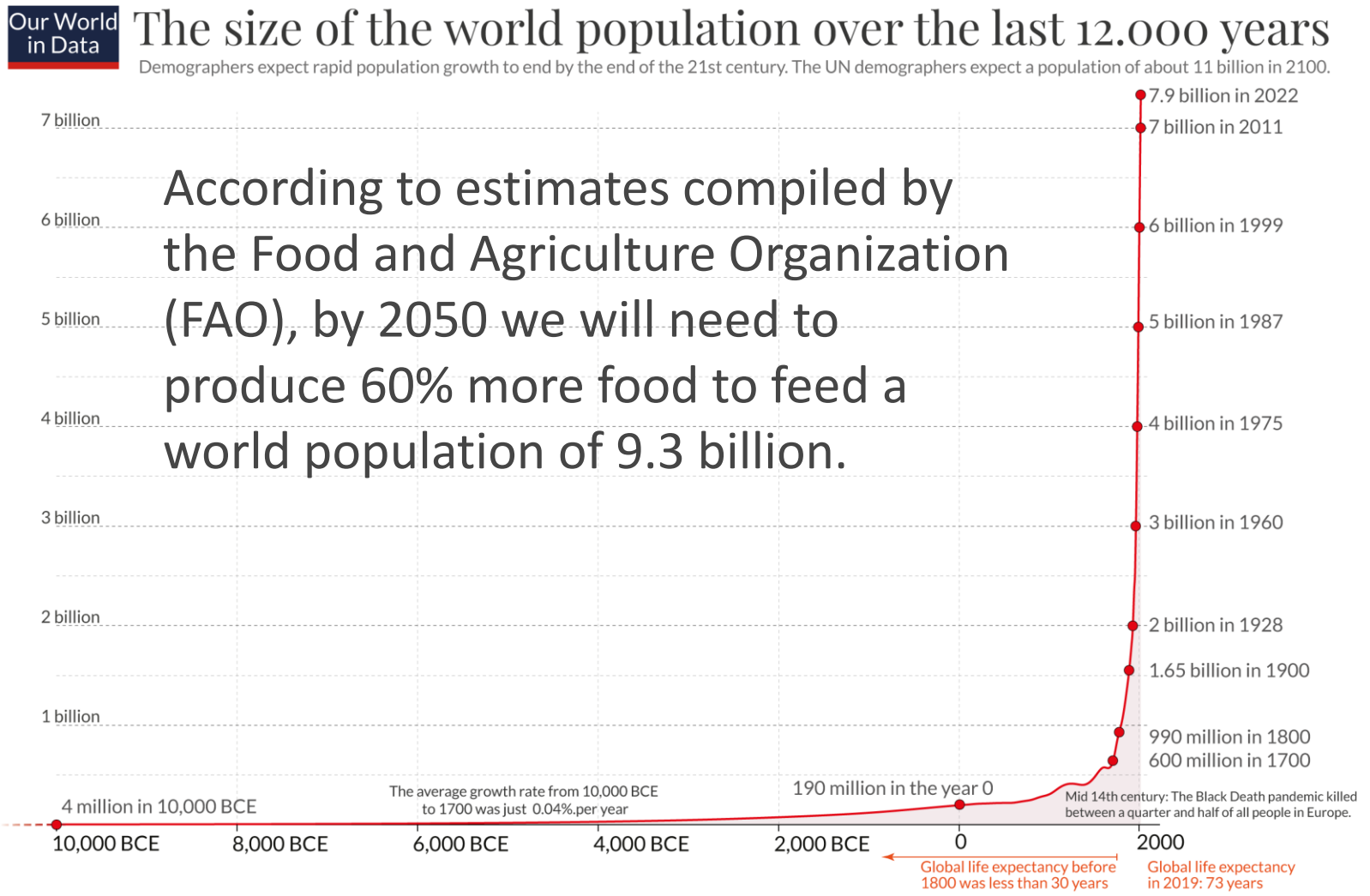
Computational Studies on Green Pesticides

Netaly Khazanov, Shahaf Kozokaro, Lina Iktelat, Omer Kaspi, Paul Clarke, Hanoch Senderowitz

Department of Chemistry, Bar Ilan University, Ramat-Gan, 5290002, Israel

XXVIII Symposium on Bioinformatics and Computer-Aided Drug Discovery
Moscow, May 2022

Why Pesticides?

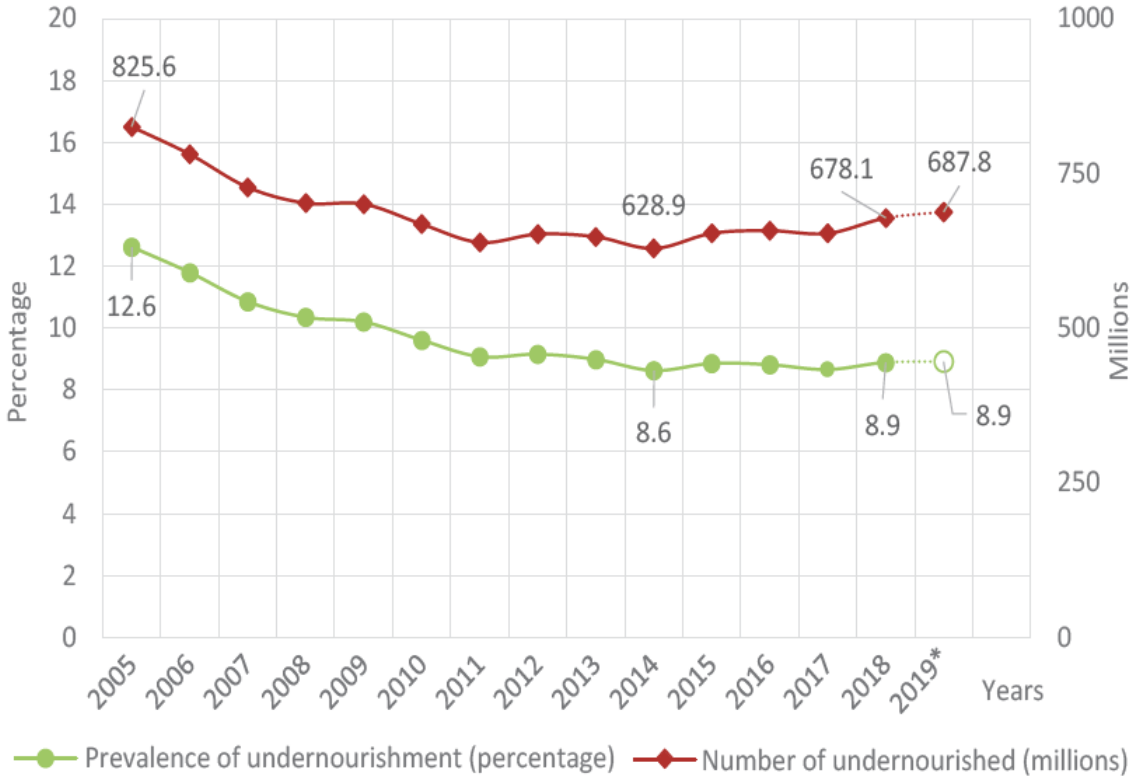


Why Pesticides?

The stall in global progress against undernourishment has been driven by many factors, including economic slowdowns, armed conflicts, humanitarian emergencies, disease outbreaks, pest infestations and adverse consequences of climate change, including drought and extreme weather events.

Figure 1

Global number and percentage of undernourished persons, 2005–2019

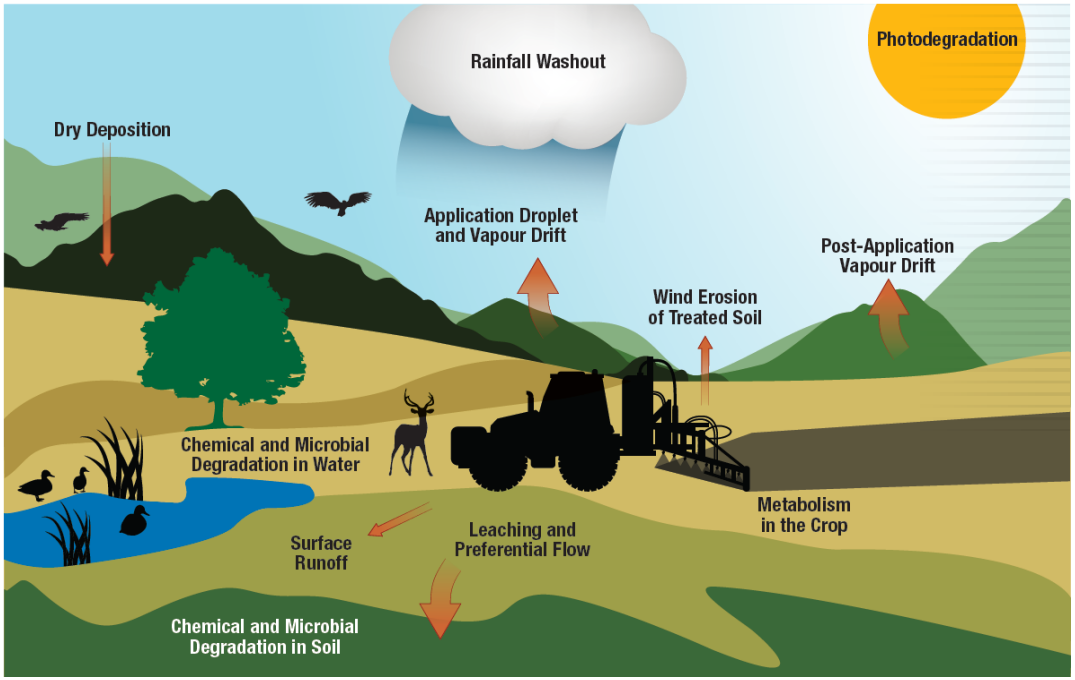


Source: Adapted from Food and Agriculture Organization of the United Nations (FAO) and others, The State of Food Security and Nutrition in the World 2020, figure 1.
Note: Values for 2019 are projected.

Why Green?

According to estimates compiled by the Food and Agriculture Organization (FAO), by 2050 we will need to produce 60 per cent more food to feed a world population of 9.3 billion. Doing that with a farming-as-usual approach would take too heavy a toll on our natural resources. Thus, we have no choice but to embark on a greener revolution.

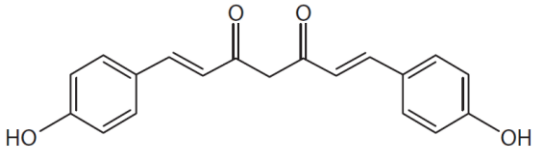
- Pesticides may be harmful to human health
- Pesticides may be harmful to the environment
- Pesticides may be harmful to the eco-system



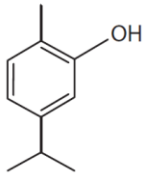
Learning from the Plants

- Plants and pathogens have co-evolved for millions of years
- Plants have developed an arsenal of tools to ward off pathogenic virulence
- Many of these compounds are poly phenolics

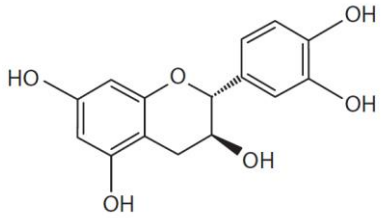
Bisdesmethoxycurcumin



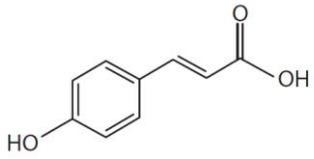
Carvacrol



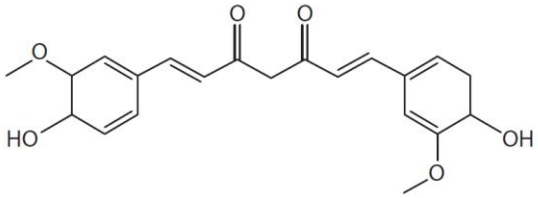
Catechin



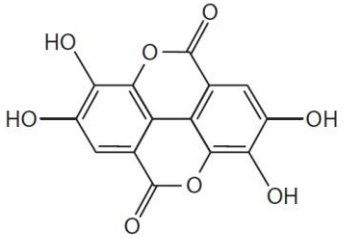
Coumaric acid



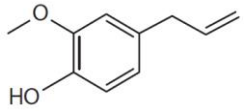
Curcumin



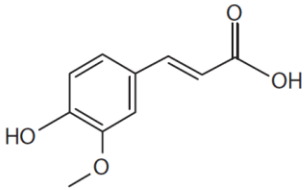
Ellagic acid



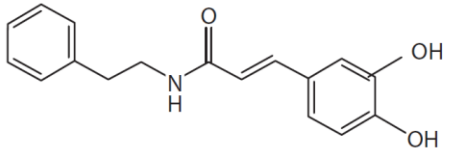
Eugenol



Ferulic acid

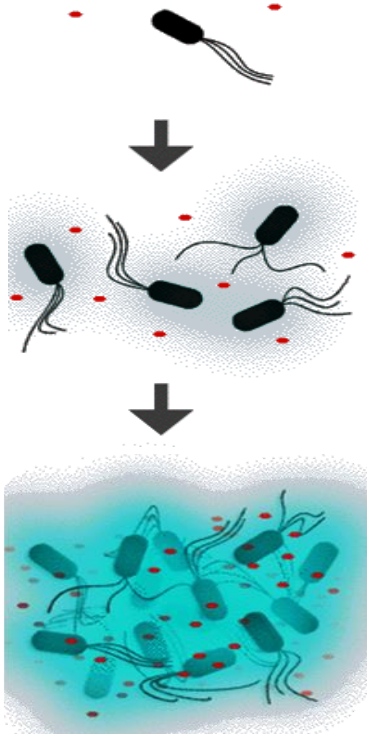


Phenylethylcinnamamide



Quorum Sensing Machinery

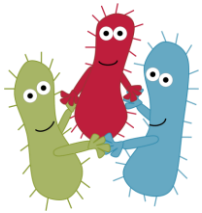
- Bacteria communicate to coordinate virulence via secreted signaling molecules called: “autoinducers”
- Acyl homoserine lactones (AHLs) are the chemical language of gram negative bacteria
- AHLs are synthesized by AHL synthases and are “read” by response regulators
- AHLs ultimately regulate the expression of genes



