Single and Multi-Objective Optimization in Drug Design

Abraham Yosipof & Hanoch Senderowitz

Dept of Business Administration, Peres Academic Center, Rehovot, Israel

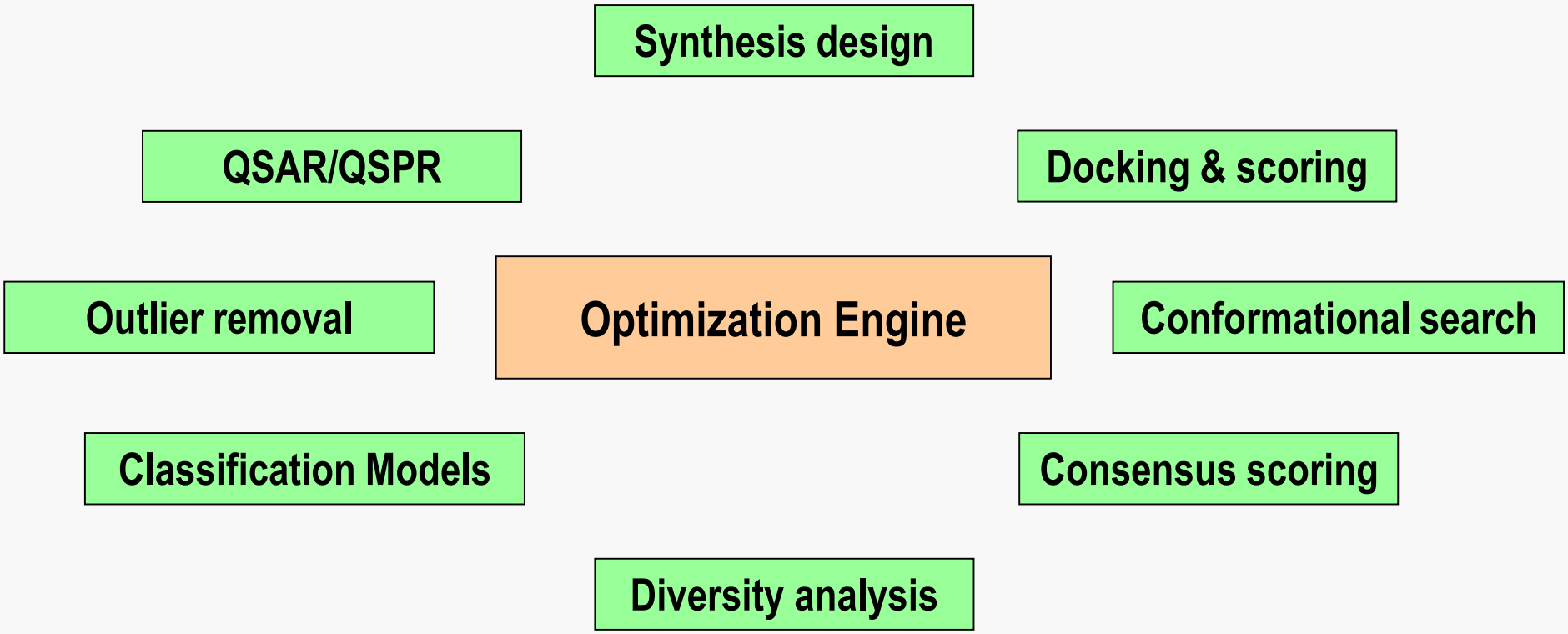
MuTaLig COST ACTION, 1st Annual meeting 2016, USI Lugano(CH), July 21-22 2016

Presentation Outline

- Drug development as an optimization problem
- Single objective optimization (SOOP) and multi objective optimization (MOOP)
- Examples
- Future directions

