CPPC 2021
First Canadian Peptide and Protein Community Virtual Symposium
27–28 MAY 2021 | ONLINE
Canadian Peptide Protein Community
biomedicines
MDPI

Discovery and design of non-hemolytic peptides using artificial intelligence

Fabien Plisson[1,*], Obed Ramírez-Sánchez[2], and Cristina Martínez-Hernández[2]

[1] CONACYT, Centro de Investigación y de Estudios Avanzados del IPN (CINVESTAV) Unidad de Genómica Avanzada Laboratorio Nacional de Genómica para la Biodiversidad (Langebio) Irapuato Guanajuato 36824 México;
[2] Centro de Investigación y de Estudios Avanzados del IPN (CINVESTAV) Unidad de Genómica Avanzada Laboratorio Nacional de Genómica para la Biodiversidad (Langebio) Irapuato Guanajuato 36824 México.

* Corresponding author: fabien.plisson@cinvestav.mx

Abstract:
Antimicrobial peptides are polypeptide sequences of 12-50 residues characterized by their charged and hydrophobic cores that were long thought to kill bacteria by a general mechanism; disrupting their membranes leading to cell lysis and death. Their direct antibacterial activities and the lack of bacterial resistance have stimulated their therapeutic avenues against antibiotic-resistant infections. Major limitations preventing AMPs from translating into clinics are their low metabolic stability, poor oral bioavailability and high toxicity. Reducing hurdles to clinical trials without compromising the therapeutic promises of peptide candidates becomes an essential step in peptide-based drug design.

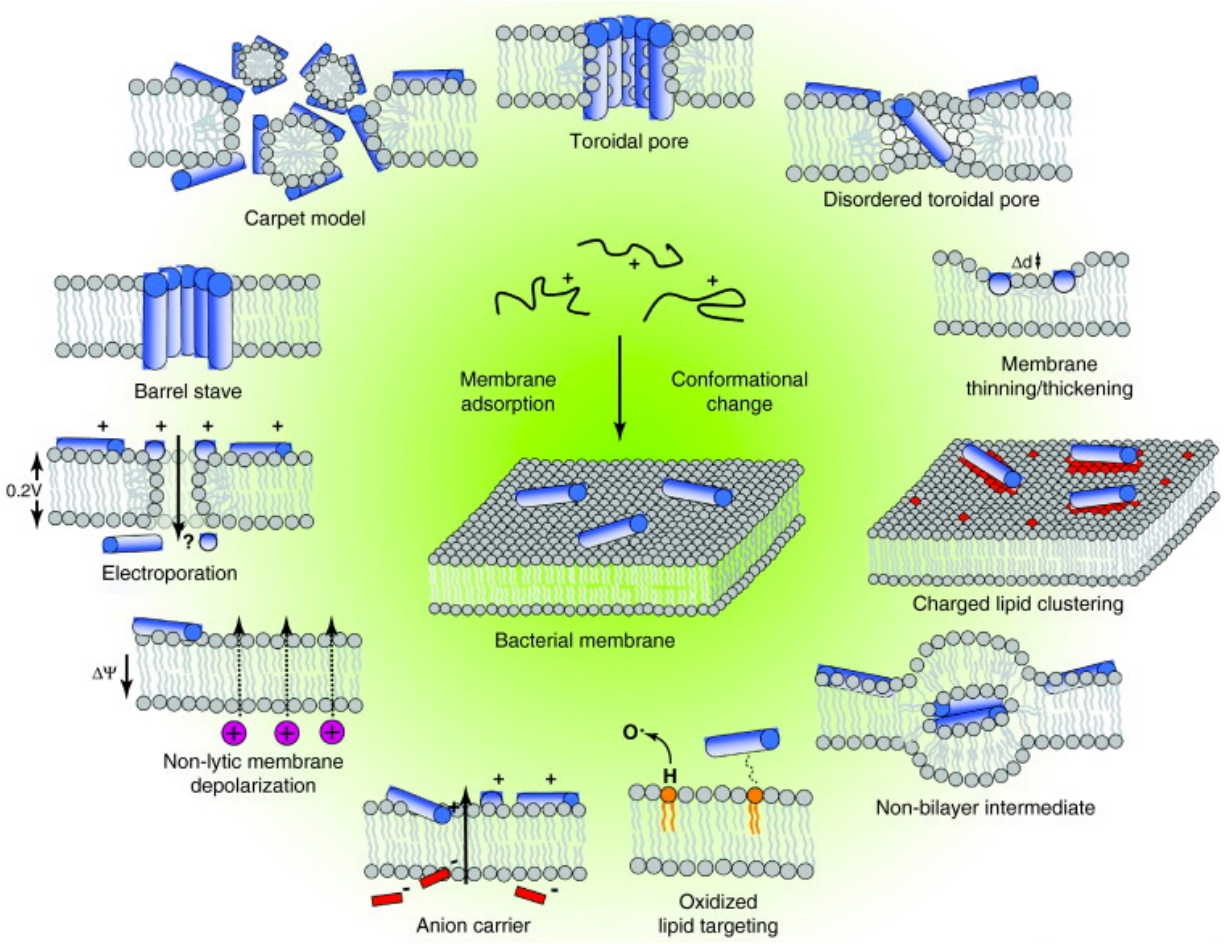
In this presentation, I will discuss the development of machine-learning models and outlier detection methods that ensure robust predictions for the discovery of AMPs and the design of novel peptides with reduced hemolytic activity. Our best models, gradient boosting classifiers, predicted the hemolytic nature from any peptide sequence with 95–97% accuracy. Nearly 70% of AMPs were predicted as hemolytic peptides. Applying multivariate outlier detection models, we found that 273 AMPs (~ 9%) could not be predicted reliably. Our combined approach led to the discovery of 34 high-confidence non-hemolytic natural AMPs, the de novo design of 507 non-hemolytic peptides, and the guidelines for non-hemolytic peptide design.

