

Digitalized Peptides: Unveiling Ocean Secrets with AI and Big Data

Byeong-Chul Kang, Ph.D

2024-09-24

INSILICOGEN

www.insilicogen.com

본 문서의 모든 콘텐츠는 저작권법의 보호를 받는 저작물로 별도의 저작권 표시 또는 다른 출처를 명시한 경우를 제외하고는 (주)인실리코젠에 저작권이 있습니다.
저작권 표시 또는 기타 소유권 표시를 삭제해서도 안되며, 당사와의 협의 또는 허락없이 무단 복제, 변경, 배포를 금지합니다.
저작권 관련 문의사항이 있으시면 info@insilicogen.com으로 연락바랍니다.
© 2024 INSILICOGEN, INC. ALL RIGHTS RESERVED.

Evolution of Various Marine Life



Dark oxygen in Earth's deep seas may hold clues about alien life

Useful peptides from evolutions

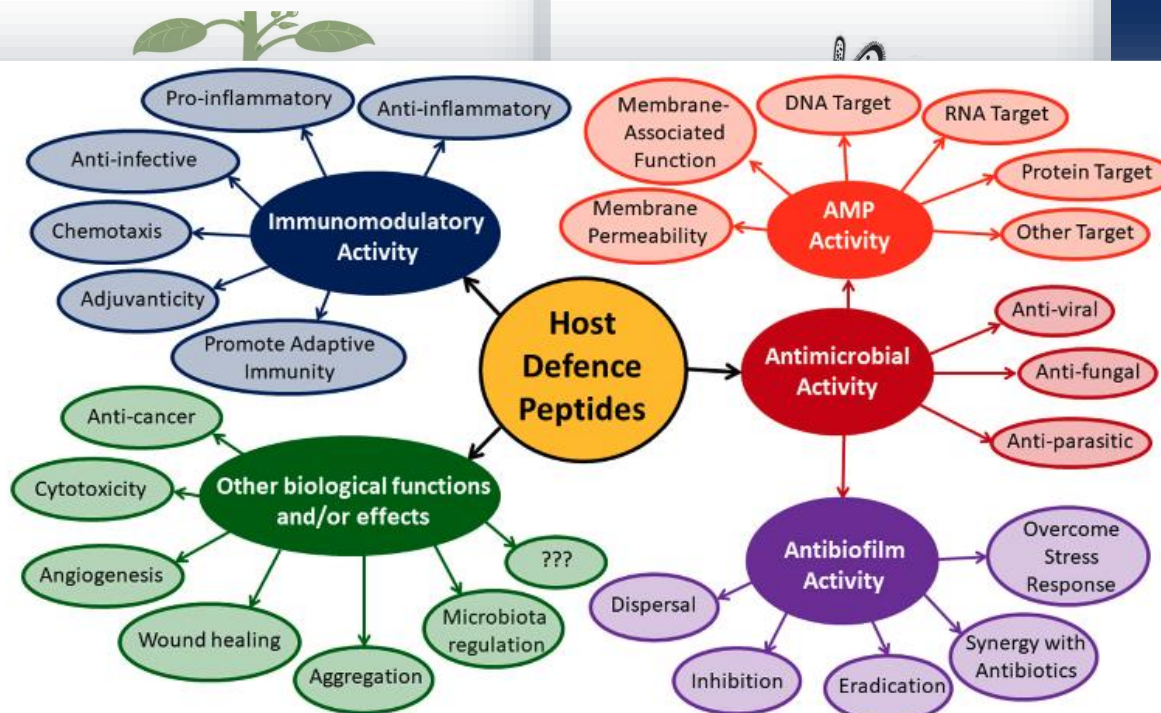
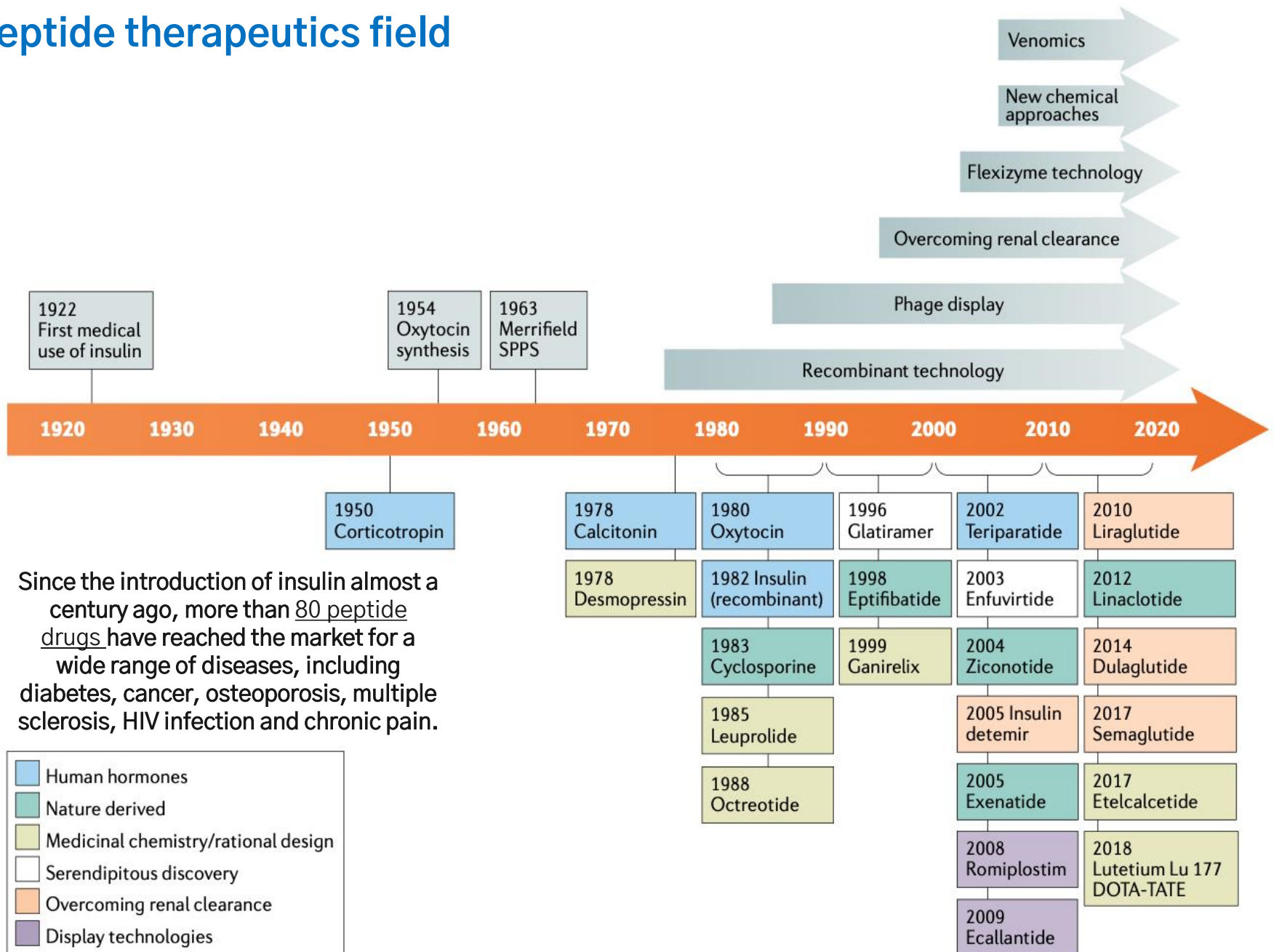


FIGURE 2 | Diversity of biological functions described for HDPs.

* Haney EF, Straus SK, Hancock REW. Reassessing the Host Defense Peptide Landscape. Front Chem. 2019 Feb 4;7:43.

- 1) Anti-Microbial Peptides
- 2) Anti-Bacterial Peptides
- 3) Anti-Bacteria-Gram-positive-peptides
- 4) Anti-Bacteria-Gram-Negative-peptides
- 5) Anti-Biofilm Peptides
- 6) Anti-Viral Peptides
- 7) Anti-HIV Peptides
- 8) Anti-Corona Peptides
- 9) Anti-Fungal Peptides
- 10) Anti-Parasitic Peptides
- 11) Quorum Sensing Peptides
- 12) Anti-Cancer Peptides
- 13) Anti-Hypertension Peptides
- 14) Cell-Penetrating Peptides
- 15) Tumor Homing Peptides
- 16) Blood Brain Barrier Peptides
- 17) Neuro Peptides
- 18) Cosmetic Collagen Peptides
- 19) Anti-Angiogenic Peptides
- 20) Anti-Inflammatory Peptides
- 21) Surface-binding Peptides
- 22) Human Protein Binding Peptides
- 23) Toxin Peptides
- 24) Hemolytic Peptides
- 25) Antioxidant Peptides

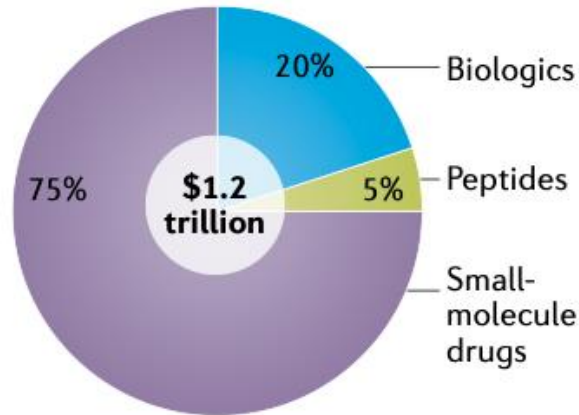
Peptide therapeutics field



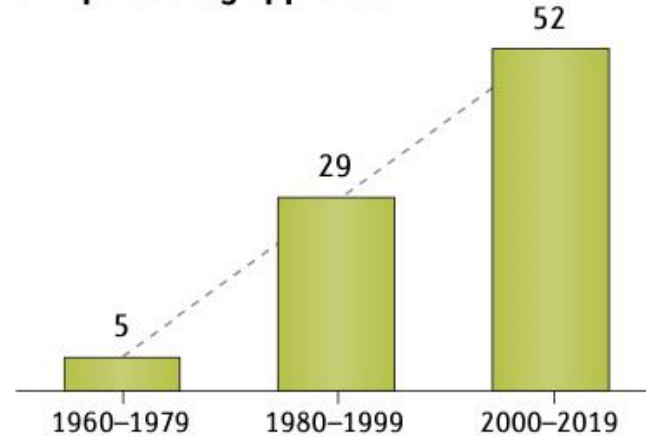
* Muttenthaler, M., King, G.F., Adams, D.J. *et al.* Trends in peptide drug discovery. *Nat Rev Drug Discov* 20, 309–325 (2021). <https://doi.org/10.1038/s41573-020-00135-8>

