Optimization of QSAR Models for Virtual Screening

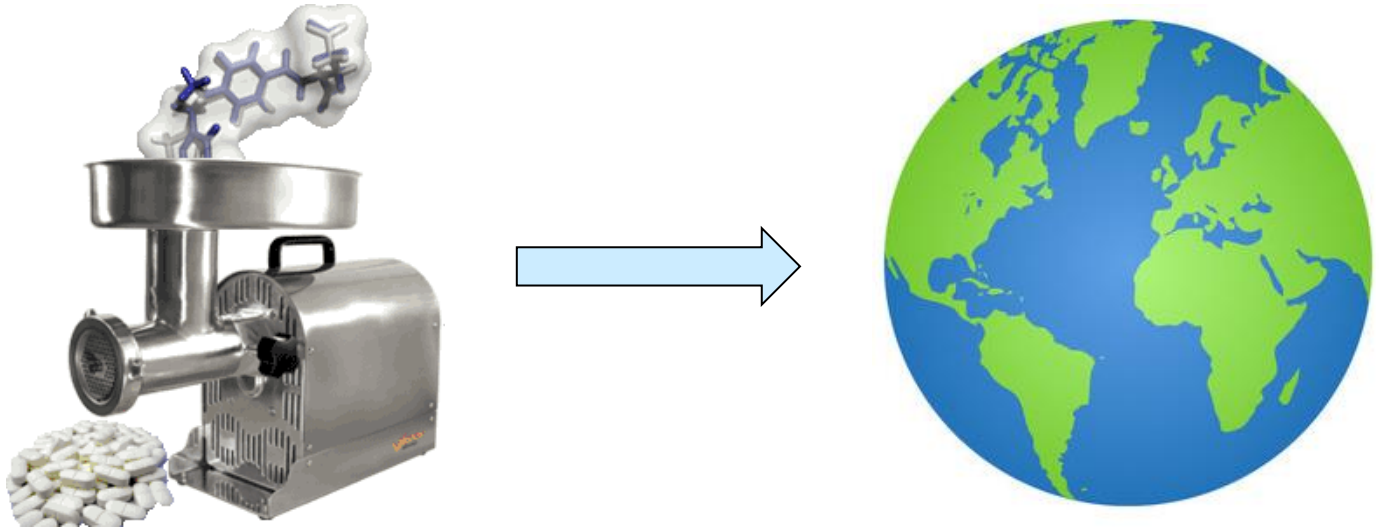
Jacob Spiegel and Hanoch Senderowitz

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

XXIV Symposium on Bioinformatics and Computer-Aided Drug Discovery
Moscow, September 2023

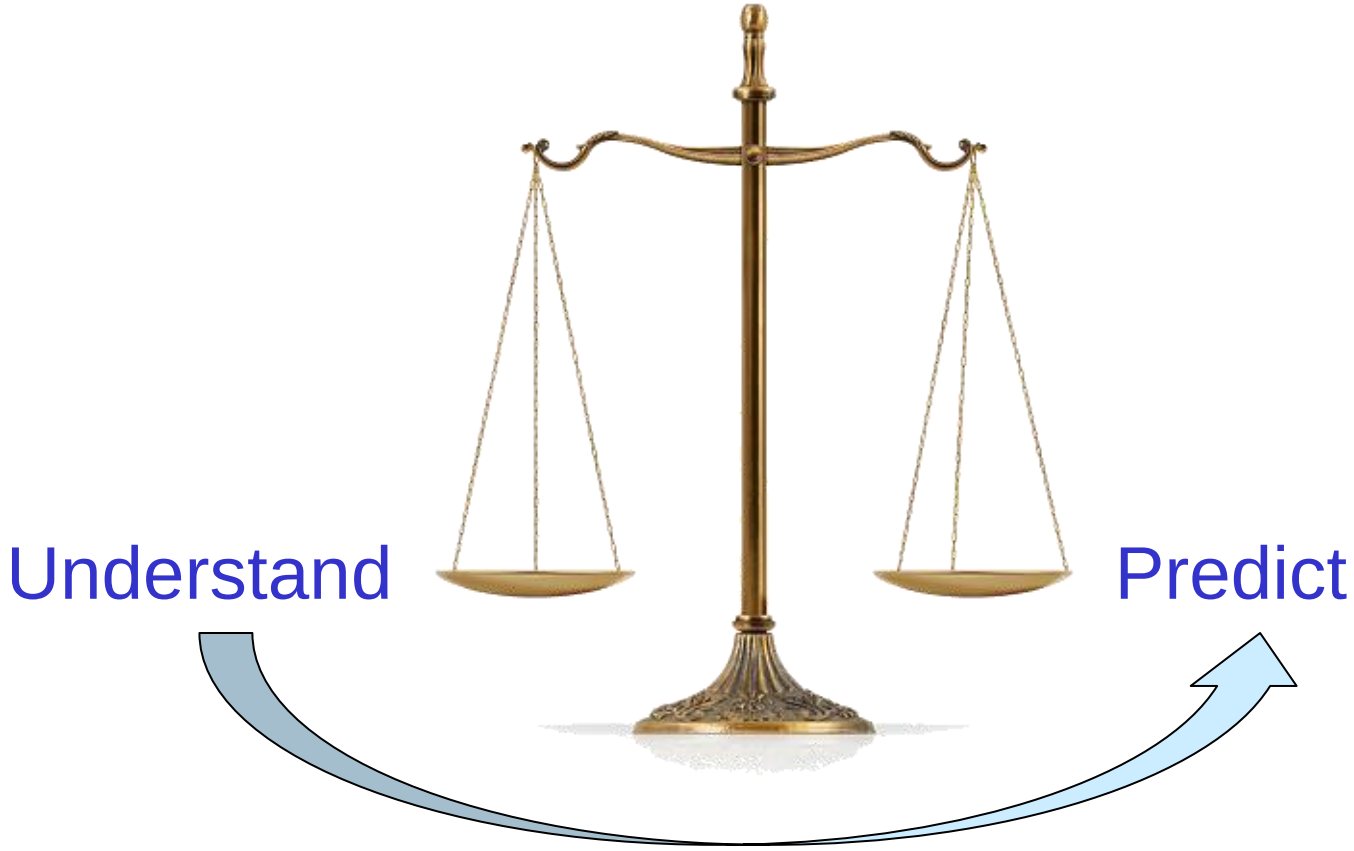
QSAR

- Correlates specific biological activity for a set of compounds with their structure-derived molecular descriptors by means of a mathematical model
- Nature of correlation, activity and descriptors are unlimited
  - BBB permeability = *f (hydrophobicity, H-bonding potential)*
  - Metabolic stability = *f (presence/absence of specific fragments)*
  - Protein crystallizability = *f (amino acid composition, secondary structure)*
- QSAR models are derived using Machine Learning (ML) techniques
- Domain of applicability: Everywhere!



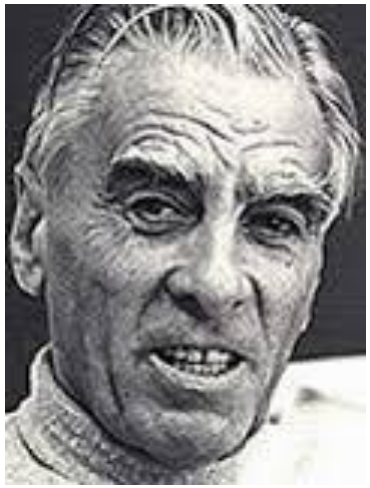
Why Derive QSAR Models?

- Understand the factors affecting molecular properties
- Predict the properties of new compounds

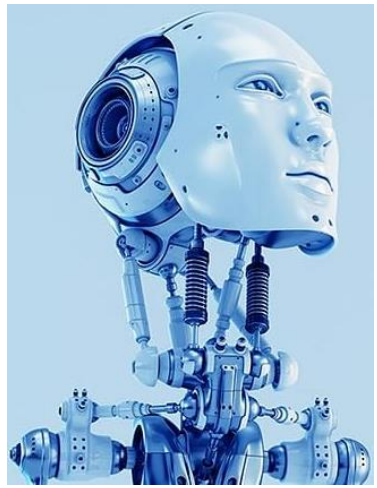


QSAR Models: The Balance

Understand



Predict



Corwin Hansch

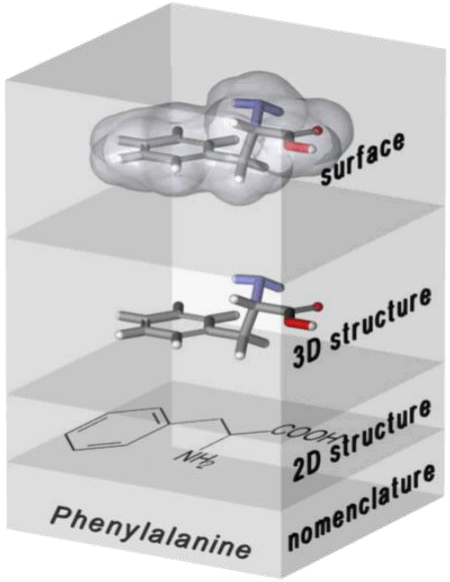
- Small congeneric series
- Few simple descriptors
- Insight into substituent effects

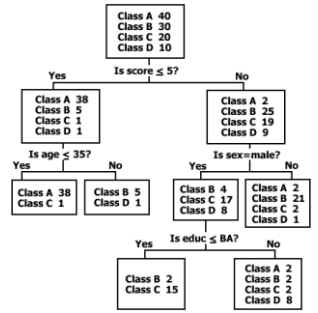
ANN

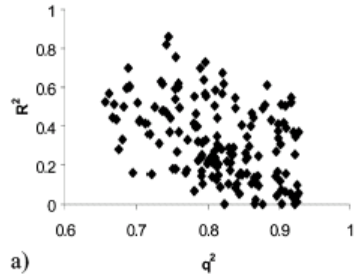
- Large diverse datasets
- Thousands of (obscure) descriptors
- Accurate (?) predictions

Designing QSAR “Experiments”

- Accurate experimental data (e.g., “activities”)
- Descriptors
  - Structure-derived (measured; calculated)
- A mathematical model
  - *e.g.*, quantitative, qualitative, linear, non-linear
- Model validation
  - Models developed on a training set and tested on an independent test set
  - Models should be simple and interpretable