Low cell density

Individual behavior

High cell density



Group behaviors

Quorum sensing

OFF

- Negative regulation of group behavior
- Virulence genes are off

ON

- Cellular attachment - biofilm
- Bioluminescence
- Virulence
- Upregulation of secretion system and toxins
- Antibiotics production
- Symbiosis
- Reduced Motility

Pectobacteria

- Gram-negative phytopathogens belonging to the Enterobacteriaceae family
- Cause soft rot in a wide range of food plants as well as ornamental crops

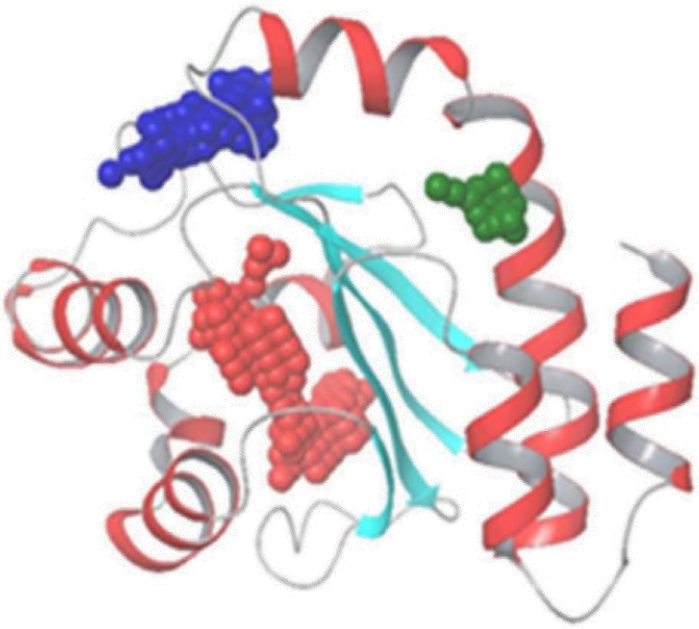
Rank	Bacterial pathogen	Author of bacterial description
1	*Pseudomonas syringae* pathovars	John Mansfield
2	*Ralstonia solanacearum*	Stéphane Genin
3	*Agrobacterium tumefaciens*	Shimpei Magori, Vitaly Citovsky
4	*Xanthomonas oryzae* pv. *oryzae*	Malinee Sriariyanum, Pamela Ronald
5	*Xanthomonas campestris* pathovars	Max Dow
6	*Xanthomonas axonopodis* pv. *manihotis*	Valérie Verdier
7	*Erwinia amylovora*	Steven V. Beer
8	*Xylella fastidiosa*	Marcos A. Machado
9	*Dickeya* (*dadantii* and *solani*)	Ian Toth
10	*Pectobacterium carotovorum* (and *P. atrosepticum*)	George Salmond

Symptoms: tissue maceration and decay, foul odor

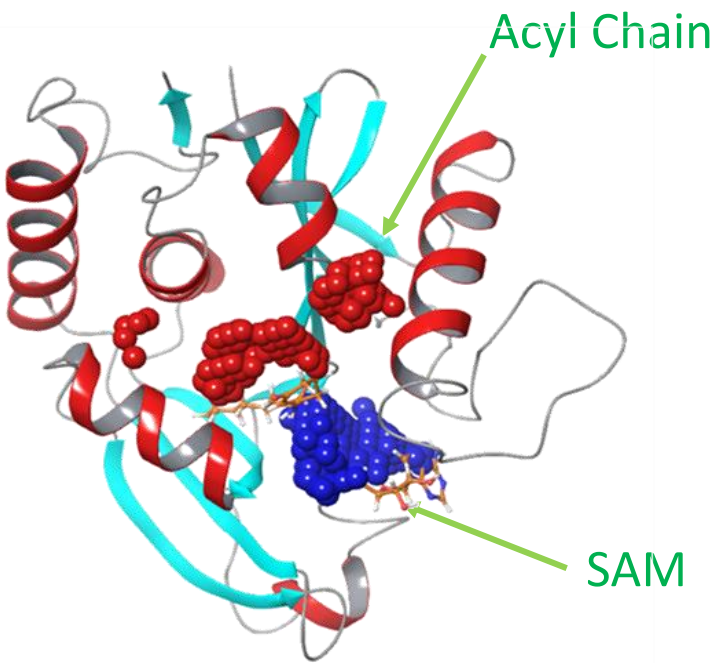
Quorum Sensing Proteins in *Pectobacterium*

- QS machinery in *Pectrobacteria* is composed from ExpI that synthesize the signaling molecule acyl-homoserine lactone (AHL) from S-adenosyl methionine (SAM) and acylated carrier protein and from ExpR that “reads” it

ExpR

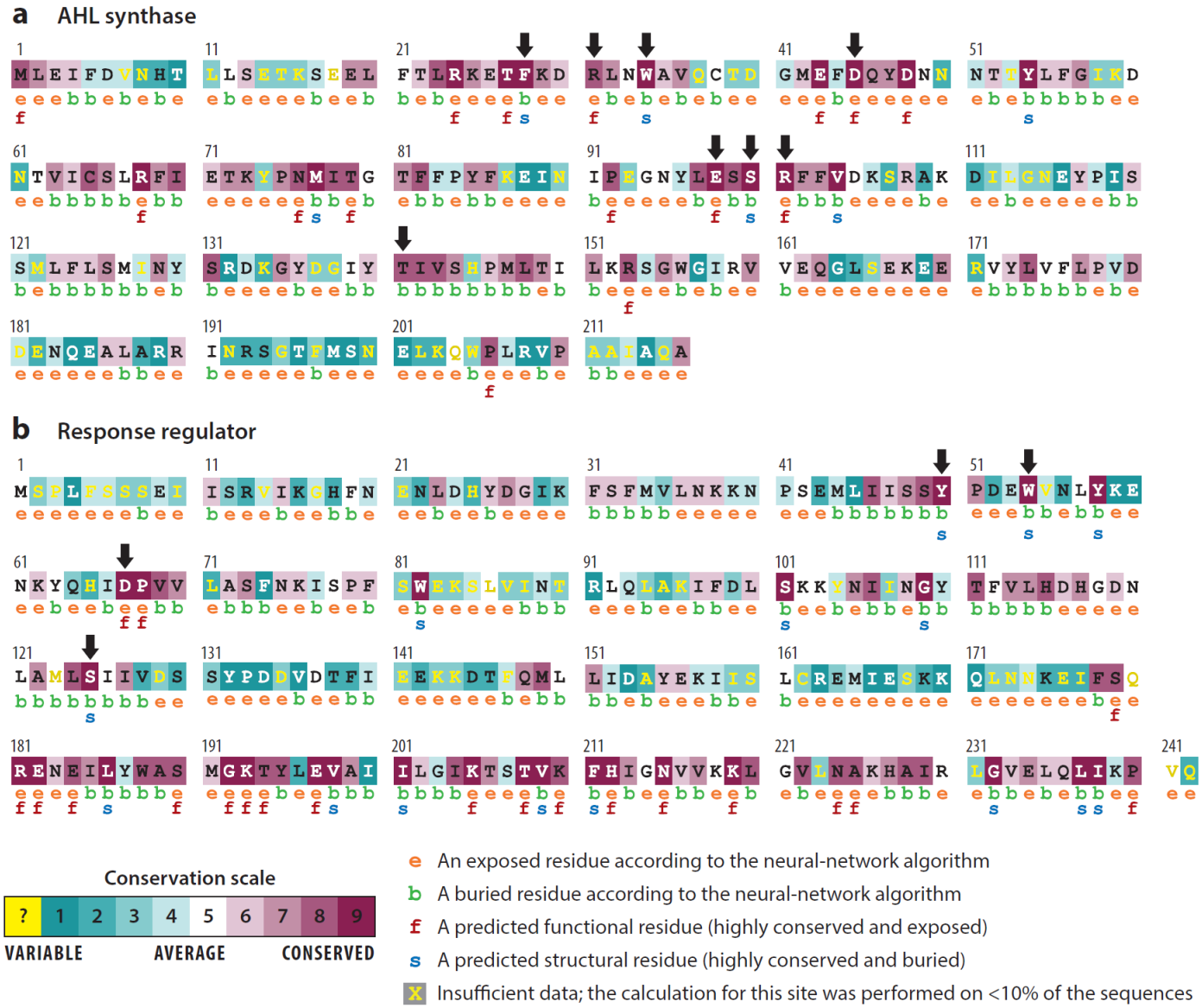


ExpI



Joshi et al., ACS Chemical Biology, 2020, 15, 7, 1883–1891, Joshi et al., Scientific Reports, 2016, 6, 38126

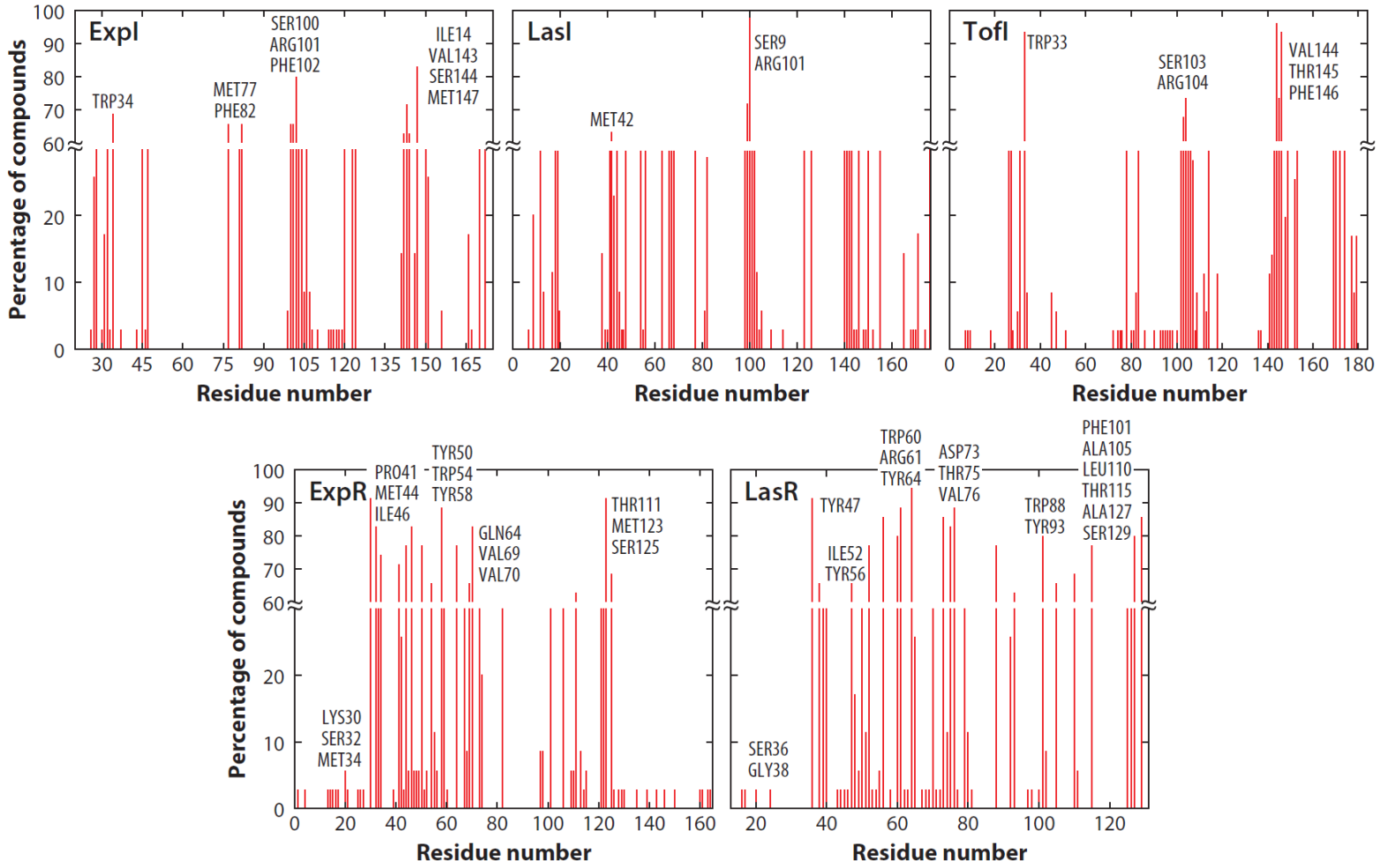
ConSurf Analysis of Quorum Sensing Proteins