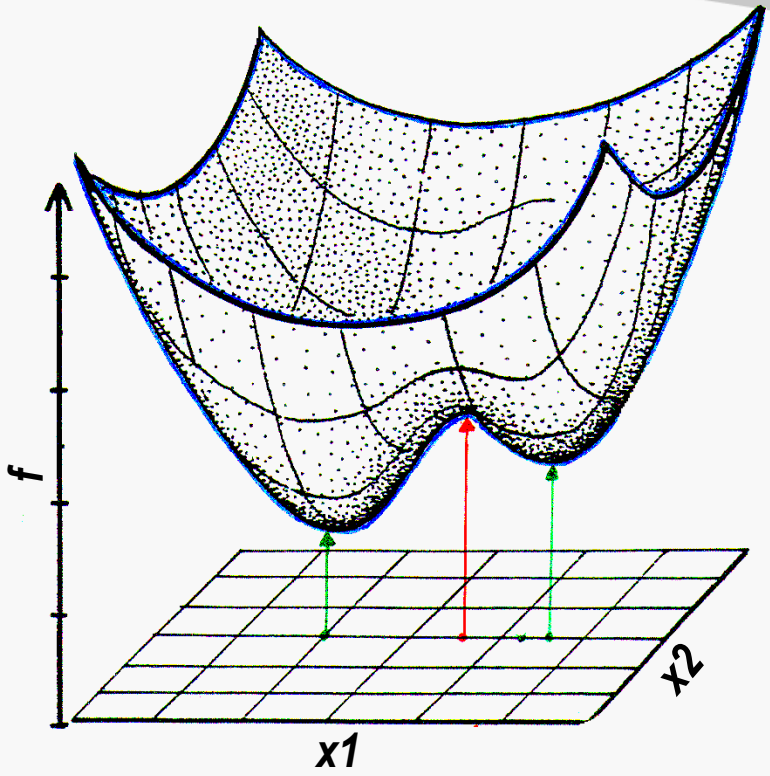
Optimization in Chemoinformatics and Drug Design

- Many problems in chemoinformatics and drug design could be formulated as (single or multi objectives) optimization problems