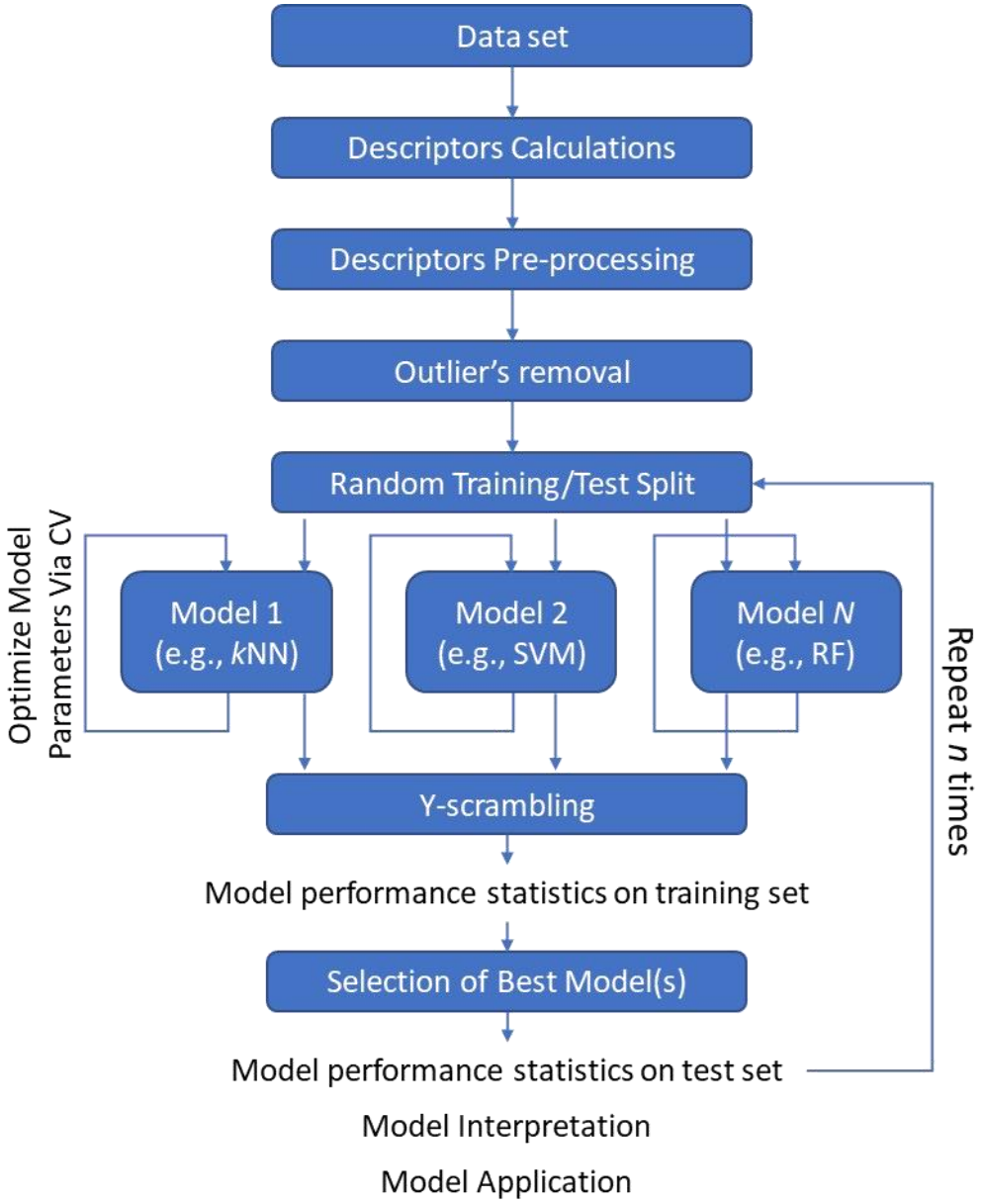






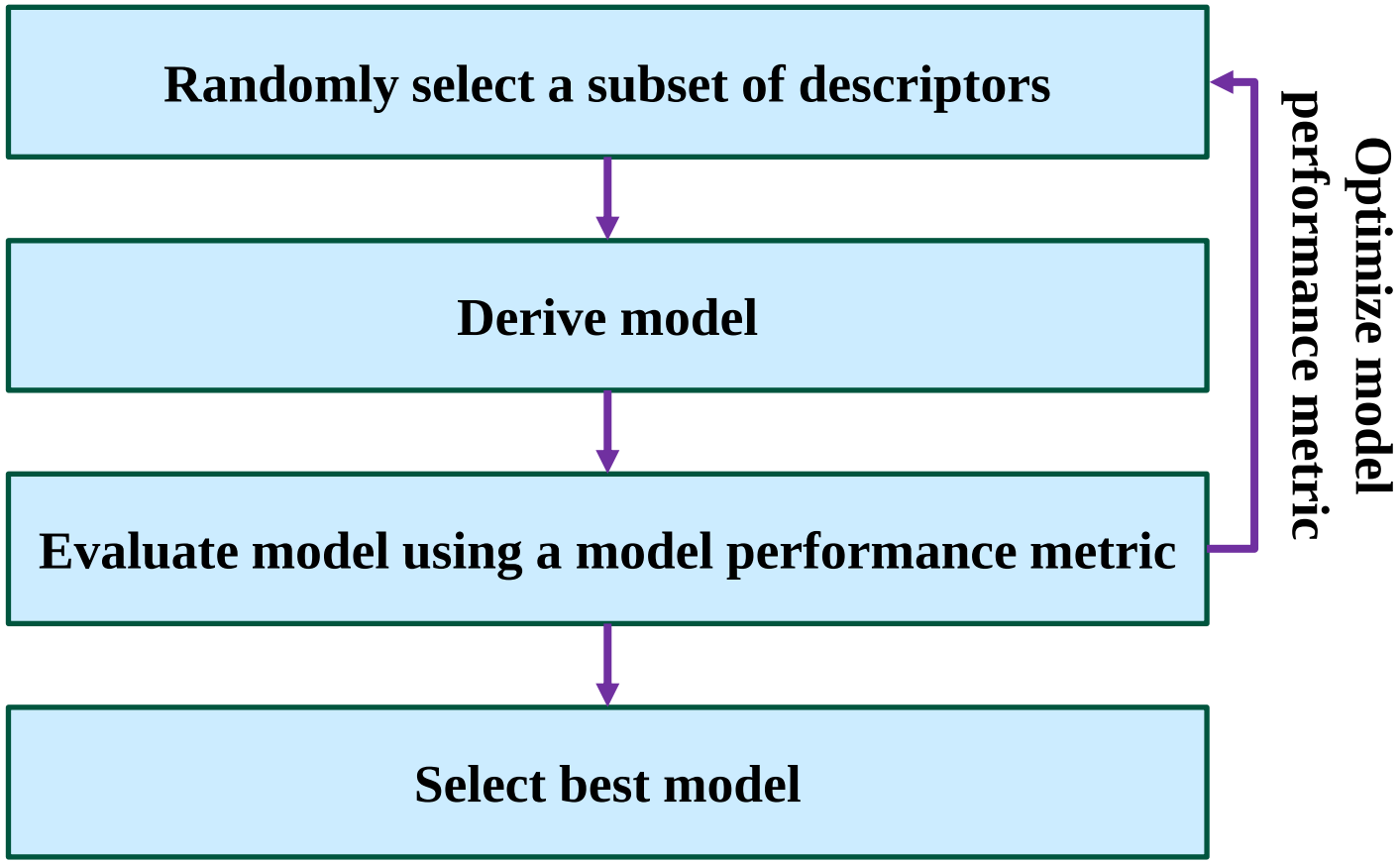
A QSAR Workflow

- “Proper” ways to derive QSAR models have been widely discussed in the literature
- This led to the development of “best practices”

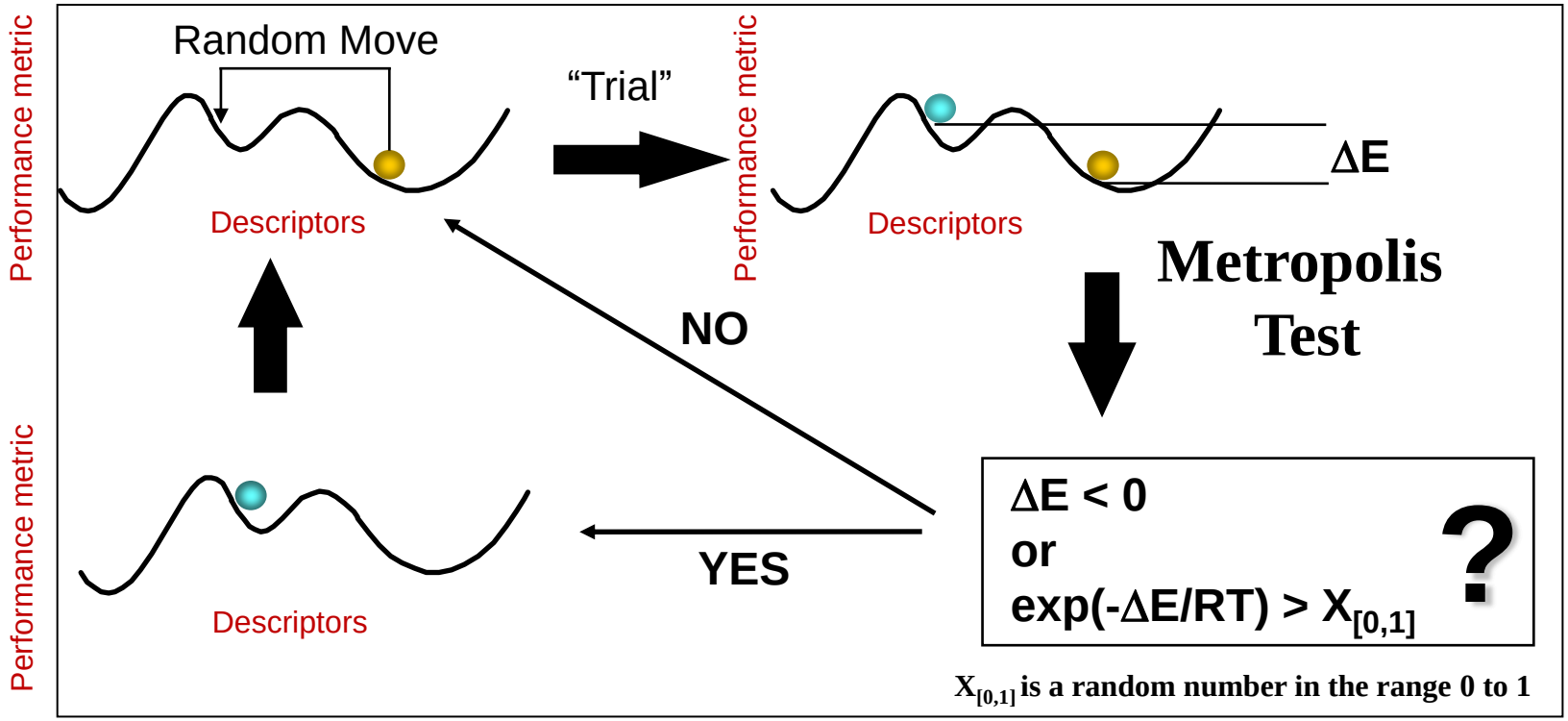


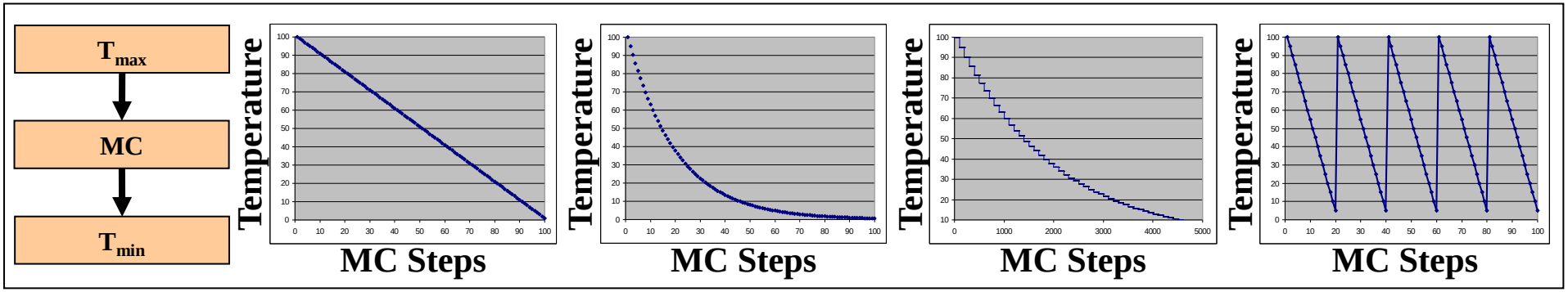
Descriptors

- Thousands of molecular descriptors (AlvaDesc, PaDEL, Canvas)
- Zillion of descriptors combinations
- Deriving QSAR models is an optimization problem in the space of the descriptors