Keywords: haemolysis, antimicrobial peptides, machine learning, gradient boosting
applicability domain

CPPC
2021

Antimicrobial Peptides (AMPs):

innate lines of defense against pathogens



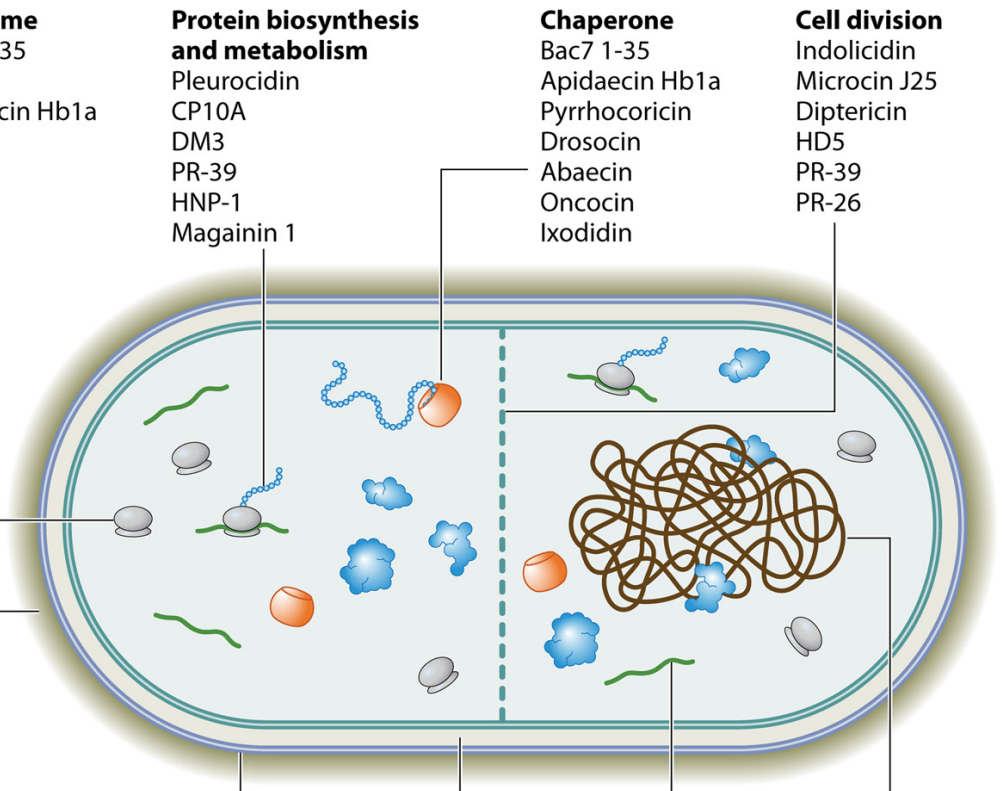
Different mechanisms of action against bacterial membranes:

Main attractive features of developing synthetic AMPs as antibiotics

Nguyen LT *et al.*, *Trends in Biotechnology* (2011)

CPPC 2021

Most AMPs **must interact** with bacterial membranes …but NOT all act to destroy them



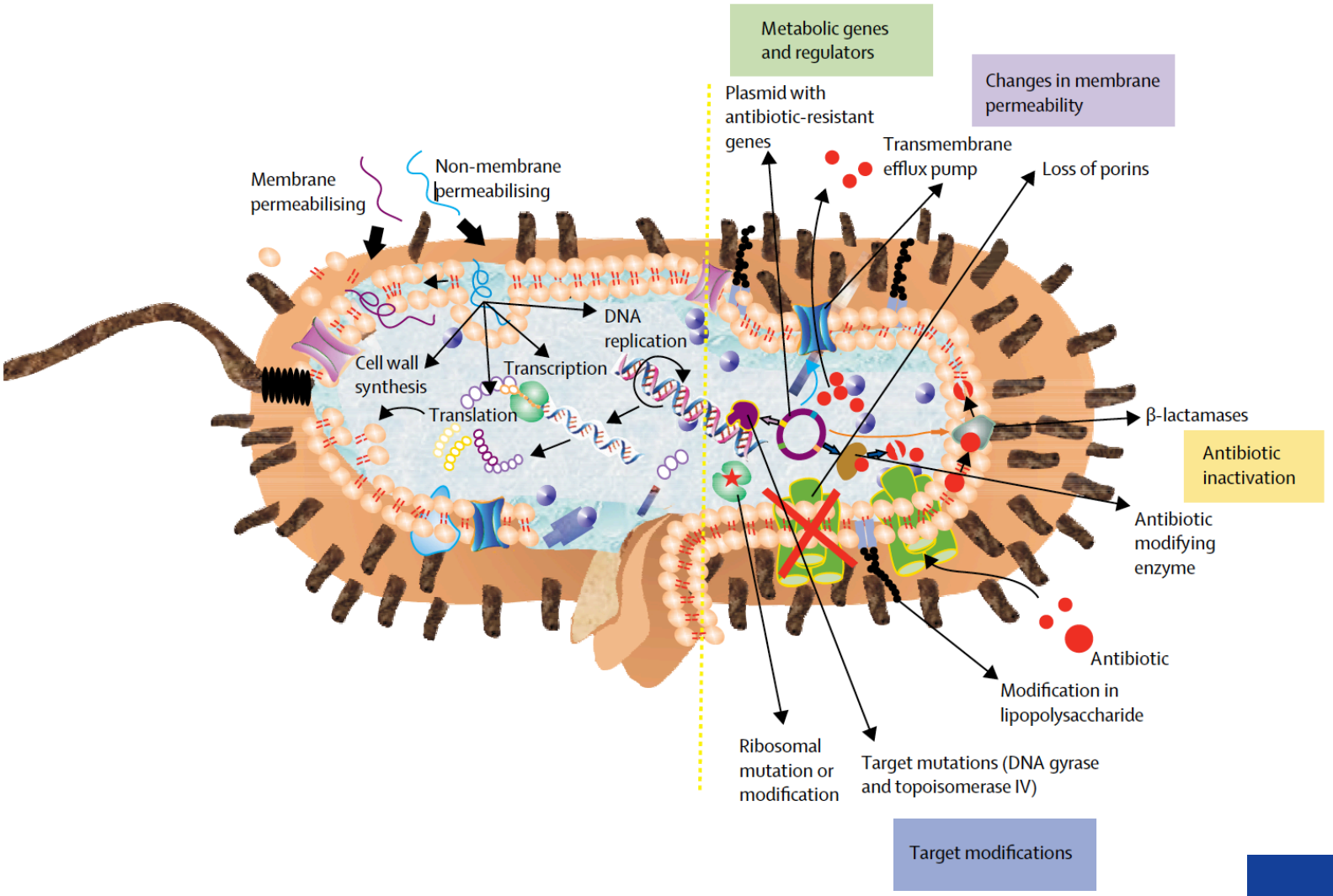
Proteases
Histatin-5
eNAP-2
Ixodidin

Other targets
Lactoferricin B
(inhibits two-component system)
Apidaecin Hb1a
(inhibits ABC transport system)
HNP-1, Pep5, tPMP-1
(activates autolytic enzyme)

Le C.-F. *et al.*, *Antimicrobial Agents and Chemotherapy* (2017)

CPPC 2021

Bacterial resistance to AMPs?