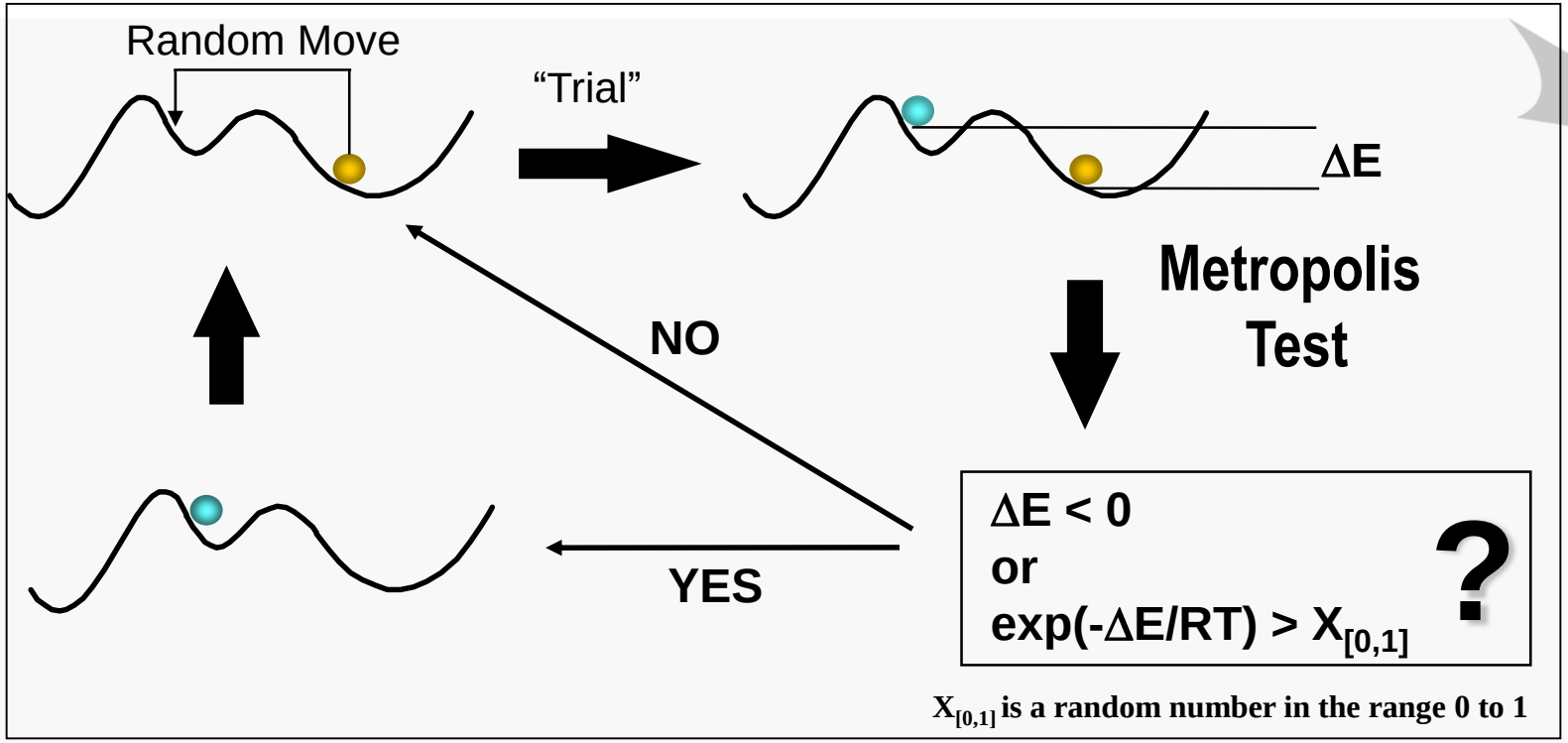


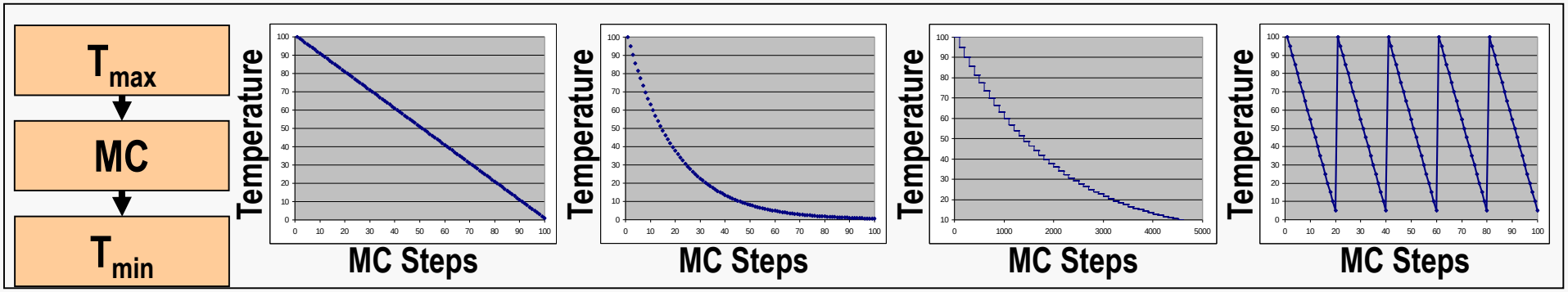
Optimization I: Target Function and Variables

- Define a target function (***f***) and corresponding variables $f = f(x_1,x_2,x_3 \ldots x_n)$
  - Target function and variables related to the scientific problem
  - Target function and variables define a multi-dimensional surface
- Target function often non-differentiable
- Location of minima is unknown
- Identify minima
  - Global minimum
  - A set of low lying minima



Optimization II: Monte Carlo/Simulated Annealing (MC/SA) Based Optimization Engine