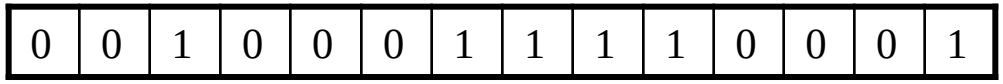
Optimization Algorithms: MC/SA



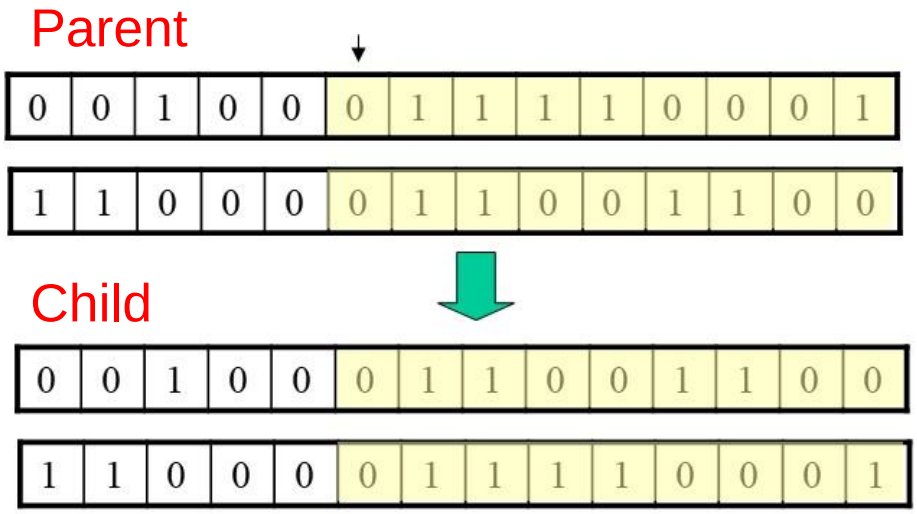


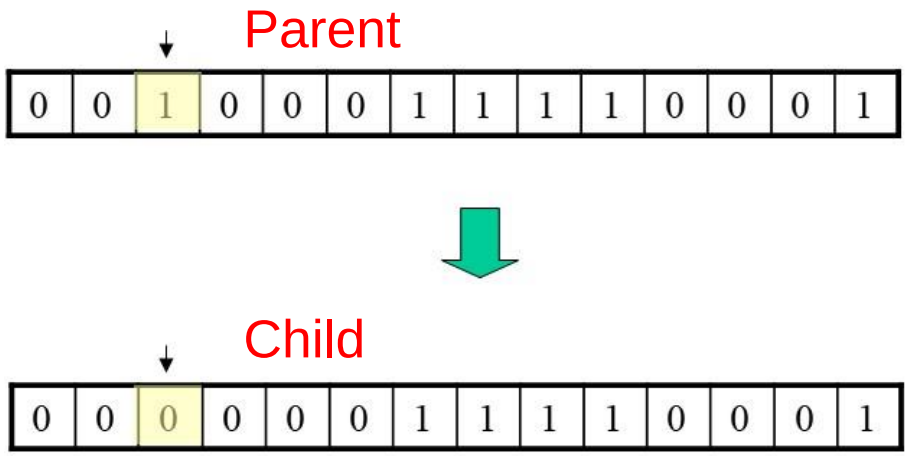
Optimization Algorithms: GA

- Create population of μ possible solutions to the optimization problem
- Code each solution by a chromosome containing a subset of descriptors



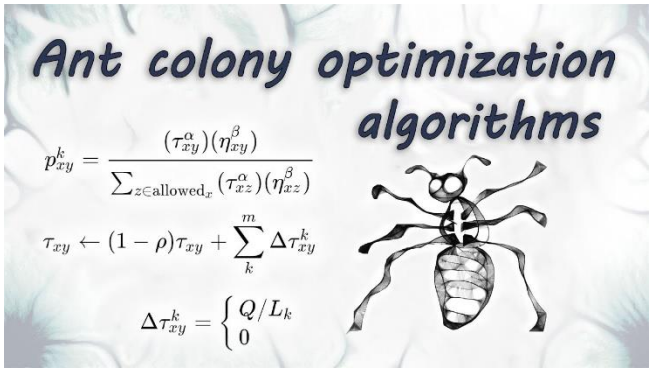
- For each chromosome calculate fitness function (model performance metric)
- Evolve population using genetic operations (selection of the fittest, cross-over, mutation) to optimize fitness function



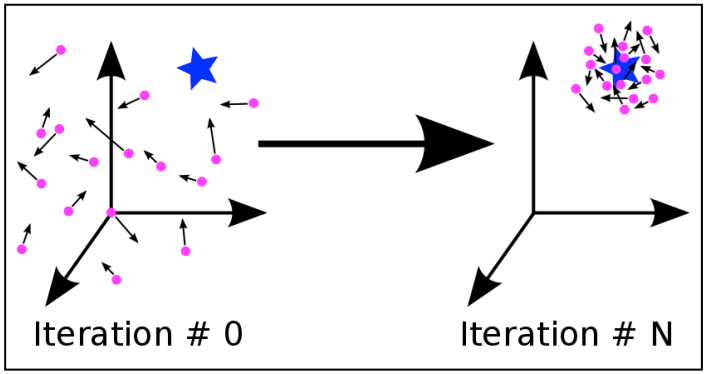


Optimization ZOO

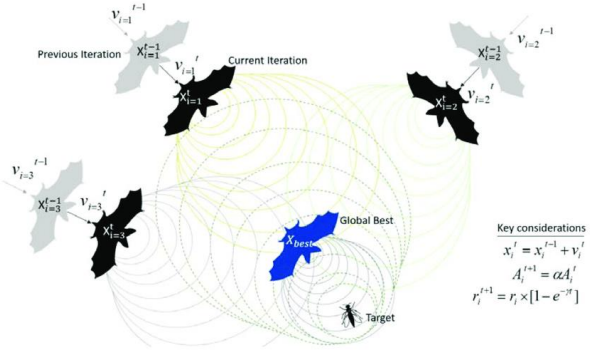
Ant Colony



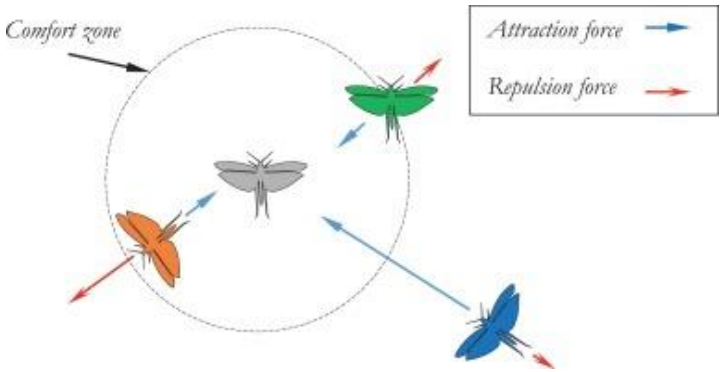
Particle Swarm



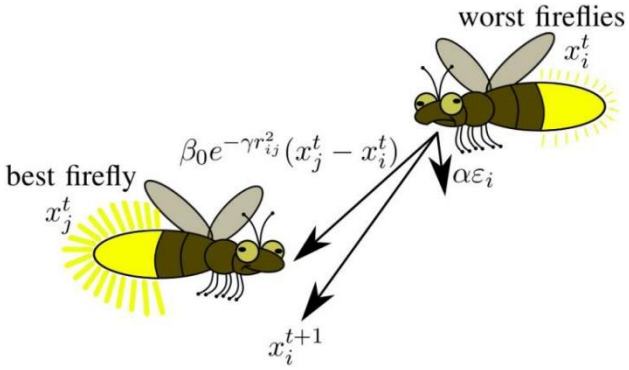
Bat Algorithm



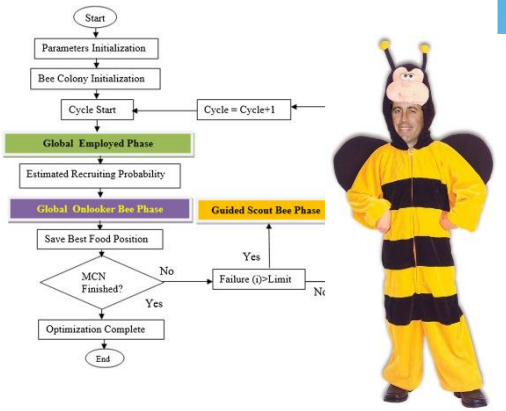
Grasshopper



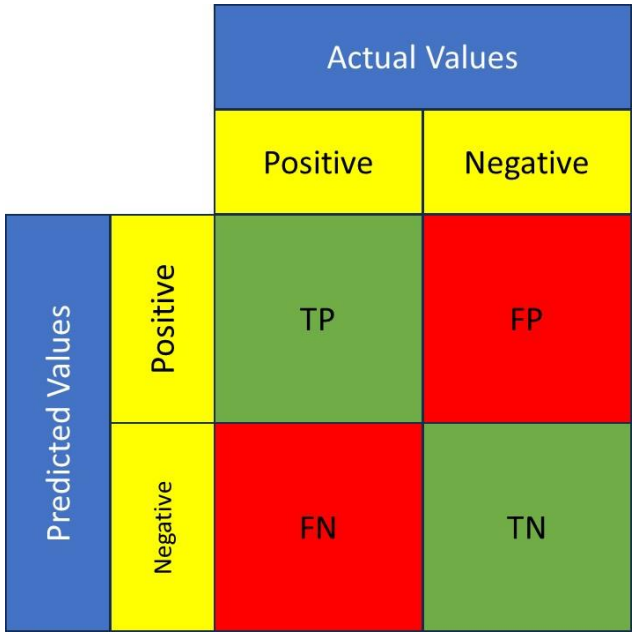
Firefly



Bee Colony



Optimizable Metrics: Classification



$$Negative\ Predicted\ Value\ (NPV) = \frac{TN}{TN + FN}$$

$$False\ Negative\ Rate\ (FNR) = \frac{FN}{FN + TP}$$

$$False\ Positive\ Rate\ (FPR) = \frac{FP}{FP + TN}$$

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Specifcity = \frac{TN}{TN + FP}$$

$$Precision = \frac{TP}{TP + FP}$$

$$F1 - score = \frac{2TP}{2TP + FP + FN}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}}$$

