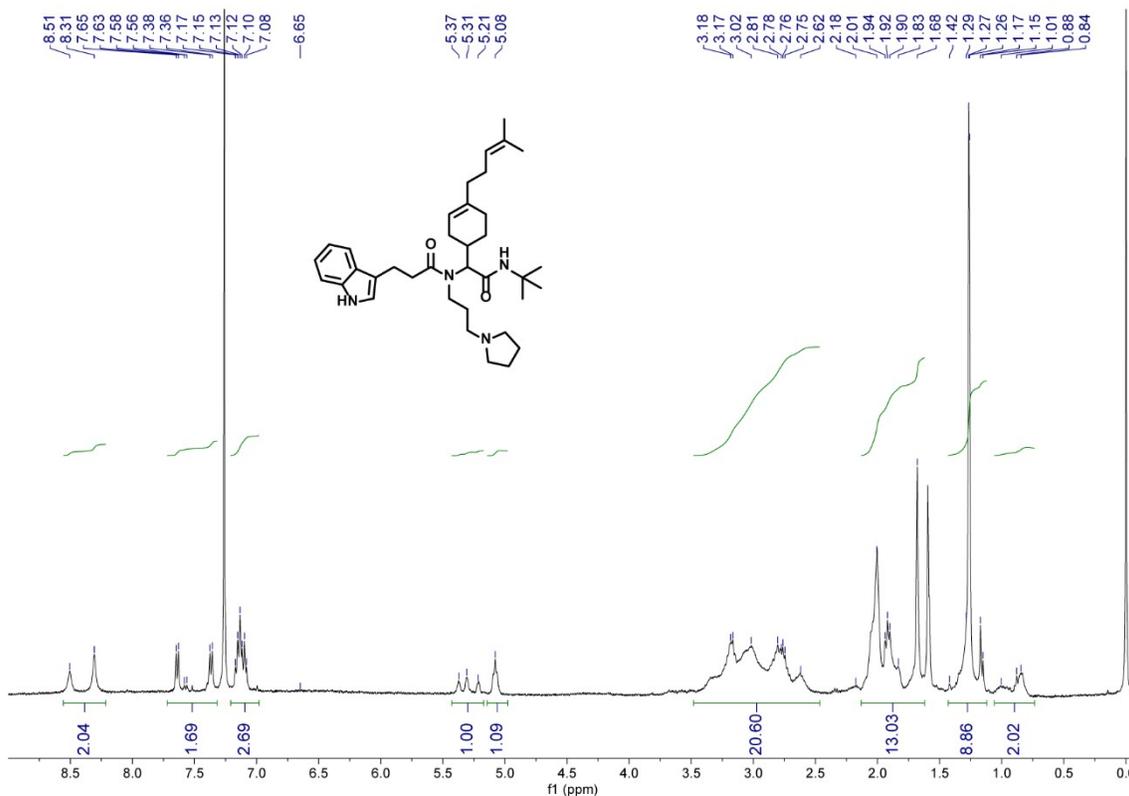


Figure S12. ^1H NMR spectrum for H3.

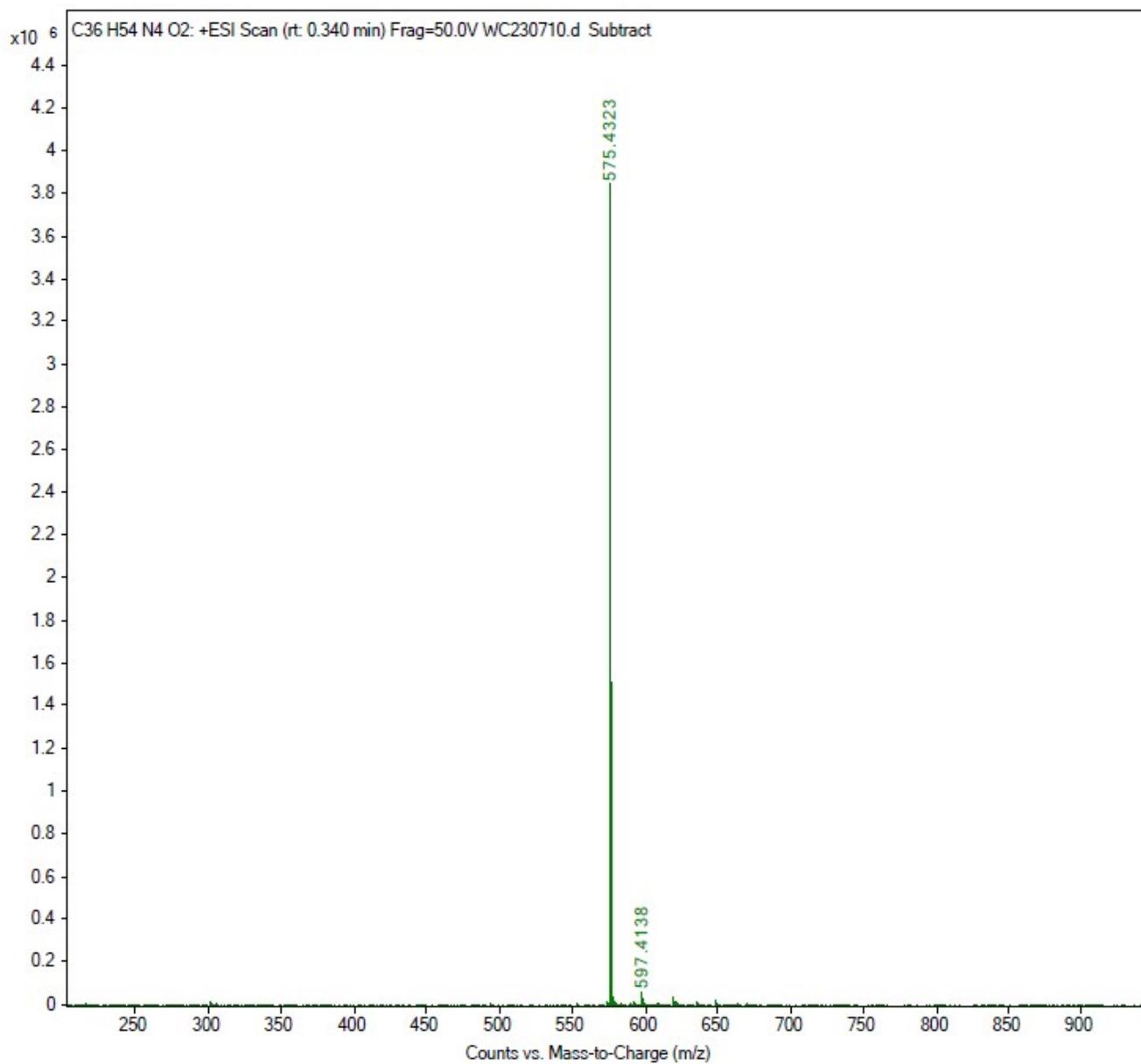


Figure S13. Mass spectrum for H3.