The peptide drug market

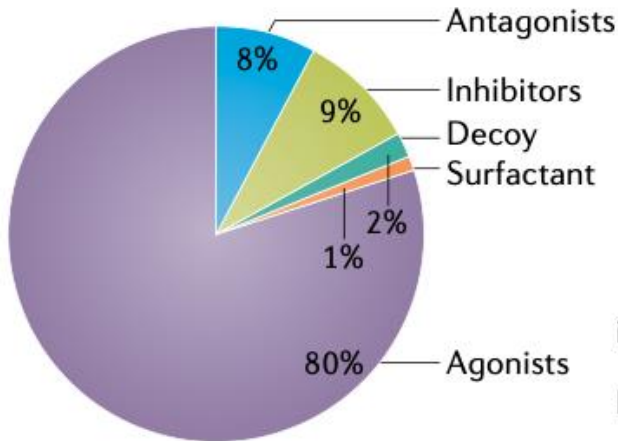
a Global pharmaceutical market (2019)



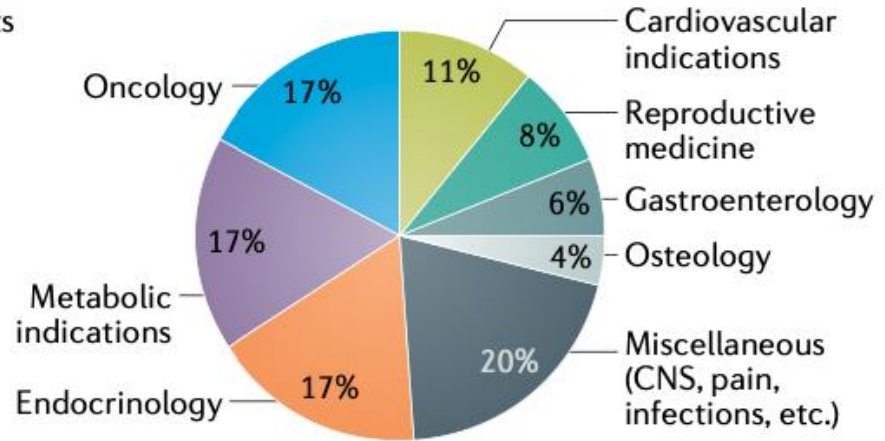
b Peptide drug approvals



c Distribution of function

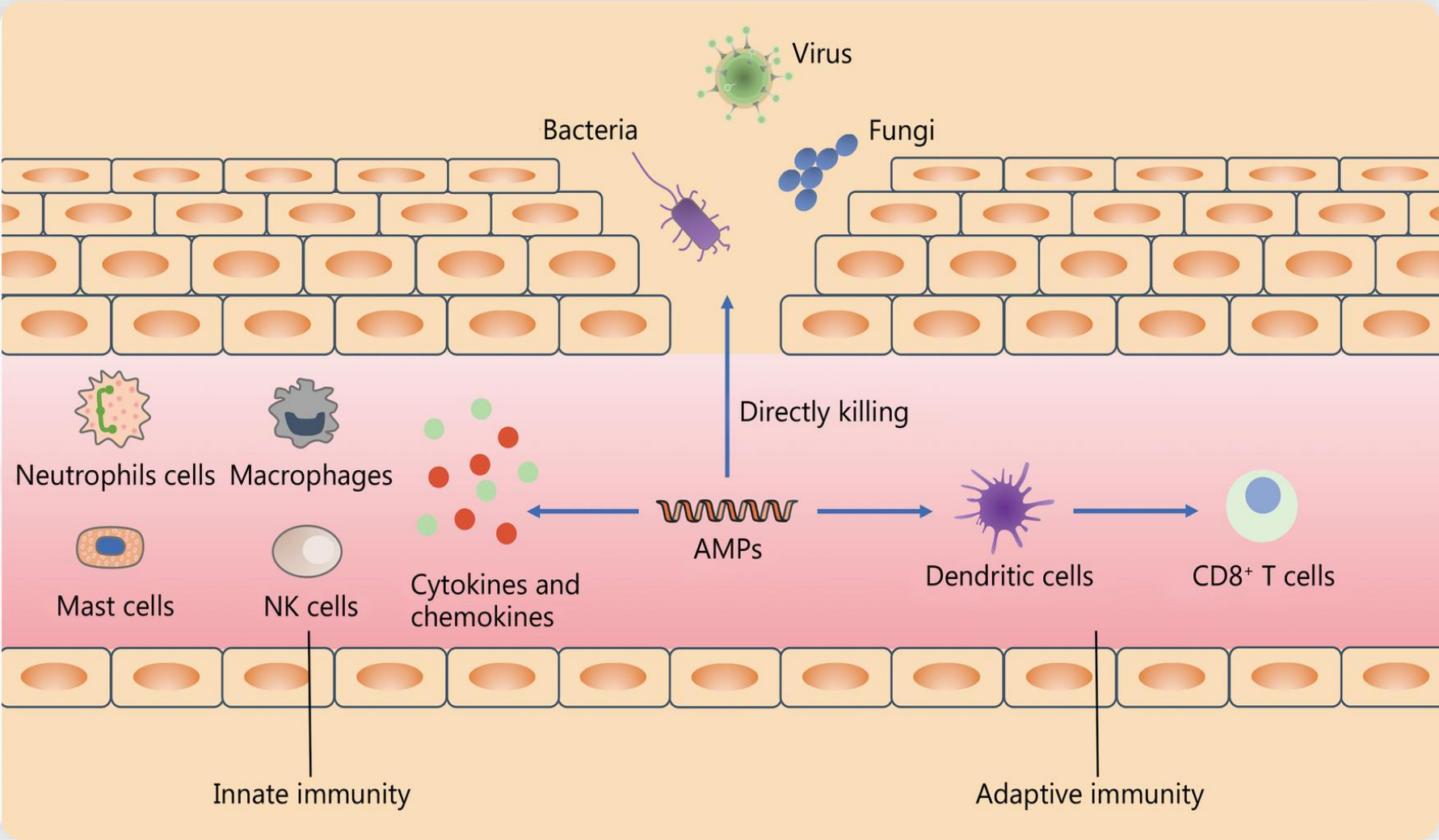


d Therapeutic indications

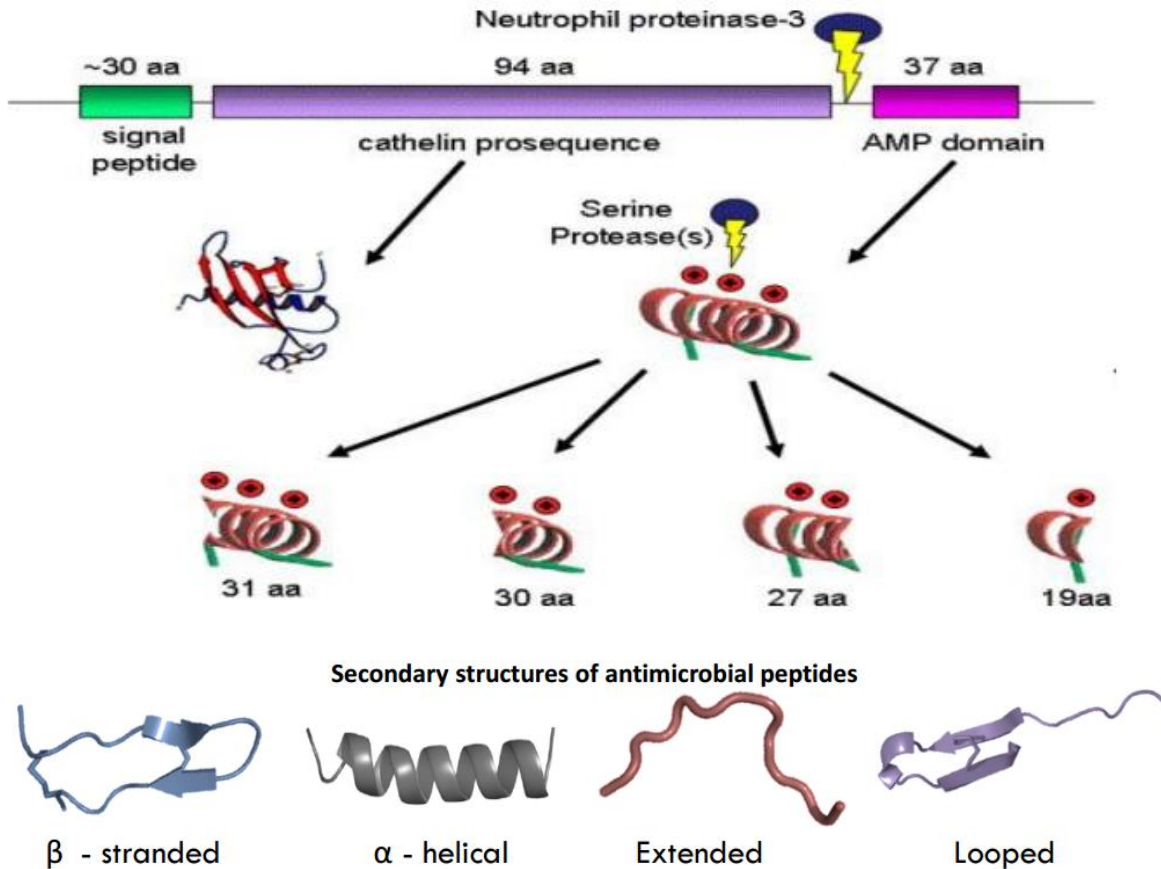


* Muttenthaler, M., King, G.F., Adams, D.J. *et al.* Trends in peptide drug discovery. *Nat Rev Drug Discov* 20, 309–325 (2021). <https://doi.org/10.1038/s41573-020-00135-8>

The mechanisms of AMPs



The characteristics of peptide



- Ubiquitous in nature,
- Described in bacteria, fungi, plants and all vertebrates,
- Known more than 1000 representatives,
- Cationic and anionic peptides.
- Less than 40 amino acids
- High content of hydrophobic domains residues.

