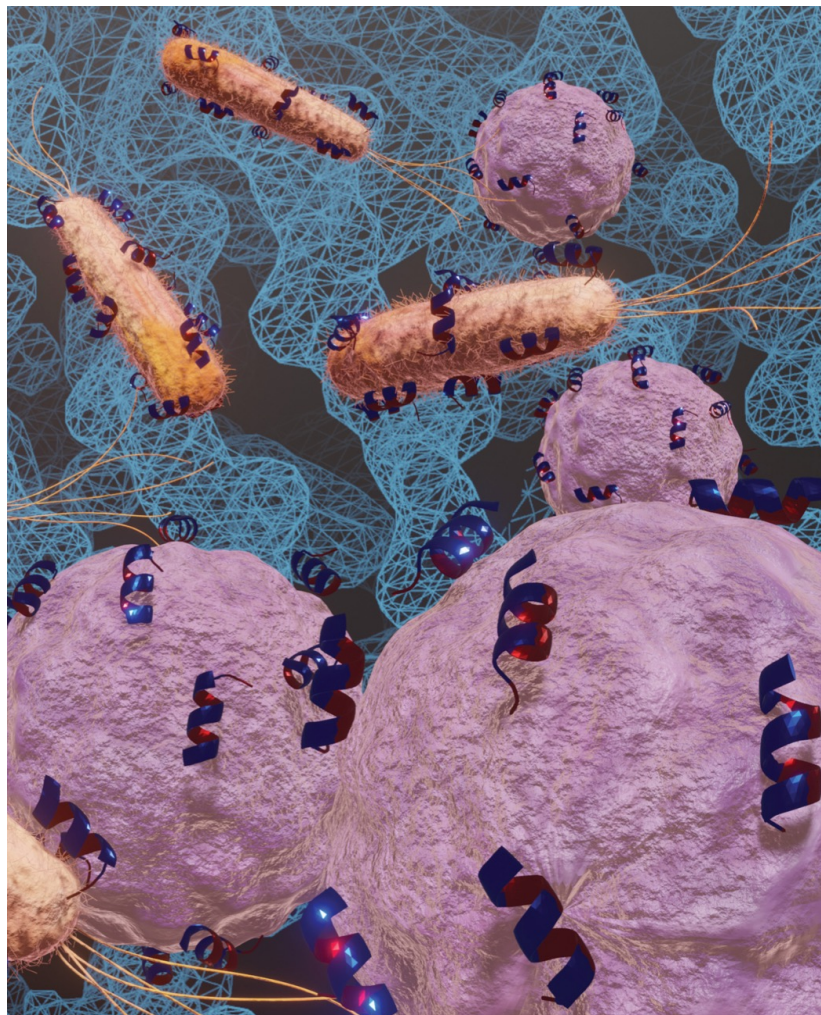


# Machine learning aided discovery of membrane disruptive, non-hemolytic anticancer peptides

**AMLD – AI & the Molecular World**

# Membrane disruptive peptides



Overcome drug resistance

Low immunogenicity

Good biocompatibility

Low metabolic toxicity

Easy synthesis

Hemolytic effects

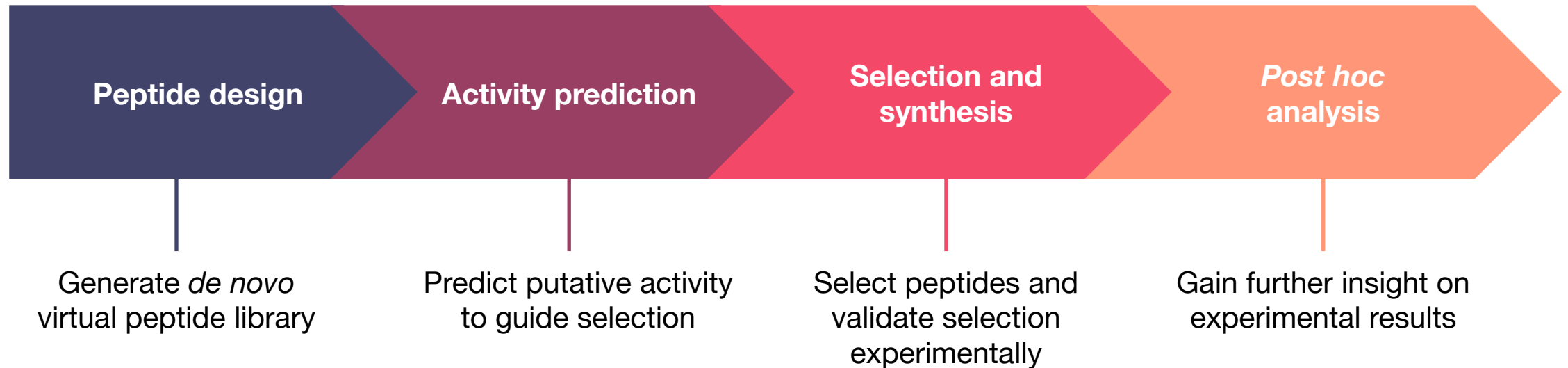