2-([1,1'-biphenyl]-4-yl)-N-(2-(butylamino)-1-(4-(4-methylpent-3-en-1-yl)cyclohex-3-en-1-yl)-2-oxoethyl)-N-(3-(pyrrolidin-1-yl)propyl)acetamide (H4):

Yellowish-brown solid. ^1H NMR (400 MHz, CDCl_3): δ 8.41 (s, 1H), 7.73 – 7.51 (m, 4H), 7.39 (dt, $J = 29.5, 6.7$ Hz, 5H), 5.54 – 5.19 (m, 1H), 5.09 (q, $J = 10.0, 6.2$ Hz, 1H), 4.06 – 2.61 (m, 10H), 2.52 – 1.87 (m, 13H), 1.84 – 1.40 (m, 9H), 1.39 – 1.01 (m, 7H), 0.99 – 0.77 (m, 3H), 0.69 (dt, $J = 11.1, 6.5$ Hz, 1H). QTOF ESI MS: m/z calculated for $\text{C}_{39}\text{H}_{55}\text{N}_3\text{O}_2$ $[\text{M}+\text{H}]^+$: 598.4367; Found 598.4369.

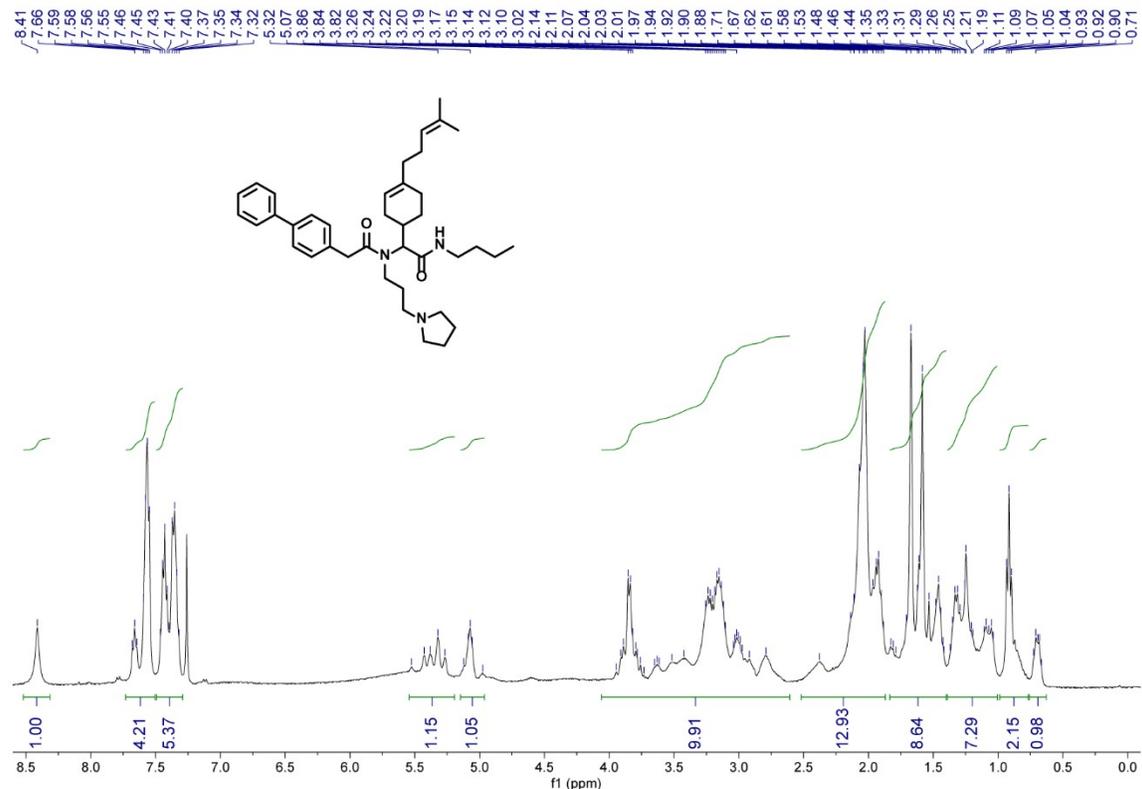


Figure S14. ^1H NMR spectrum for H4.