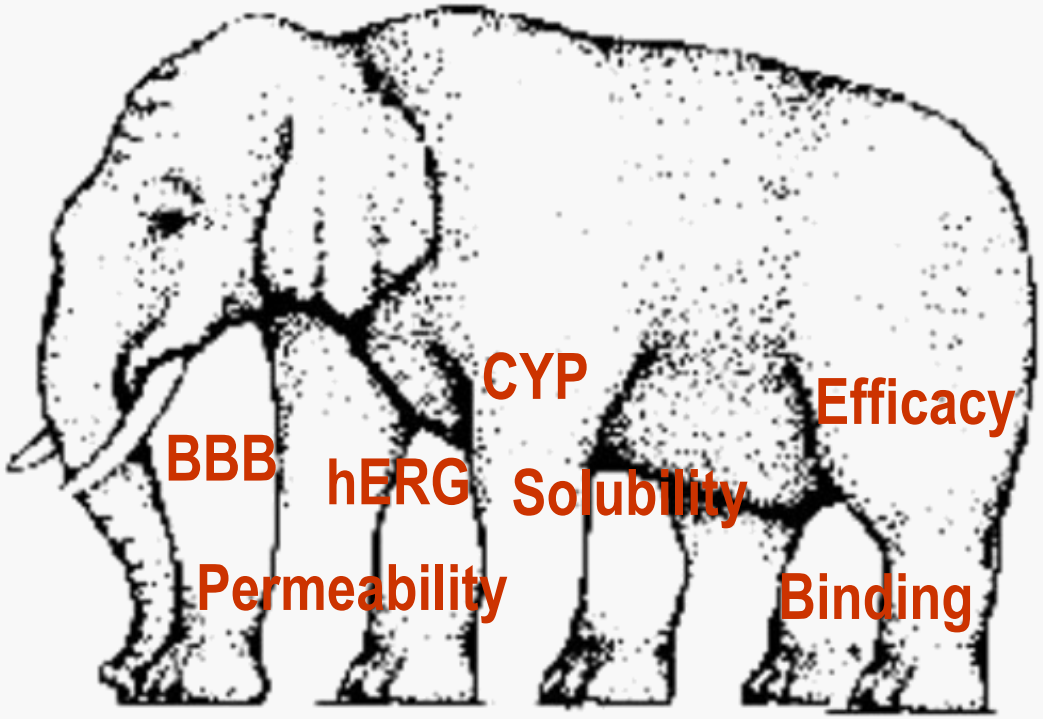


Optimization Problems

- Single objective
  - A global minimum exists (although it may be difficult to locate it)
- Multi objective
  - To single solution
  - A set of equally good (non-dominated) solutions exist that representing various compromises among the objectives

Lead Optimization: The Art of Balance

- Successful drug candidates necessarily represent a compromise between numerous, sometimes competing objectives
- Drug design is a multi-objective optimization problem

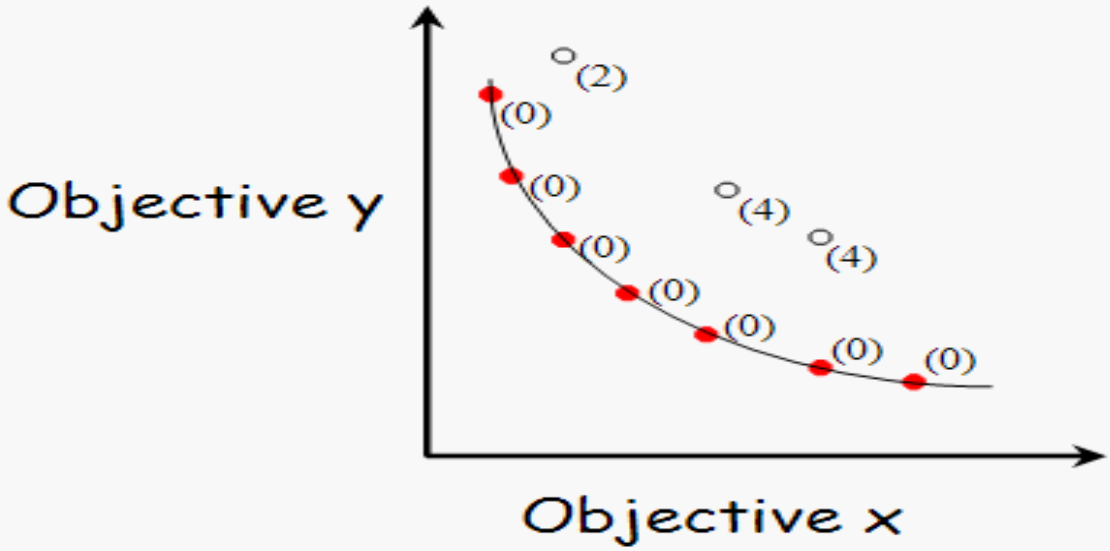


Strategies for Solving MOOP

- Transferring MOOP into SOOP

$$f(n)=w_1(Objective_1)+ w_2(Objective_2)+\ldots+ w_n(Objective_n)$$

- Pareto Optimization



  - A set of equally good (non-dominated) solutions, that representing various compromises among the objectives