Baeriswyl S. *et al.* A Mixed Chirality  $\alpha$ -Helix in a Stapled Bicyclic and a Linear Antimicrobial Peptide Revealed by X-Ray Crystallography. *RSC Chem. Biol.* 2021, 2 (6), 1608–1617



Dr. Alice Capecchi



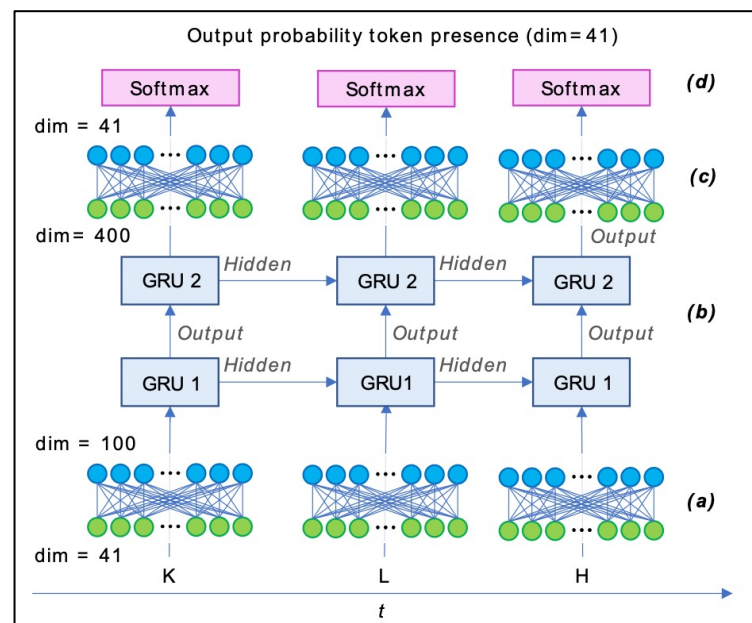
Elena Zakharova



# Peptide design

## Recurrent Neural Network (RNN)

- Trained on 9'548 linear sequences from DBAASP
- Two layers of gated recurrent unit (GRU) cells
- Transfer learning with 53 peptides active against HeLa cells