Optimizable Metrics: Regression

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i^{fitted}\right)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \equiv 1 - \frac{RSS}{TSS}$$ Training set

$$Q^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i^{pred}\right)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \equiv 1 - \frac{PRESS}{TSS}$$ Cross-Validation

$$Q_{F1}^2 = 1 - \frac{\sum_{i=1}^{n_{EXT}}\left(y_i - \hat{y}_i^{pred}\right)^2}{\sum_{i=1}^{n_{EXT}}(y_i - \bar{y}_{TR})^2} \equiv 1 - \frac{PRESS}{TSS_{\bar{y}_{TR}}}$$ Test set

$$Q_{F2}^2 = 1 - \frac{\sum_{i=1}^{n_{EXT}}\left(y_i - \hat{y}_i^{pred}\right)^2}{\sum_{i=1}^{n_{EXT}}(y_i - \bar{y}_{EXT})^2} \equiv 1 - \frac{PRESS}{TSS_{\bar{y}_{EXT}}}$$ Test set

$$Q_{F3}^2 = 1 - \frac{\sum_{i=1}^{n_{EXT}}\left(y_i - \hat{y}_i^{pred}\right)^2 / n_{EXT}}{\sum_{i=1}^{n_{TR}}(y_i - \bar{y}_{TR})^2 / n_{TR}} \equiv 1 - \frac{PRESS/n_{EXT}}{TSS/n_{TR}}$$ Test set

Optimizable Metrics: Regression

$$CCC = \frac{2 \times \sum_{i=1}^{n_{EXT}} (y_i - \bar{y})\,(\hat{y}_i - \bar{\hat{y}})}{\sum_{i=1}^{n_{EXT}} (y_i - \bar{y})^2 + \sum_{i=1}^{n_{EXT}} \left(\hat{y}_i - \bar{\hat{y}}\right)^2 + n_{EXT}(\bar{y} - \bar{\hat{y}})^2}$$

Concordance Correlation Coefficient

$$r_m^2 = r^2\left(1 - \sqrt{r^2 - r_0^2}\right) > 0.5;\ \overline{r_m^2} = \frac{\left(r_m^2 - r_m^{'\,2}\right)}{2};\ \Delta r_m^2 = \left|r_m^2 - r_m^{'\,2}\right| < 0.2$$

$$R^2 = \frac{\sum_{i=1}^{n_{EXT}} (y_i - \bar{y})\,(\hat{y}_i - \bar{\hat{y}})}{\sum_{i=1}^{n_{EXT}} (y_i - \bar{y})^2 \sum_{i=1}^{n_{EXT}} \left(\hat{y}_i - \bar{\hat{y}}\right)^2} \approx 1$$

$$R_0^2 = 1 - \frac{\sum_{i=1}^{n_{EXT}} \left(\hat{y}_i - y_i^{r_0}\right)^2}{\sum_{i=1}^{n_{EXT}} \left(\hat{y}_i - \bar{\hat{y}}\right)^2} \approx R^2;\ R_0^{'\,2} = 1 - \frac{\sum_{i=1}^{n_{EXT}} \left(y_i - \hat{y}_i^{\,r_0}\right)^2}{\sum_{i=1}^{n_{EXT}} (y_i - \bar{y})^2} \approx R^2$$

$$k = \frac{\sum_{i=1}^{n_{EXT}} y_i \hat{y}_i}{\sum_{i=1}^{n_{EXT}} y_i \hat{y}_i^2} \approx 1;\ k' = \frac{\sum_{i=1}^{n_{EXT}} y_i \hat{y}_i}{\sum_{i=1}^{n_{EXT}} y_i y_i^2} \approx 1;\ y_i^{r_0} = k\hat{y}_i;\ \hat{y}_i^{\,r_0} = k' y_i$$

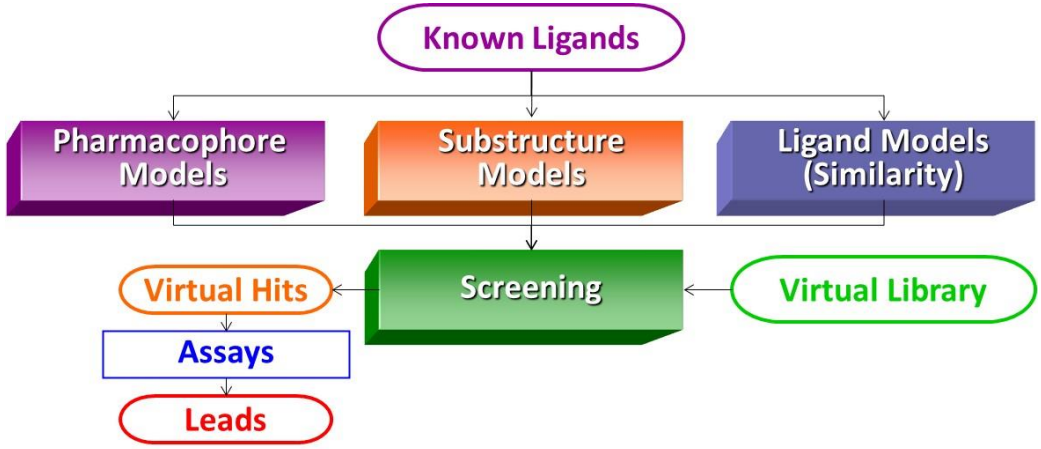
QSAR Models for Virtual Screening

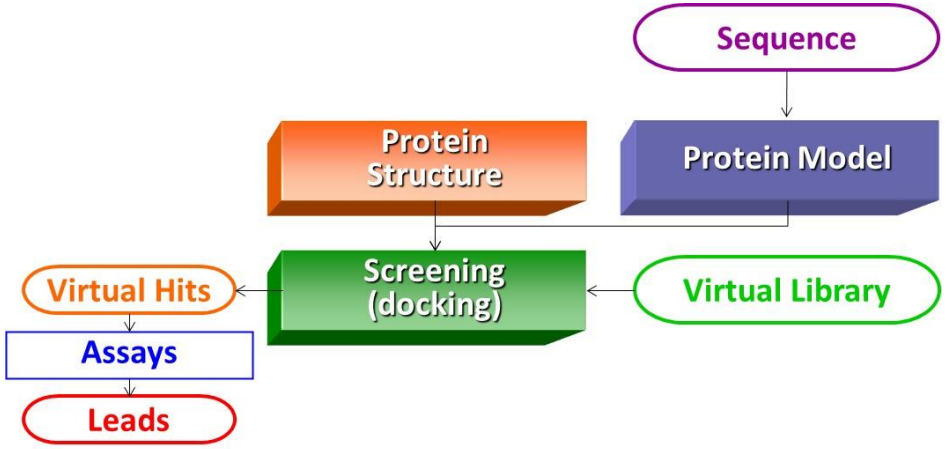
Ligand-based

- Faster
- No binding mode
- No binding free energy
- More “local”

Structure-based

- Slower
- Binding modes provided
- Binding free energy estimated
- More “global”

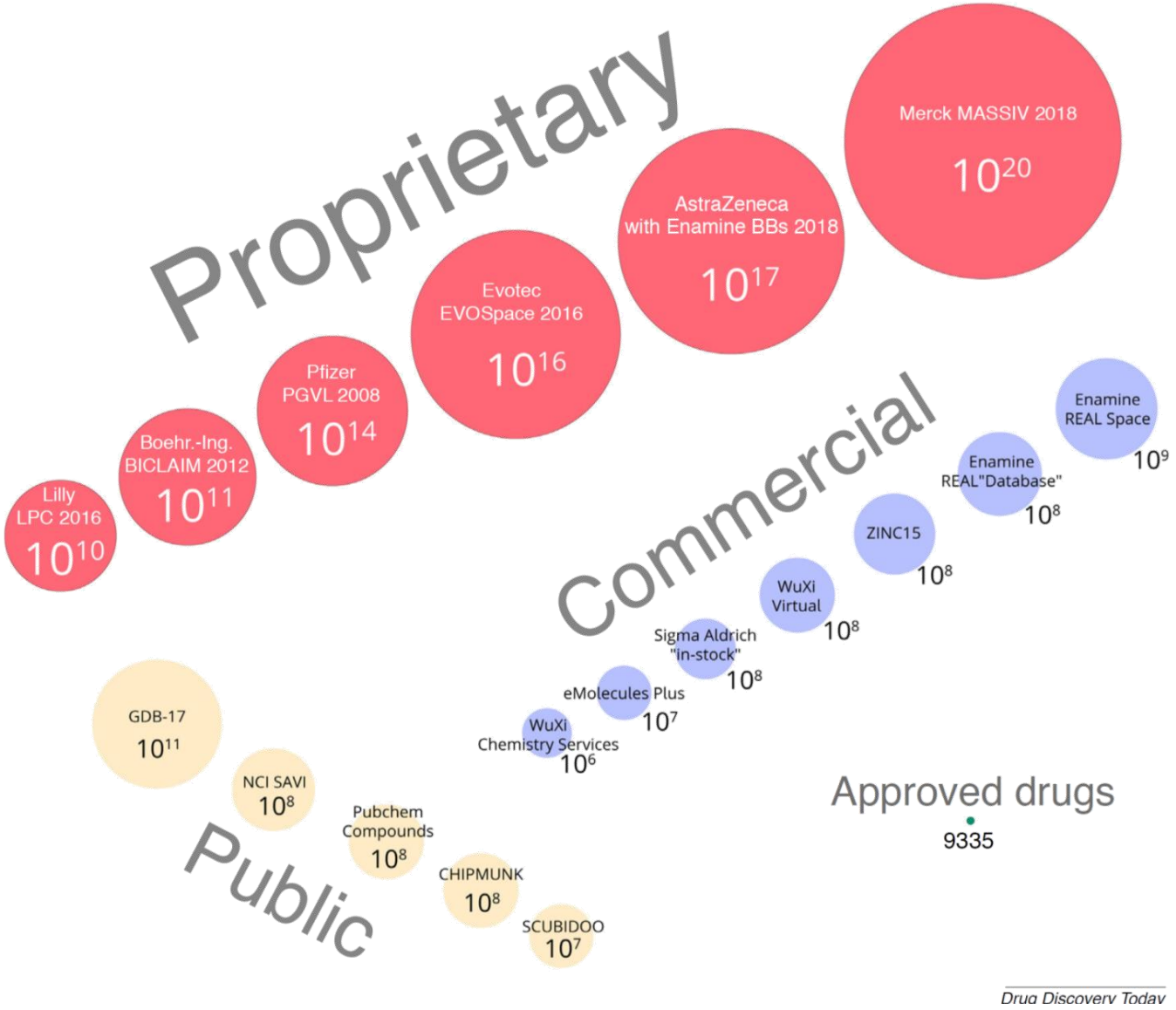




QSAR models could be used for virtual screening

Why should we use QSAR models for virtual screening?