Magana M. *et al.*, *Lancet Infectious Diseases* (2020)

CPPC 2021

Alternatives to conventional antibiotics

AMP	Description	Phase	Indication	Administration	Clinical trial identifier if available
Pexiganan (MSI-78)	Analog of magainin (skin of African clawed frog)	Phase III	Infected diabetic foot ulcers	Topical cream	NCT00563394, NCT00563433
Omiganan	Derived from indolicidin (bovine)	Phase II/III	Catheter infections and rosacea	Topical gel	NCT00231153, NCT01784133
Lytixar (LTX-109)	Synthetic antimicrobial peptidomimetic	Phase I/II	Uncomplicated Gram-positive skin infections, impetigo, and nasal colonization with *S. aureus*	Topical hydrogel	NCT01223222, NCT01803035, NCT01158235
hLF1-11	Derived from lactoferricin (human)	Phase I/II	Bacteraemia and fungal infections in immunocompromized haematopoetic stem cell transplant recipients	Intravenous treatment (in saline)	NCT00509938
Novexatin (NP-213)	Derived from defensins (human)	Phase II	Onychomycosis (fungal nail infection)	Topical brush-on-treatment	
CZEN-002	Dimeric octamer derived from α-MSH (human)	Phase IIb	Vaginal candidiasis	Vaginal gel	
LL-37	LL-37 (human)	Phase I/II	Hard-to-heal venous leg ulcers	Polyvinyl alcohol-based solution for administration in the wound bed	
PXL01	Derived from lactoferricin (human)	Phase II	Prevention of post-surgical adhesion formation in hand surgery	Hyaluronic acid-based hydrogel for administration at the surgical site	NCT01022242
Iseganan (IB-367)	Derived from protegrin 1 (porcine leukocytes)	Phase III	Oral mucositis in patients receiving radiotherapy for head and neck malignancy	Oral solution	NCT00022373
PAC-113	Derived from histatin 3 (human saliva)	Phase II	Oral candidiasis in HIV seropositive patients	Mouthrinse	NCT00659971

α-MSH, α-melanocyte-stimulating hormone.

Mahlapuu M. *et al., Frontiers Cell. Infect. Microbiology* (2016)

CPPC 2021

Defines toxicity

The toxicity of peptides can be broadly classified into 3 categories:
- cytotoxicity
- haemotoxicity (lysis of red blood cells)
- immunotoxicity

Various methods exist to predict cytotoxicity, immunotoxicity/allergenicity of peptides but almost none for predicting peptides with haemolytic capacity

1. Half maximum Effective Concentration (EC_{50}) or Hazardous Concentration (HC_{50}) $\leq$100 µM
2. Minimum Hemolytic Concentration (MHC) $\leq$ 250 µg/ml
3. $>$10% hemolytic activity upto 100 µM

Chaudhary K., *et al. Scientific Reports* (2016)

CPPC 2021

The rise of online predictors for haemolytic activity

A Web Server and Mobile App for Computing Hemolytic Potency of Peptides

Kumardeep Chaudhary, Ritesh Kumar, Sandeep Singh, Abhishek Tuknait, Ankur Gautam, Deepika Mathur, Priya Anand, Grish C. Varshney & Gajendra P. S. Raghava

Scientific Reports **6**, Article number: 22843 (2016) | Cite this article

FUTURE MEDICINAL CHEMISTRY, VOL. 9, NO. 3 | RESEARCH ARTICLE

HemoPred: a web server for predicting the hemolytic activity of peptides

Thet Su Win, Aijaz Ahmad Malik, Virapong Prachayasittikul, Jarl E S Wikberg, Chanin Nantasenamat & Watshara Shoombuatong

Published Online: 17 Feb 2017 | https://doi.org/10.4155/fmc-2016-0188

JOURNAL ARTICLE OPEN ACCESS

A Method for Predicting Hemolytic Potency of Chemically Modified Peptides From Its Structure

Kumar V, Kumar R, Agrawal P et al. See more

Frontiers in Pharmacology (2020) 11

DOI: 10.3389/fphar.2020.00054

HLPpred-Fuse: improved and robust prediction of hemolytic peptide and its activity by fusing multiple feature representation

Md Mehedi Hasan, Nalini Schaduangrat, Shaherin Basith, Gwang Lee, Watshara Shoombuatong, Balachandran Manavalan

Bioinformatics, Volume 36, Issue 11, June 2020, Pages 3350–3356, https://doi.org/10.1093/bioinformatics/btaa160

Published: 14 April 2020 **Article history**

HAPPENN is a novel tool for hemolytic activity prediction for therapeutic peptides which employs neural networks