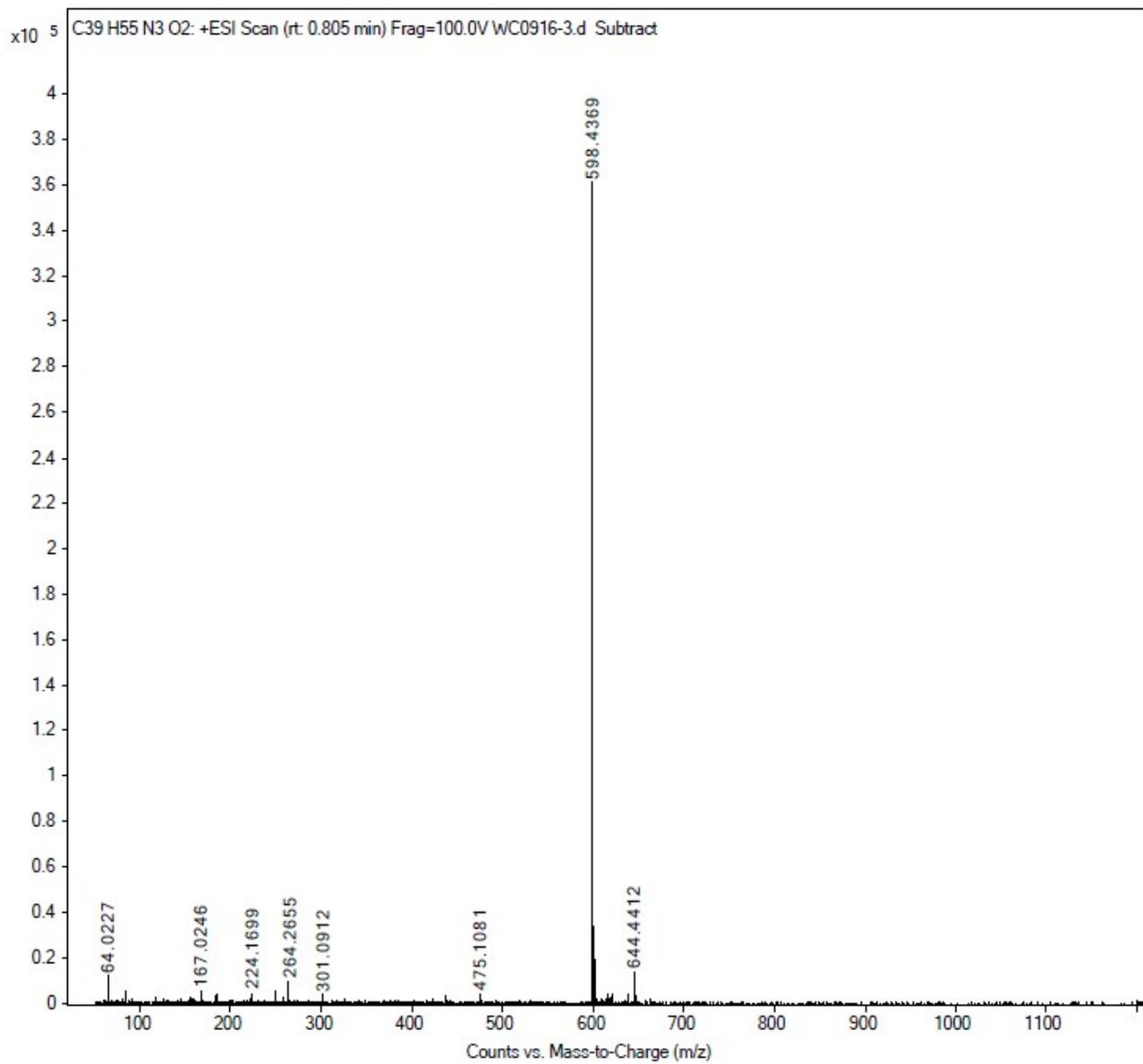


Figure S15. Mass spectrum for H4.

***N*-benzyl-2-(2-(4-isopropylphenyl)-*N*-(3-(pyrrolidin-1-yl)propyl)acetamido)-2-(4-(4-methylpent-3-en-1-yl)cyclohex-3-en-1-yl)acetamide (H5):**

Reddish brown solid. ¹H NMR (400 MHz, CDCl₃): δ 8.26 (s, 1H), 7.35 – 7.27 (m, 3H), 7.24 – 6.98 (m, 6H), 5.54 – 5.25 (m, 1H), 5.13 – 4.99 (m, 1H), 4.54 – 3.56 (m, 5H), 3.54 – 2.37 (m, 5H), 2.20 – 1.77 (m, 13H), 1.73 – 1.42 (m, 6H), 1.38 – 0.96 (m, 11H), 0.92 – 0.74 (m, 3H). QTOF ESI MS: m/z calculated for C₃₉H₅₅N₃O₂ [M+H]⁺: 598.4367; Found 598.4367.

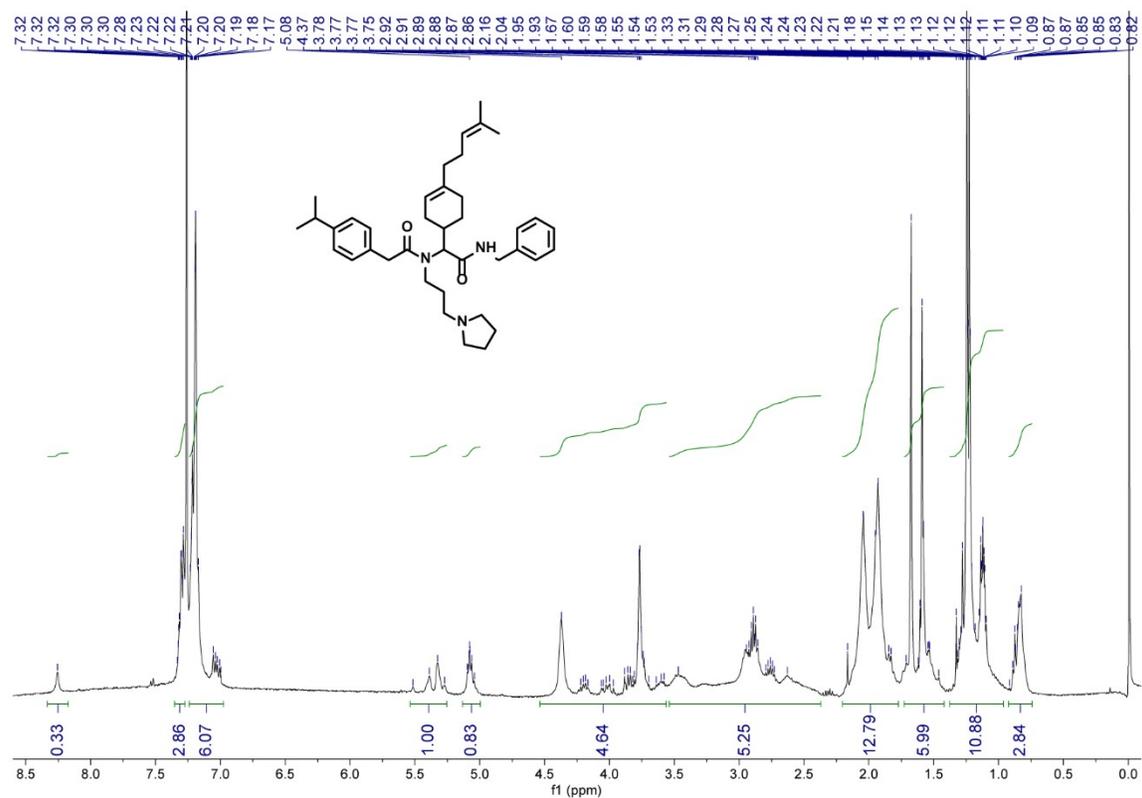


Figure S16. ¹H NMR spectrum for H5.