Joshi et al., Annual Review of Phytopathology, 2021, 59, 153-190

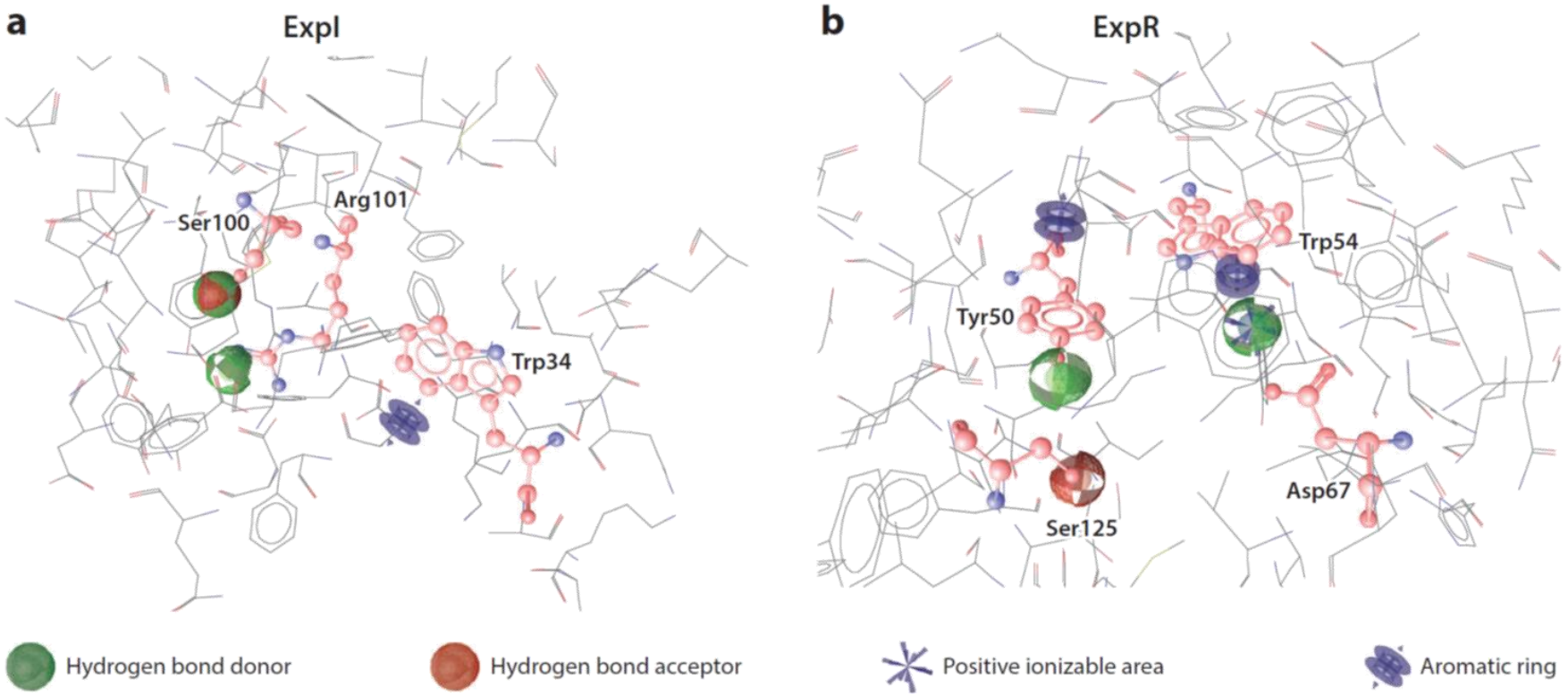
Structural Analysis of Quorum Sensing Complexes

- Based on the docking of 35 ligands known to affect bacterial QS machinery into 5 relevant crystal structures / homology models



Joshi et al., Annual Review of Phytopathology, 2021, 59, 153-190

Global Pharmacophore Models

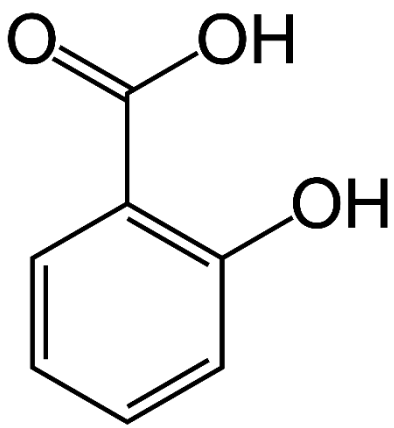


In *pectobacteria*, ExpI is a more relevant target for QS inhibition

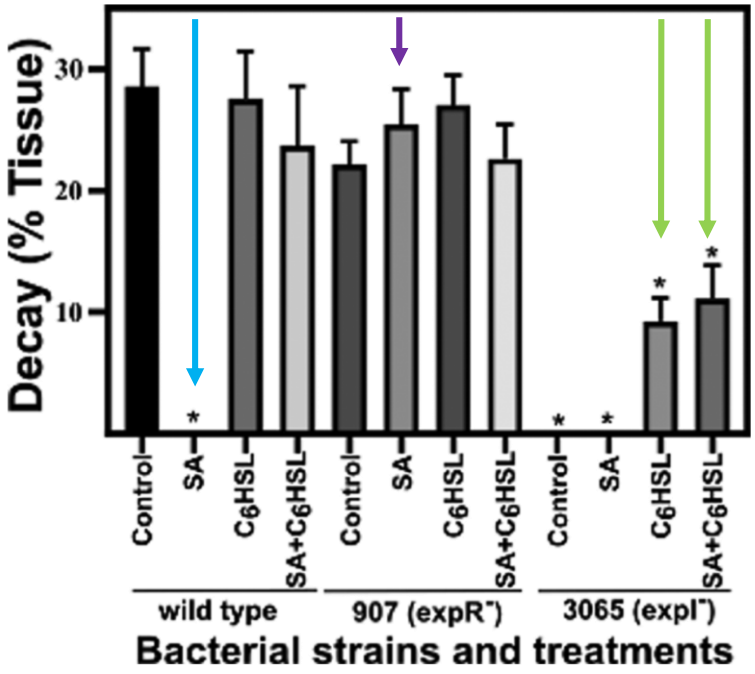
Joshi et al., Annual Review of Phytopathology, 2021, 59, 153-190

Binding of Salicylic Acid to ExpI in *Pectobacterium*

- Salicylic acid (SA) is a hormone that mediates systemic acquired resistance in plants
- Can SA interfere with QS by directly binding to ExpI?

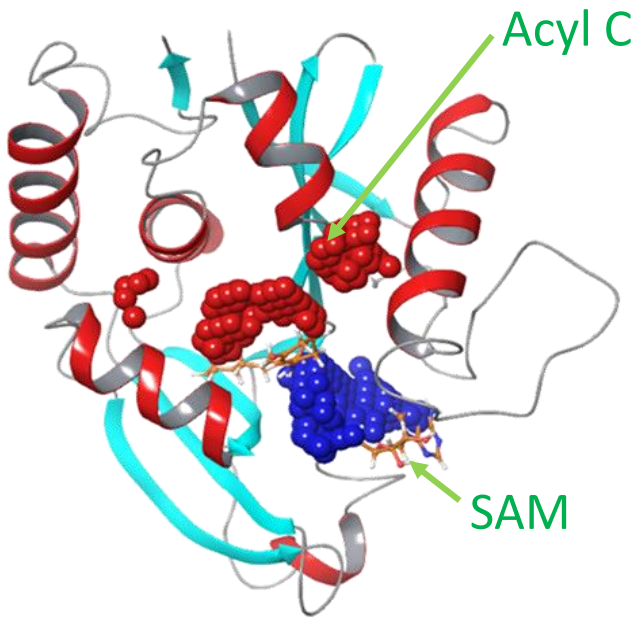


- SA reduced virulence of the WT construct
- Virulence of a mutant lacking ExpI was restored by exogenous AHL and was not abolished by addition of SA
- SA did not affect virulence of a mutant lacking ExpR
- SA operates via ExpI

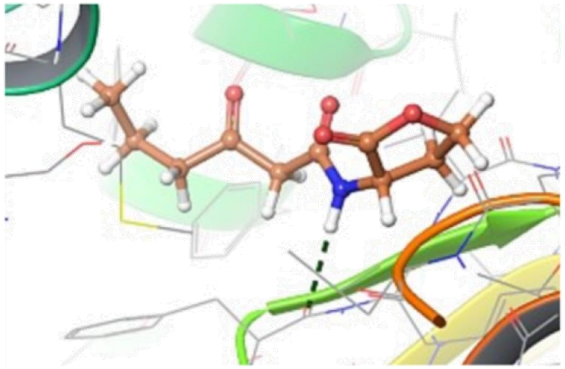


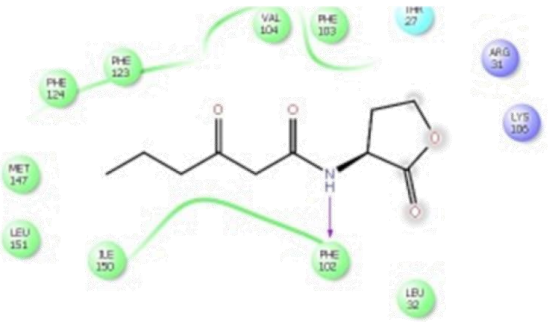
Joshi et al., ACS Chemical Biology, 2020, 15, 7, 1883–1891

Binding of Salicylic Acid to ExpI in *Pectobacterium*

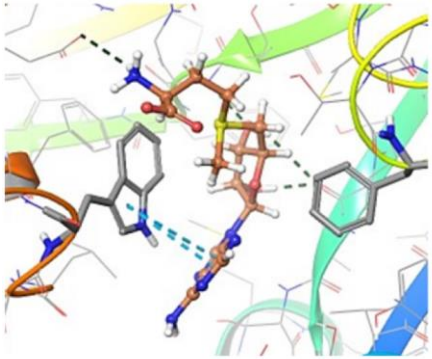


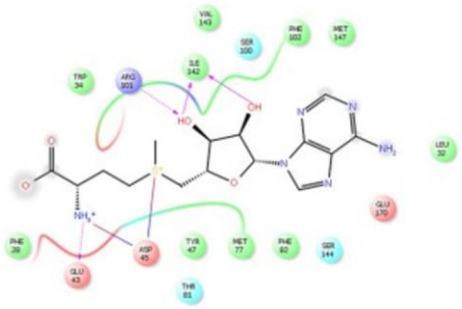
C_6HSL docks to the Acyl chain part of the site



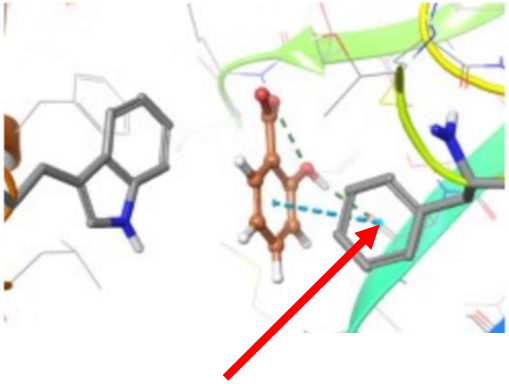


SAM docks to the SAM part of the site





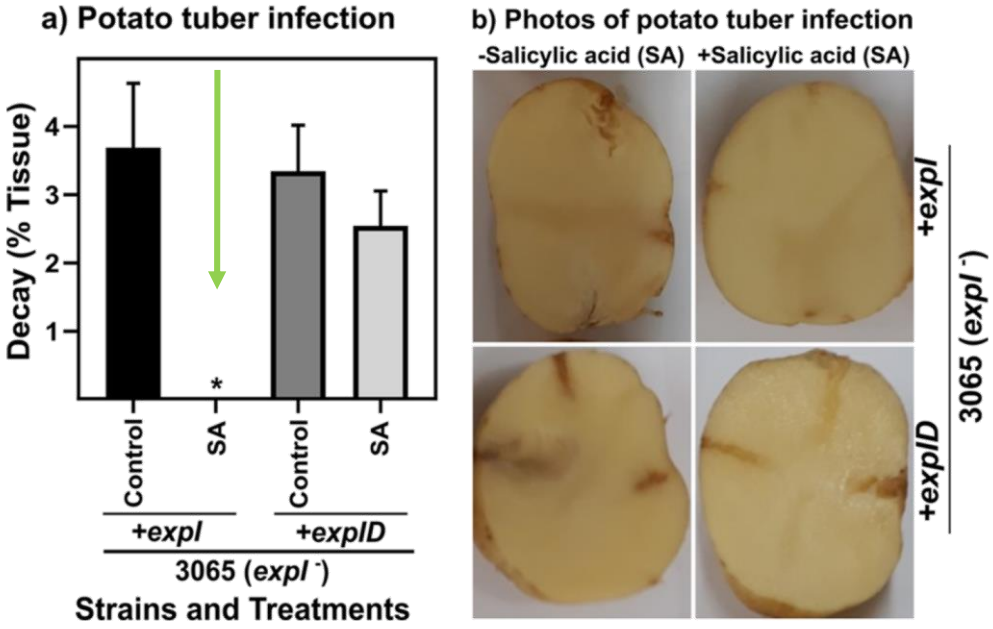
SA docks to the SAM part of the site



Binding of Salicylic Acid to ExpI in *Pectobacterium*

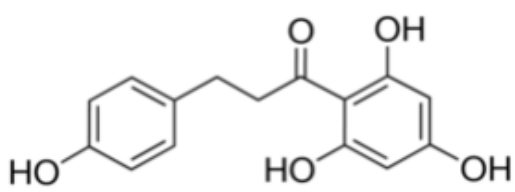
In Silico Designed

Protein/Ligand	Glide-XP (Kcal/mol)	ITC (Kcal/mol)
ExpI-C_6HSL	-6.4	-12.48±0.4
ExpI-SA	-5.3	-4.01±0.16
F82A-ExpI-SA	-4.3	-3.5±0.44