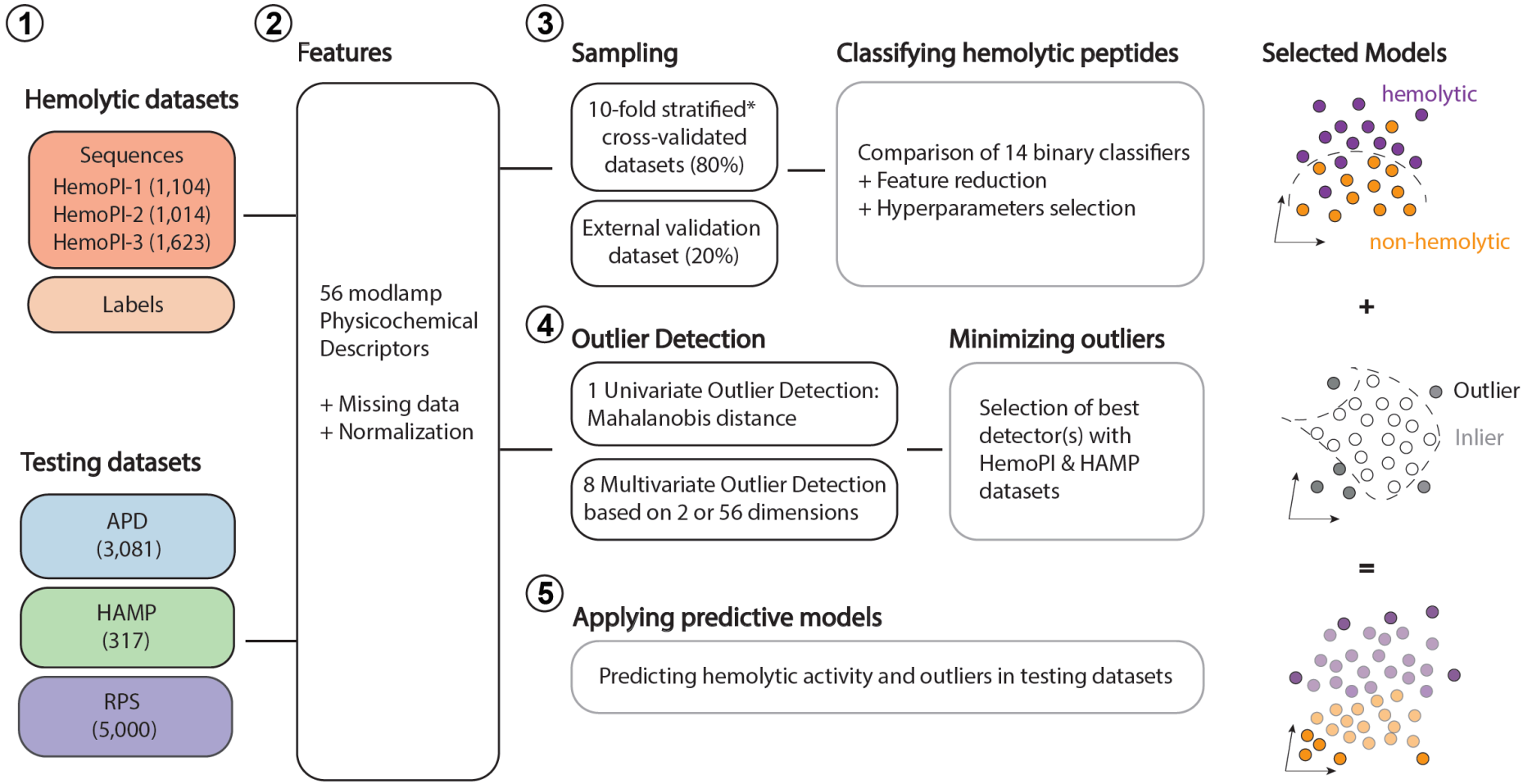
Timmons P, Hewage C

Scientific Reports (2020) 10(1)

DOI: 10.1038/s41598-020-67701-3

CPPC 2021

ML-guided discovery and design of non-hemolytic AMPs



Plisson F., *et al. Scientific Reports* (2020)

CPPC 2021

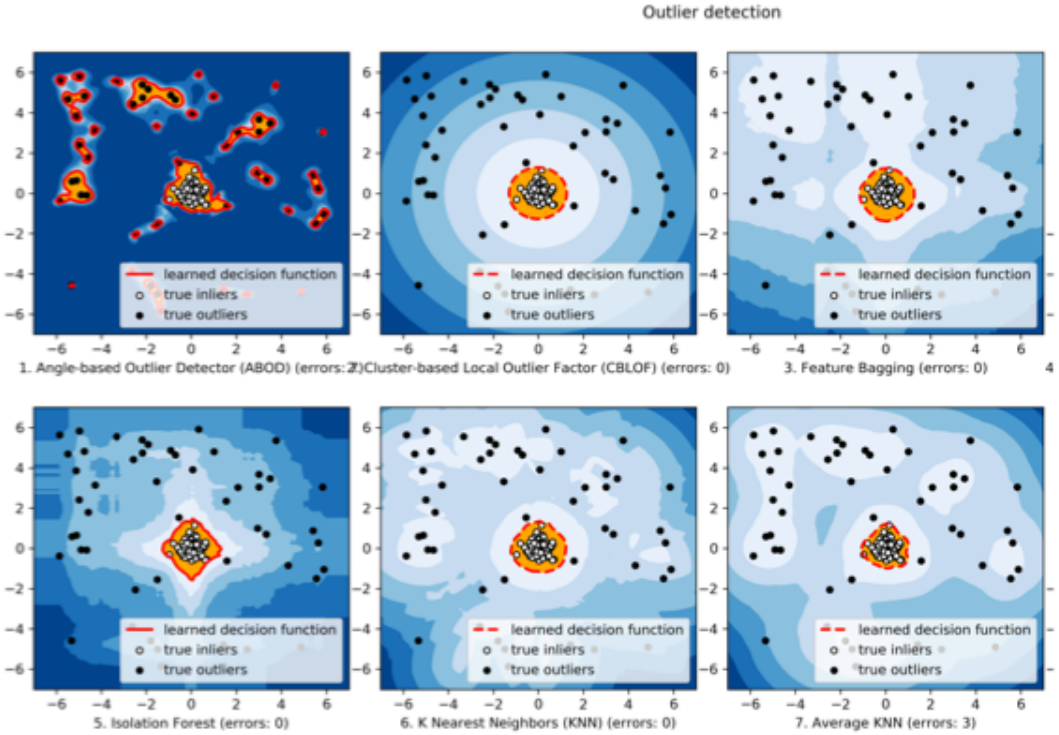
Binary classifiers to predict hemolytic nature & activity