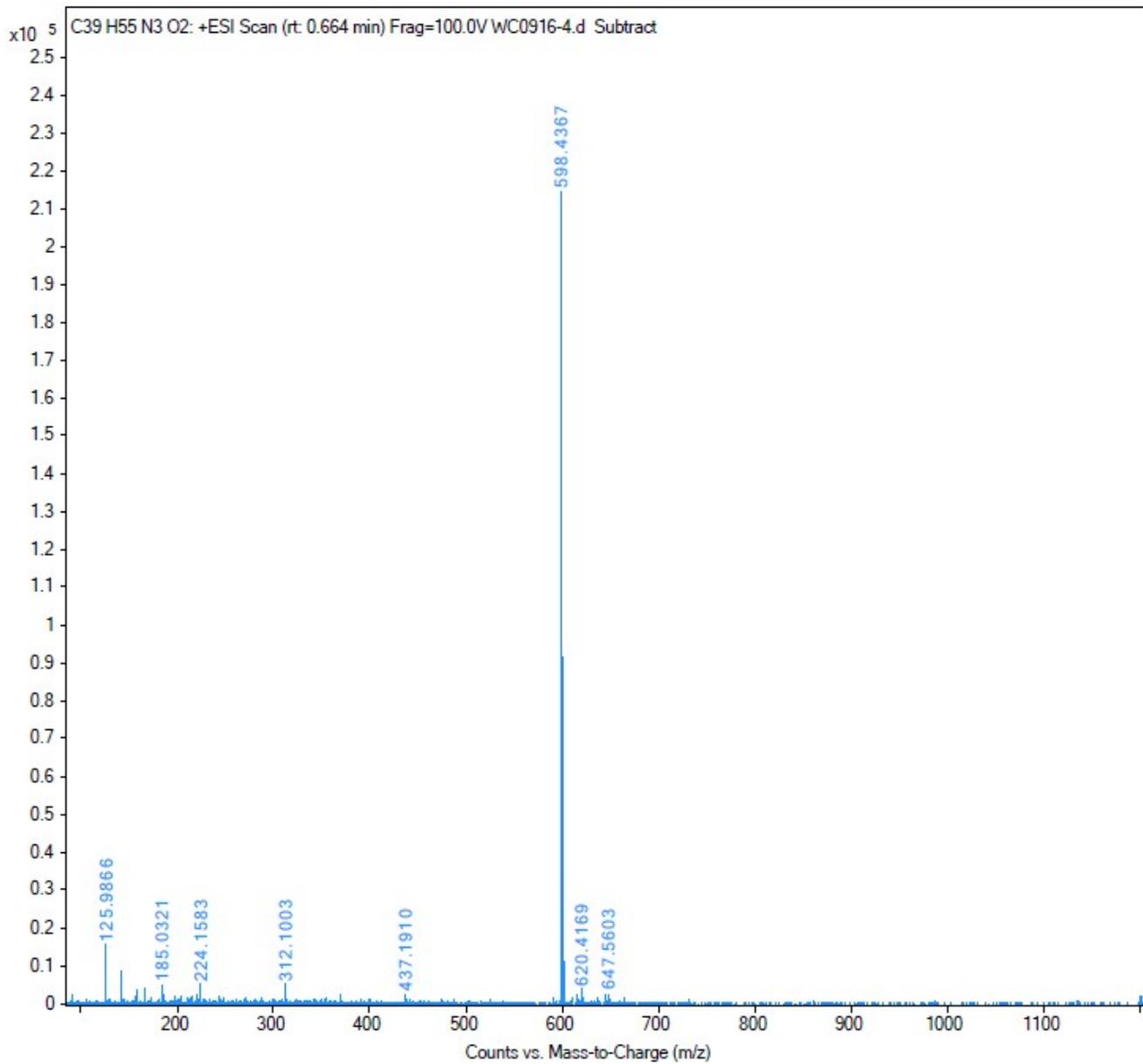


Figure S17. Mass spectrum for H5.

***N*-butyl-2-(2-(4-isopropylphenyl)-*N*-(3-(pyrrolidin-1-yl)propyl)acetamido)-2-(4-(4-methylpent-3-en-1-yl)cyclohex-3-en-1-yl)acetamide (H6):**

Yellowish-brown solid. ¹H NMR (400 MHz, CDCl₃): δ 8.34 (s, 1H), 7.22 (d, *J* = 30.3 Hz, 6H), 5.34 (q, *J* = 23.9, 20.3 Hz, 1H), 5.08 (s, 1H), 4.49 (d, *J* = 94.7 Hz, 3H), 3.98 – 2.58 (m, 10H), 2.50 – 1.75 (m, 11H), 1.75 – 1.38 (m, 9H), 1.37 – 0.99 (m, 11H), 0.97 – 0.71 (m, 4H). QTOF ESI MS: *m/z* calculated for C₃₆H₅₇N₃O₂ [M+H]⁺: 564.4524; Found 564.4525.