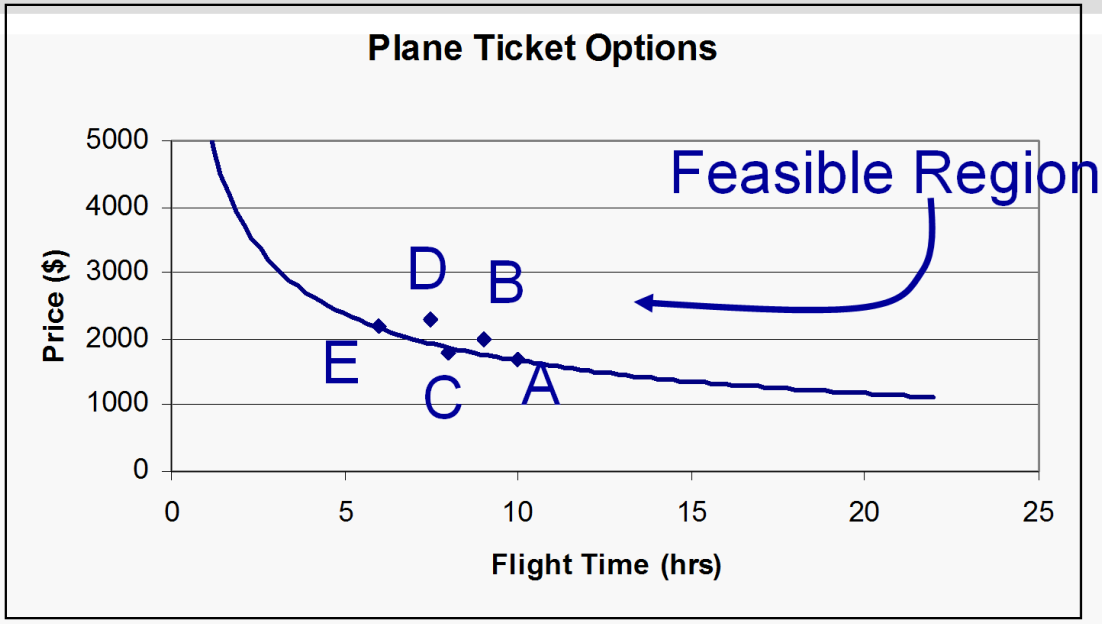
Pareto Optimization: Example

- Suppose you need to fly on a long trip: Should you choose the cheapest ticket (more connections) or shortest flying time (more expensive)?

Ticket	Time of travel (hr.)	Cost ($)
A	10	1700
B	9	2000
C	8	1800
D	7.5	2300
E	6	2200

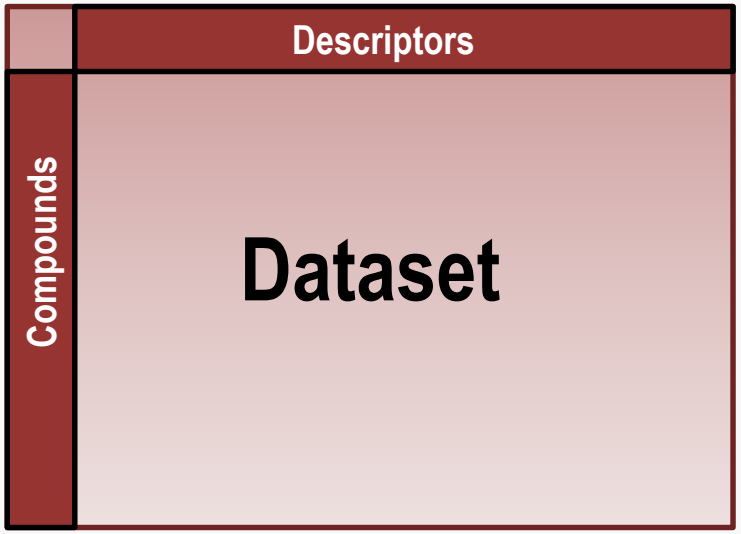
- If we compare tickets A & B, we can't say that either is superior without knowing the relative importance of Travel Time vs. Price
- However, comparing tickets B & C shows that C is better than B in both objectives, so we can say that C *"dominates"* B
- So, as long as C is a feasible option, there is no reason we would choose B
- If we finish the comparisons, we also see that D is dominated by E
- The rest of the options (A, C, & E) have a trade-off associated with Time vs. Price, so none is clearly superior to the others.
- We call this the "non-dominated" set of solutions become none of the solutions are dominated