Classifiers		Hyperparameters	Feature reduction	N	Acc. (%)	Prec. (%)	MCC statistic	CK statistic	AUC ROC
HemoPI-1 model and validation datasets									
1.1	LDA	'solver': 'svd', 'tol': 0.0001	RFECV	18	95.1	92.6	0.903	0.903	0.951
					94.6	92.5	0.891	0.891	0.946
1.2	GBC	'max_depth': 4, 'max_features': 'sqrt', 'min_samples_leaf': 10, 'n_estimators': 240	MC (0.75)	26	96.5	95.0	0.930	0.930	0.965
					92.7	89.6	0.855	0.855	0.927
1.3	GBC	'max_depth': 4, 'max_features': 'sqrt', 'min_samples_leaf': 10, 'n_estimators': 208	None	56	96.0	94.6	0.921	0.921	0.960
					92.3	89.2	0.846	0.846	0.923
HemoPI-2 model and validation datasets									
2.1	GBC	'max_depth': 4, 'max_features': 'sqrt', 'min_samples_leaf': 2, 'n_estimators': 112	None	56	77.7	74.0	0.549	0.549	0.774
					74.3	70.4	0.479	0.476	0.736
2.2	GBC	Default	RFECV	15	77.8	74.2	0.552	0.552	0.775
					73.2	69.8	0.459	0.482	0.728
2.3	GBC	Default	None	56	76.7	72.9	0.529	0.528	0.763
					72.3	68.9	0.439	0.437	0.717
HemoPI-3 model and validation datasets									
3.1	GBC	'max_depth': 18, 'max_features': 'log2', 'min_samples_leaf': 10, 'n_estimators':192	None	56	80.0	76.4	0.597	0.597	0.796
					71.7	68.3	0.427	0.425	0.711
3.2	GBC	'max_depth': 12, 'max_features': 'sqrt', 'min_samples_leaf': 8, 'n_estimators': 160	RFECV	40	78.2	74.4	0.559	0.558	0.777
					74.5	70.8	0.483	0.482	0.740
3.3	GBC	'max_depth': 20, 'max_features': 'log2', 'min_samples_leaf': 8, 'n_estimators': 96	MC (0.75)	28	78.0	74.4	0.556	0.556	0.777
					72.6	69.0	0.445	0.443	0.719

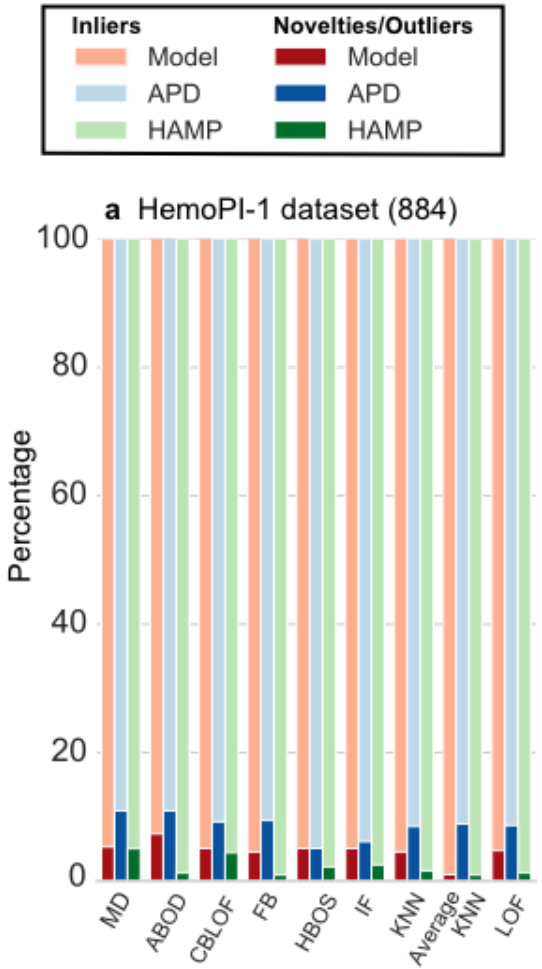
Optimal number N of descriptors were determined using multicollinearity, RFECV: tenfold cross-validated recursive feature extraction or BE: backward elimination.

Plisson F., *et al. Scientific Reports* (2020)

Maximising Inliers (o) and minimising novelties/outliers (•)



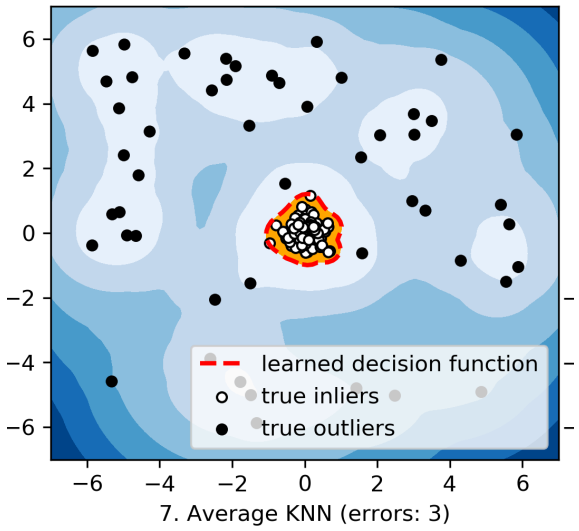
https://raw.githubusercontent.com/yzhao062/pyod/master/examples/ALL.png