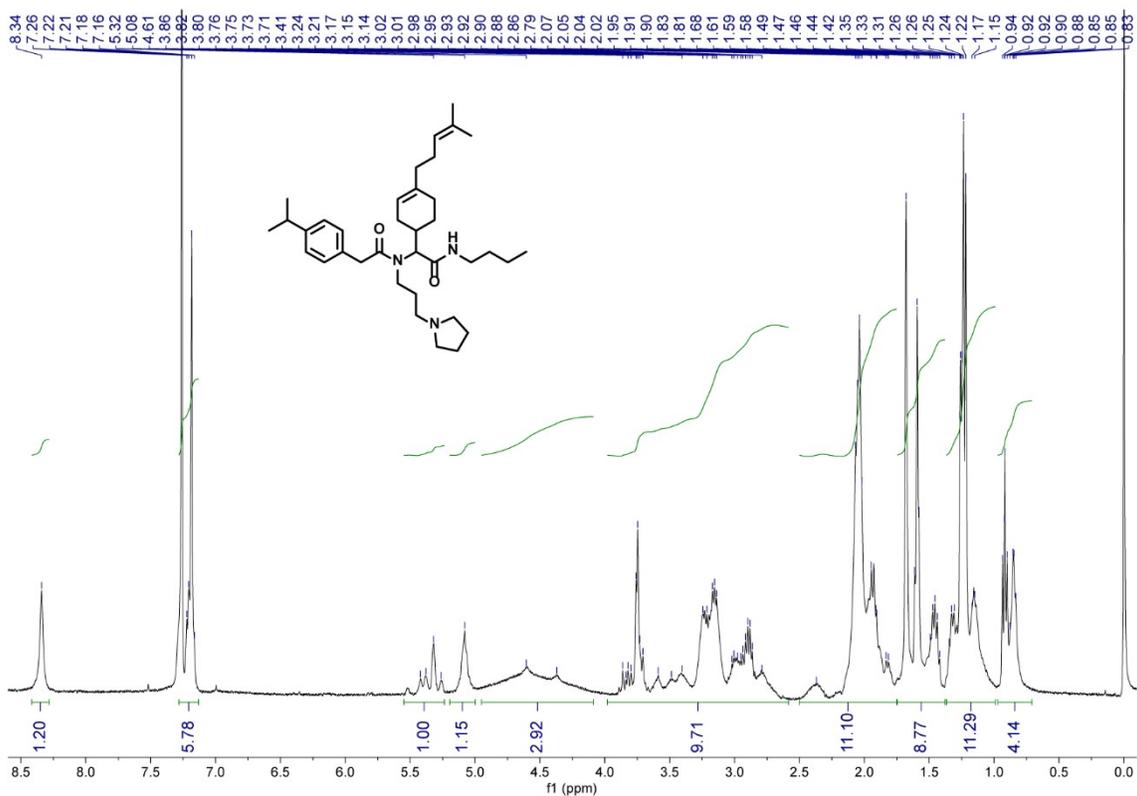


Figure S18. ¹H NMR spectrum for H6.

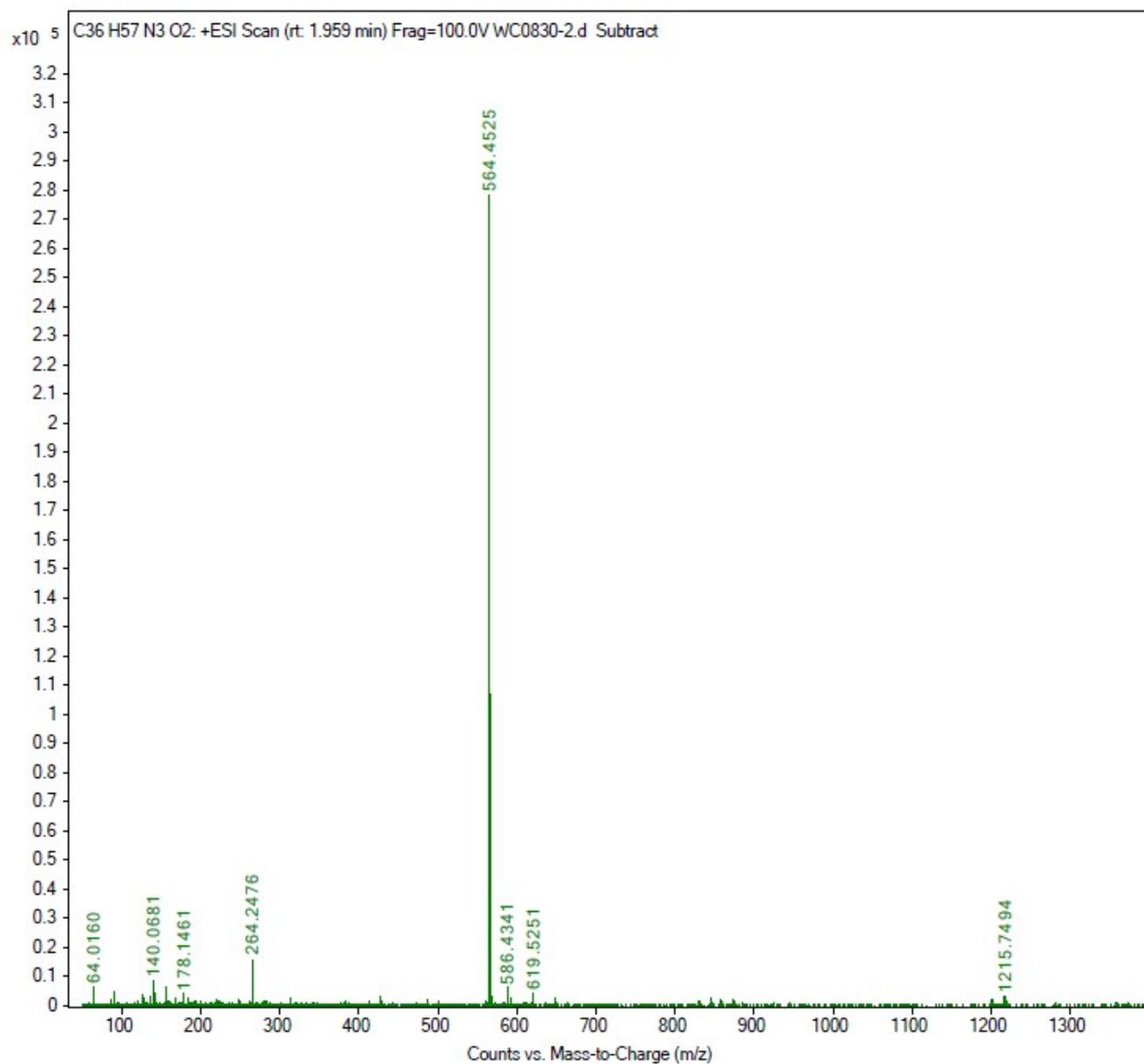


Figure S19. Mass spectrum for H6.