Pareto Optimization: Example



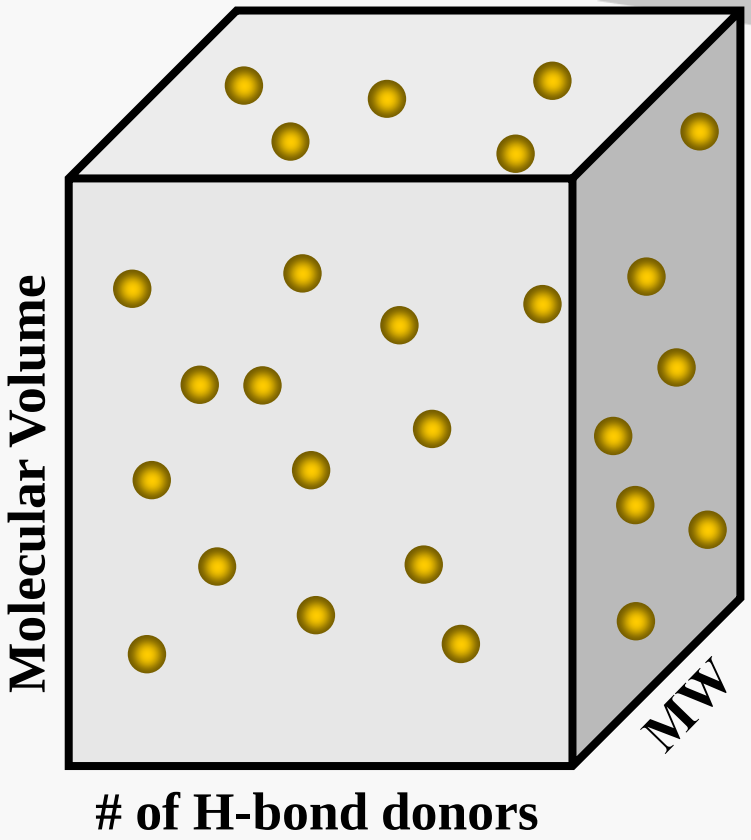
- Solutions that lie along the line are non-dominated solutions while those that lie inside the line are dominated because there is always another solution on the line that has at least one objective that is better
- The line is called the Pareto front and solutions on it are called Pareto-optimal
- All Pareto-optimal solutions are non-dominated
- Thus, it is important in MOO to find the solutions as close as possible to the Pareto front & as far along it as possible

Dataset



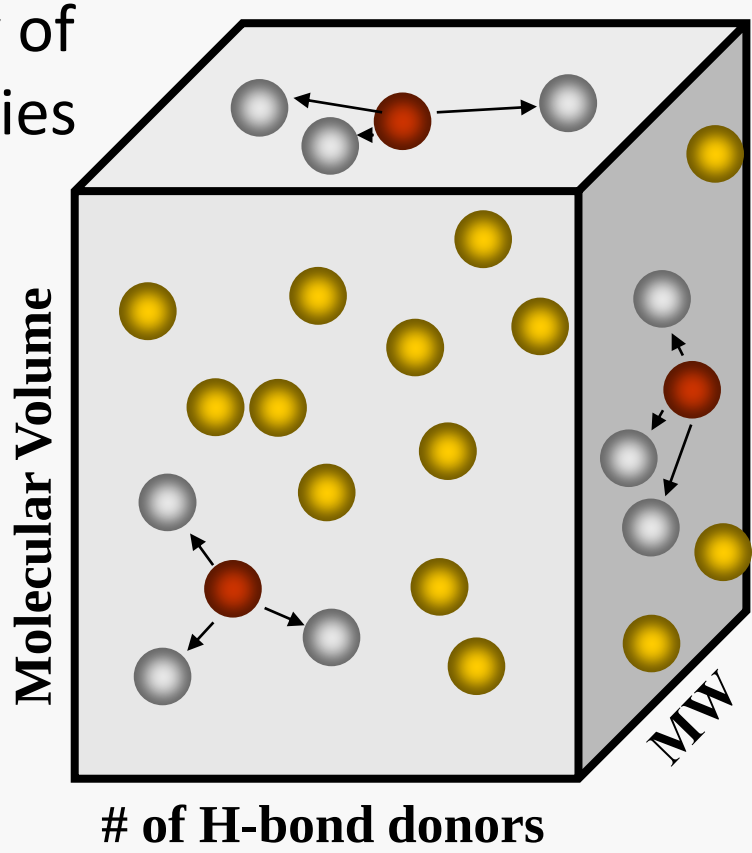
The Concept of a Property Space

- Each axes describes a molecular property (descriptor).
- Each molecule is represented by a point
- Euclidean distance is calculated in terms of the normalized descriptors
- The distance between any two points represents the degree of similarity between the corresponding molecules in terms of the selected descriptors.