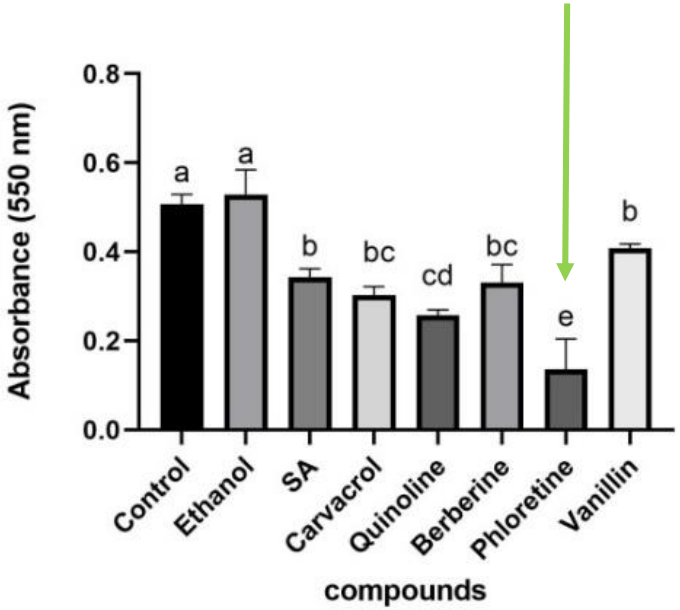


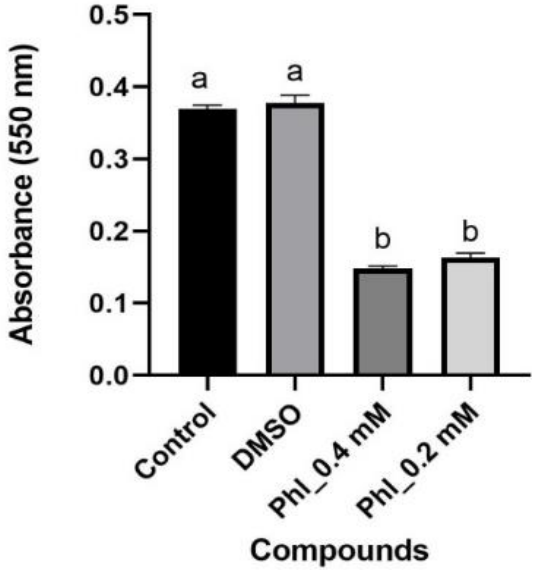
Phloretin Interferes with AHL Synthesis

- *Erwinia amylovora* is the cause of fire blight on apple and pear
- The phytoalexin phloretin accumulates in apple leaves in response to *E. amylovora*



Phloretin Interferes with Biofilm Formation



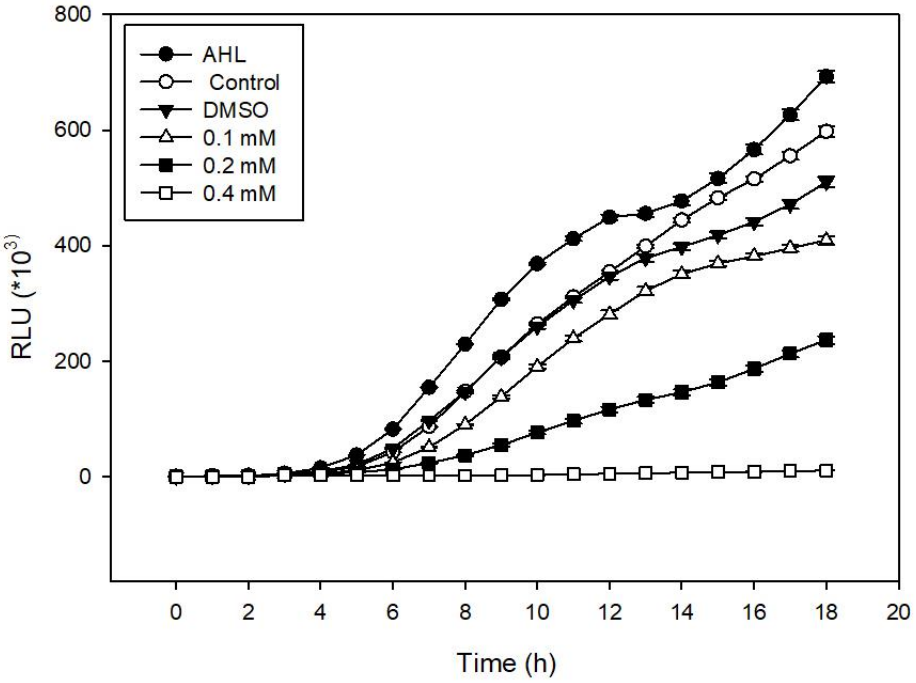


Pun et al., Frontiers in Plant Sciences, 2021, 12, 671807

Phloretin Interferes with AHL Synthesis

Phloretin Interferes with AHL synthesis

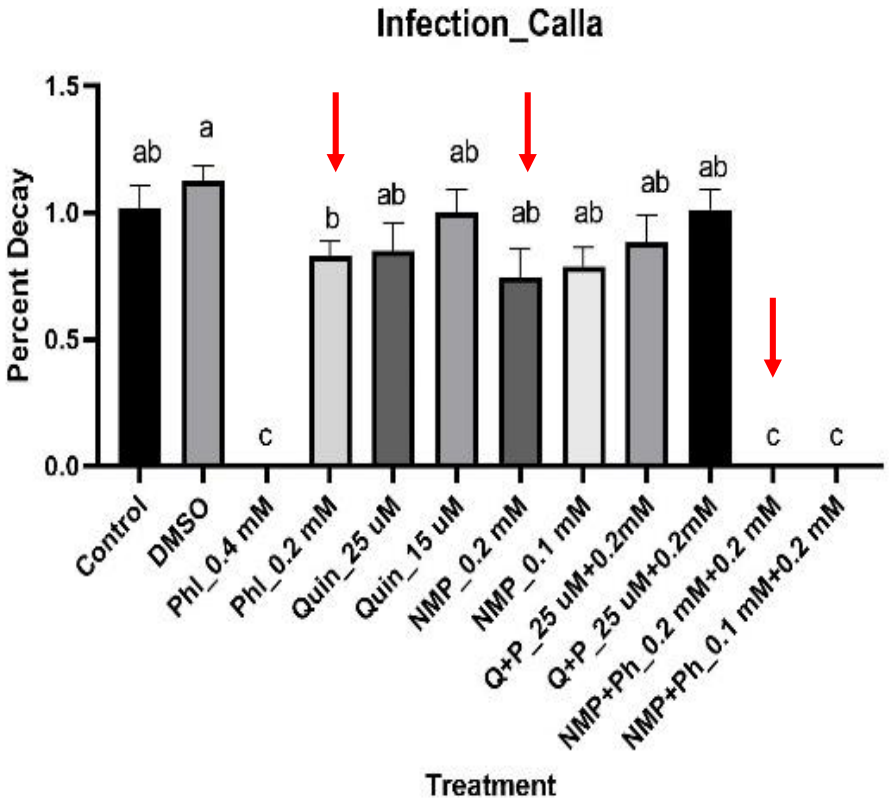
Luminescence assay

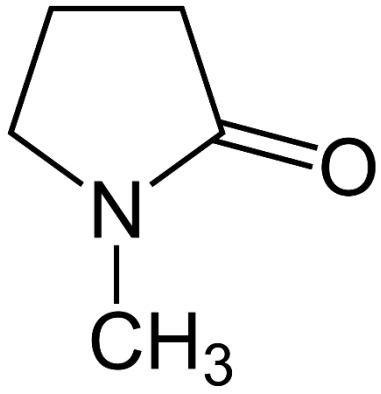


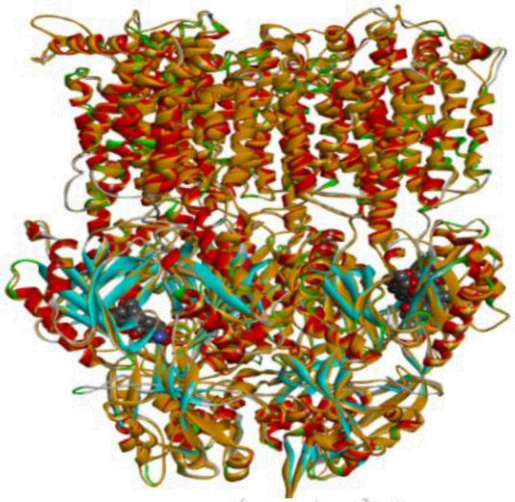
Ligand	Glide-XP score Expl kcal/mol
OC_6HSL	-6.4
SAM	-6.2
Salicylic acid	-5.3
Carvacrol	-6.2
Phloretin	-5.4

Phloretin is a Substrate of the AcrAB/TolC Efflux Pump

Simultaneous application of Phloretin and a inhibitor does wonders!



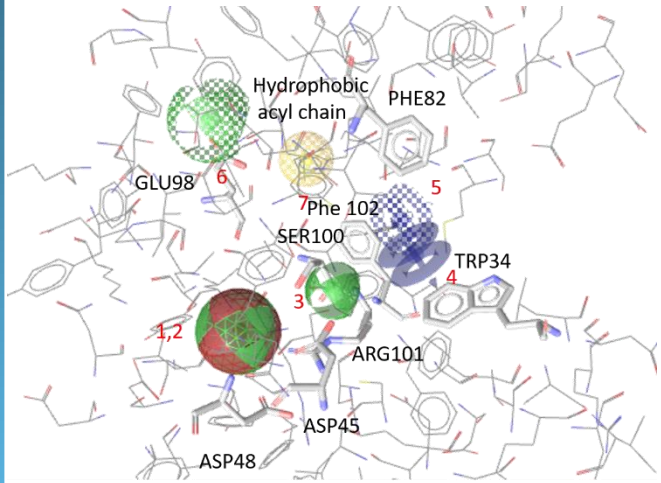




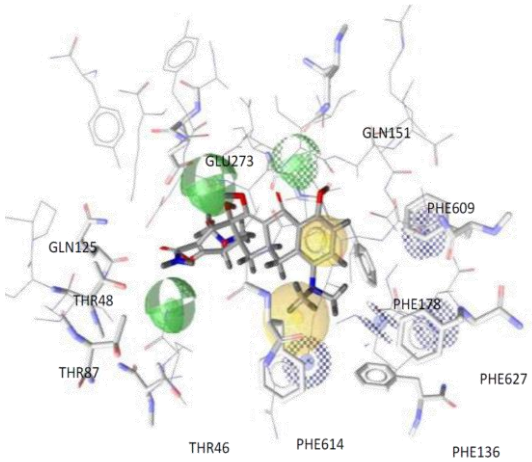
A Poly-Pharmacological Approach

Expi and AcrAB/TolC-1 are viable targets for virulence control

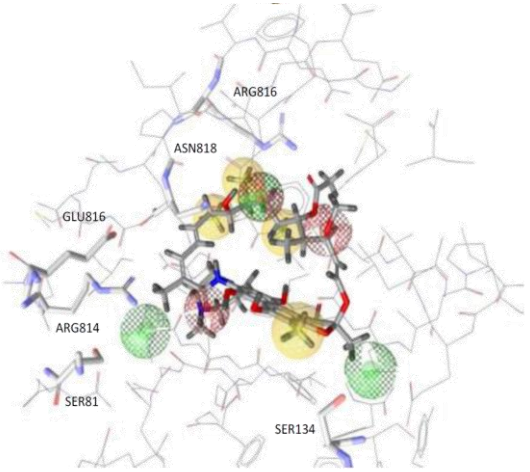
ExpI



AcrAB/TolC - 1



AcrAB/TolC - 2



Pharmacophore-Based VS
- ZINC (~13.5M)
- Enamine (~25M)
- MolPort (~8M)