The methods of peptide screening

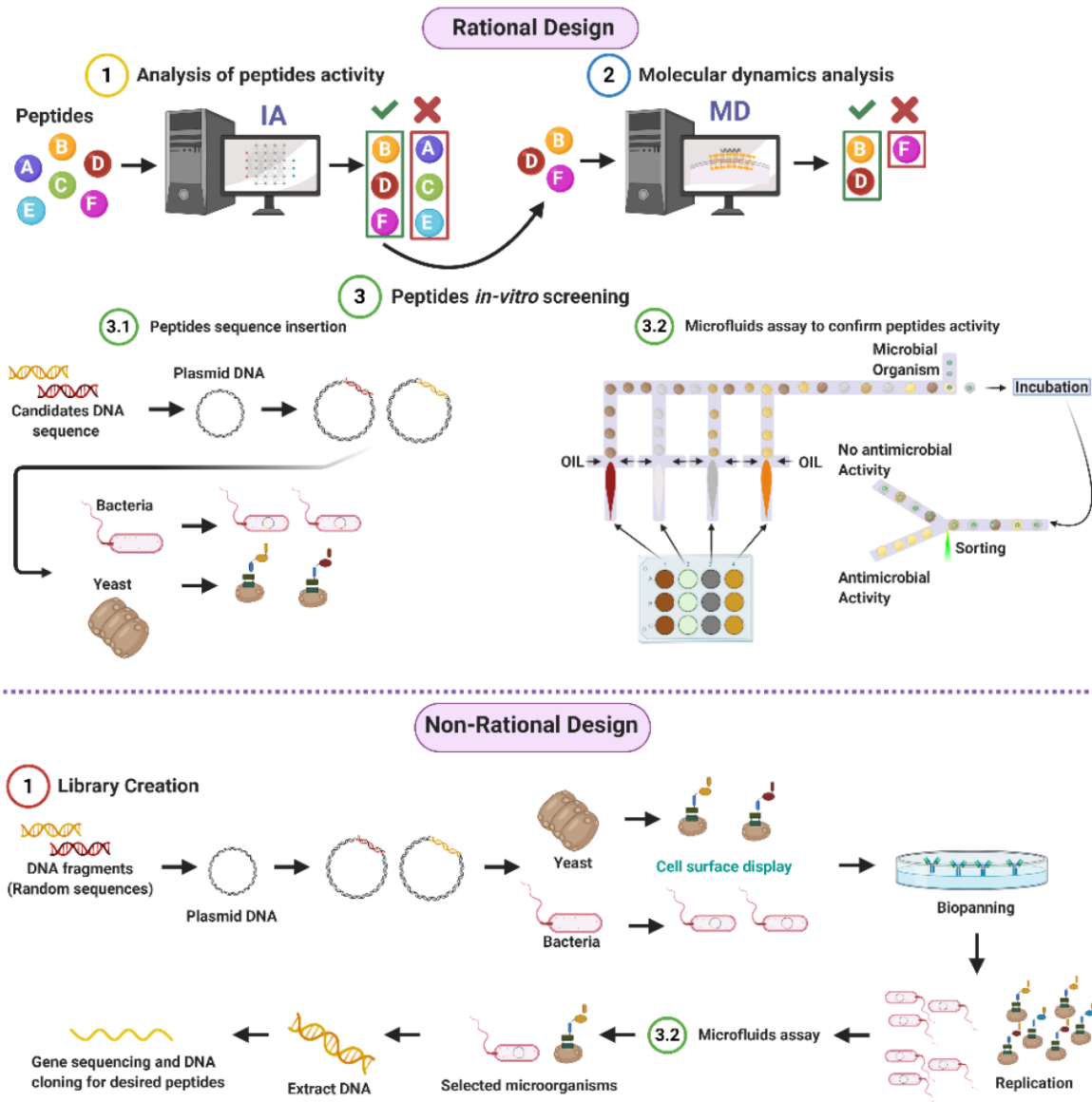
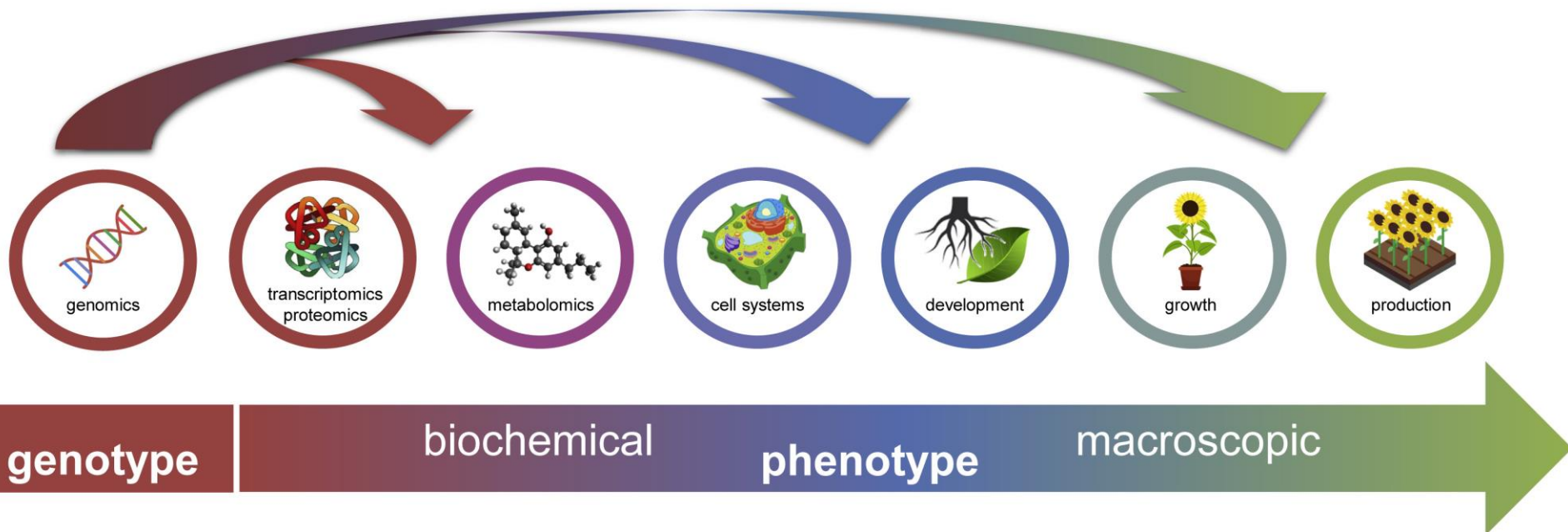


Figure 1

Antimicrobial Peptides (AMPs) discovery framework. Rational design steps: (I) Deep learning techniques identify sequences with potential antimicrobial activity, (II) membrane-disruption capabilities of selected sequences are analyzed via molecular dynamics (MD), (III) the host cell is modified, and sequences are inserted, finally (IV) antimicrobial activity is corroborated by a microfluidic system. Non-rational design steps: (I) Random sequences are expressed on host cells through cell surface display, (II) modified microorganisms are analyzed by a microfluidics system to obtain AMPs candidates, and (III) DNA is extracted, sequenced, and cloned (Created with BioRender).