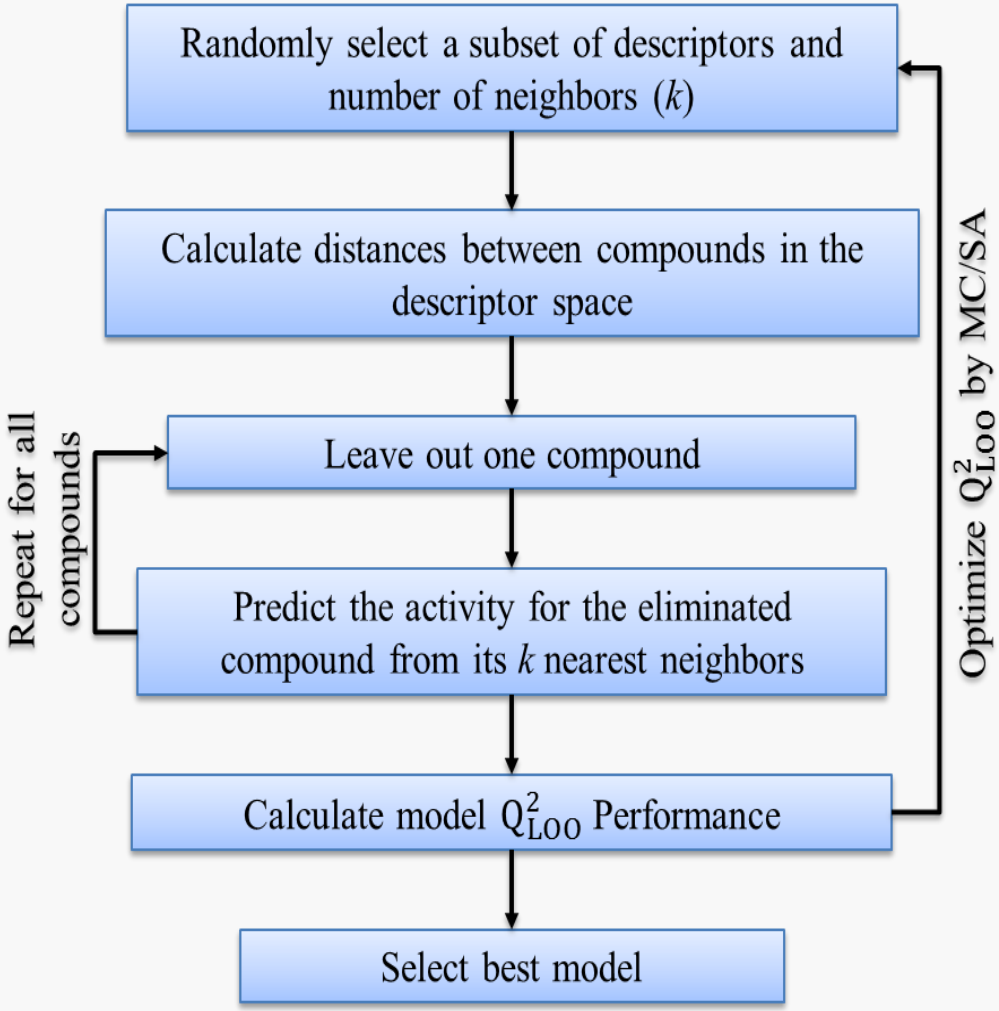


QSAR Modeling as a SOOP: *k* Nearest Neighbors (*k*NN)

- *The idea:* Similar compounds have similar activities.
- *The method: kNN* predicts the activity of a compound from the averaged activities of its *k* nearest neighbors
- *The challenge:* Identify the relevant descriptors space
- *Advantages:* Non-linear