→ Docking →

Analog Search
- ZINC (~13.5M)
- Enamine (~25M)
- MolPort (~8M)
- Enamine Real Space (~21B)

To date, this procedure has been applied to ExpI leading to several compounds with anti-virulence activity

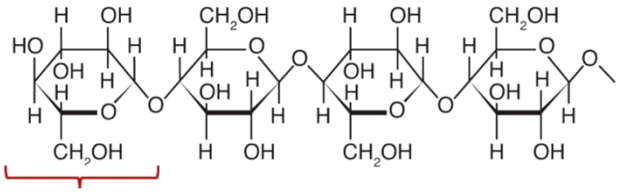
Oomycete

- Fungus-like eukaryotic microorganisms, some of which are severe crop pathogens
- *Phytophthora infestans*, the agent of potato late blight, was responsible for the Irish potato famine in the 19th century
- *Phytophthora capsici* attacks and rots pepper, cucumber, watermelon and tomato
- *Phytophthora ramorum* is responsible for sudden oak and larch death diseases in Europe and North America
- *Pythium ultimum* causes damping off and root rot on of vegetables and ornamental plants in nurseries and greenhouses
- *Plasmopora viticola* is the agent of grapevine downy mildew, a disease of high importance for viticulture globally

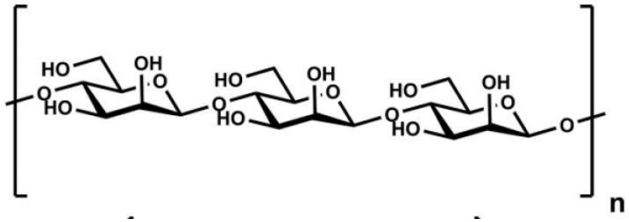
The Cell Wall of Oomycetes

- The cell wall of oomycetes is primarily composed of Cellulose, β-1,3 and β-1,6 glucans, and small amount of chitin in some species