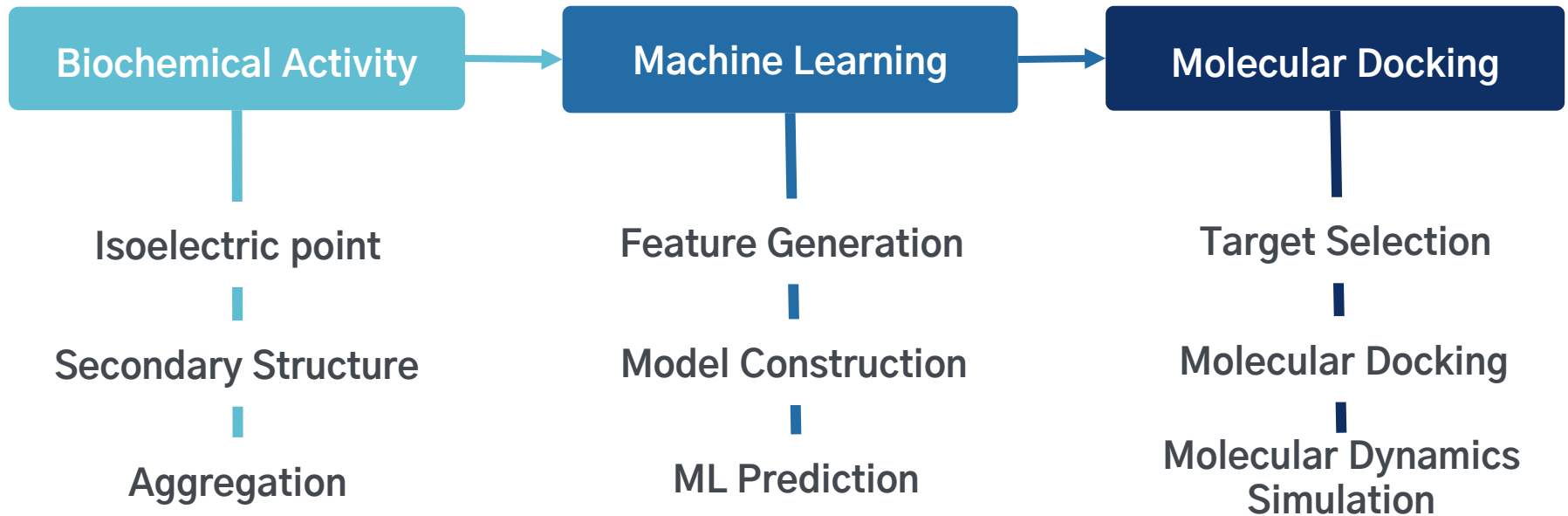
Discovery of latent paths



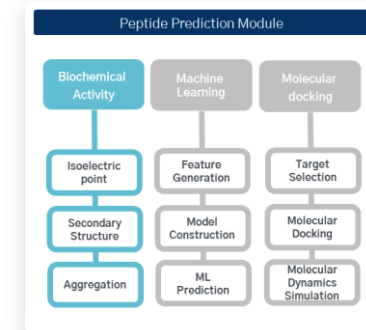
Dijk et al., 2021

iPeptide: Pipeline of Prediction Peptide Functions by AI and Bigdata

Peptide Prediction Module



Biochemical Activity Module



PEPSTATS

pI 8-12

PEPSTATS

≤80mer

TANGO

0 ≤ HELIX ≤ 25

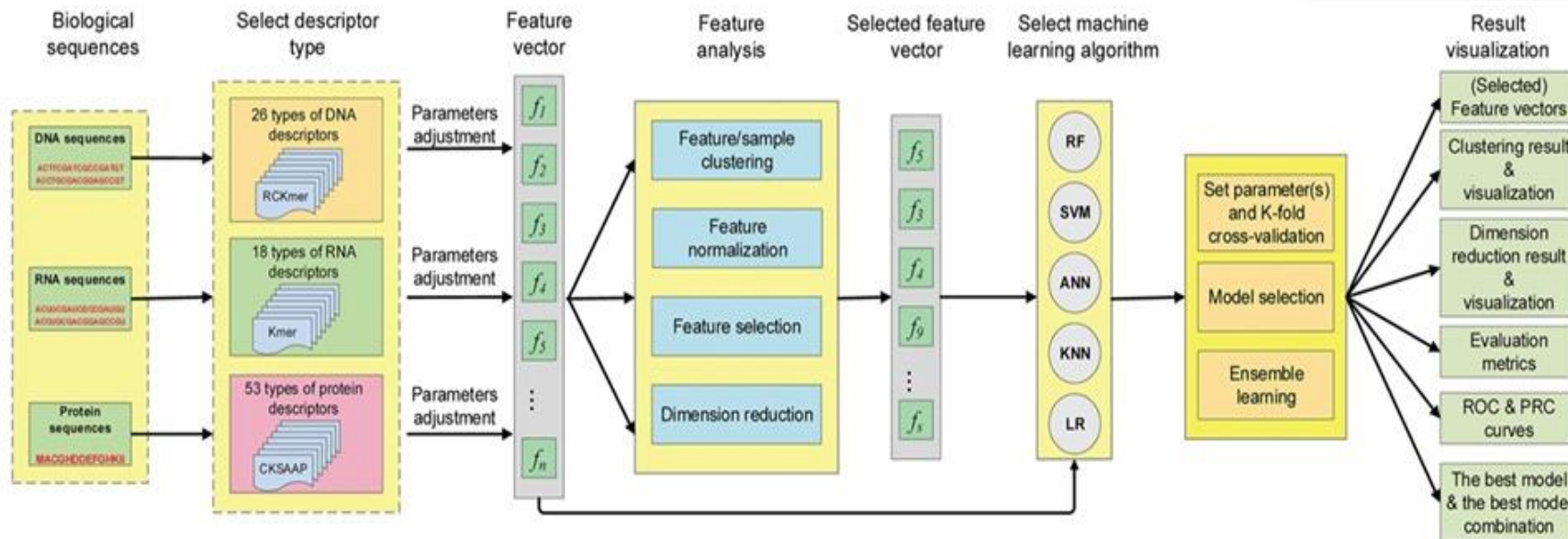
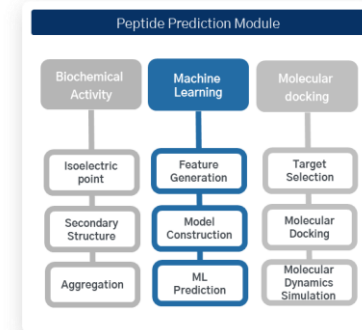
TANGO

25 ≤ BETA ≤ 100

| Peptide ID | Protein ID | Protein Length | Peptide Length | Charge | pI | AMPA | AGG | HELIX | BETA | Na4vSS |
|--------------|----------------------|----------------|----------------|--------|---------|------|----------|----------|---------|--------|
| lc HAC006522 | c160080_g2_i1_orf_1 | 2820 | 14 | 6 | 11.3903 | 1 | 0 | 0 | 73.6688 | -11.8 |
| lc HAC025999 | c250256_g1_i1_orf_1 | 101 | 21 | 4.5 | 10.8574 | 1 | 0 | 0 | 25.4228 | -26.9 |
| lc HAC014527 | c166686_g2_i2_orf_1 | 111 | 30 | 10 | 10.5844 | 1 | 0 | 13.3799 | 61.6417 | -24.1 |
| lc HAC001193 | c125962_g1_i2_orf_1 | 138 | 42 | 8.5 | 11.0963 | 1 | 1.5712 | 0.732012 | 99.3862 | -17.3 |
| lc HAC001951 | c135784_g2_i1_orf_1 | 187 | 22 | 6.5 | 10.6722 | 1 | 0 | 0 | 55.357 | -35.8 |
| lc HAC011964 | c164819_g3_i1_orf_1 | 307 | 37 | 5.5 | 9.8963 | 1 | 0 | 10.8876 | 92.6157 | -19 |
| lc HAC025533 | c233785_g1_i1_orf_1 | 121 | 16 | 5 | 11.2543 | 1 | 0 | 0 | 26.3324 | -19.7 |
| lc HAC014476 | c166654_g3_i10_orf_1 | 486 | 13 | 4.5 | 11.1319 | 1 | 0 | 0 | 50.2313 | -10 |
| lc HAC004349 | c154779_g1_i3_orf_1 | 248 | 46 | 10.5 | 11.1669 | 1 | 0 | 8.23356 | 81.6385 | -18 |
| lc HAC011635 | c164585_g7_i2_orf_1 | 320 | 24 | 6 | 10.9599 | 1 | 6.81118 | 5.86831 | 70.5187 | -4.5 |
| lc HAC011671 | c164614_g2_i10_orf_1 | 532 | 28 | 3.5 | 9.9102 | 1 | 328.1 | 0 | 98.8156 | 18.3 |
| lc HAC010521 | c163722_g3_i2_orf_1 | 143 | 26 | 5 | 9.1286 | 1 | 0 | 4.05635 | 68.4868 | -7.5 |
| lc HAC013679 | c166093_g2_i1_orf_1 | 119 | 15 | 4 | 11.2782 | 1 | 0 | 0 | 30.1323 | -28.7 |
| lc HAC014464 | c166649_g11_i2_orf_1 | 190 | 40 | 9 | 11.4361 | 1 | 0 | 1.54979 | 69.0337 | -24.2 |
| lc HAC001589 | c131275_g1_i1_orf_1 | 267 | 17 | 5 | 11.6003 | 1 | 0 | 0 | 33.2872 | -29.5 |
| lc HAC026432 | c4313_g1_i1_orf_1 | 124 | 12 | 4 | 11.432 | 1 | 0 | 0 | 36.6491 | -36.5 |
| lc HAC027085 | c74463_g1_i2_orf_1 | 185 | 23 | 8 | 11.9329 | 1 | 24.7006 | 2.71787 | 79.4181 | -0.5 |
| lc HAC024432 | c198501_g1_i1_orf_1 | 107 | 16 | 5.5 | 9.578 | 1 | 0 | 0 | 59.4511 | -27.1 |
| lc HAC007954 | c161578_g8_i6_orf_1 | 241 | 12 | 3 | 11.8231 | 1 | 0 | 0 | 41.7054 | 30.1 |
| lc HAC005302 | c157987_g1_i1_orf_1 | 112 | 17 | 6 | 11.7349 | 1 | 2.43328 | 0 | 76.7368 | 12 |
| lc HAC013532 | c165993_g1_i1_orf_1 | 498 | 45 | 4.5 | 9.8132 | 1 | 0.985743 | 1.51299 | 69.6379 | -3.1 |
| lc HAC011144 | c164200_g2_i1_orf_1 | 752 | 29 | 5 | 10.506 | 1 | 0 | 1.26229 | 76.6215 | 11.3 |
| lc HAC023305 | c173566_g2_i13_orf_1 | 622 | 29 | 8 | 11.2625 | 1 | 0 | 0.900093 | 76.0684 | -37.8 |
| lc HAC007474 | c161062_g9_i1_orf_1 | 108 | 24 | 6 | 11.0304 | 1 | 4.02987 | 15.0029 | 31.1093 | 5.1 |
| lc HAC026025 | c251589_g1_i1_orf_1 | 103 | 18 | 4.5 | 11.1319 | 1 | 0 | 1.41064 | 53.2283 | -13.1 |
| lc HAC018594 | c169543_g3_i1_orf_1 | 133 | 45 | 9 | 11.6297 | 1 | 0 | 2.65262 | 90.7872 | -16.1 |
| lc HAC012562 | c165268_g4_i4_orf_1 | 154 | 36 | 13 | 11.6155 | 1 | 0 | 5.51249 | 63.216 | -35.8 |
| lc HAC004574 | c155627_g3_i2_orf_1 | 170 | 14 | 3 | 11.8231 | 1 | 422.985 | 0 | 80.8755 | 25.3 |
| lc HAC013886 | c166238_g7_i3_orf_1 | 578 | 22 | 2.5 | 10.5273 | 1 | 1.17633 | 0 | 82.4955 | 8.5 |
| lc HAC002628 | c143437_g2_i1_orf_1 | 299 | 30 | 7 | 11.8225 | 1 | 0 | 18.3266 | 26.5701 | -39.7 |
| lc HAC025474 | c231775_g1_i1_orf_1 | 112 | 23 | 5 | 9.6583 | 1 | 0 | 2.10845 | 87.2242 | -25.9 |

| Propensity | Methods | Descriptions/Parameters | Cutoff/Filters | # of Sequences |
|------------------------|--------------|--|-------------------|----------------|
| | Raw Sequence | Total Given Proteins Sequences | | 28,103 |
| | Raw Sequence | Small Proteins/Peptides | ≤ 100 | 447 |
| | EPESTFIND | Proteins contains potential protease cleavage site | l= 0 | 5,418 |
| | AMPA | Antimicrobial Spots | l= 0 | 1,434 |
| | Pepstats | Total Peptides | ≤ 100 | 1,434 |
| Molecular | Pepstats | Peptide Length | ≤ 50 | 432 |
| | Pepstats | Small Proteins which have AMP peptides | ≤ 50 | 0 |
| | Pepstats | Charge | > 0 | 1,298 |
| | Pepstats | Isoelectric Point (pI) | 8 ≤ pI ≤ 12 | 1,050 |
| | Tango | AGG | ≤ 500 | 1,136 |
| Aggregation (In-vivo) | Tango | Helix | 0 ≤ Helix ≤ 25 | 741 |
| | Tango | Beta | 25 ≤ Beta ≤ 100 | 353 |
| Aggregation (In-vitro) | Aggrescan | Na4vSS | -40 ≤ Na4vSS ≤ 60 | 1,152 |
| | Allerdicator | Predictions | Non-Allergen | 1,430 |
| Homologous (CAMP) | Blast | Novel | No Blast Hits | 1,397 |
| | Blast | Known | Blast Hits | 37 |
| total | | | | 74 |