Peptide design

Activity prediction

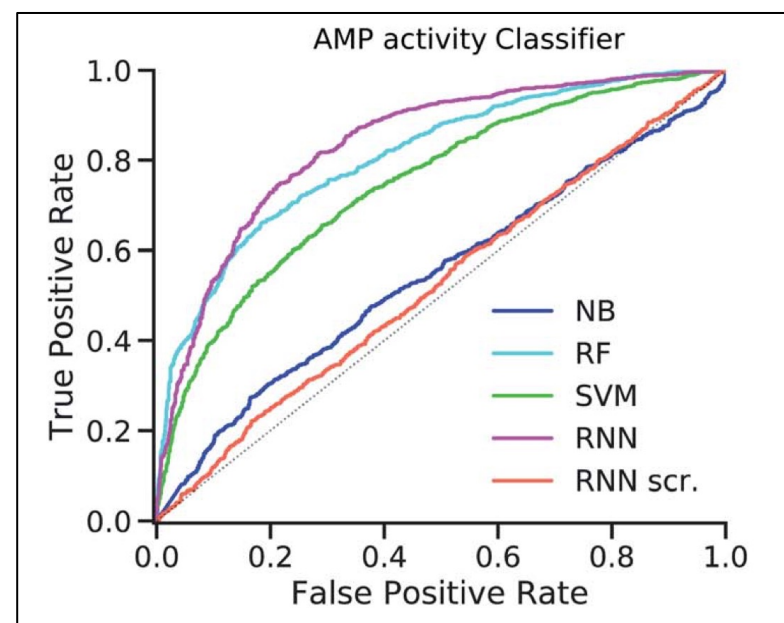
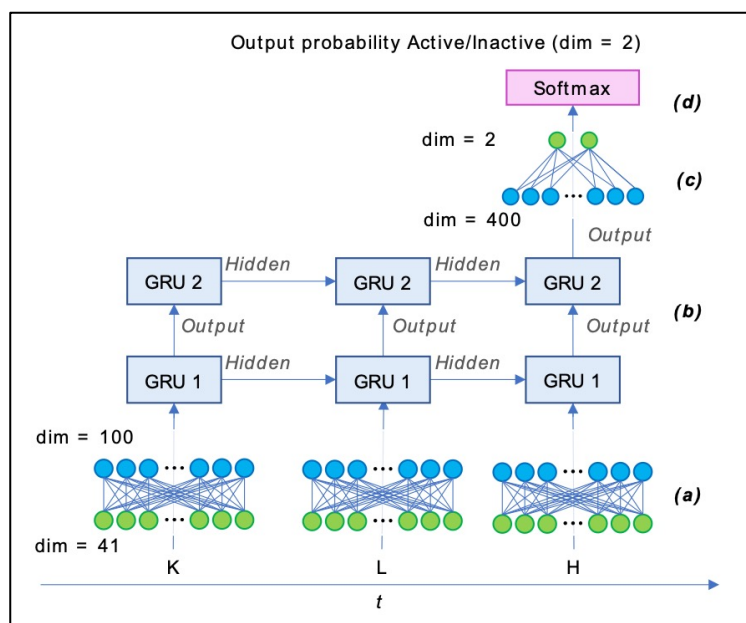
Selection and synthesis

Post hoc analysis

# Activity prediction

## Activity classifier

- Trained on DBAASP activity data of 6'641 sequences
- Two layers of GRU cells



Peptide design

Activity prediction

Selection and synthesis

Post hoc analysis

# Selection and synthesis

## Selection filters

- LD > 5 from training sets and LD > 4 from test set
- Discarded sequences with LD > 8 (90% quantile)
- Pred. helicity > 0.8; hydrophobic moment > 0.3
- Max. length = 15 AA; L-AA only
- Butina clustering centroids

Name	Sequence	HeLa (cancer)	PAO1 (bacterial)	A. Baumannii (bacterial)	Helix, %
A-1	FAKKFFKKFAKFVK	Active	Active	Active	73.7
A-2	WFKRILKYLKCLV	Active	Active	Active	65.5
A-3	WLNALKKILGHLIRH	Active	Active	Active	78.5
A-4	KYLYLVRLVGRLYR	Active	Active	Active	68.3
A-5	WKRIIRIIRWIRKYY	Active	Inactive	Inactive	74.1
A-6	FAARILRAWFRFLRR	Active	Inactive	Active	75.4
A-7	SISRLWHSLLRHLLH	Active	Inactive	Active	75.9
A-8	KNFKKLMKKVASVL	Inactive	Active	Active	50.7
A-9	KWRSKIKKIMRTFK	Inactive	Active	Active	45.9
A-10	GLLGRLLAKLLANS	Inactive	Active	Active	49.3
A-11	SFSKWMGKLNIFKK	Inactive	Active	Active	50.0
A-12	LLRHCLRRIRDRLV	Inactive	Active	Active	69.5
A-13	VFRQWQKIMRRLVRR	Inactive	Inactive	Active	48.8