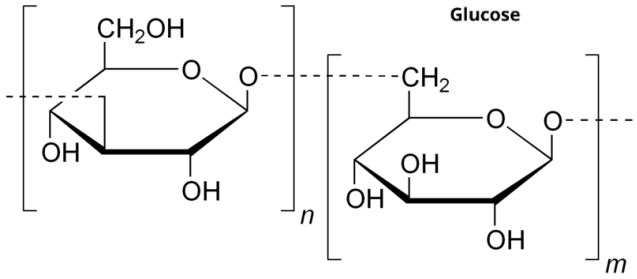
Cellulose



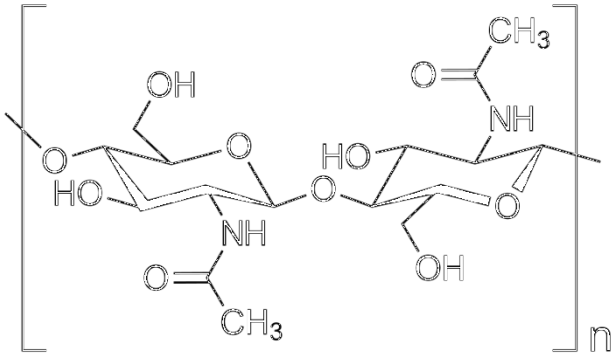
Mannan

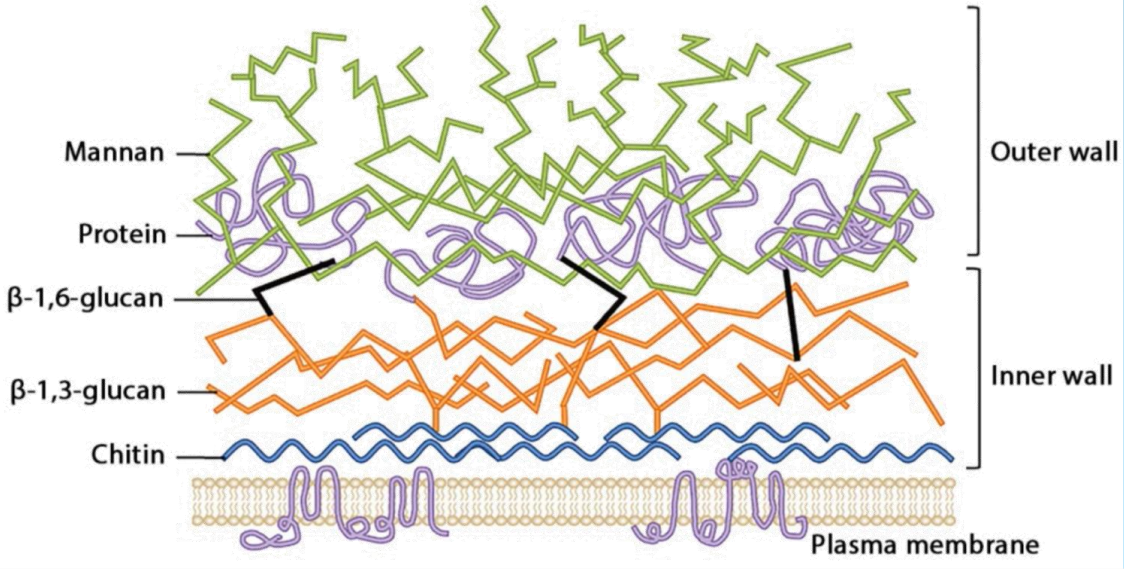


Glucan



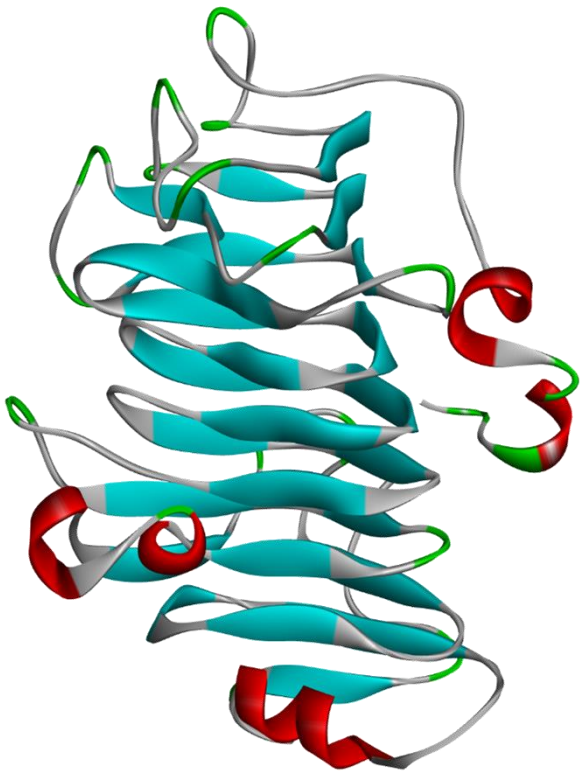
Chitin



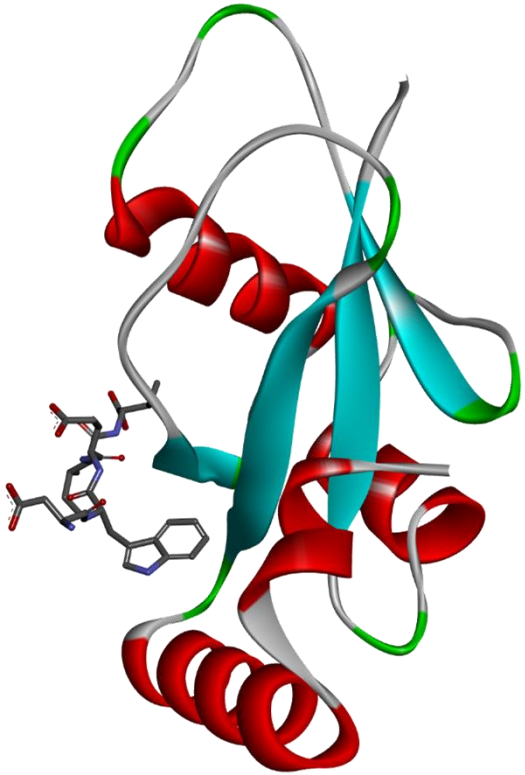


(Some) Proteins that Participate in Oomycetes Cell-Wall Construction

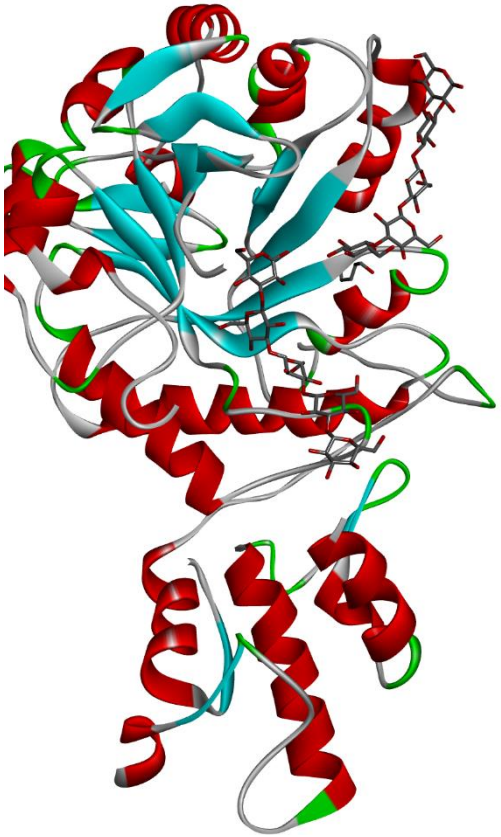
Pectinesterase



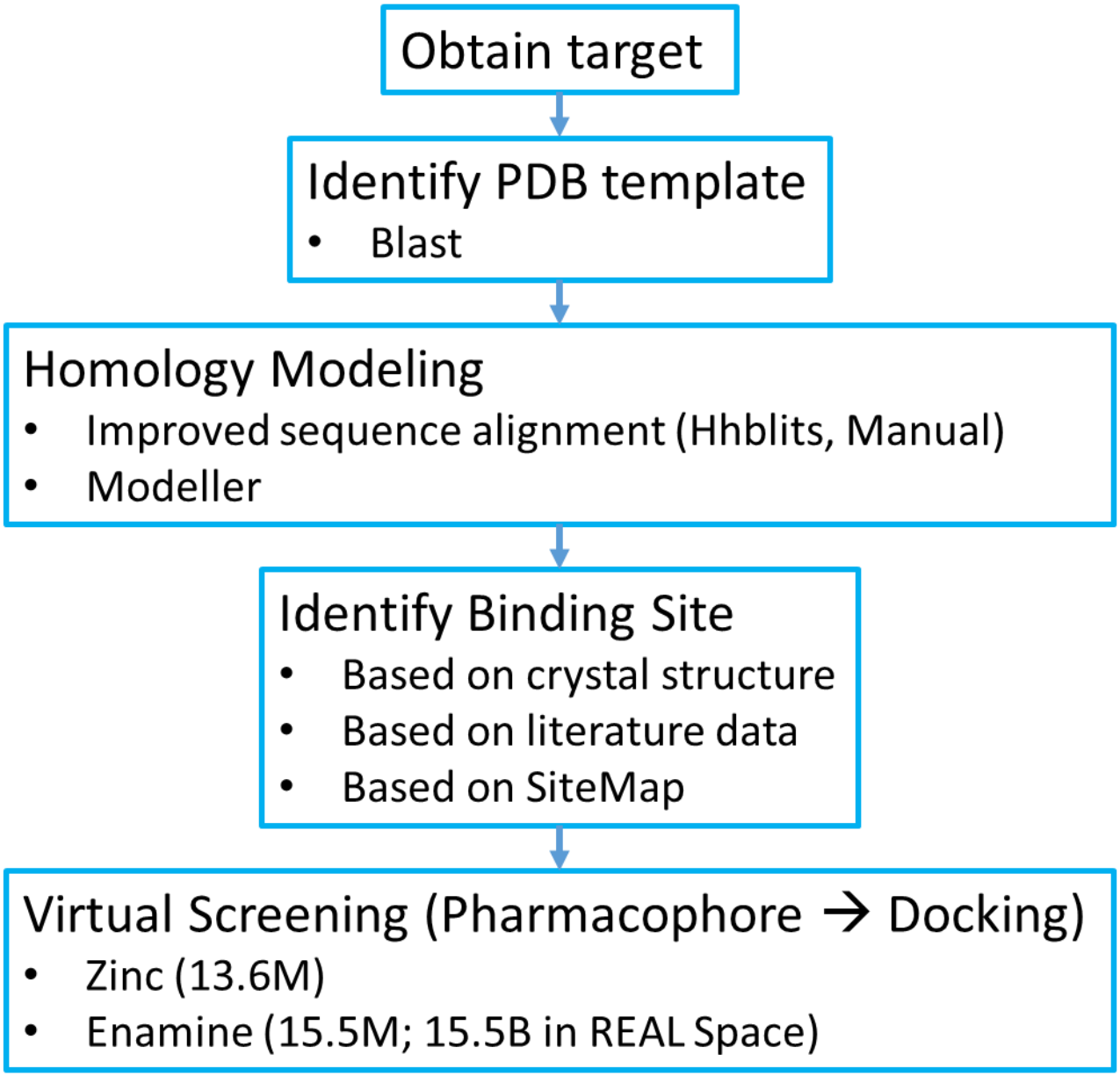
PexRD54-ATG8



1,3-beta-glucanosyltransferase

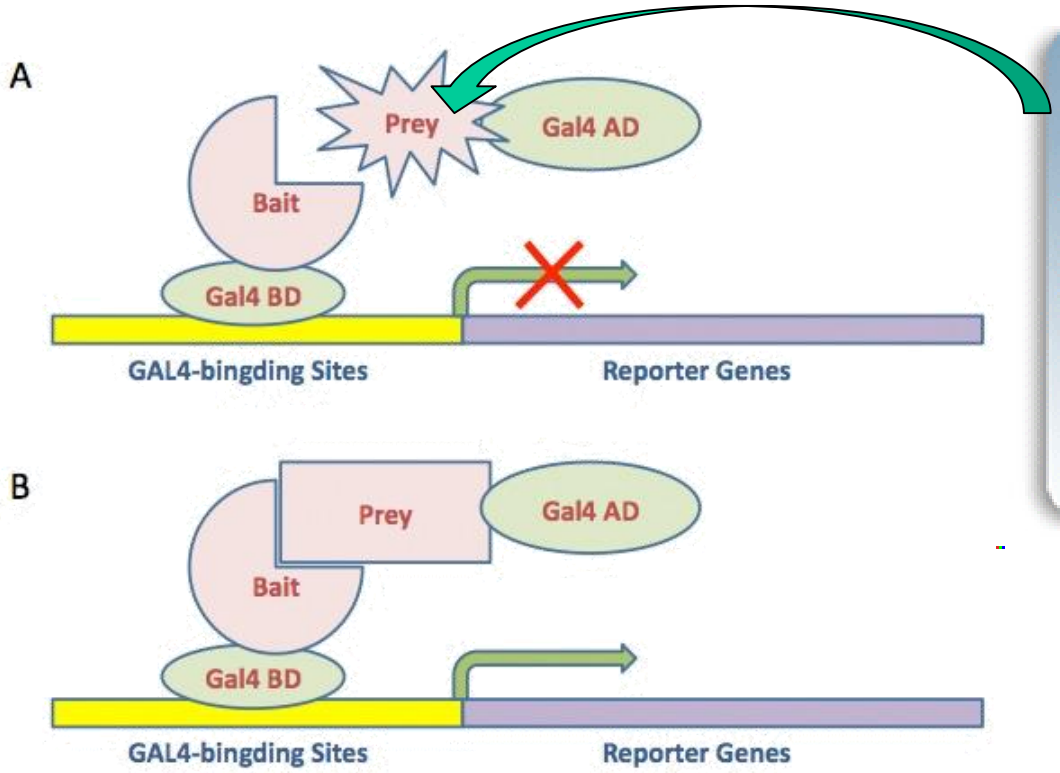


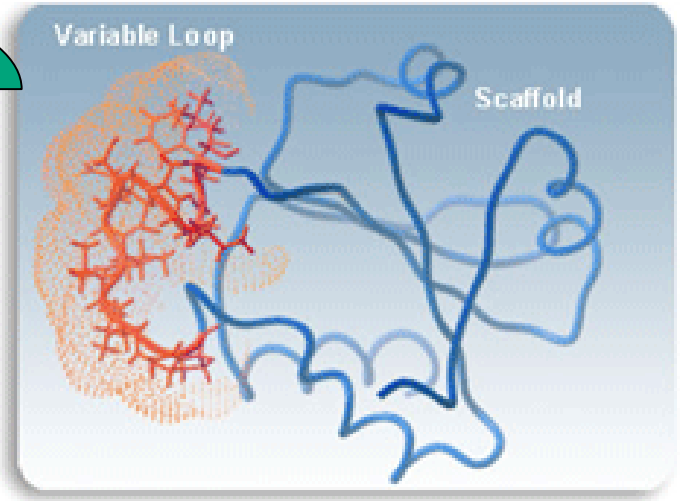
Modeling Workflow



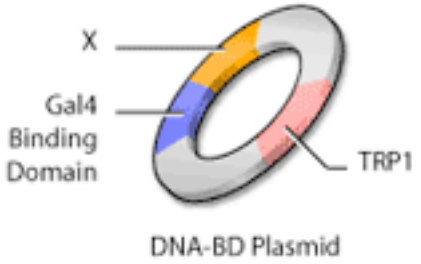
The Yeast-2 Hybrid System

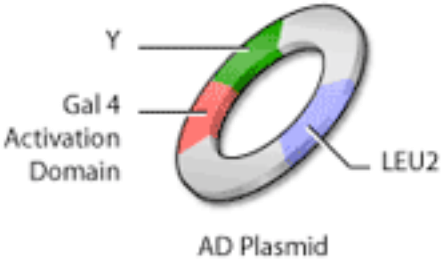
- Identify linear or cyclic peptide aptamers that inhibit surfaced exposed, vital enzymes involved in oomycete cell-wall formation and cell stability
- Upon binding of the Prey (peptide) to the Bait (target), the two components of the Gal4 transcription factor come together, a reporter gene is activated and an appropriate readout is made possible

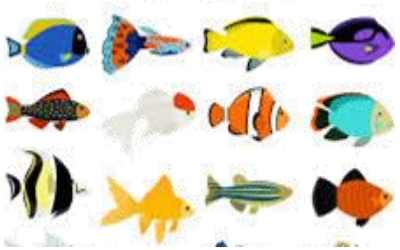




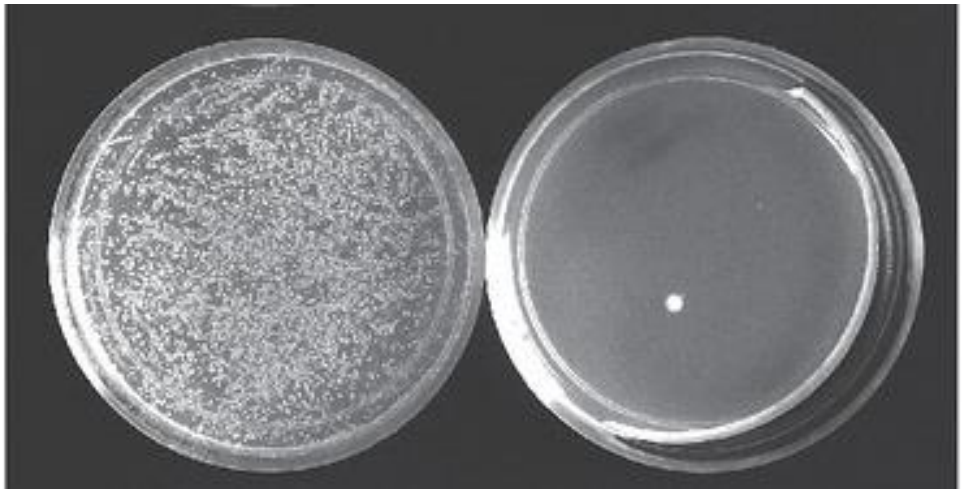
The Yeast-2 Hybrid System







How to find the interactor



Challenges in Modeling the Data

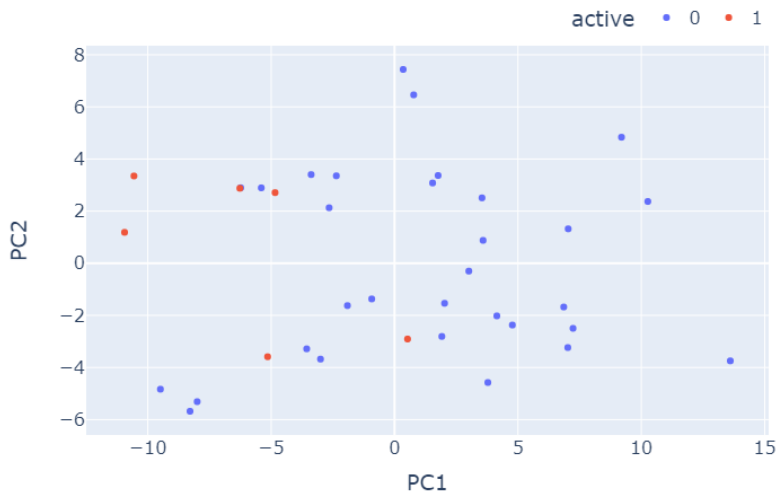
- HTS data are
  - Noisy (FP, FN)
  - Imbalanced (More inactives than actives)
  - Represent multiple MOA
- And for peptides
  - Global vs. AA-based descriptors
  - 2D vs. 3D descriptors
  - Sequence dependent descriptors
- And for this dataset
  - Overall small number of peptides
  - Actives and inactives unseperable
  - Sparse coverage of descriptors space
- Which means:
  - Classification (RF)

The Data

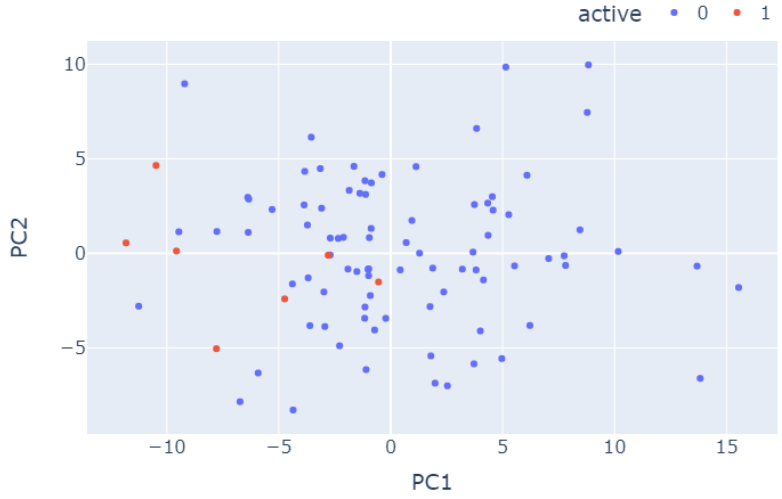
Dataset	No. Actives	No. Inactives (1)	No. of Inactives (2)
PiEPIC2B	42	61	61+40+30+12=143
PiAVR3a	40	61	61+42+30+12=145
PvCesA2	30	61	61+42+40+12=155
AtRGL2	12	61	61+42+40+30=173
Total	124	61	

The Data

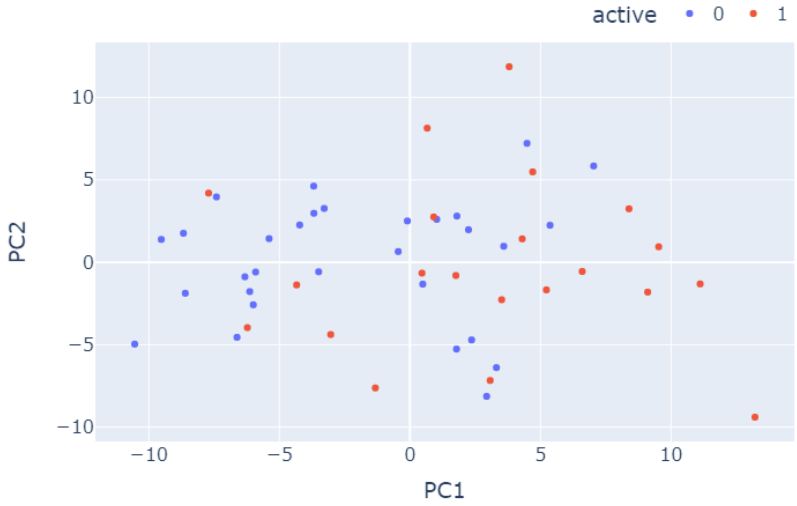
AtRGL2 Training Sets:
VS Completely Inactive