Machine Learning Module



| Feature | Accuracy | AUC | Recall | Prec. | F1 | Kappa | MCC | Model |
|-----------------|----------|------|--------|-------|------|-------|------|---------------------------------|
| AAC | 0.87 | 0.94 | 0.83 | 0.9 | 0.86 | 0.73 | 0.73 | Extra Trees Classifier |
| AAindex | 0.86 | 0.94 | 0.83 | 0.88 | 0.85 | 0.71 | 0.72 | Light Gradient Boosting Machine |
| Kgap | 0.84 | 0.92 | 0.81 | 0.87 | 0.84 | 0.69 | 0.69 | CatBoost Classifier |
| TPC | 0.83 | 0.91 | 0.84 | 0.83 | 0.83 | 0.66 | 0.66 | Extra Trees Classifier |
| DPC | 0.83 | 0.92 | 0.75 | 0.89 | 0.81 | 0.65 | 0.66 | Extra Trees Classifier |
| MappingProtein | 0.8 | 0.89 | 0.76 | 0.83 | 0.8 | 0.61 | 0.61 | CatBoost Classifier |
| MappingClass | 0.8 | 0.85 | 0.76 | 0.83 | 0.79 | 0.6 | 0.6 | CatBoost Classifier |
| FourierClasses4 | 0.72 | 0.72 | 0.66 | 0.75 | 0.7 | 0.43 | 0.44 | Ridge Classifier |
| FourierClasses5 | 0.71 | 0.78 | 0.64 | 0.76 | 0.69 | 0.43 | 0.44 | Gradient Boosting Classifier |
| ComplexNetwork | 0.69 | 0.69 | 0.62 | 0.73 | 0.67 | 0.39 | 0.39 | Ridge Classifier |
| Tsallis | 0.65 | 0.71 | 0.54 | 0.7 | 0.61 | 0.3 | 0.31 | Gradient Boosting Classifier |
| TsallisEntrop | 0.62 | 0.65 | 0.47 | 0.68 | 0.56 | 0.24 | 0.25 | Extreme Gradient Boosting |

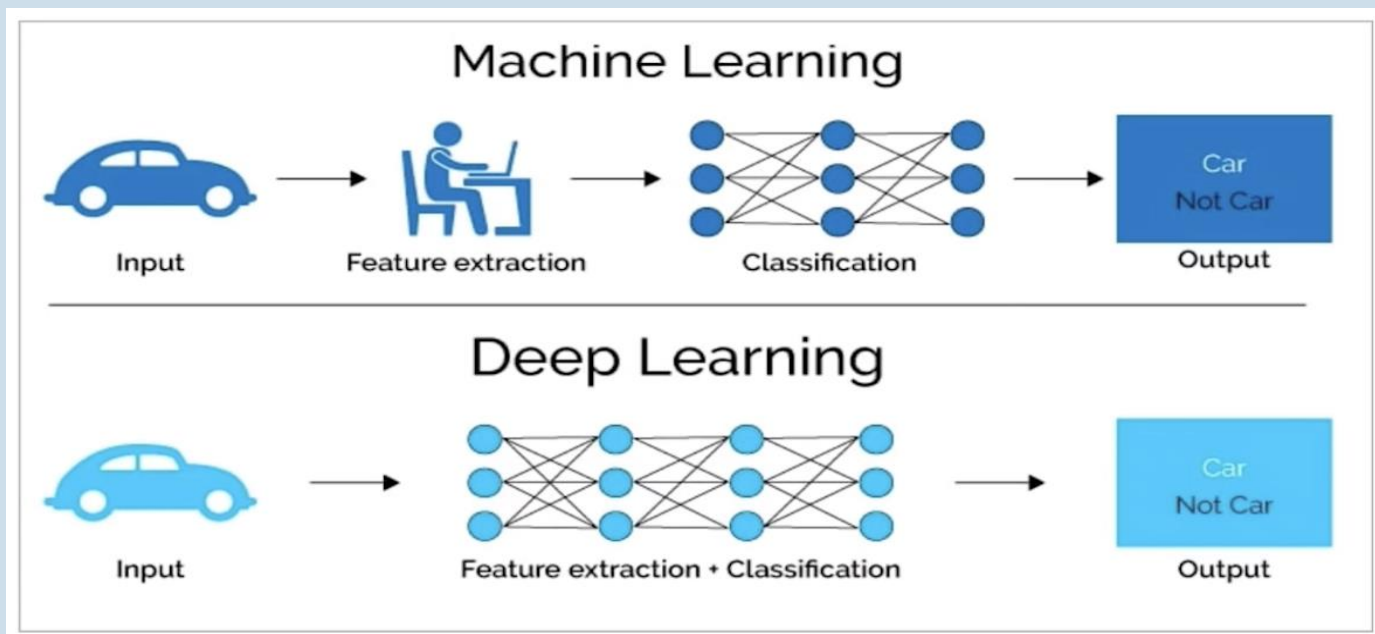
- 13 features
- 15 algorithms

Brief Bioinform, Volume 21, Issue 3, May 2020, Pages 1047–1057

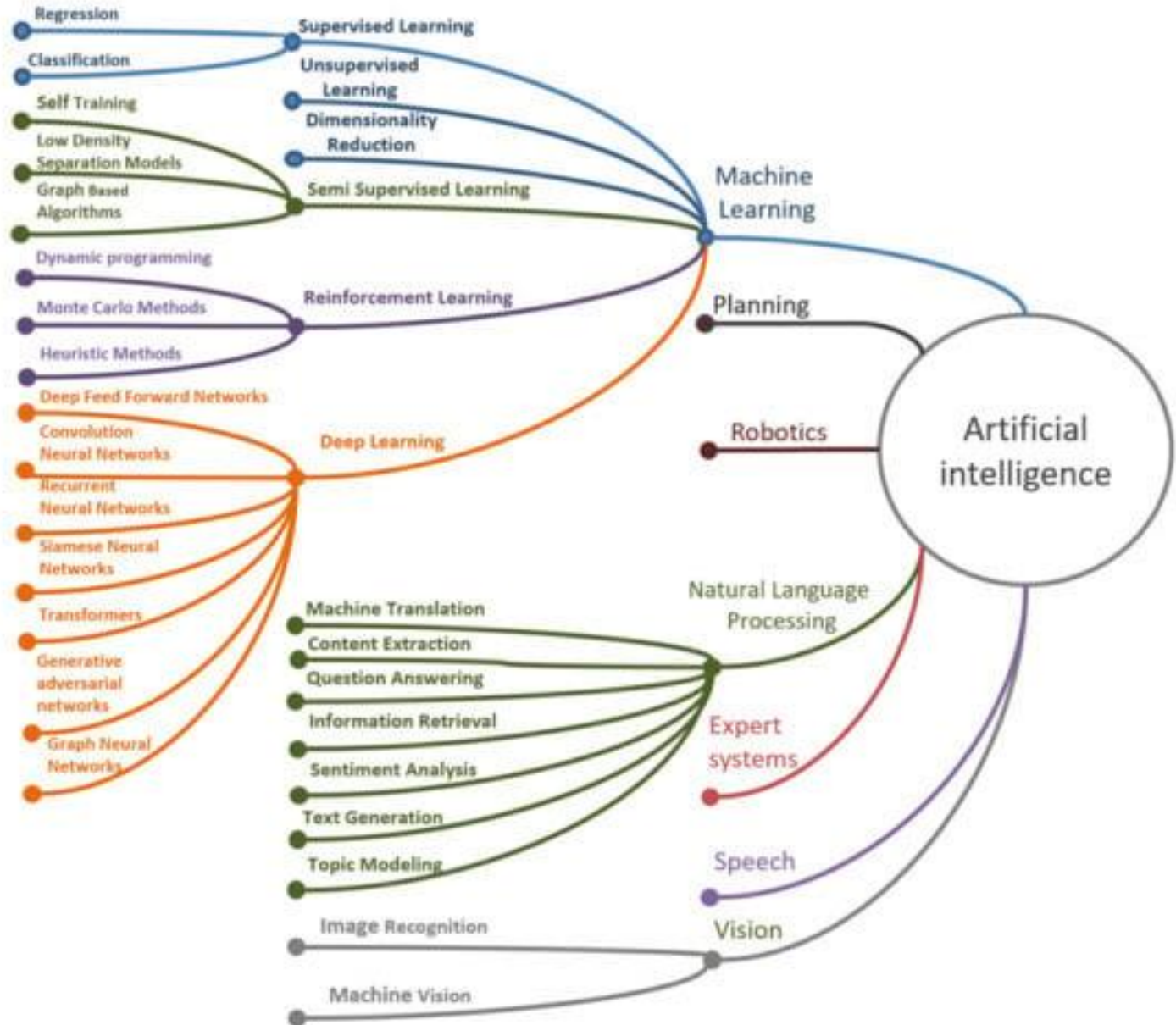
Beyond statistics

Statistics (linear, $y = ax + b$) vs ML (non-linear, $y = f(x)$)

$$y = f(x)$$

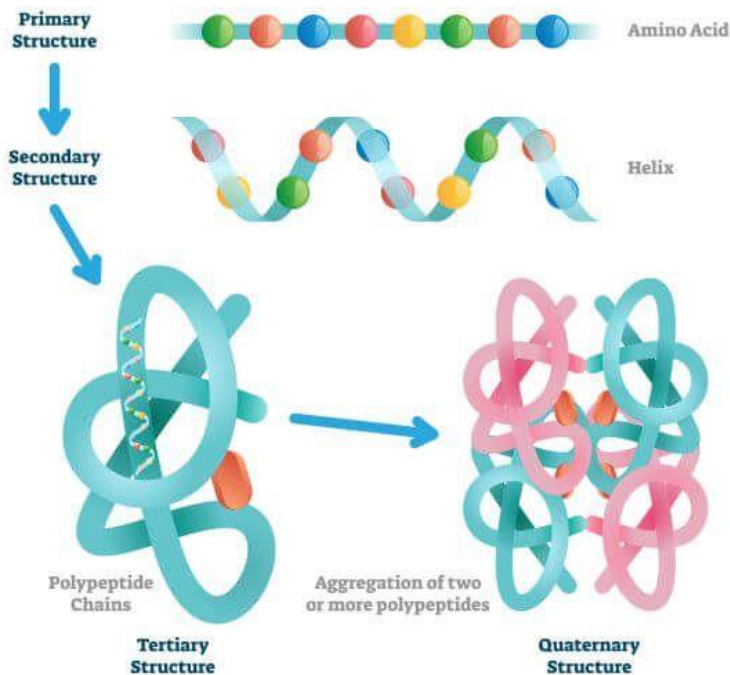


Algorithms



Structural and physicochemical features used in ML Module

PROTEIN STRUCTURE



Descriptor groups

Amino Acid Composition (AAC)

Enhanced Amino Acid Composition (EAAC)

Composition of k-spaced Amino Acid Pairs (CKSAAP)

Tri-Peptide Composition (TPC)

Grouped Amino Acid Composition (GAAC)

Enhanced GAAC (EGAAC)

Composition of k-Spaced Amino Acid Group Pairs (CKSAAGP)

Grouped Tri-Peptide Composition (GTPC)

Binary (binary)

Moran correlation (Moran)

Geary correlation (Geary)

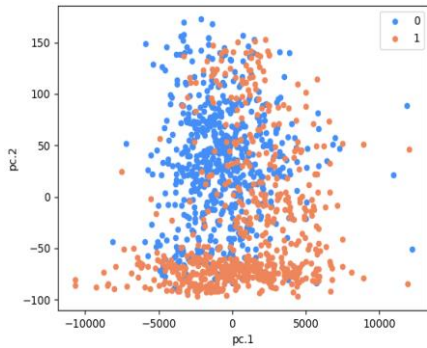
Normalized Moreau-Broto Autocorrelation (NMBroto)

Composition/Transition/Distribution (CTD)

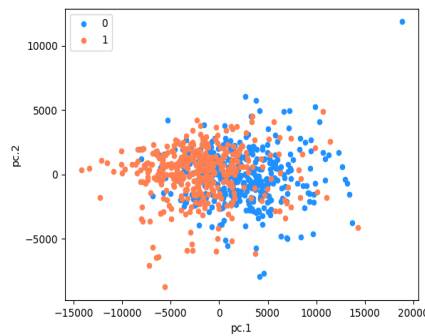
(200 features)

Why we need AI?