The Accessible Chemical Space

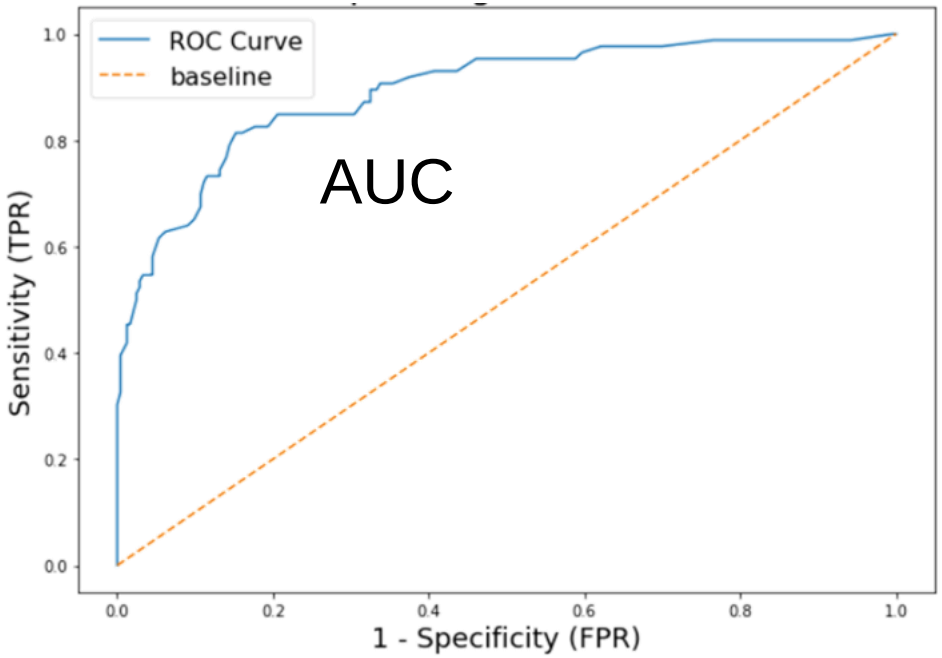


Hoffmann et al., Drug Discovery Today 2019, 24 1148

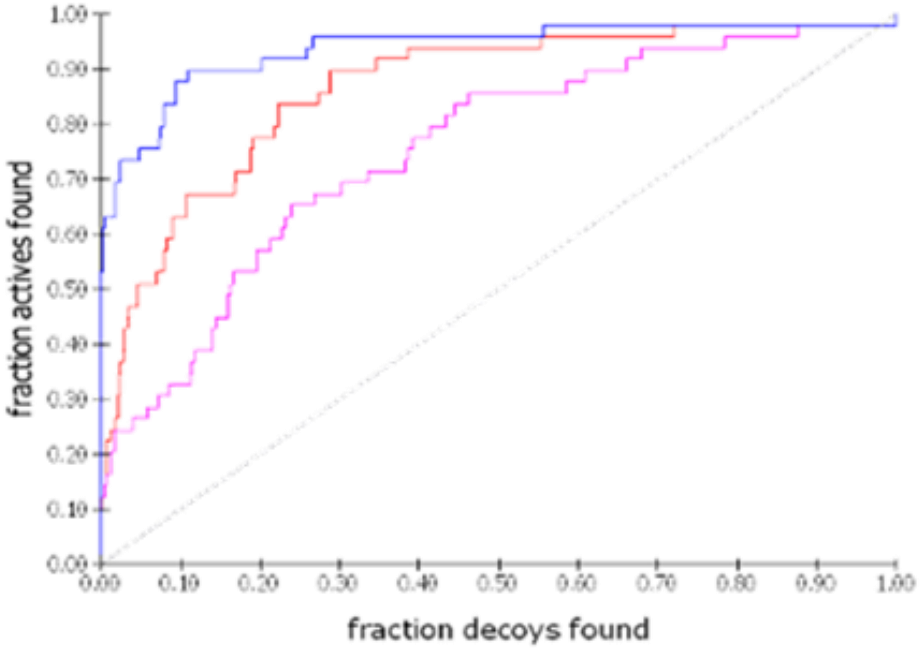
Evaluation of QSAR Models for VS

- No *a priori* reason to expect QSAR models characterized by favorable R^2/Q^2 values to perform well in VS (and vice versa)
  - The task of identifying somewhat active compounds from within a pool of inactive compound is different from the task of classifying/predicting activity of compounds series
- QSAR models for Virtual Screening should be developed and evaluated using a VS-aware metric

AUC under the ROC curve



Enrichment



Datasets

- Experimental data obtained from ChEMBL/literature; Decoys from ZINC
- Ligands prepared with Schrodinger's LigPrep
- 3 MLR models pre target (7, 10, 13 descriptors) derived by Schrodinger's Canvas by optimizing the regression standard deviation
- Validation and test sets only differ in the actives/inactives ratios

Dataset	# descriptors	Training Set			Validation Set			Test set		
		# Actives	# Inactives	Maximal Enrichment	# Actives	# Inactives	Maximal Enrichment	# Actives	# Random	Maximal Enrichment
$5HT_{2C}$	7, 10, 13	50	87	2.7	50	87	2.7	50	5141	103.8
M2	7, 10, 13	50	84	2.7	50	84	2.7	50	5141	103.8
H1	7, 10, 13	50	90	2.8	50	90	2.8	50	5141	103.8
hERG	7, 10, 13	100	600	7.0	100	600	7.0	100	5141	52.4
M3	7, 10, 13	75	75	2.0	75	75	2.0	75	5141	69.5
D1	7, 10, 13	58	58	2.0	58	58	2.0	58	5141	89.6
$Alpha_{2C}$	7, 10, 13	57	57	2.0	57	57	2.0	57	5141	91.2

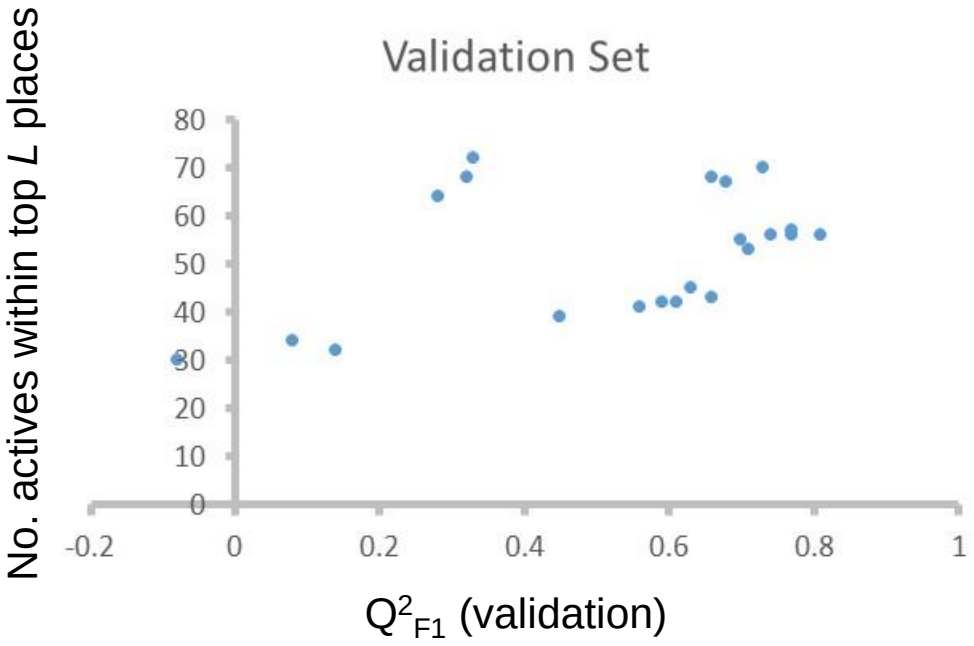
Spiegel et al., IJMS 2020, 21, 7828

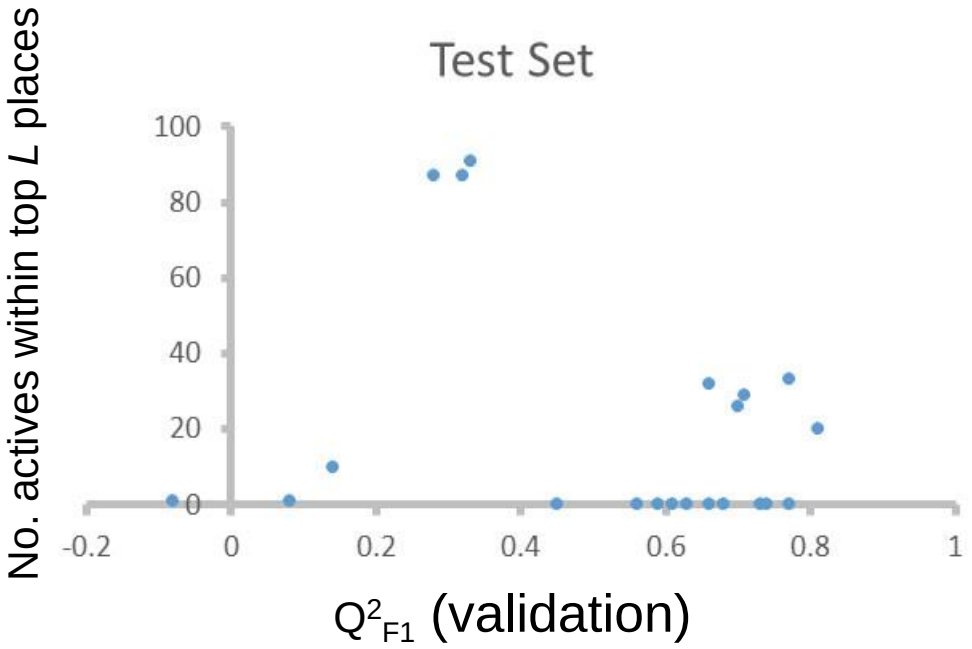
Resulting MLR Models

Set	# actives = L	# Descriptors	Train		Validation				Test
			R^2	# actives among L top places (enrichment)	Q^2_{F1}	Q^2_{F2}	Q^2_{F3}	# actives among L top places (enrichment)	# actives among L top places (enrichment)
M2	50	13	0.82	48 (2.6)	0.61	0.61	0.62	42 (2.3)	0 (0)
H1	50	13	0.82	49 (2.6)	0.56	0.56	0.60	41 (2.2)	0 (0)
$5HT_{2C}$	50	13	0.70	45 (2.5)	-0.08	-0.08	-0.03	30 (1.7)	1 (2.1)
hERG	100	13	0.39	71 (5.0)	0.33	0.33	0.24	72 (5.0)	91 (47.7)
M3	75	13	0.91	75 (2.0)	0.73	0.73	0.75	70 (1.9)	0 (0)
D1	58	13	0.88	58 (2.0)	0.74	0.74	0.72	56 (1.9)	0 (0)
$Alpha_{2C}$	57	13	0.83	54 (1.9)	0.71	0.71	0.72	53 (1.9)	29 (46.4)

- Single point Enrichment based on *L*
- Reasonable enrichment for validation set
- Poor enrichment for test set (more representative of VS)