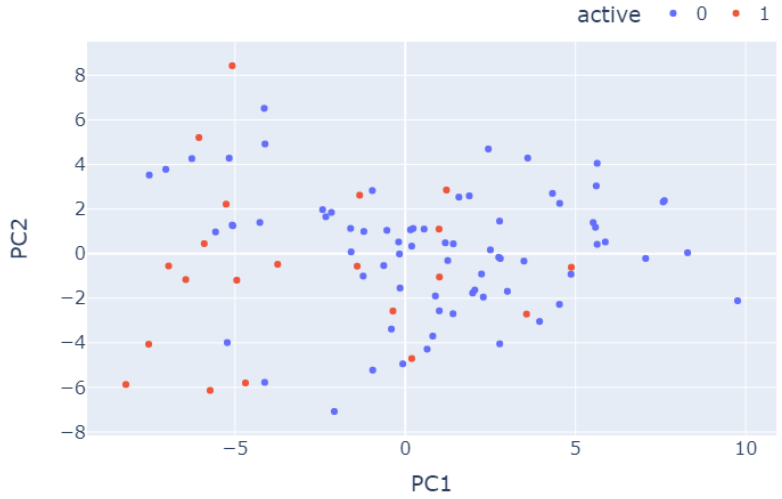
VS Inactive and Other Actives



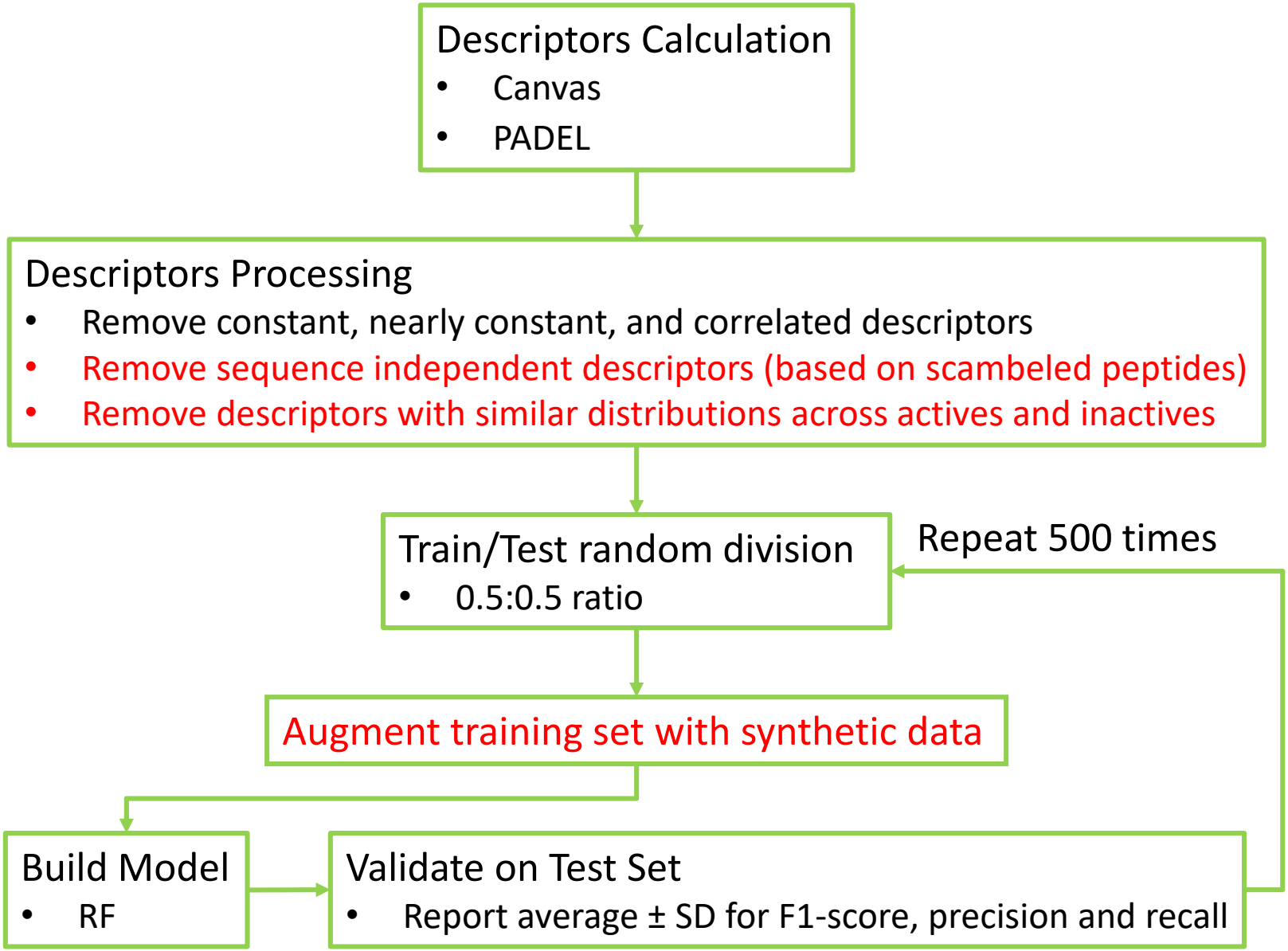
PiEPIC2B Training Sets
VS Completely Inactive



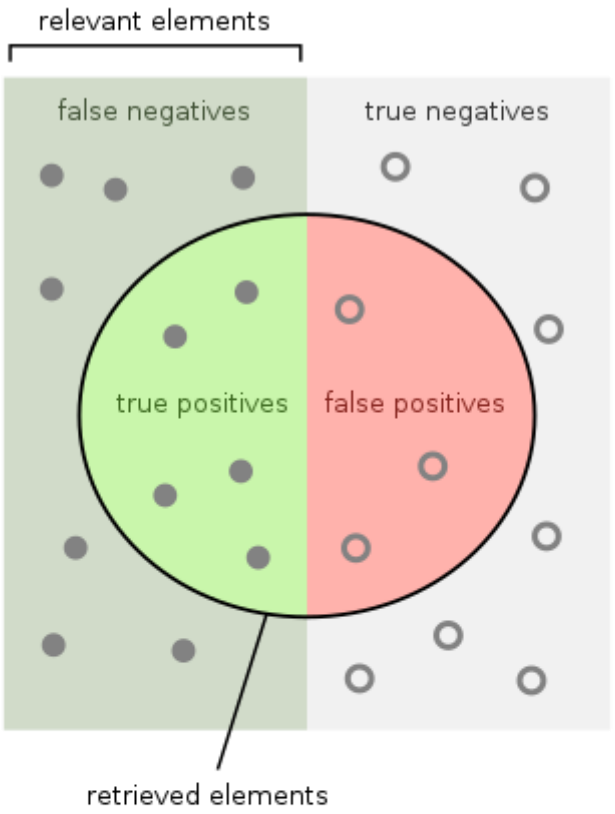
VS Inactive and Other Actives



Computational Workflow



Definitions

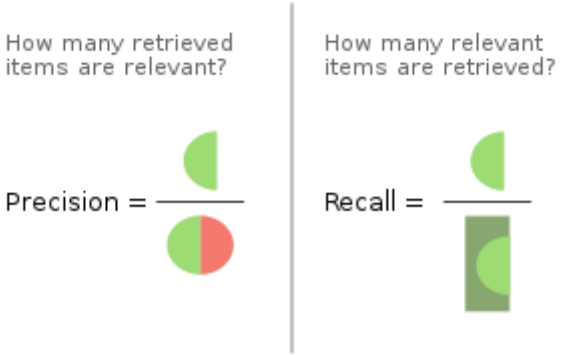


- Precision (PPV): The fraction of relevant instances among the retrieved instances

$$Precision(A) = \frac{Samples\ in\ class\ A \cap Samples\ predicted\ as\ class\ A}{All\ samples\ perdicted\ as\ class\ A} = \frac{TP}{TP + FP}$$

- Recall (sensitivity): The fraction of relevant instances that were retrieved

$$Recall(A) = \frac{Samples\ in\ class\ A \cap Samples\ predicted\ as\ class\ A}{All\ samples\ actually\ in\ class\ A} = \frac{TP}{TP + FN}$$



- F1-score: Harmonic mean of precision and recall

$$F1 = \left(\frac{Precision * Recall}{Precision + Recall}\right) * 2$$

Computational Workflow

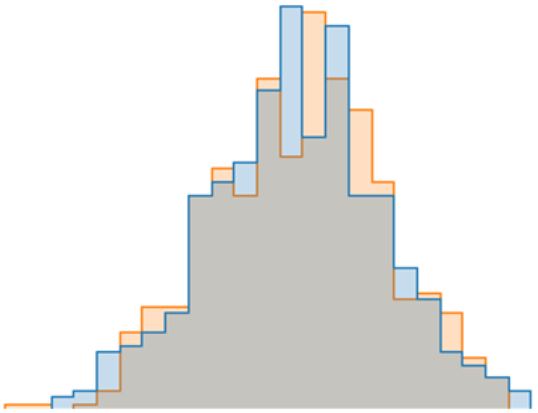
Peptide	Desc1	Desc2	...	Desc100	active
0	0.44	347	...	1.12	0
1	0.97	500	...	4.15	1
2	0.12	783	...	2.14	1
3	0.36	245	...	0.89	0
4	0.88	108	...	3.45	0
5	0.20	790	...	1.09	1

Peptide	Desc1	active
1	0.97	1
2	0.12	1
5	0.20	1

Peptide	Desc1	active
0	0.44	0
3	0.36	0
4	0.88	0

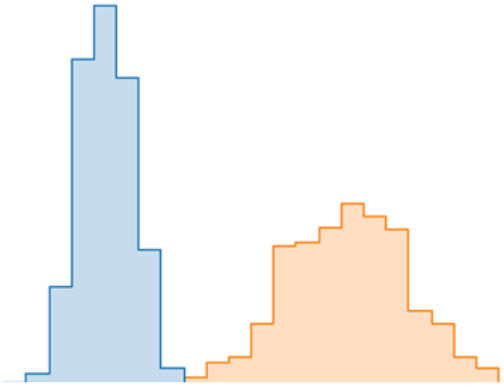
If p-value < 0.05:

Discard Desc1



Else:

Keep Desc1



Computational Workflow

Reduced set

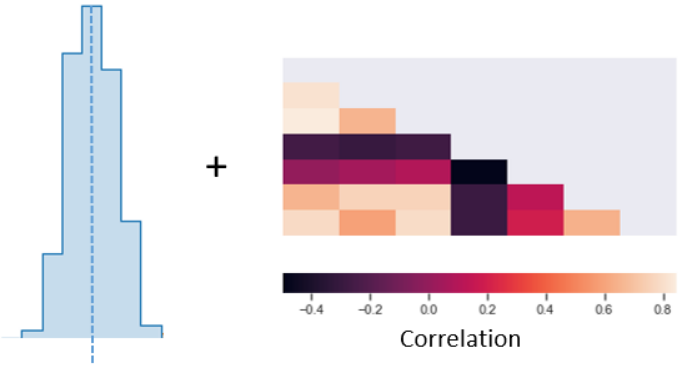
Peptide	Desc2	Desc7	...	Desc50	active
0	347	0.314	...	18.0	0
1	500	0.00	...	40.7	1
2	783	0.00	...	35.6	1
3	245	0.267	...	20.0	0
4	108	0.108	...	22.7	0
5	790	0.01	...	45.3	1

mean

covariance matrix

Random sampling

+



Peptide	Desc2	Desc7	...	Desc50	active
1	500	0.00	...	40.7	1
2	783	0.00	...	35.6	1
5	790	0.01	...	45.3	1

Generate Synthetic Data

SYN_0	657	0.012	...	37	1

Repeat for all descriptors

Repeat until:
No. Actives = No. Inactives

Overall Results: Set vs. Neg.

Dataset	Data Source	F1-score ± SD	Precision ± SD	Recall ± SD
PiEPIC2B (42/61)	Original	0.66 ± 0.05	0.67 ± 0.05	0.64 ± 0.06
	Original + Synthetic	0.64 ± 0.05	0.65 ± 0.06	0.65 ± 0.06
PiAVR3a (40/61)	Original	0.64 ± 0.05	0.66 ± 0.06	0.66 ± 0.05
	Original + Synthetic	0.63 ± 0.06	0.64 ± 0.06	0.64 ± 0.06
PvCesA2 (30/61)	Original	0.71 ± 0.05	0.72 ± 0.06	0.73 ± 0.05
	Original + Synthetic	0.70 ± 0.06	0.71 ± 0.06	0.70 ± 0.06
AtRGL2 (12/61)	Original	0.82 ± 0.05	0.84 ± 0.07	0.85 ± 0.04
	Original + Synthetic	0.83 ± 0.04	0.84 ± 0.06	0.84 ± 0.06
All (124/61)	Original	0.67 ± 0.04	0.68 ± 0.04	0.70 ± 0.04
	Original + Synthetic	0.67 ± 0.04	0.68 ± 0.04	0.66 ± 0.05

- Reasonably good models
- No large differences between models based on the original data and models based on the original + synthetic data

Overall Results: Set vs. Neg. for Actives

Dataset	Data Source	F1-score ± SD	Precision ± SD	Recall ± SD
PiEPIC2B (42/61)	Original	0.54 ± 0.08	0.62 ± 0.09	0.49 ± 0.10
	Original + Synthetic	0.55 ± 0.07	0.57 ± 0.08	0.56 ± 0.11
PiAVR3a (40/61)	Original	0.49 ± 0.09	0.60 ± 0.11	0.43 ± 0.10
	Original + Synthetic	0.52 ± 0.08	0.54 ± 0.09	0.51 ± 0.11
PvCesA2 (30/61)	Original	0.51 ± 0.11	0.65 ± 0.14	0.43 ± 0.12
	Original + Synthetic	0.55 ± 0.09	0.55 ± 0.10	0.58 ± 0.13
AtRGL2 (12/61)	Original	0.36 ± 0.20	0.65 ± 0.34	0.28 ± 0.18
	Original + Synthetic	0.52 ± 0.15	0.56 ± 0.18	0.53 ± 0.19
All (124/61)	Original	0.80 ± 0.03	0.73 ± 0.02	0.88 ± 0.05
	Original + Synthetic	0.74 ± 0.04	0.77 ± 0.03	0.72 ± 0.07