Peptide design

Activity prediction

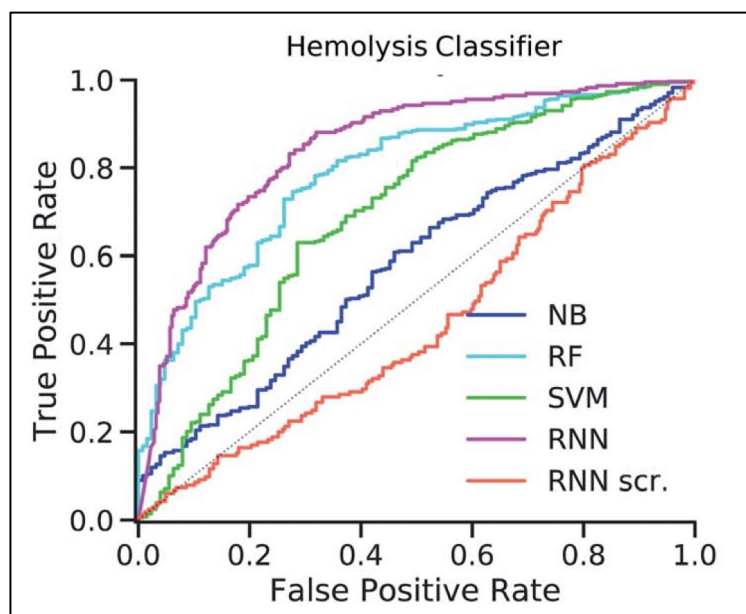
Selection and synthesis

Post hoc analysis

# Post hoc analysis

## Hemolysis classifier

- Trained on DBAASP hemolysis data of 2'262 sequences
- Architecture similar to activity classifier
- Only one layer of GRU cells



Name	Sequence	Predicted	Experimental	Outcome
A-1	FAKKFFKKFAKF	Non-hemolytic	Non-hemolytic	True
A-2	WFKRILKYLKLV	Hemolytic	Hemolytic	True
A-3	WLNALKKILGHLIRH	Hemolytic	Hemolytic	True
A-4	KYLKYLVRVLRGRLYR	Non-hemolytic	Hemolytic	False
A-5	WKRIVRIIRWIRKYY	Hemolytic	Hemolytic	True
A-6	FAARILRAWFRFLRR	Hemolytic	Hemolytic	True
A-7	SISRLWHSLLRHLLH	Hemolytic	Hemolytic	True
A-8	KNFKKLMKKVASVL	Non-hemolytic	Non-hemolytic	True
A-9	KWRSKIKKIMRTFK	Hemolytic	Non-hemolytic	False
A-10	GLLRLAKLLANS	Hemolytic	Non-hemolytic	False
A-11	SFSKWMGKLNIFKK	Hemolytic	Non-hemolytic	False
A-12	LLRHCLRRIRDRLV	Non-hemolytic	Non-hemolytic	True
A-13	VFRQWQIMRRLVRR	Hemolytic	Non-hemolytic	False