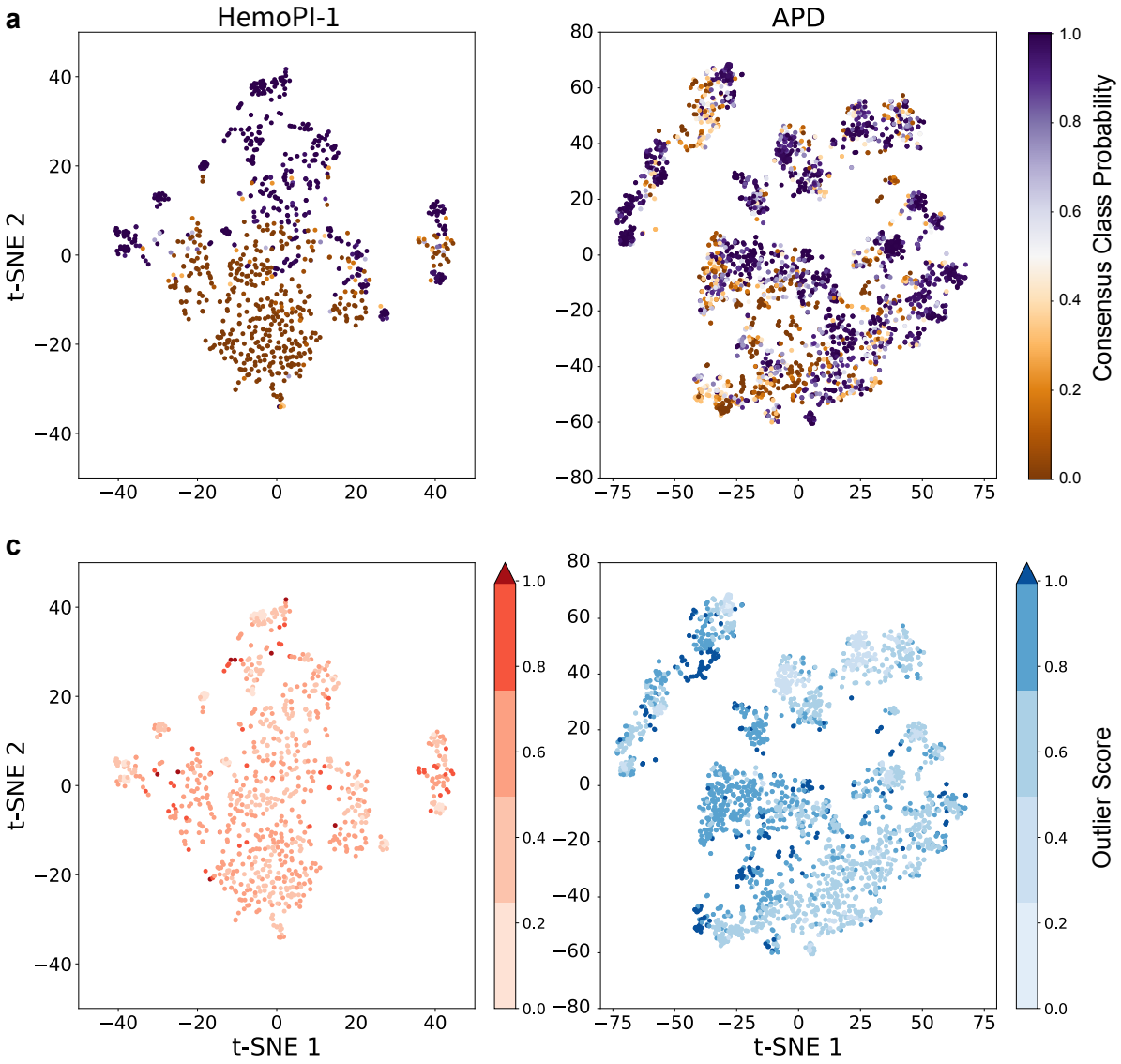
Average KNN is the best outlier detector with HemoPI datasets

Plisson F., *et al. Scientific Reports* (2020)