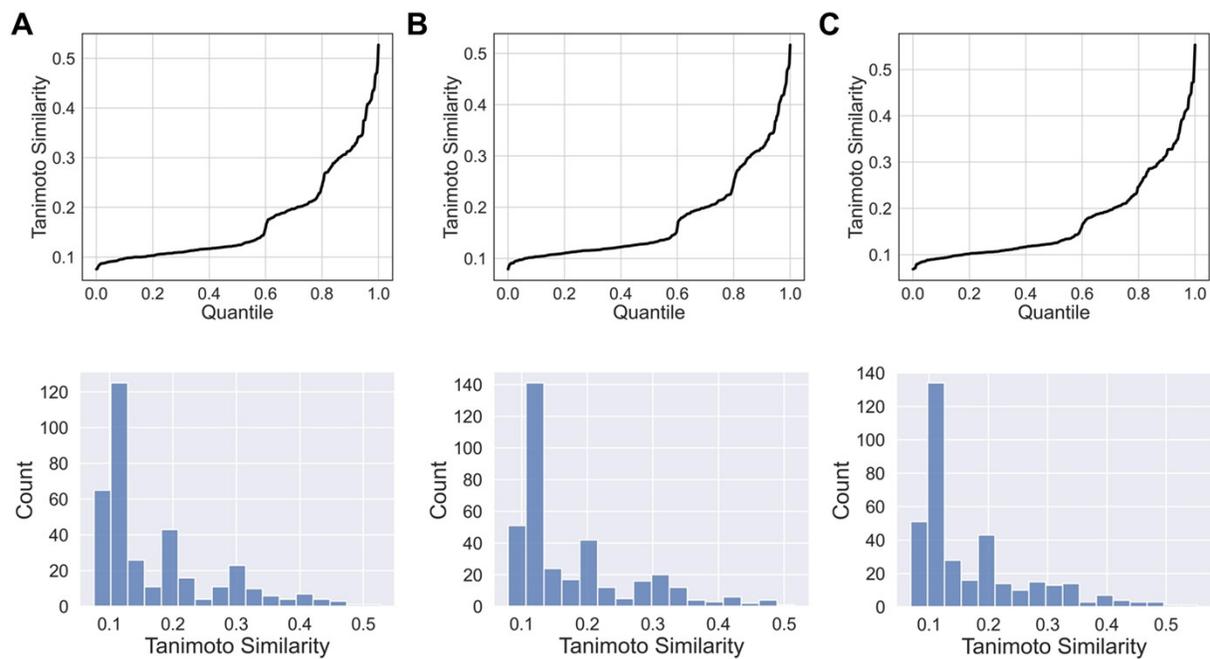


Figure S20. Quantile plots and histograms of Tanimoto similarity of A) H4, B) H5, and C) H6 to the preliminary dataset. The three selected compounds showed less than 0.5 similarity to most compounds in the preliminary dataset, which indicates they have good structural diversity⁴.

Table S1. Reagents for the library construction.

Serial number	Name	CAS	Smiles	Producer
a1	4-Carboxyphenylboronic acid	14047-29-1	<chem>OB(C1=CC=C(C(O)=O)C=C1)O</chem>	Energy Chemical
a2	Lactobionic acid	96-82-2	<chem>O[C@@H]1[C@@H](O)[C@@H](CO)O[C@H]1O</chem>	Energy Chemical
a3	Acetic acid	64-19-7	<chem>CC(O)=O</chem>	Sinopharm chemical reagent
a4	Benzoylformic acid	611-73-4	<chem>O=C(C(C1=CC=CC=C1)=O)O</chem>	Aladdin
a5	Acrylic acid	79-10-7	<chem>C=CC(O)=O</chem>	Energy Chemical
a6	Linoleic acid	60-33-3	<chem>CCCCC/C=C\C/C=C\C\CCCC(O)=O</chem>	Sigma
a7	Tretinoin	302-79-4	<chem>OC(/C=C(C)/C=C/C=C(C)/C=C/C1=C(C)CCCC1(C)C)=O</chem>	Energy Chemical
a8	Pentadecafluorooctanoic acid	335-67-1	<chem>O=C(O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F</chem>	Aladdin
a9	4-(Phenylazo)benzoic acid	1562-93-2	<chem>O=C(O)C1=CC=C(N=NC2=CC=CC=C2)C=C1</chem>	Aladdin
a10	Stearic acid	57-11-4	<chem>CCCCCCCCCCCCCCCCC(O)=O</chem>	Macklin
a11	Oleic acid	112-80-1	<chem>CCCCCCCC/C=C\C\CCCC(O)=O</chem>	Macklin
a12	4-Morpholineacetic acid	3235-69-6	<chem>O=C(O)CN1CCOCC1</chem>	Aladdin
a13	4-Biphenylacetic acid	5728-52-9	<chem>O=C(O)CC1=CC=C(C2=CC=CC=C2)C=C1</chem>	Aladdin
a14	2-(4-Isopropylphenyl)acetic acid	4476-28-2	<chem>O=C(O)CC1=CC=C(C(C)C)C=C1</chem>	Aladdin
a15	Trichloroacetic acid	76-03-9	<chem>O=C(O)C(Cl)(Cl)Cl</chem>	Macklin
a16	Palmitic acid	57-10-3	<chem>CCCCCCCCCCCCCCCC(O)=O</chem>	Macklin
a17	Trifluoroacetic acid	76-05-1	<chem>O=C(O)C(F)(F)F</chem>	Chembee
a18	Benzoic acid	65-85-0	<chem>O=C(O)C1=CC=CC=C1</chem>	Sinopharm chemical reagent
a19	3-Bromo-2-(bromomethyl)propionic acid	41459-42-1	<chem>O=C(O)C(CBr)CBr</chem>	TCI
a20	2-Naphthylacetic acid	581-96-4	<chem>O=C(O)CC1=CC=C2C=CC=CC2=C1</chem>	TCI
a21	2-(4-Hydroxybenzeneazo)benzoic acid	1634-82-8	<chem>O=C(O)C1=CC=CC=C1N=NC2=CC=C(O)C=C2</chem>	Sigma
a22	2-Bromo-2-methylpropionic acid	2052-01-9	<chem>CC(C)(Br)C(O)=O</chem>	Sigma