Peptide design

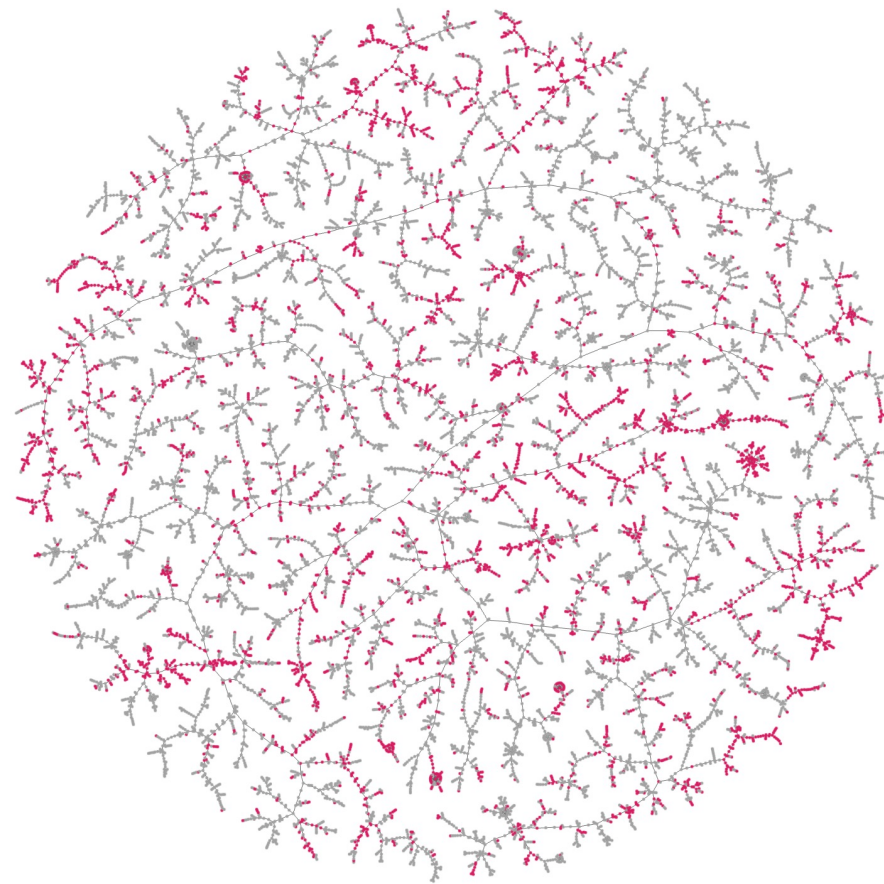
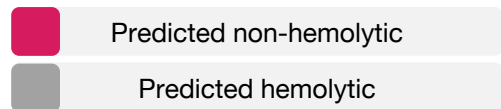
Activity prediction

Selection and synthesis

Post hoc analysis



MAP4  
UMAP



MAP4  
TMAP

Peptide design

Activity prediction

Selection and synthesis

Post hoc analysis

Probst D. and Reymond J.-L. Visualization of Very Large High-Dimensional Data Sets as Minimum Spanning Trees. J Cheminform 2020, 12 (1), 12.



# Thank you for your attention!

## Special thanks to:

Prof. Dr. Jean-Louis Reymond

Dr. Alice Capecchi  
Elena Zakharova

The Reymond Group

Organizers of the AMLD AI &  
the Molecular World track



European Research Council  
Established by the European Commission

*u<sup>b</sup>*

b  
UNIVERSITÄT  
BERN



markus.orsi@unibe.ch



@MarkusOrsi



Markus Orsi



**Table S1.** Performance on the test of the NB, RF, SVM, RNN, and RNN with scrambled labels (RNN scr.) models for the AMP activity (act.) and hemolysis (hem.) classification tasks.

<b>Classifier</b>	<b>ROC AUC</b>	<b>Accuracy <sup>a)</sup></b>	<b>Precision <sup>a)</sup></b>	<b>Recall <sup>a)</sup></b>	<b>F1 score <sup>a)</sup></b>	<b>MCC <sup>a)</sup></b>
NB act.	0.55	0.55	0.59	0.32	0.42	0.11
SVM act.	0.75	0.68	0.68	0.68	0.68	0.36
RF act.	0.81	0.71	0.70	0.75	0.73	0.44
<b>RNN act.</b>	<b>0.84</b>	<b>0.76</b>	<b>0.74</b>	<b>0.80</b>	<b>0.77</b>	<b>0.53</b>
RNN scr. act.	0.51	0.49	0.35	0.03	0.05	-0.06
NB hem.	0.58	0.56	0.48	<b>0.76</b>	0.59	0.19
SVM hem.	0.69	0.73	0.72	0.58	0.65	0.44
RF hem.	0.80	<b>0.77</b>	<b>0.81</b>	0.60	0.69	<b>0.53</b>
<b>RNN hem.</b>	<b>0.87</b>	<b>0.76</b>	0.70	<b>0.76</b>	<b>0.73</b>	<b>0.52</b>
RNN scr. hem.	0.45	0.61	0.41	0.05	0.10	0.01

a) The probabilistic prediction values were converted into binary classification values using a threshold of 0.5. The best values and the selected classifiers are reported in bold.