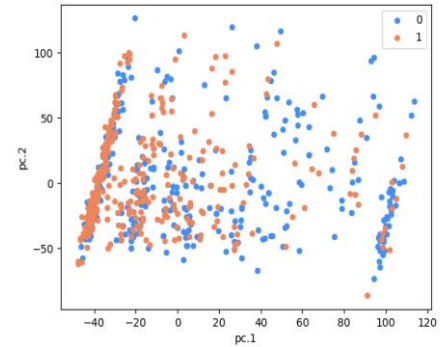
Each peptide with 200 features (1: Positive, 0: Negative)



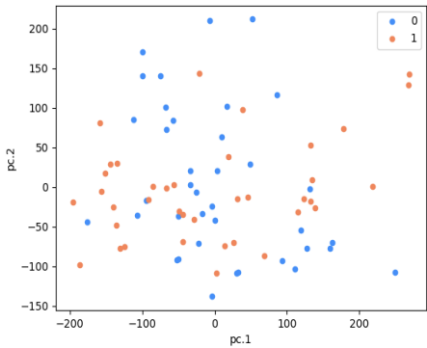
Hemolytic Peptides



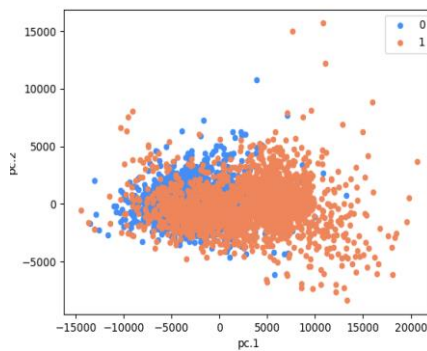
Cell-Penetrating Peptides



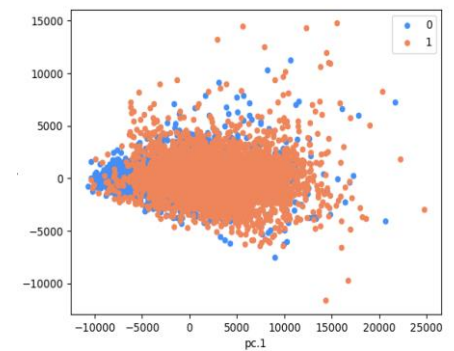
Anti-HIV Peptides



Bitter Taste Peptides



Toxin Peptides



Anti-Microbial Peptides

Algorithm Selection for Optimization of AI-Based Peptide Prediction

AMP: FourierClasses4 & CatBoost Classifier



AFP: FourierClasses4 & Extra Trees Classifier



AIP: Tballis & Gradient Boosting Classifier

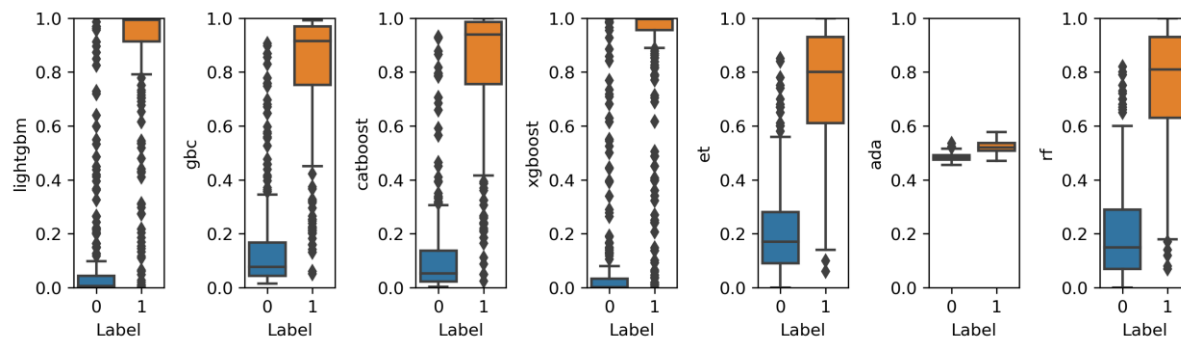
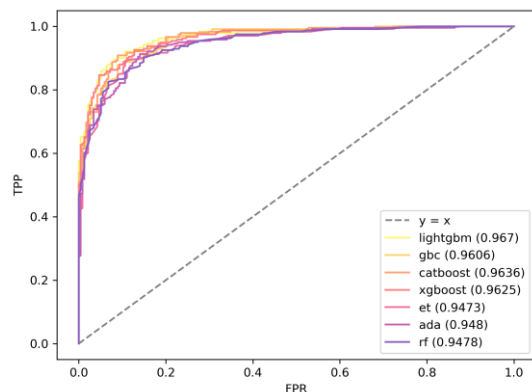


- 15 models : Logistic Regression, Ridge Classifier, Linear Discriminant Analysis, Random Forest Classifier, Naïve Bayes, CatBoost Classifier, Gradient Boosting Classifier, Ada Boost Classifier, Extra Trees Classifier, Quadratic Discriminant Analysis, Light Gradient Boosting Machine, K Neighbors Classifier, Decision Tree Classifier, Extreme Gradient Boosting, SVM-Linear Kern

Anti-Bacterial Peptide

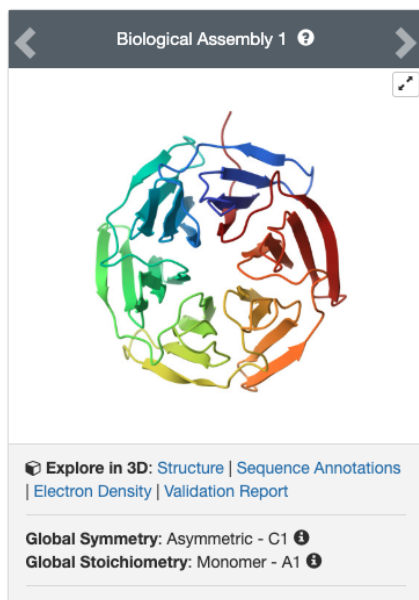
- Accuracy

- Version1.0 : 84.5% → Version2.0 : **90.6%**



| Model | Accuracy | AUC | Recall | Prec. | F1 | Kappa | MCC |
|----------|---------------------------------|-------|--------|-------|-------|-------|-------|
| lightgbm | Light Gradient Boosting Machine | 0.906 | 0.967 | 0.880 | 0.930 | 0.904 | 0.812 |
| gbc | Gradient Boosting Classifier | 0.891 | 0.961 | 0.871 | 0.909 | 0.890 | 0.782 |
| catboost | CatBoost Classifier | 0.887 | 0.964 | 0.842 | 0.927 | 0.883 | 0.777 |
| xgboost | Extreme Gradient Boosting | 0.887 | 0.963 | 0.867 | 0.905 | 0.886 | 0.774 |
| et | Extra Trees Classifier | 0.878 | 0.947 | 0.847 | 0.907 | 0.876 | 0.758 |
| ada | Ada Boost Classifier | 0.872 | 0.948 | 0.855 | 0.888 | 0.871 | 0.744 |
| rf | Random Forest Classifier | 0.866 | 0.948 | 0.838 | 0.890 | 0.863 | 0.733 |

Molecular docking Module



7K2L

Kelch domain of human KEAP1 bound to Nrf2 cyclic peptide, c[BAL-NPETGE]

PDB DOI: <https://doi.org/10.2210/pdb7K2L/pdb>

Classification: **PROTEIN BINDING**

Organism(s): Homo sapiens, synthetic construct

Expression System: Escherichia coli

Mutation(s): No ⓘ

Deposited: 2020-09-08 Released: 2021-04-07

Deposition Author(s): Muellers, S.N., Allen, K.N.

Experimental Data Snapshot

Method: X-RAY DIFFRACTION

Resolution: 1.98 Å

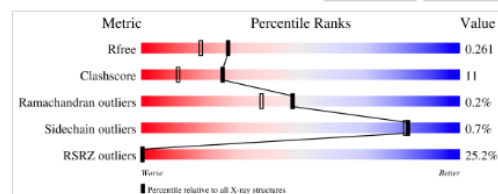
R-Value Free: 0.261

R-Value Work: 0.221

R-Value Observed: 0.222

wwPDB Validation ⓘ

[3D Report](#) [Full Report](#)



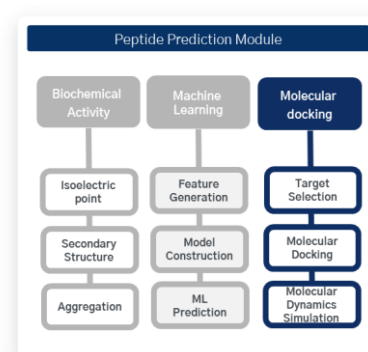
Find similar proteins by: [Sequence](#) | [3D Structure](#) ⓘ

Entity ID: 3

| Molecule | Chains ⓘ | Sequence Length | Organism | Details | Image |
|-----------------------------------|----------------------------|-----------------|-------------------------------------|------------------|-------|
| Nrf2 cyclic peptide,c[BAL-NPETGE] | C [auth P] | 7 | synthetic construct | Mutation(s): 0 ⓘ | |



Overall, cyclizing the linear 7-mer, fixing the ϕ angle at position 78, and stepwise optimization of the linker length and structure together gave a ~200-fold increase in binding affinity, resulting in a K_D for KEAP1 of 20 nM.