Serial number	Name	CAS	Smiles	Producer
a23	Gallic acid	149-91-7	<chem>OC1=CC(C(O)=O)=CC(O)=C1O</chem>	Aladdin
a24	Levulinic acid	123-76-2	<chem>O=C(C)CCC(O)=O</chem>	Energy Chemical
a25	3-Indolepropionic acid	830-96-6	<chem>O=C(O)CCC1=CNC2=CC=CC=C2</chem>	Aladdin
a26	3,4-Dihydroxybenzoic acid	99-50-3	<chem>O=C(O)C1=CC=C(O)C(O)=C1</chem>	Duoxi Biological
a27	3-Chloropropionic acid	107-94-8	<chem>O=C(O)CCCl</chem>	Energy Chemical
a28	2-Naphthoxyacetic acid	120-23-0	<chem>O=C(O)COC1=CC=C2C=CC=CC2=C1</chem>	Aladdin
a29	3-Bromopropionic acid	590-92-1	<chem>O=C(O)CCBr</chem>	Chembee
a30	1-Adamantaneacetic acid	4942-47-6	<chem>O=C(O)CC12CC3CC(C2)CC(C3)C1</chem>	Energy Chemical
a31	1-Adamantanecarboxylic acid	828-51-3	<chem>O=C(C12CC3CC(C2)CC(C3)C1)O</chem>	Energy Chemical
a32	Bromopyruvic acid	1113-59-3	<chem>O=C(CBr)C(O)=O</chem>	Energy Chemical
a33	3-(2-Hydroxyphenyl)propionic acid	495-78-3	<chem>O=C(O)CCC1=CC=CC=C1O</chem>	Alfa Aesar
a34	4-Azidobenzoic acid	6427-66-3	<chem>O=C(O)C1=CC=C(N=[N+]=[N-])C=C1</chem>	Aladdin
a35	Chloroambucil	305-03-3	<chem>OC(CCCC1=CC=C(N(CCC1)CCCl)C=C1)O</chem>	Aladdin
a36	3-Phosphonopropionic acid	5962-42-5	<chem>O=C(O)CCP(O)(O)=O</chem>	TCI
a37	Caffeic acid	331-39-5	<chem>OC1=CC(C=CC(O)=O)=CC=C1O</chem>	Aladdin
a38	Tazobactam acid	89786-04-9	<chem>O=C1C[C@]2([H])N1[C@@H](C(O)=O)[C@@](CN3C=CN=N3)(C)S2(=O)=O</chem>	Aladdin
b1	2-(1 <i>H</i> -1,2,3-triazol-1-yl)ethanamine	4320-94-9	<chem>NCCN1N=NC=C1</chem>	Leyan
b2	5-(Dimethylamino)amylamine	3209-46-9	<chem>NCCCCCN(C)C</chem>	Bidepharm
b3	(1-Ethyl-4-piperidiny)l)methanamine	21168-71-8	<chem>NCC1CCN(CC)CC1</chem>	Bidepharm
b4	1-(3-Aminopropyl)pyrrolidine	23159-07-1	<chem>NCCCCN1CCCC1</chem>	Aladdin
b5	2-Aminopentane	63493-28-7	<chem>CC(N)CCC</chem>	Aladdin
b6	DL-Alaninol	6168-72-5	<chem>NC(C)CO</chem>	Aladdin
b7	1-Aminoundecane	7307-55-3	<chem>CCCCCCCCCCCN</chem>	Macklin
b8	2-Amino-4-chlorophenol-6-sulfonic Acid	88-23-3	<chem>OC1=C(S(=O)(O)=O)C=C(Cl)C=C1N</chem>	Macklin
b9	Sulfanilic acid	121-57-3	<chem>O=S(O)(C1=CC=C(N)C=C1)O</chem>	Macklin
b10	1-Naphthylamine-5-sulfonic acid	84-89-9	<chem>NC1=CC=C(C(S(=O)(O)=O)C2=CC=C1)O</chem>	Aladdin