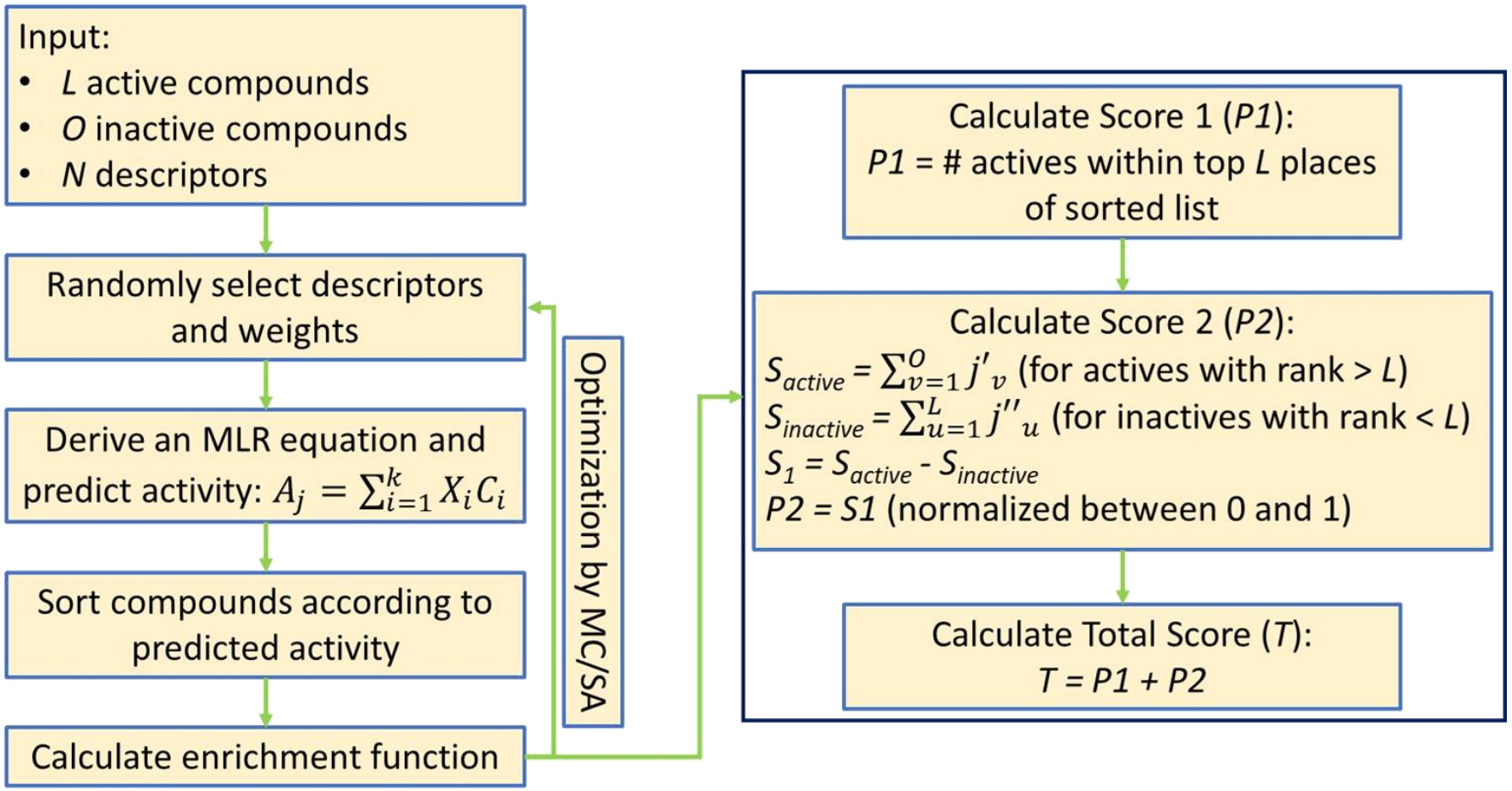
MLR Models





Claim 1: No *a priori* reason to expect QSAR models characterized by favorable R^2/Q^2 values to perform well in VS

Enrichment Optimization Algorithm (EOA)



- Primary Score: Number of actives within the first L places (redundant)
- Secondary score: Designed to remove redundancy
- Ranges [0-1] to ensure the suprimicy of the primary score

EOA Models

Set	# Descriptors	# Actives = *L*	# actives among *L* top places in best model (enrichment)		
			Train	Validation	Test1
M2	7	50	47 (2.5)	40 (2.1)	40 (83.1)
M2	10	50	47 (2.5)	44 (2.4)	39 (81.0)
M2	13	50	47 (2.5)	42 (2.3)	38 (78.9)
H1	7	50	48 (2.6)	37 (2.1)	31 (64.4)
H1	10	50	49 (2.6)	42 (2.3)	32 (66.4)
H1	13	50	48 (2.6)	38 (2.0)	32 (66.4)
$5HT_{2C}$	7	50	45 (2.5)	34 (1.9)	0 (0)
$5HT_{2C}$	10	50	45 (2.5)	32 (1.8)	1 (2.1)
$5HT_{2C}$	13	50	47 (2.6)	33 (1.8)	0 (0)
hERG	7	100	67 (4.7)	60 (4.2)	86 (45.1)
hERG	10	100	75 (5.3)	61 (4.3)	89 (46.6)
hERG	13	100	77 (5.4)	59 (4.1)	87 (45.6)
M3	7	75	74 (2.0)	65 (1.7)	49 (45.4)
M3	10	75	74 (2.0)	67 (1.8)	0 (0)
M3	13	75	74 (2.0)	70 (1.9)	57 (52.9)
D1	7	58	56 (1.9)	54 (1.9)	29 (44.8)
D1	10	58	57 (2.0)	55 (1.9)	20 (30.9)
D1	13	58	57 (2.0)	54 (1.9)	41 (63.4)
$Alpha_{2C}$	7	57	49 (1.7)	47 (1.6)	25 (40.0)
$Alpha_{2C}$	10	57	49 (1.7)	45 (1.6)	25 (40.0)
$Alpha_{2C}$	13	57	56 (2.0)	52 (1.8)	15 (24.0)

- 15 EOA models per target (5 repeats X 7, 10, 13 descriptors)
- Best results over 5 optimization runs are presented
- Better consistency between Train/Validation/Test results in terms of enrichment:
  - MLR: $r^2_{(train/test)} = 0.16$
  - MLR: $r_{2(validation/test)} = 0.25$
  - EOA: $r^2_{(train/test)} = 0.37$ (0.63)
  - EOA: $r_{2(validation/test)} = 0.28$ (0.51)
- Outperforms MLR for 12 models
- Similar to MLR for 2 models
- Outperformed by MLR for 7 models

EOA Models as Binary Classifiers

- Virtual Screening is akin to a binary classification problem
- Compare the performances of EOA with RF and SVM

Dataset	# descriptors	Training Set			Validation Set			Test Set		
		# Actives	# Inactives	Maximal Enrichment	# Actives	# Inactives	Maximal Enrichment	# Actives	# Inactives	Maximal Enrichment
$5HT_{2C}$	7, 10, 13	50	450	10.0	50	450	10.0	50	5141	103.8
M2	7, 10, 13	50	450	10.0	50	450	10.0	50	5141	103.8
H1	7, 10, 13	50	450	10.0	50	450	10.0	50	5141	103.8
hERG	7, 10, 13	50	450	10.0	50	450	10.0	50	5141	103.8
M3	7, 10, 13	50	450	10.0	50	450	10.0	50	5141	103.8
D1	7, 10, 13	50	450	10.0	50	450	10.0	50	5141	103.8
$Alpha_{2C}$	7, 10, 13	50	450	10.0	50	450	10.0	50	5141	103.8

EOA Models as Binary Classifiers

- For each dataset, 12 models were derived (4 random selections for training and validation and three sets of descriptors)
- Results of best (across 3 descriptors sets) models are shown

Set	Run	# actives among *L* top places (enrichment; MCC)		
		EOA-Test	RF-Test	SVM-Test
M2	1	50 (103.8; 1.00)	39 (81.0; 0.88)	36 (74.8; 0.85)
	2	48 (99.7; 0.96)	37 (76.8; 0.86)	37 (76.8; 0.86)
	3	47 (97.6; 0.94)	42 (87.2; 0.92)	37 (76.8; 0.86)
	4	48 (99.7; 0.96)	40 (83.1; 0.89)	38 (78.9; 0.87)
H1	1	38 (78.9; 0.76)	29 (60.2; 0.73)	33 (68.5; 0.81)
	2	47 (97.6; 0.94)	33 (68.5; 0.78)	34 (70.6; 0.82)
	3	42 (87.2; 0.84)	39 (81.0; 0.70)	31 (64.4; 0.79)
	4	39 (81.0; 0.78)	27 (56.1; 0.52)	27 (56.1; 0.73)
$5HT_{2C}$	1	50 (103.8; 1.00)	50 (103.8; 1.00)	48 (99.7; 0.98)
	2	50 (103.8; 1.00)	50 (103.8; 1.00)	50 (103.8; 1.00)
	3	50 (103.8; 1.00)	50 (103.8; 1.00)	50 (103.8; 1.00)
	4	50 (103.8; 1.00)	50 (103.8; 1.00)	50 (103.8; 1.00)
hERG	1	42 (87.2; 0.84)	23 (47.8; 0.31)	19 (39.5; 0.61)
	2	38 (78.9; 0.76)	26 (54.0; 0.33)	19 (39.5; 0.61)
	3	45 (93.4; 0.90)	32 (66.4; 0.64)	22 (45.7; 0.66)
	4	47 (97.6; 0.94)	19 (39.5; 0.55)	18 (37.4; 0.60)
D1	1	44 (91.4; 0.88)	33 (68.5; 0.69)	44 (91.4; 0.94)
	2	46 (95.5; 0.92)	31 (64.4; 0.55)	35 (72.7; 0.84)
	3	45 (93.4; 0.90)	37 (76.8; 0.71)	36 (74.8; 0.85)
	4	45 (93.4; 0.90)	40 (83.1; 0.87)	37 (76.8; 0.86)
M3	1	49 (101.7 0.98)	50 (103.8; 1.00)	45 (93.4; 0.95)
	2	48 (99.7; 0.96)	35 (72.7; 0.84)	35 (72.7; 0.84)
	3	48 (99.7; 0.96)	32 (66.4; 0.80)	33 (68.5; 0.81)
	4	44 (91.4; 0.88)	37 (76.8; 0.86)	37 (76.8; 0.86)
$Alpha_{2c}$	1	37 (76.8; 0.74)	35 (72.7; 0.84)	33 (68.5; 0.81)
	2	44 (91.4; 0.88)	38 (78.9; 0.87)	34 (70.6; 0.82)
	3	43 (89.3; 0.86)	38 (78.9; 0.79)	34 (70.6; 0.82)
	4	42 (87.2; 0.84)	40 (83.1; 0.89)	41 (85.1; 0.90)

Claim 2: QSAR models for Virtual Screening should be developed and evaluated using a VS-aware metric

EOA Models vs. Docking

- Datasets retrieved from the DUD and DUD-E databases
- Ligand/proteins prepared with Schrodinger's LigPrep/protein preparation wizard
- Descriptors calculated by Schrodinger's Canvas and pre-processed
- 12 EOA models per target (4 different subsets *X* 7, 10, 13 descriptors)
- Docking (GOLD, Glide, AutoDock Vina) performed with default settings; AUC and $EF_{1\%}$ calculated using consensus across two crystal structures
- EOA models and docking algorithms evaluated on same test sets

Dataset	PDB Codes	# Active	# Decoy	Training		Validation		Test	
				# Active	# Decoy	# Active	# Decoy	# Active	# Decoy
ACES	1e66, 1acj	643	24161	430	3333	106	832	107	19996
HIVPR	1xl2, 2pwc	1366	35071	912	3333	227	832	227	30906
MK14	2qd9, 3o8t	911	34896	608	3333	151	832	152	30731
UROK	1sqt, 4fue	298	9262	200	1666	49	416	49	7180
TRY1	2ayw, 3rxl	755	24760	504	3333	125	832	126	20595

