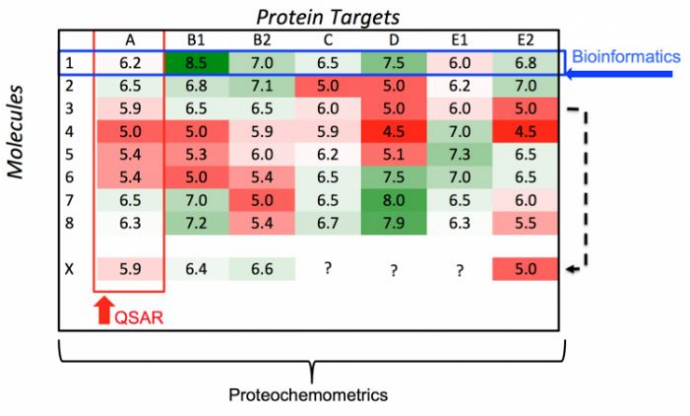
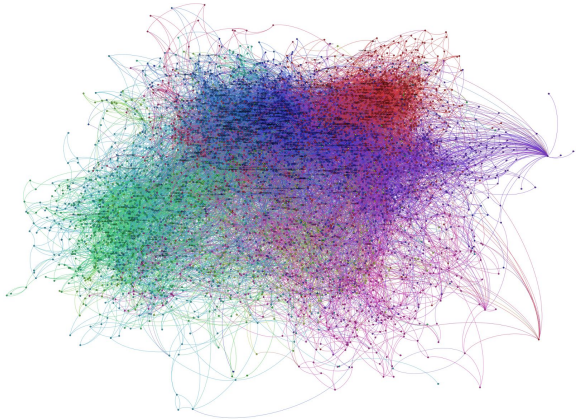
Proteochemometrics and Systems Biology for Drug Repositioning

Fraunhofer SCAI-Bio
September 18th, 2018
Charles Tapley Hoyt

What do Chemicals do in Biological Systems?





Proteochemometrics

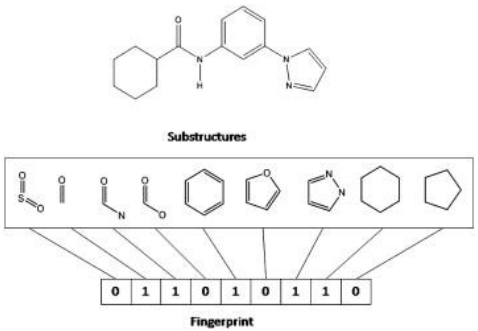
Which targets (proteins) do chemicals modulate?

Systems Biology

What are the downstream effects of proteins?

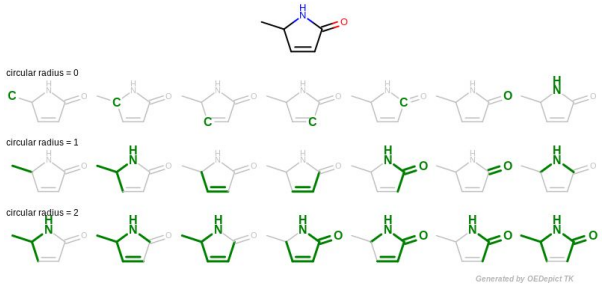
Image from https://www.universiteitleiden.nl/en/research/research-projects/science/lacdr-proteochemometrics

Feature Engineering for Chemicals



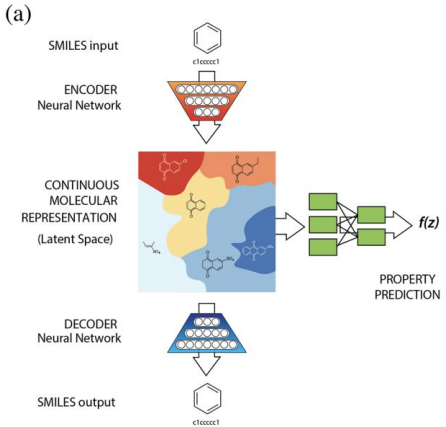
Structural fingerprints (e.g., MACCS)

https://doi.org/10.1016/j.bbapap.2013.05.010



Circular fingerprints (e.g., ECFP)

https://docs.eyesopen.com/toolkits/python/graphsimtk/fingerprint.html



Representation learning (variational autoencoders with CNN or RNN)

https://pubs.acs.org/doi/pdf/10.1021/acscentsci.7b00572

Proteochemometrics

- ~11K known drugs over ~5K targets
- Millions of drug-like molecules that have been tested over ~10K targets
- 10^{20}-10^{30} possible drug like molecules and ~20K possible targets

Data

- Sparse (chemical *x* protein) matrix
- Sequence-based chemical-chemical and protein-protein similarities

Multi-task learning used for regression task of predicting new chemical-protein assay values

Systems Biology and Drug Repurposing

- Networks are commonly used to represent biological systems in Systems biology
- Use predictions for chemicals-target values to calculate aggregate downstream effects via biological networks (e.g., heat diffusion)
- Identify drugs or combinations of drugs exerting perturbations desirable in therapeutic context by connecting proteins to biological processes and higher-order phenotypes