Serial number	Name	CAS	Smiles	Producer
b11	Aminomethanesulfonic acid	13881-91-9	<chem>O=S(CN)(O)=O</chem>	Aladdin
b12	Sodium 2-[(2-aminoethyl)amino]ethane sulphonate	34730-59-1	<chem>NCCNCCS(=O)([O-])=O.[Na+]</chem>	Macklin
b13	Taurine	107-35-7	<chem>OS(=O)(CCN)=O</chem>	Aladdin
b14	3-Amino-1-propanesulfonic acid	3687-18-1	<chem>O=S(CCCN)(O)=O</chem>	Aladdin
b15	2-Amino-3-methoxyphenol	40925-69-7	<chem>OC1=CC=CC(OC)=C1N</chem>	Bidepharm
b16	3-Amino-5-methylphenol	76619-89-1	<chem>OC1=CC(C)=CC(N)=C1</chem>	Macklin
b17	5-Amino-2-methoxyphenol	1687-53-2	<chem>OC1=CC(N)=CC=C1OC</chem>	Aladdin
b18	3-Morpholinopropylamine	123-00-2	<chem>NCCCN1CCOCC1</chem>	Bidepharm
b19	Tryptamine	61-54-1	<chem>NCCC1=CNC2=CC=CC=C12</chem>	Macklin
b20	Benzylamine	100-46-9	<chem>NCC1=CC=CC=C1</chem>	Aladdin
b21	Hexylamine	111-26-2	<chem>NCCCCCC</chem>	Macklin
c1	1-Ethyl-1 <i>H</i> -pyrazole-4-carbaldehyde	304903-10-4	<chem>O=CC1=CN(CC)N=C1</chem>	Macklin
c2	Benzotriazole-1-carbaldehyde	72773-04-7	<chem>O=CN1N=NC2=CC=CC=C21</chem>	Leyan
c3	2-[1,2,4]Triazol-1-yl-benzaldehyde	138479-53-5	<chem>O=CC1=CC=CC=C1N2N=CN=C2</chem>	Macklin
c4	1,2-Dimethylbenzimidazole-5-carbaldehyde	4597-21-1	<chem>O=CC1=CC=C2N(C)C(C)=NC2=C1</chem>	Bidepharm
c5	3-(Dimethylamino)-2-methyl-2-propenal	19125-76-9	<chem>O=CC(C)=CN(C)C</chem>	Macklin
c6	1-[(Diethylamino)methyl]cyclopentanecarboxaldehyde	208349-33-1	<chem>O=CC1(CN(CC)CC)CCC1</chem>	Bidepharm
c7	p-Dimethylaminobenzaldehyde	100-10-7	<chem>O=CC1=CC=C(N(C)C)C=C1</chem>	Aladdin
c8	4-(4-Methylpent-3-en-1-yl)cyclohex-3-ene-1-carbaldehyde	37677-14-8	<chem>O=CC1CC=C(CC/C=C(C)/C)CC1</chem>	Aladdin
c9	4-Methyl-2-phenyl-2-pentenal	26643-91-4	<chem>CC(C)C=C(C1=CC=CC=C1)C=O</chem>	Aladdin
c10	4-Formylbenzene-1,3-disulfonic acid	88-39-1	<chem>O=S(C1=CC=C(C=O)C(S(=O)(O)=O)=C1)(O)=O</chem>	Bidepharm
c11	3,5-Bis(1,1-dimethylethyl)-2-hydroxy-benzaldehyde	37942-07-7	<chem>O=CC1=CC(C(C)C)C(C)=CC(C(C)C)C=C1O</chem>	Macklin
c12	4-Hydroxy benzaldehyde	123-08-0	<chem>O=CC1=CC=C(O)C=C1</chem>	Aladdin
c13	3,5-Dichloro-4-hydroxybenzaldehyde	2314-36-5	<chem>O=CC1=CC(Cl)=C(O)C(Cl)=C1</chem>	Energy Chemical

Serial number	Name	CAS	Smiles	Producer
c14	3-Phenylpropionaldehyde	104-53-0	<chem>O=CCCC1=CC=CC=C1</chem>	Macklin
c15	2-Methylvaleraldehyde	123-15-9	<chem>CCCC(C)C=O</chem>	Macklin
c16	α -Methylcinnamaldehyde	101-39-3	<chem>O=CC(C)=CC1=CC=CC=C1</chem>	Aladdin
c17	Valeraldehyde	110-62-3	<chem>CCCCC=O</chem>	Macklin
c18	2-Thiazolecarboxaldehyde	10200-59-6	<chem>O=CC1=NC=CS1</chem>	Macklin
c19	Trans-Cinnamaldehyde	14371-10-9	<chem>O=C/C=C/C1=CC=CC=C1</chem>	Aladdin
c20	Salicylaldehyde	90-02-8	<chem>OC1=CC=CC=C1C=O</chem>	Macklin
d1	Benzyl Isocyanide	10340-91-7	<chem>[C-]#[N+]CC1=CC=CC=C1</chem>	Energy Chemical
d2	2-Isocyanonaphthalene	10124-78-4	<chem>[C-]#[N+]C1=CC=C2C=CC=CC2=C1</chem>	Energy Chemical
d3	1-Isocyanobutane	2769-64-4	<chem>CCCC[N+]#[C-]</chem>	Acme
d4	Isocyanocyclohexane	931-53-3	<chem>[C-]#[N+]C1CCCCC1</chem>	Aladdin
d5	<i>Tert</i> -Butyl isocyanide	7188-38-7	<chem>CC(C)([N+]#[C-])C</chem>	Aladdin
d6	Methyl isocyanoacetate	39687-95-1	<chem>O=C(OC)C[N+]#[C-]</chem>	Macklin
d7	Tosylmethyl isocyanide	36635-61-7	<chem>O=S(C[N+]#[C-])(C1=CC=C(C)C=C1)=O</chem>	Macklin

Table S2. Yields and MIC values of H1-6.

Label	H1	H2	H3	H4	H5	H6
Yield	52.6%	89.0%	57.4%	40.0%	60.5%	30.3%
MIC (μM)	24	24	24	12	12	12

References

1. M. Zhou, Y. Qian, J. Xie, W. Zhang, W. Jiang, X. Xiao, S. Chen, C. Dai, Z. Cong, Z. Ji, N. Shao, L. Liu, Y. Wu and R. Liu, Poly(2-Oxazoline)-Based Functional Peptide Mimics: Eradicating MRSA Infections and Persisters while Alleviating Antimicrobial Resistance, *Angew. Chem. Int. Ed.*, 2020, **59**, 6412-6419.
2. H. Zhang, Q. Chen, J. Xie, Z. Cong, C. Cao, W. Zhang, D. Zhang, S. Chen, J. Gu, S. Deng, Z. Qiao, X. Zhang, M. Li, Z. Lu and R. Liu, Switching from membrane disrupting to membrane crossing, an effective strategy in designing antibacterial polypeptide, *Sci Adv*, 2023, **9**, eabn0771.
3. L. Zou, X. Li, Y. Huang, C. Wang, Y. Fang, J. Zhao, Q. Jin and J. Ji, Raspberry-like gold nanozyme-hybrid liposomes for hypoxia-enhanced biofilm eradication, *Nano Today*, 2023, **50**, 101828.
4. F. Wong, E. J. Zheng, J. A. Valeri, N. M. Donghia, M. N. Anahtar, S. Omori, A. Li, A. Cubillos-Ruiz, A. Krishnan, W. Jin, A. L. Manson, J. Friedrichs, R. Helbig, B. Hajian, D. K. Fiejtek, F. F. Wagner, H. H. Soutter, A. M. Earl, J. M. Stokes, L. D. Renner and J. J. Collins, Discovery of a structural class of antibiotics with explainable deep learning, *Nature*, 2024, **626**, 177-185.