Spiegel et al., IJMS 2022, 23, 43

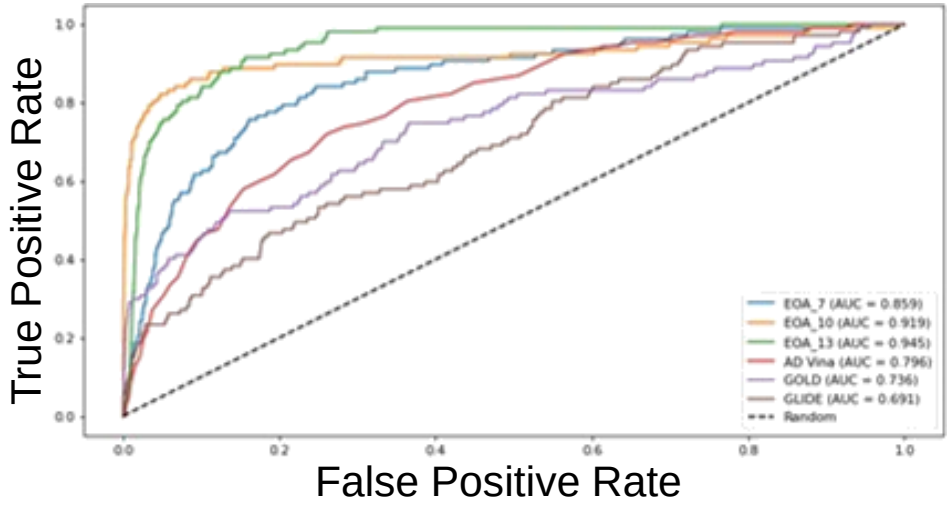
Results

Set	Method	AUC					$EF_{1\%}$				
		ACES	HIVPR	MK14	UROK	TRY1	ACES	HIVPR	MK14	UROK	TRY1
1	EOA-7	0.862	0.775	0.905	0.997	0.979	36.449	3.965	40.132	77.551	73.810
	EOA-10	0.886	0.946	0.947	0.997	0.996	26.168	20.705	39.474	81.633	80.159
	EOA-13	0.899	0.977	0.927	0.996	0.986	58.879	59.471	25.658	81.633	58.730
	AD Vina	0.764	0.747	0.737	0.749	0.806	10.280	7.048	10.526	4.082	4.762
	GOLD	0.739	0.726	0.676	0.830	0.861	30.841	13.656	12.500	32.653	15.873
	Glide	0.735	0.678	0.743	0.801	0.847	14.953	11.454	15.789	40.816	35.714
2	EOA-7	0.808	0.813	0.902	0.986	0.958	8.411	14.097	20.395	69.388	75.397
	EOA-10	0.885	0.955	0.914	0.978	0.982	43.925	21.586	41.447	81.633	79.365
	EOA-13	0.921	0.927	0.925	0.987	0.966	9.346	19.383	37.500	73.469	71.429
	AD Vina	0.760	0.754	0.753	0.766	0.795	12.150	3.965	8.553	2.041	4.762
	GOLD	0.719	0.687	0.686	0.785	0.849	28.972	12.335	9.868	28.571	23.810
	Glide	0.693	0.612	0.729	0.816	0.832	14.019	8.370	18.421	36.735	42.857
3	EOA-7	0.896	0.918	0.894	0.955	0.982	42.991	9.692	25.000	79.592	76.984
	EOA-10	0.860	0.946	0.891	0.956	0.980	7.477	34.802	23.684	75.510	79.365
	EOA-13	0.895	0.882	0.918	0.959	0.981	21.495	31.718	27.632	89.796	73.016
	AD Vina	0.762	0.769	0.768	0.777	0.786	10.280	7.048	9.868	6.122	3.968
	GOLD	0.710	0.715	0.686	0.866	0.849	27.103	17.621	9.868	34.694	26.190
	Glide	0.687	0.667	0.753	0.840	0.857	14.953	10.132	23.026	42.857	44.444
4	EOA-7	0.859	0.915	0.892	0.980	0.983	14.019	16.300	21.711	61.224	80.159
	EOA-10	0.919	0.922	0.934	0.986	0.982	58.879	13.656	23.026	67.347	76.190
	EOA-13	0.945	0.978	0.934	0.983	0.983	13.084	51.542	36.184	79.592	79.365
	AD Vina	0.796	0.740	0.766	0.765	0.809	8.411	5.727	5.921	6.122	7.143
	GOLD	0.736	0.722	0.663	0.818	0.854	28.972	13.656	7.895	26.531	25.397
	Glide	0.691	0.646	0.728	0.818	0.860	10.280	12.335	16.447	40.816	46.825

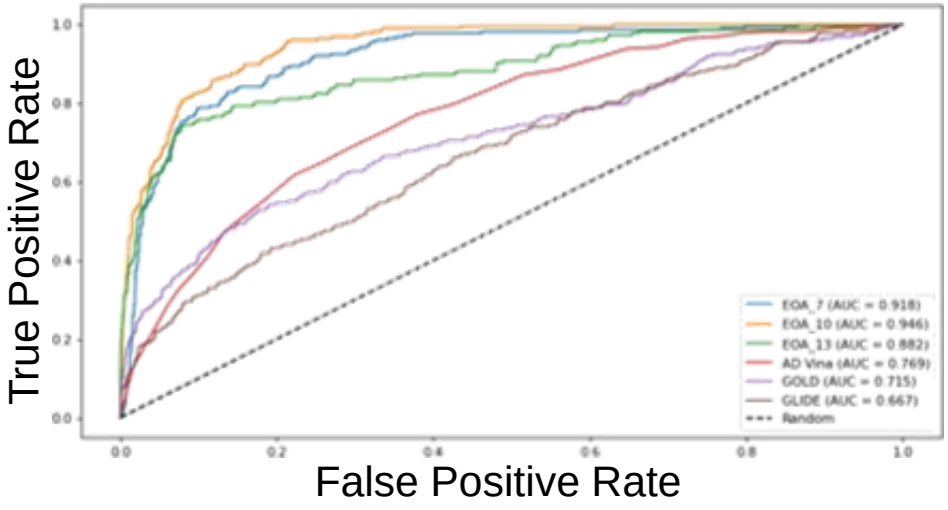
- AUC: Measure of global success
- $EF_{1\%}$: Measure of success at early stages of screening

Results

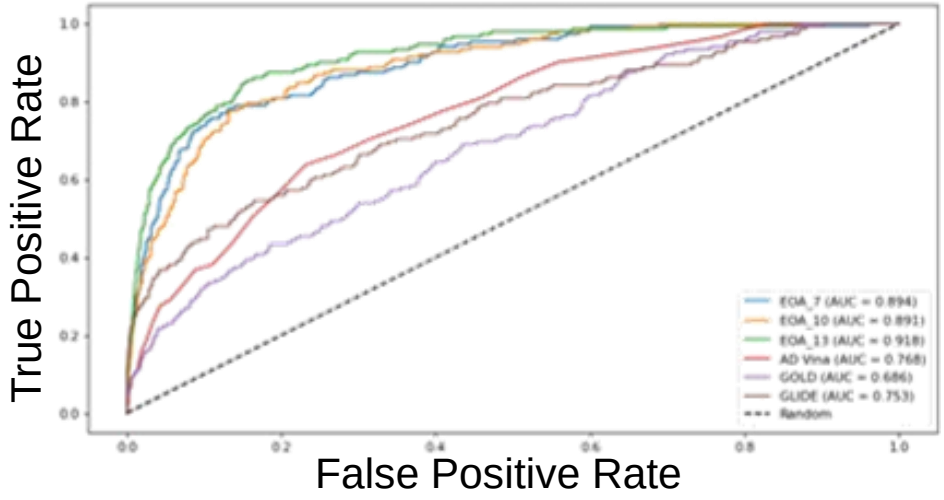
Acetylcholinesterase



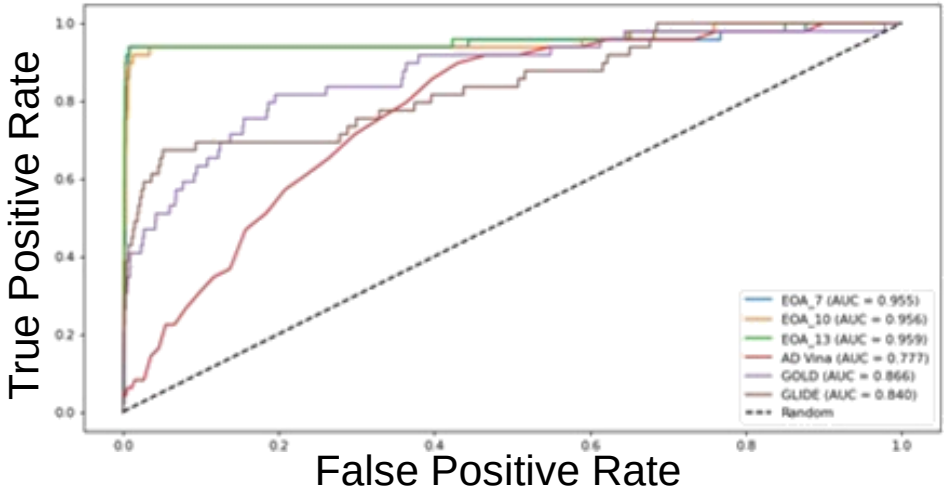
Human immunodeficiency virus type 1 protease



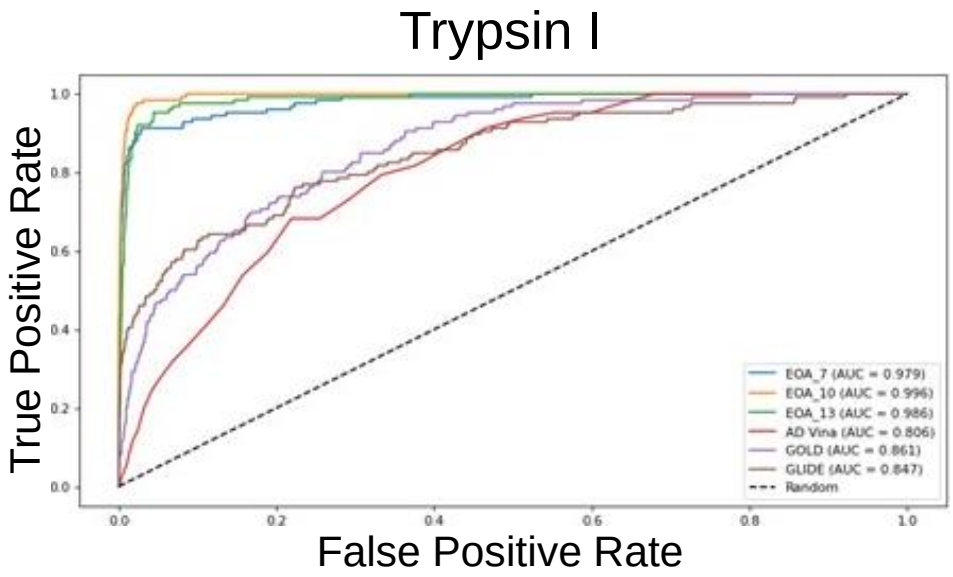
MAP kinase p38 alpha



Urokinase-type plasminogen activator



Results



EOA outperformed docking in all cases

Key Descriptors

Descriptor	Description
PEOE3	Charge descriptor
HBD	H-bond descriptor
MR1	Molar refractivity
ALOGP2	Water-Octanol partition descriptor
ALOGP7	Water-Octanol partition descriptor
ssCH2_Cnt	Count descriptor
aaCH_Cnt	Count descriptor