- F1 increases
- Precision decreases
- Recall increases

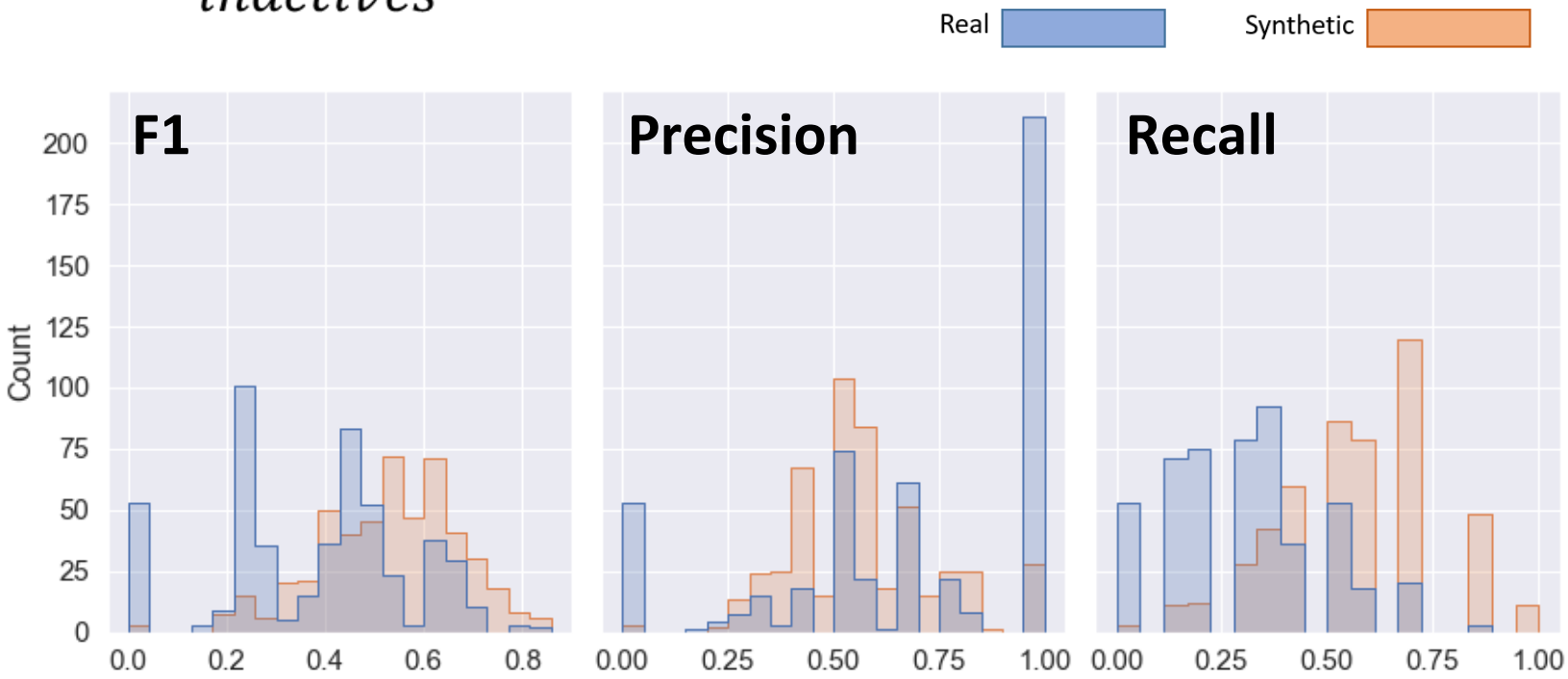
Overall Results: Set vs. Neg. for Inactives

Dataset	Data Source	F1-score ± SD	Precision ± SD	Recall ± SD
PiEPIC2B (42/61)	Original	0.74 ± 0.05	0.70 ± 0.04	0.79 ± 0.09
	Original + Synthetic	0.70 ± 0.05	0.70 ± 0.05	0.71 ± 0.09
PiAVR3a (40/61)	Original	0.74 ± 0.05	0.69 ± 0.04	0.81 ± 0.08
	Original + Synthetic	0.70 ± 0.06	0.70 ± 0.05	0.71 ± 0.10
PvCesA2 (30/61)	Original	0.81 ± 0.04	0.76 ± 0.04	0.88 ± 0.07
	Original + Synthetic	0.77 ± 0.06	0.79 ± 0.05	0.76 ± 0.10
AtRGL2 (12/61)	Original	0.92 ± 0.02	0.87 ± 0.03	0.97 ± 0.04
	Original + Synthetic	0.90 ± 0.03	0.90 ± 0.04	0.90 ± 0.06
All (124/61)	Original	0.43 ± 0.07	0.58 ± 0.10	0.35 ± 0.08
	Original + Synthetic	0.52 ± 0.06	0.49 ± 0.05	0.57 ± 0.10

- F1 decreases
- Precision increases
- Recall decreases

AtRGL2 Results: Set vs. Neg.

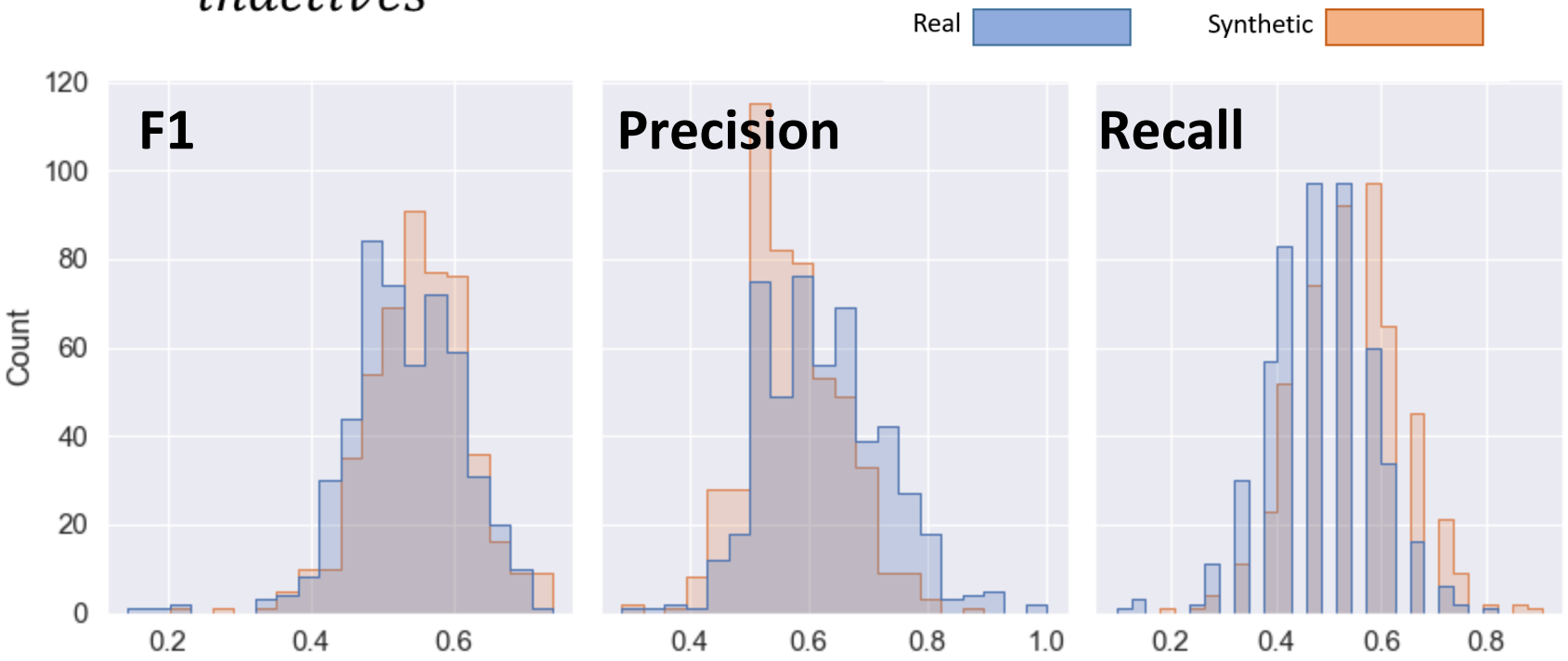
Where $\frac{actives}{inactives}$ is small (<<1)



- Increase in **active** F1 score
- Large gain in Recall
- Smaller cost in Precision

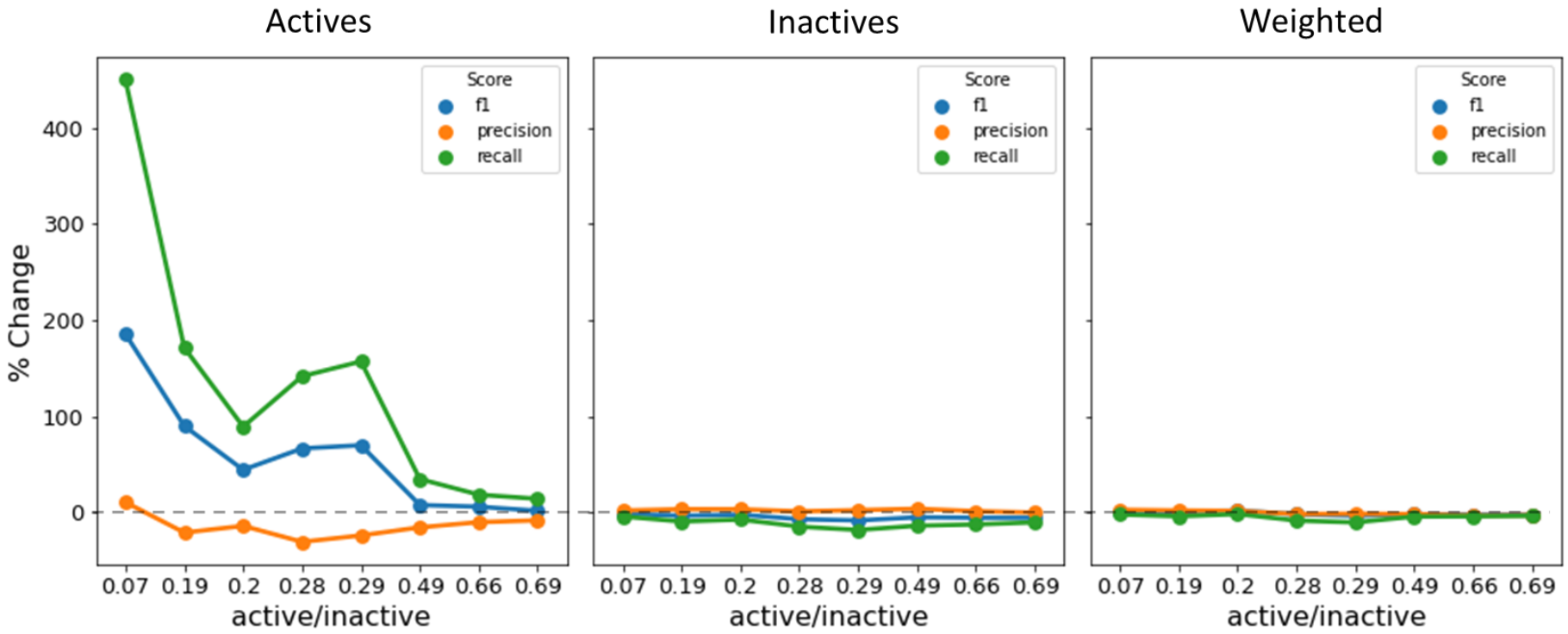
PiEPIC2B Results: Set vs. Neg.

where $\frac{actives}{inactives}$ is close to 1



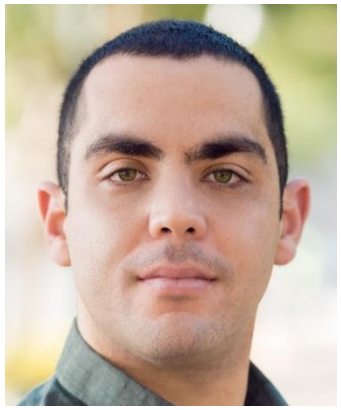
- Small increase in active F1 score
- Gain in recall is reduced
- Same cost to precision

Summary



- Synthetic data lead to more stable F1-score and precision for actives
- The best improvement in model performance upon adding synthetic data is obtained for active compounds when the datasets are the most biased
- Lack of overall improvement attributable to data inseparability
- Hypothesis: Workflow will work for biased yet separable sets

Acknowledgments



Omer Kaspi

Netaly Khazanov

Lina Iktelat

Shahaf Kozokaro

Paul Clarke

Collaborators

- Iris Yedidia
- Paolo Pesaresi
- Simona Masiero
- Vincent Bulone
- Vaibhav Srivastava

Funding

- EU
- BARD Foundation
- Israel Ministry of Agriculture