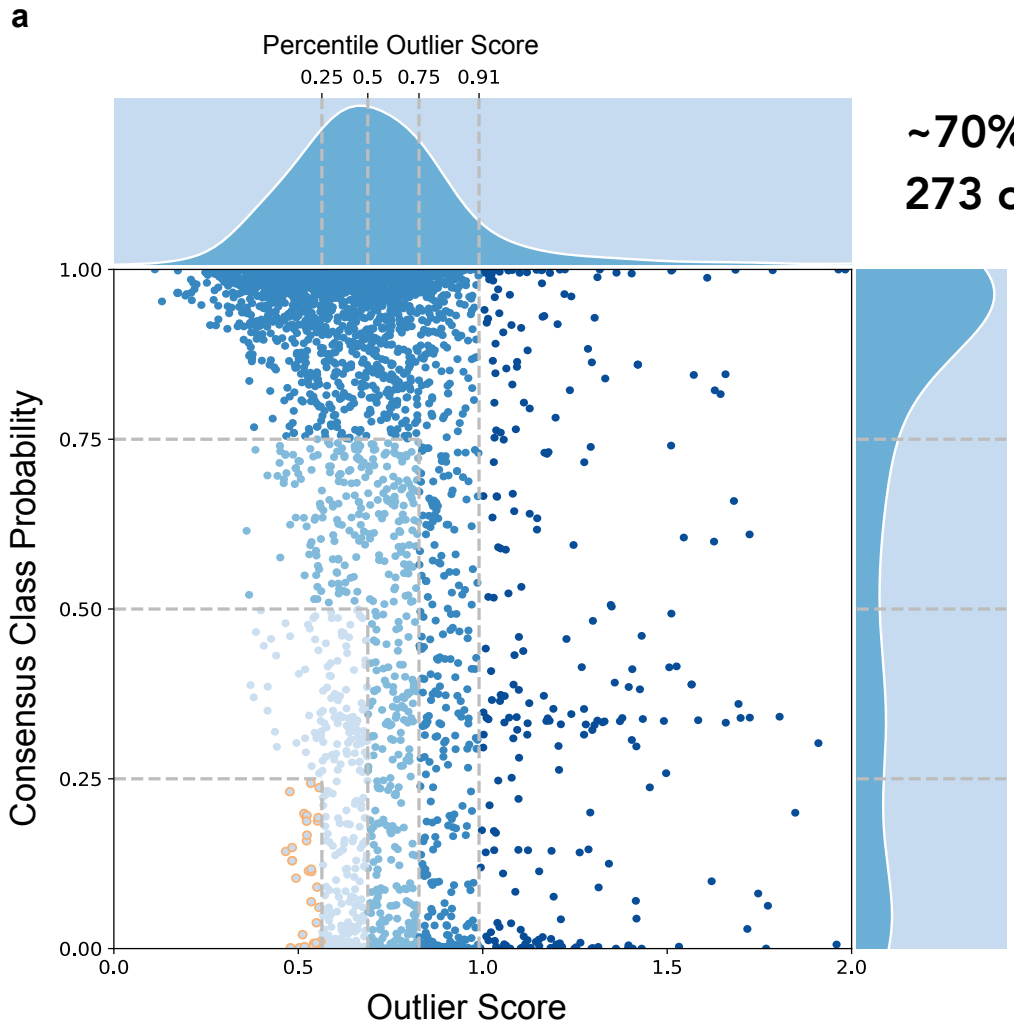
CPPC 2021



https://raw.githubusercontent.com/yzhao062/pyod/master/examples/ALL.png



Discovery of novel (non-)hemolytic AMPs

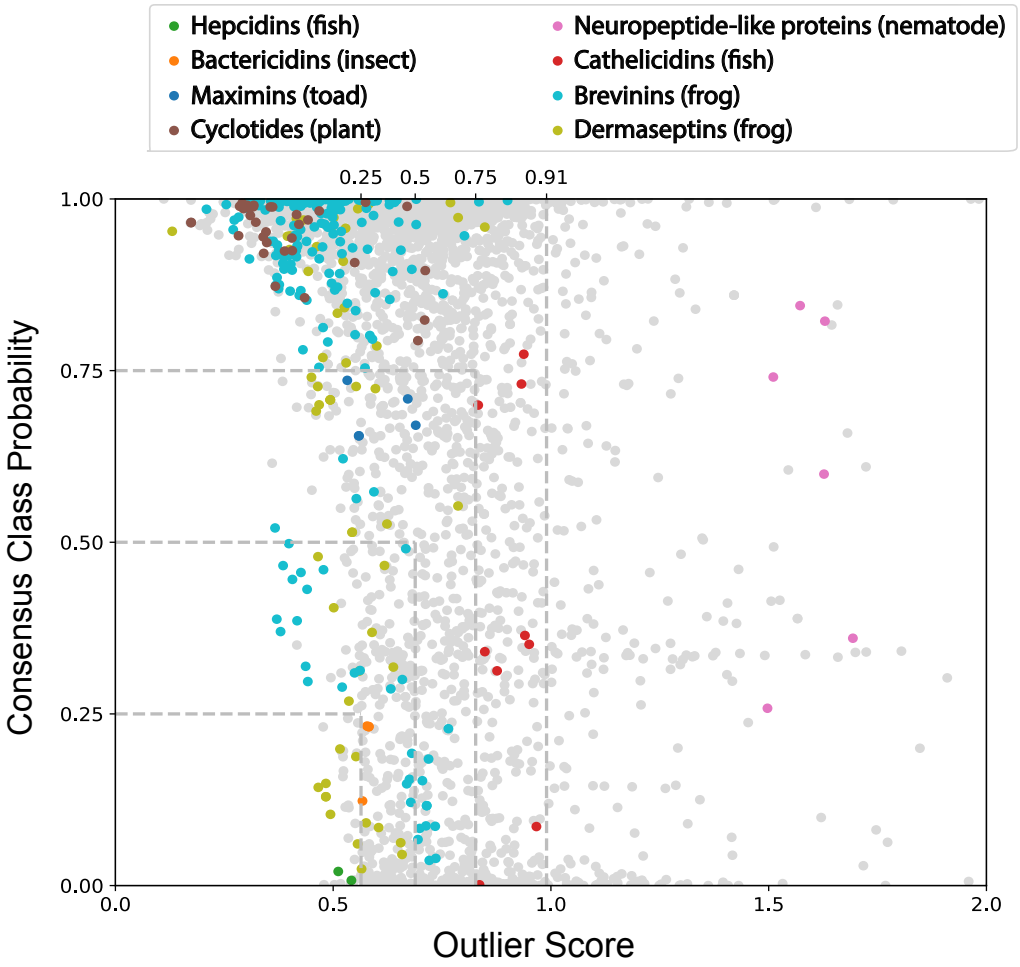


~70% AMPs were predicted haemolytic.
273 out 3,081 (~9%) AMPs are outliers

Plisson F., *et al. Scientific Reports* (2020)

CPPC 2021

300 out of 317 known haemolytic peptides

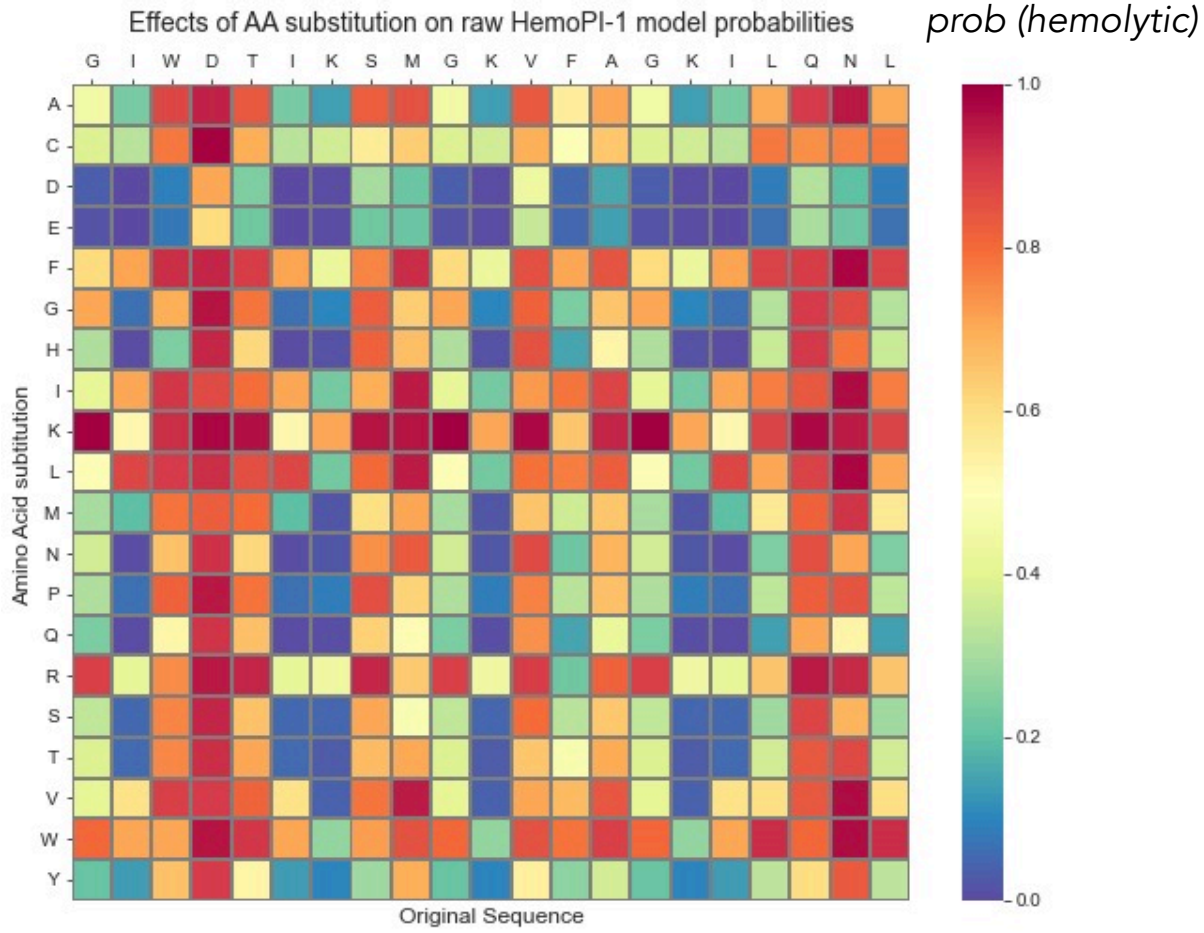


P27-Seminalplasmin,
Odorranain-M1,
Ranatuerin-2PRb,
Ocellatin-PT6,
Maximin 45,
rt-CATH-1b/2a