Dataset	AUC	$EF_{1\%}$
ACES	0.825	24.3
HIVPR	0.860	18.9
MK14	0.807	13.1
UROK	0.897	20.4
TRY1	0.932	24.6

- Key descriptors “make sense” (related to protein/ligand interactions/ADME
- Yet EOA models developed from key descriptors have lower performances

Performances/Ease of Separation Correlation

Target	Set	# Descriptors	Euclidean Distances (Average ± Standard Deviation)	AUC
ACES	4	10	4.34±1.41	0.885
HIVPR	2	10	4.10±1.23	0.955
MK14	1	10	4.90±1.59	0.947
UROK	1	10	4.59±1.30	0.997
TRY1	1	10	4.72±1.32	0.996

No correlation between actives/inactives distances and AUC

Target-Specific Scoring Functions for VS

- EOA outperformed docking in all cases
- EOA models are target-specific while docking is not
  - Docking tools designed to be as general as possible
  - Many (empirical) scoring function are in fact QSAR models

GlideScore

$$\begin{aligned}\Delta G_{\text{bind}} = {} & C_{\text{lipo-lipo}} \sum f(r_{\text{lr}}) + \\ & C_{\text{hbond-neut-neut}} \sum g(\Delta r)\, h(\Delta\alpha) + \\ & C_{\text{hbond-neut-charged}} \sum g(\Delta r)\, h(\Delta\alpha) + \\ & C_{\text{hbond-charged-charged}} \sum g(\Delta r)\, h(\Delta\alpha) + \\ & C_{\text{max-metal-ion}} \sum f(r_{\text{lm}}) + C_{\text{rotb}} H_{\text{rotb}} + \\ & C_{\text{polar-phob}} V_{\text{polar-phob}} + C_{\text{coul}} E_{\text{coul}} + \\ & C_{\text{vdW}} E_{\text{vdW}} + \text{solvation terms} \quad (2)\end{aligned}$$

Use EOA to derive target-specific scoring functions for VS by re-deriving the weights associated with the different energy terms

Target Specific Scoring Function for GOLD

- Output raw data
- User-defined functions for both scoring and docking could be implemented
- ChemPLP function:

$$ChemPLP = 1.0 \times S(PLP) - 3.0 \times S(Hbond) - 6.0 \times S(metal) - 3.0 \times S(cho)$$

$$Score = ChemPLP - 1.0 \times S(clash) - 2.0 \times S(tor)$$

  - S(PLP): Piecewise linear potential
  - S(Hbond): Distance-dependent H-bonds
  - S(cho): Angle-dependent H-bonds
  - S(metal): Metal-ligand interactions
  - S(clash): VdW clashes
  - S(tor): Torsional energy

Simple but not flexible enough

Datasets

- Datasets retrieved from the DUD-E. For all datasets poor AUC/enrichment were reported in DUD-E
- Ligand/proteins prepared with Schrodinger's LigPrep/protein preparation wizard
- Descriptors are the components that make up GOLD's ChemPLP function
- Four EOA models (corresponding to 4 subsets of the data) were derived

Target	PDB	Total # of Actives	Total # of Decoys	Train		Validation		Test	
				Actives	Decoys	Actives	Decoys	Actives	Decoys
PGH1	2oyu	244	10,286	164	1333	40	332	40	8621
CYP2C9	1r9o	172	6,777	116	1333	28	332	28	5112
ANDR	2am9	513	13,509	343	2000	85	499	85	11,010
PRGR	3kba	439	14,993	294	2000	72	499	73	12,494
GCR	3bqd	522	13,792	349	2000	86	499	87	11,293
HIVRT	3lan	634	14,314	424	2000	105	499	105	11,815

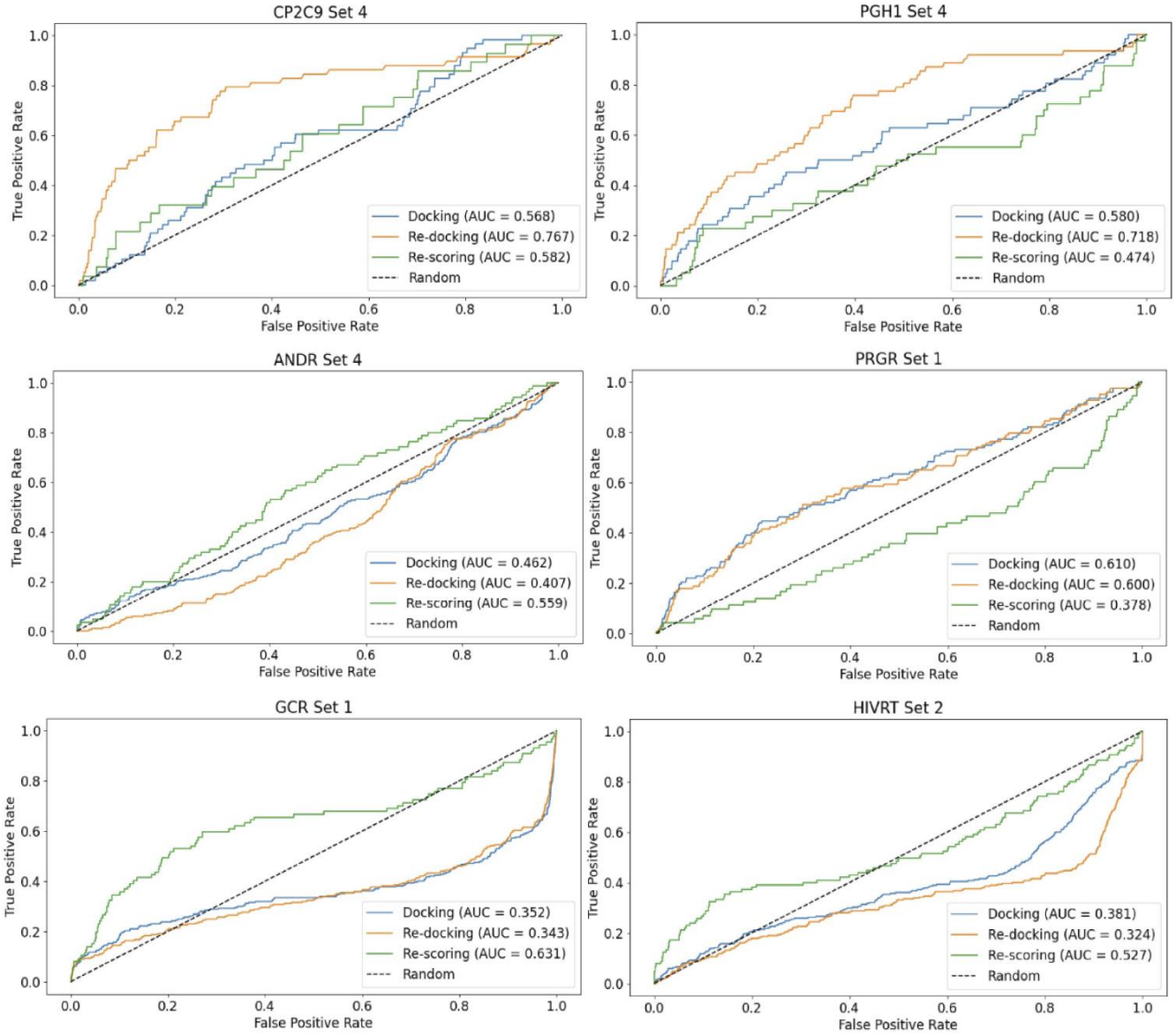
Spiegel et al., Molecular Informatics, 2022, 41, 2200034

Results

- Dock ligands with GOLD and calculate AUC and EF1% values
- For each target, use EOA to rederive GOLD's ChemPLP function by assigning new weights to its components (S(PLP), S(Hbond), S(cho), S(metal), S(clash), S(tor))
  - New weights were kept similar to GOLD's original weights
- Use target-specific functions for rescoring and redocking

Target	EOA-derived functions					Original GOLD
	Set1	Set2	Set3	Set4	Average	
PGH1	2.06	5.58	6.12	2.49	4.06	7.52
CYP2C9	1.88	3.01	1.75	3.15	2.45	6.22
ANDR	0.82	0.81	0.80	0.80	0.81	0.82
PRGR	0.85	0.92	0.84	0.67	0.82	0.99
GCR	0.56	0.83	0.95	0.56	0.73	0.94
HIVRT	0.59	0.60	0.39	0.47	0.51	0.35

Results


CP2C9 Set 4
True Positive Rate
False Positive Rate
Docking (AUC = 0.568)
Re-docking (AUC = 0.767)
Re-scoring (AUC = 0.582)
Random
PGH1 Set 4
Docking (AUC = 0.580)
Re-docking (AUC = 0.718)
Re-scoring (AUC = 0.474)
Random
ANDR Set 4
Docking (AUC = 0.462)
Re-docking (AUC = 0.407)
Re-scoring (AUC = 0.559)
Random
PRGR Set 1
Docking (AUC = 0.610)
Re-docking (AUC = 0.600)
Re-scoring (AUC = 0.378)
Random
GCR Set 1
Docking (AUC = 0.352)
Re-docking (AUC = 0.343)
Re-scoring (AUC = 0.631)
Random
HIVRT Set 2
Docking (AUC = 0.381)
Re-docking (AUC = 0.324)
Re-scoring (AUC = 0.527)
Random

Results

Target	Original GOLD		Rescoring		Re-docking	
	AUC	$EF_{1\%}$	AUC	$EF_{1\%}$	AUC	$EF_{1\%}$
PGH1	0.57	5.60	0.52	0.63	0.66	5.84
CYP2C9	0.61	3.27	0.66	3.57	0.72	9.74
ANDR	0.46	2.81	0.51	0.39	0.43	1.77
PRGR	0.56	2.90	0.45	2.74	0.53	1.63
GCR	0.35	4.68	0.63	5.75	0.35	4.70
HIVRT	0.39	3.84	0.49	5.95	0.35	3.14

Results inconclusive!!!

Conclusions

- QSAR models should be evaluated in a context-aware manner
- The EOA is a first step towards realizing this goal
  - Outperformed MLR/RF/SVM/Docking in VS experiments
  - Could be used to derive target-specific scoring functions
- Limitations
  - EOA requires large data set for model derivation (but can use qualitative data)
  - EOA *per se* does not provide information on binding modes (but could be combined with docking)
- Future improvements (are necessary)
  - Optimize a better model performance metric (AUC)
  - Use a better optimization algorithm
  - Apply to additional VS experiments/docking tools

Acknowledgments

Horizon 2020
European Union Funding
for Research & Innovation

Kobi Spiegel

רשות החדשנות
Israel Innovation
Authority

BARD

ELA
ASSOCIATION EUROPÉENNE
CONTRE LES LEUCODYSTROPHIES

arimnet
Coordination of Agricultural Research
in the Mediterranean Area

INNI
Israel National Nanotechnology Initiative

BSF
United States – Israel
Binational Science Foundation

משרד החקלאות ופיתוח הכפר

CYSTIC FIBROSIS FOUNDATION
THERAPEUTICS

Group Members

- Netaly Khazanov
- Anat Berliner
- Omer Kaspi
- Paul Clarke
- Ava Yuan Xiu
- Lina Iktelat
- Shahaf Kozokaro
- Lior Lublin
- Dina Khasanova
- Alexandra Olshanova
- Anastasiia Krokhina
- Ana Juarez
- Shir Zubeada