Step 1 – Target Selection

ProteinTargets

File Edit View Insert Format Data Tools Extensions Help

100% 123 Default... 10 B I A

A1 S.No

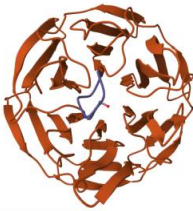
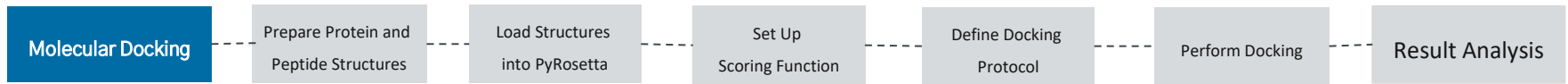
| A | B | C | D | E | F | G | H |
|----|---|---|---|-----------------------|------------------|-------------------------|--|
| 1 | S.No Target Name | Alternative Target | Drug Control | Treatment targ | PDB ID | Reference | Role & Fuction |
| 2 | 1 Tumor Necrosis Factor-alpha (TNF-α) | TNFR1 (TNFR1 is generally more implicated in ir IL-6R) | Infliximab, adalimumab | Downregulate | 7JRA, 1TNR (TNF | Article | A key cytokine in inflammation, TNF-α initiates inflammatory r |
| 3 | 2 Interleukin-6 (IL-6) | | Tocilizumab | Downregulate | 1ALU, 1P9M (IL-6 | Article | Involved in inflammation and infection responses and plays a n IL-1β is a potent mediator of the inflammatory response and car |
| 4 | 3 Interleukin-1 Beta (IL-1β) | | Anakinra | Downregulate | | | A transcription factor that regulates genes involved in inflamma |
| 5 | 4 Nuclear Factor-kappa B (NF-κB) | | Disulfiram, olmesartan and dithiocarbam | Downregulate | | | An enzyme that synthesizes prostaglandins, leading to inflamma |
| 6 | 5 Cyclooxygenase-2 (COX-2) | | Celecoxib | Downregulate | | | Responsible for releasing arachidonic acid, a precursor of pro-inf |
| 7 | 6 Phospholipase A2 (PLA2) | | Varespladib | Downregulate | | | Kinases that play roles in cell proliferation, differentiation, and r |
| 8 | 7 Mitogen-Activated Protein Kinases (MAPKs) | | Trametinib (Mekinist) | Downregulate | | | Enzymes that transmit signals from cytokine receptors, influenci |
| 9 | 8 Janus Kinases (JAKs) | | Tofacitinib | Downregulate | | | Involved in the recruitment of leukocytes to sites of inflamma |
| 10 | 9 Adhesion Molecules (e.g., ICAM-1, VCAM-1) | | CD80 and CD86 (Belatacept (Nulojix)) | Downregulate | | | A pro-inflammatory cytokine produced by Th17 cells, important |
| 11 | 10 Interleukin-17 (IL-17) | | Secukinumab | Downregulate | | | |

Anti – inflammation

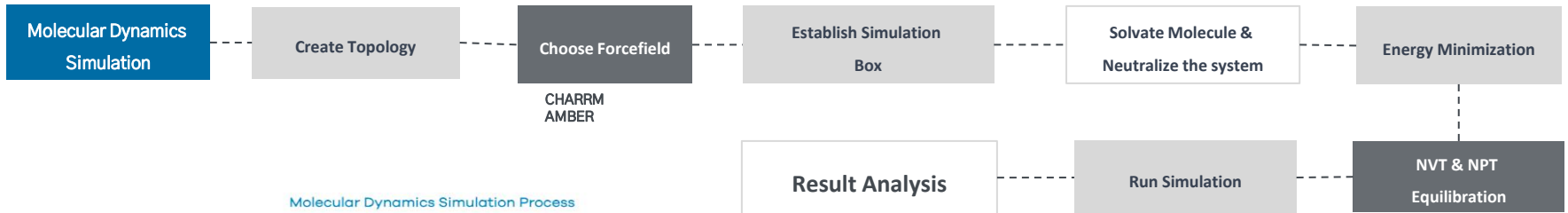
| A | B | C | D | E | F | G | H |
|----|---|--|-------------------------|-------------------------|---------------|-------------------------|---|
| 1 | S.No Target Name | Drug Control | Peptide Control | Treatment target | PDB ID | Reference | Role & Fuction |
| 2 | 1 Ribosomal Subunits (30S and 50S) | Tetracyclines (30S) & macrolides (50S) | Peptide | Inhibition | | Article | Essential for protein synthesis |
| 3 | 2 DNA Gyrase | Fluoroquinolones | Peptide | Inhibition | | Article | DNA replication |
| 4 | 3 Topoisomerase IV | Fluoroquinolones | Peptide | Inhibition | | | DNA replication |
| 5 | 4 Penicillin-Binding Proteins (PBPs) | penicillins | Peptide | Inhibition | | | cell wall synthesis machinery |
| 6 | 5 Dihydrofolate Reductase (DHFR) | Trimethoprim, sulfonamides | Peptide | Inhibition | | | part of the folic acid synthesis pathway |
| 7 | 6 RNA Polymerase | Rifamycin | Peptide | Inhibition | | | RNA synthesis |
| 8 | 7 Beta-Lactamases | clavulanic acid, sulbactam, tazobactam | Peptide | Inhibition | | | defense mechanism against beta-lactam antibiotics - antibiotic resistance |
| 9 | 8 Efflux Pumps | NA | Peptide | Inhibition | | | resistance against a wide array of antibiotics |
| 10 | | | | | | | |
| 11 | Example | | | | | | |
| 12 | 1 Urinary Tract Infections (UTIs) - Escherichia coli (most common), Staphylococcus saprophyticus | | | | | | |
| 13 | Penicillin-binding proteins | Trimethoprim/sulfamethoxazole, Fluoroquinolones, Nitrofurant | | Inhibition | | | Cell wall synthesis enzymes |
| 14 | DNA gyrase | Trimethoprim/sulfamethoxazole, Fluoroquinolones, Nitrofurant | | Inhibition | | | folate biosynthesis |
| 15 | | | | | | | |
| 16 | 2 Respiratory Tract Infections | | | | | | |
| 17 | Streptococcal Pharyngitis - Streptococcus pyogenes | | | Inhibition | | | |
| 18 | Penicillin-binding proteins | Penicillin or amoxicillin | | Inhibition | | | Cell wall synthesis enzymes |
| 19 | | | | | | | |
| 20 | 3 Pneumonia | | | | | | |
| 21 | Streptococcus pneumoniae, Haemophilus influenzae | | | Inhibition | | | Cell wall synthesis enzymes |
| 22 | Ribosomal subunits | Macrolides, penicillins, cephalosporins | | Inhibition | | | Essential for protein synthesis |

Anti – Bacterial

Step 2 – Molecular docking/ Step 3 – Molecular Dynamics Simulation

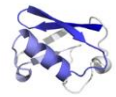


Overall, cyclizing the linear 7-mer, fixing the ϕ angle at position 78, and stepwise optimization of the linker length and structure together gave a **~200-fold increase in binding affinity**, resulting in a KD for KEAP1 of 20 nM.