Plisson F., *et al. Scientific Reports* (2020)

CPPC 2021

Deep mutational scanning to find key residues

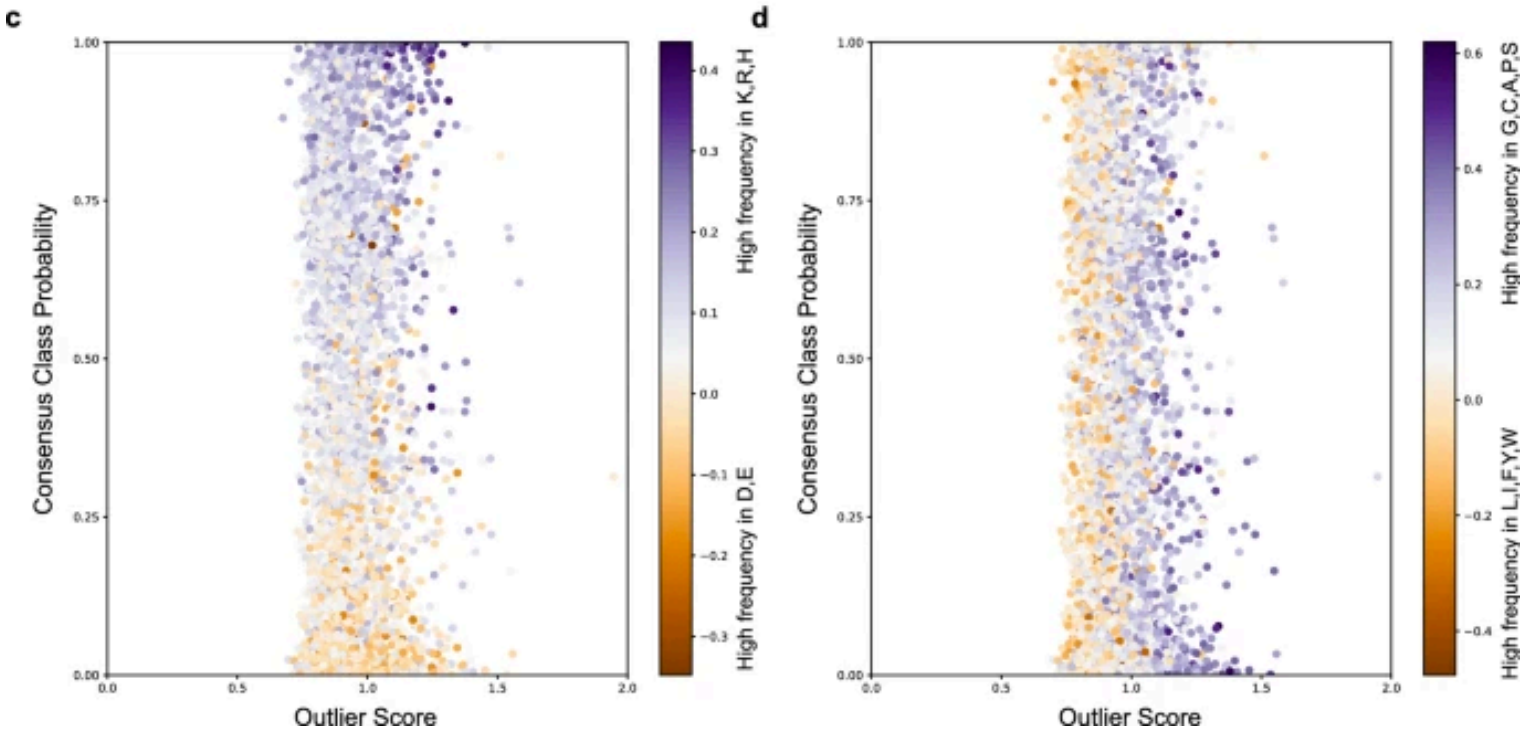


Fowler, D. & Fields, S. *Nature Methods* (2014) - ligand binding, protein stability
Rollins, N. *et al.*, *Nature Genetics* (2019) - double mutants / 3D protein structure

CPPC 2021

De novo design of 5,000 random peptide sequences

507 (~10%) were non-haemolytic peptides and low outlier scores.



"To design non-hemolytic peptides; peptide sequences should be neutral or slightly charged sequences with ~ 20% positively and negatively charged residues (ratio 3:1), an equal proportion (~ 30%) of aromatic/aliphatic residues (ratio 1:2) and small amino acids in random peptide sequences to insure robust hemolytic predictions."

Plisson F., *et al. Scientific Reports* (2020)

CPPC 2021

Conclusions

Predictive models (Gradient Boosting) were able to classify 95-97% (HemoPI-1) the hemolytic nature of any peptide sequence.

External validation with hemolytic AMPs (HAMPs) identified 300 out of 317 known hemolytic peptides.

First application of multivariate outlier detection to peptide predictive modelling. Average KNN was identified as the best outlier detector across all HemoPI databases.

<u>Applications</u>
~70% AMPs of the Antimicrobial Peptide Database (3081) were predicted as hemolytic peptides.

Deep mutational scanning (single mutants) can help identifying key residues associated with a biological activity (fitness function).

De novo design amplified the key characteristics (properties, AAs) found in (non-)hemolytic peptides.

CPPC
2021

Acknowledgments

CONACYT CÁTEDRAS

Cinvestav - Irapuato
Langebio

Thank you for your attention!

CPPC 2021