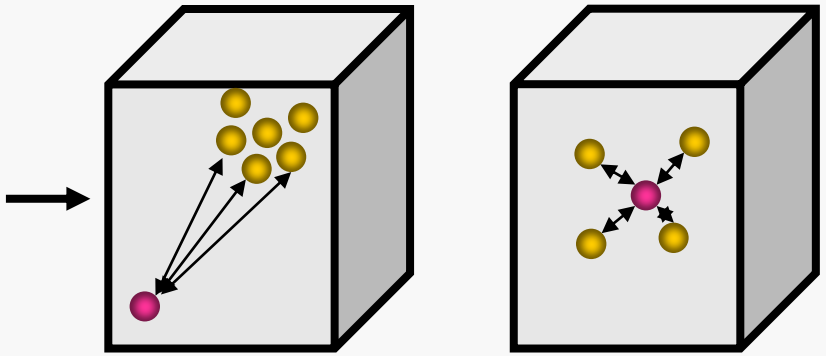


QSAR Modeling as a SOOP: *k* Nearest Neighbors (*k*NN)



k Nearest Neighbors (*k*NN) optimization based outlier removal

- Remove outliers for a better description of the bulk
- Based on the similar properties principle, predictions over a long distance are less accurate than predictions over a short distance
- Remove the compound which provides the largest increase in Q^2_{LOO} upon its removal from the data set.
- SOOP => maximizes Q^2_{LOO}
- MOOP => to minimize the number of compounds to be removed and maximizes *k*NN-derived Q^2_{LOO}.



stopping criteia $Q^2_{LOO} \geq 0.85$

DataBase	*N*	*SOOP*	*MOOP*	*%*
LogBBB	152	19	13	32%
F7	355	22	20	9%
DHFR	673	87	75	14%
		AVERAGE		18%

QSAR Modeling as a MOOP

Define biological targets

↓

Obtain experimental data (e.g., CHEMBL)

↓

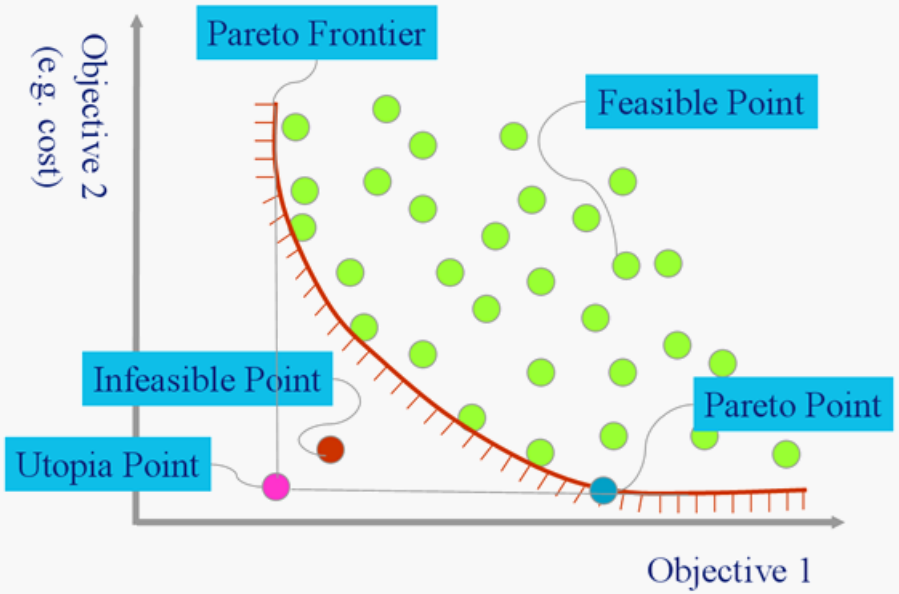
Develop predictive QSAR models

↓

Virtually screen databases using models

↓

Select compounds based on multi-objective optimization



Future Directions

- Introducing MOOP as the method of choice for the optimization of multi-target ligands
- Building a database of QSAR models relevant to the COST action
- Virtual screening using multiple QSAR models (MOOP)

Acknowledgments

Prof.Hanoch Senderowitz

- Dr. Oren Nahum
- Omer Kaspi

INNI Israel National Nanotechnology Initiative


THANK YOU