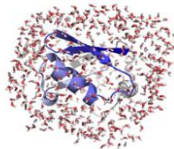


CHARRM
AMBER

Molecular Dynamics Simulation Process



Structure Check
Protonation State
Solvation
Adding Salts



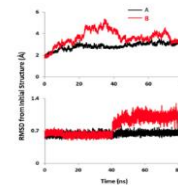
2. Preparation of the Simulation System

Select Simulation Protocol



3. Simulation Run

Post-processing



4. Trajectory Analysis

The results of Peptide Prediction

| Peptide ID | Protein ID | Protein Length | Peptide Length | Charge | pI | AMPA | AGG | HELIX | BETA | Na4vSS | SVM | RF | ANN | DA | Consensus | Toxin |
|---------------|----------------------|----------------|----------------|--------|---------|------|----------|----------|---------|--------|------|------|------|------|-----------|-----------|
| lcl HAC006522 | c160080_g2_i1_orf_1 | 2820 | 14 | 6 | 11.3903 | 1 | 0 | 0 | 73.6688 | -11.8 | AMP | AMP | AMP | AMP | 4 | Non-Toxin |
| lcl HAC025999 | c250256_g1_i1_orf_1 | 101 | 21 | 4.5 | 10.8574 | 1 | 0 | 0 | 25.4228 | -26.9 | NAMP | AMP | AMP | AMP | 3 | Non-Toxin |
| lcl HAC014527 | c166686_g2_i2_orf_1 | 111 | 30 | 10 | 10.5844 | 1 | 0 | 13.3799 | 61.6417 | -24.1 | AMP | AMP | AMP | AMP | 4 | Toxin |
| lcl HAC001193 | c125962_g1_i2_orf_1 | 138 | 42 | 8.5 | 11.0963 | 1 | 1.5712 | 0.732012 | 99.3862 | -17.3 | AMP | AMP | AMP | AMP | 4 | Toxin |
| lcl HAC001951 | c135784_g2_i1_orf_1 | 187 | 22 | 6.5 | 10.6722 | 1 | 0 | 0 | 55.357 | -35.8 | NAMP | AMP | AMP | AMP | 3 | Toxin |
| lcl HAC011964 | c164819_g3_i1_orf_1 | 307 | 37 | 5.5 | 9.8963 | 1 | 0 | 10.8876 | 92.6157 | -19 | AMP | AMP | AMP | AMP | 4 | Toxin |
| lcl HAC025533 | c233785_g1_i1_orf_1 | 121 | 16 | 5 | 11.2543 | 1 | 0 | 0 | 26.3324 | -19.7 | AMP | AMP | AMP | NAMP | 3 | Non-Toxin |
| lcl HAC014476 | c166654_g3_i10_orf_1 | 486 | 13 | 4.5 | 11.1319 | 1 | 0 | 0 | 50.2313 | -10 | NAMP | AMP | AMP | AMP | 3 | Non-Toxin |
| lcl HAC004349 | c154779_g1_i3_orf_1 | 248 | 46 | 10.5 | 11.1669 | 1 | 0 | 8.23356 | 81.6385 | -18 | AMP | NAMP | AMP | AMP | 3 | Non-Toxin |
| lcl HAC011635 | c164585_g7_i2_orf_1 | 320 | 24 | 6 | 10.9599 | 1 | 6.81118 | 5.86831 | 70.5187 | -4.5 | AMP | AMP | AMP | AMP | 4 | Toxin |
| lcl HAC011671 | c164614_g2_i10_orf_1 | 532 | 28 | 3.5 | 9.9102 | 1 | 328.1 | 0 | 98.8156 | 18.3 | AMP | AMP | AMP | AMP | 4 | Toxin |
| lcl HAC010521 | c163722_g3_i2_orf_1 | 143 | 26 | 5 | 9.1286 | 1 | 0 | 4.05635 | 68.4868 | -7.5 | AMP | AMP | AMP | AMP | 4 | Toxin |
| lcl HAC013679 | c166093_g2_i1_orf_1 | 119 | 15 | 4 | 11.2782 | 1 | 0 | 0 | 30.1323 | -28.7 | AMP | NAMP | AMP | AMP | 3 | Non-Toxin |
| lcl HAC014464 | c166649_g11_i2_orf_1 | 190 | 40 | 9 | 11.4361 | 1 | 0 | 1.54979 | 69.0337 | -24.2 | AMP | AMP | AMP | AMP | 4 | Toxin |
| lcl HAC001589 | c131275_g1_i1_orf_1 | 267 | 17 | 5 | 11.6003 | 1 | 0 | 0 | 33.2872 | -29.5 | AMP | NAMP | AMP | AMP | 3 | Toxin |
| lcl HAC026432 | c4313_g1_i1_orf_1 | 124 | 12 | 4 | 11.432 | 1 | 0 | 0 | 36.6491 | -36.5 | AMP | AMP | AMP | NAMP | 3 | Toxin |
| lcl HAC027085 | c74463_g1_i2_orf_1 | 185 | 23 | 8 | 11.9329 | 1 | 24.7006 | 2.71787 | 79.4181 | -0.5 | AMP | AMP | AMP | AMP | 4 | Non-Toxin |
| lcl HAC024432 | c198501_g1_i1_orf_1 | 107 | 16 | 5.5 | 9.578 | 1 | 0 | 0 | 59.4511 | -27.1 | NAMP | AMP | AMP | AMP | 3 | Toxin |
| lcl HAC007954 | c161578_g8_i6_orf_1 | 241 | 12 | 3 | 11.8231 | 1 | 0 | 0 | 41.7054 | 30.1 | AMP | AMP | AMP | AMP | 4 | Toxin |
| lcl HAC005302 | c157987_g1_i1_orf_1 | 112 | 17 | 6 | 11.7349 | 1 | 2.43328 | 0 | 76.7368 | 12 | AMP | AMP | AMP | AMP | 4 | Toxin |
| lcl HAC013532 | c165993_g1_i1_orf_1 | 498 | 45 | 4.5 | 9.8132 | 1 | 0.985743 | 1.51299 | 69.6379 | -3.1 | AMP | AMP | AMP | AMP | 4 | Non-Toxin |
| lcl HAC011144 | c164200_g2_i1_orf_1 | 752 | 29 | 5 | 10.506 | 1 | 0 | 1.26229 | 76.6215 | 11.3 | AMP | AMP | AMP | AMP | 4 | Toxin |
| lcl HAC023305 | c173566_g2_i13_orf_1 | 622 | 29 | 8 | 11.2625 | 1 | 0 | 0.900093 | 76.0684 | -37.8 | NAMP | AMP | AMP | AMP | 3 | Non-Toxin |
| lcl HAC007474 | c161062_g9_i1_orf_1 | 108 | 24 | 6 | 11.0304 | 1 | 4.02987 | 15.0029 | 31.1093 | 5.1 | AMP | AMP | AMP | AMP | 4 | Non-Toxin |
| lcl HAC026025 | c251589_g1_i1_orf_1 | 103 | 18 | 4.5 | 11.1319 | 1 | 0 | 1.41064 | 53.2283 | -13.1 | AMP | NAMP | AMP | AMP | 3 | Non-Toxin |
| lcl HAC018594 | c169543_g3_i1_orf_1 | 133 | 45 | 9 | 11.6297 | 1 | 0 | 2.65262 | 90.7872 | -16.1 | AMP | AMP | AMP | AMP | 4 | Non-Toxin |
| lcl HAC012562 | c165268_g4_i4_orf_1 | 154 | 36 | 13 | 11.6155 | 1 | 0 | 5.51249 | 63.216 | -35.8 | AMP | AMP | AMP | AMP | 4 | Toxin |
| lcl HAC004574 | c155627_g3_i2_orf_1 | 170 | 14 | 3 | 11.8231 | 1 | 422.985 | 0 | 80.8755 | 25.3 | AMP | AMP | AMP | AMP | 4 | Non-Toxin |
| lcl HAC013886 | c166238_g7_i3_orf_1 | 578 | 22 | 2.5 | 10.5273 | 1 | 1.17633 | 0 | 82.4955 | 8.5 | AMP | AMP | NAMP | AMP | 3 | Non-Toxin |
| lcl HAC002628 | c143437_g2_i1_orf_1 | 299 | 30 | 7 | 11.8225 | 1 | 0 | 18.3266 | 26.5701 | -39.7 | AMP | AMP | AMP | AMP | 4 | Non-Toxin |
| lcl HAC025474 | c231775_g1_i1_orf_1 | 112 | 23 | 5 | 9.6583 | 1 | 0 | 2.10845 | 87.2242 | -25.9 | AMP | AMP | AMP | AMP | 4 | Toxin |

Case Study

De Novo Genome project and AMP Analysis of the King Centipede

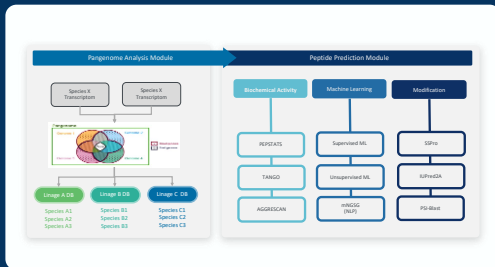
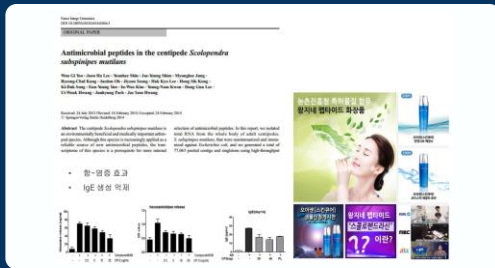
- 내용** Identification of Peptides Effective for Human Autoimmune Diseases from Toxicity-Related Genes of the King Centipede (*Scolopendra subspinipes mutilans*), which has been used as a herbal medicine for various diseases including arthritis
- 연구기간** 2014~2017
- 논문화** Yoo et al., Functional & integrative genomics 2014 [IF: 3.54]; Park et al., Scientific Reports, 2018 [IF:4.996]
- 산업화** Development of Cosmetics for Atopic Dermatitis Improvement Using Scolopendrasin I Derived from the King Centipede

Comparative Transcriptome Analysis and AMP Analysis of Leeches

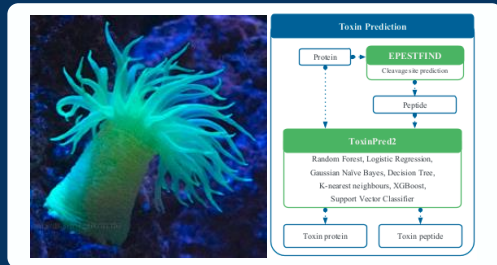
- 내용** The comparative transcriptome analysis with *Hirudo nipponia* and *Haemadipsa rjukjuana* and screening of anti-inflammatory peptide
- 연구기간** 2022~2023
- 논문화** On going
- 산업화** On going

Acquisition of Omics Big Data and Peptide Materials Derived from Island Wildlife

- 내용** The goal is to collect over 35,000 omics big data from more than 80 species of island wildlife and analyze peptide utility information using AI to secure basic utility information such as antiviral, antifungal, anticancer, antibacterial, and immune-enhancing properties.
- 연구기간** 2023~2026
- 논문화** On going
- 산업화** On going

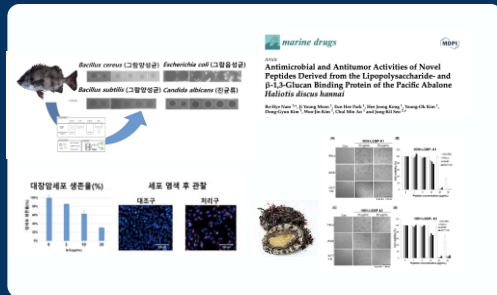


Case Study



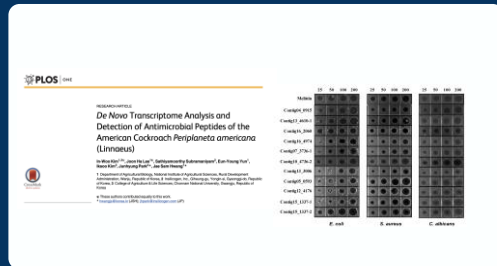
Identification of Toxic and Useful Genes in Five Species of Island-Derived Wildlife

- 내용** 말미잘 (*Halcurias carlgreni*) 등 섬 유래 야생생물 5종에서 기계학습을 이용한 독성 단백질을 예측
 의학적, 산업적으로 활용 가능한 신규 유용 유전자 발굴 목적
- 연구기간** 2021
- 논문화** On going
- 산업화** On going



Exploration of Useful Peptides Derived from Aquatic Organisms

- 내용** Discovery of Useful Peptides from Aquatic Organisms through Machine Learning Predictions
 [National Institute of Fisheries Science / 15 Species of Aquatic Organisms]
 Antimicrobial and Antitumor Functional Substances
- 연구기간** 2014~2021
- 논문화** Scientific reports, etc.
- 산업화** Anti-cancer materials



Exploration of Useful Genes from Cockroach Transcriptome

- 내용** Exploration of Useful Peptides from the Transcriptome of Cockroach (*Periplaneta americana*)
 Aimed at Discovering New Useful Genes with Medical and Industrial Applications in More Than Four Insect Species
- 연구기간** 2016
- 논문화** Kim et al., PLOSone 2017 [IF: 3.24]
- 산업화** Auto-immune disease materials

Implications

AI as an Interpreters: marine originated peptides to human health

The Role of AI—Bridging Marine Peptides and Human Healthcare

| | | |
|---|---|---|
| Interpreter of Marine Peptides | Complexity of Marine Peptides | Peptides derived from marine organisms possess unique structures and functions. |
| | AI as a Mediator | AI analyzes the complex “language” of marine peptides, translating it into applications for human health. |
| | Data Integration and Analysis | Rapid identification of promising peptides through multi-omics data and machine learning. |
| Accelerating Applications of Marine Peptides | Faster Drug Development: | Utilizing AI to discover marine peptide-based drug candidates. |
| | Personalized Therapies: | Potential to develop peptide treatments tailored to individual disease characteristics. |
| | Disease Prevention and Management: | Establishing new prevention strategies using antimicrobial and antiviral peptides. |

Implications of Digitalized *Marine* Peptides

A Future Unveiled by the Fusion of Marine Life and AI

- **A New Treasure Trove of Resources:** Unlocking infinite resources through the combination of marine biology and AI.
- **Enhancing Human Health:** Applying marine peptides in human healthcare to improve quality of life.
- **Collaboration and Advancement:** Emphasizing the need for ongoing research and cooperation between marine science and AI fields.



Thank you!



Contact Info.

| | |
|---------|--|
| OFFICE | 경기도 용인시 기흥구 흥덕1로 13 흥덕IT밸리 타워 A동 2901-2903호 |
| EMAIL | info@insilicogen.com |
| PHONE # | 031 278 0061 |
| FAX | 031 278 0062 |

