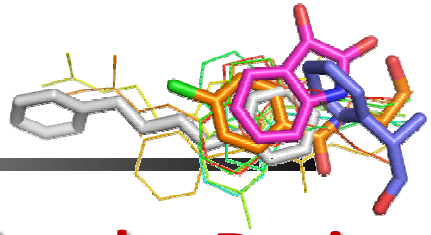
GOETHE UNIVERSITÄT FRANKFURT AM MAIN

De Novo Design



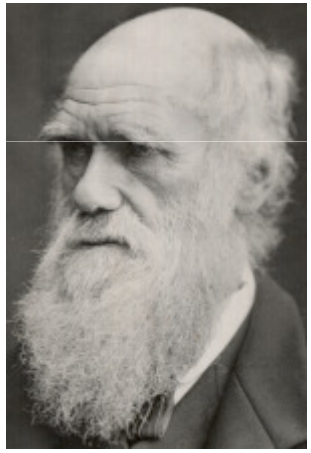
Adaptive Systems for Molecular Design

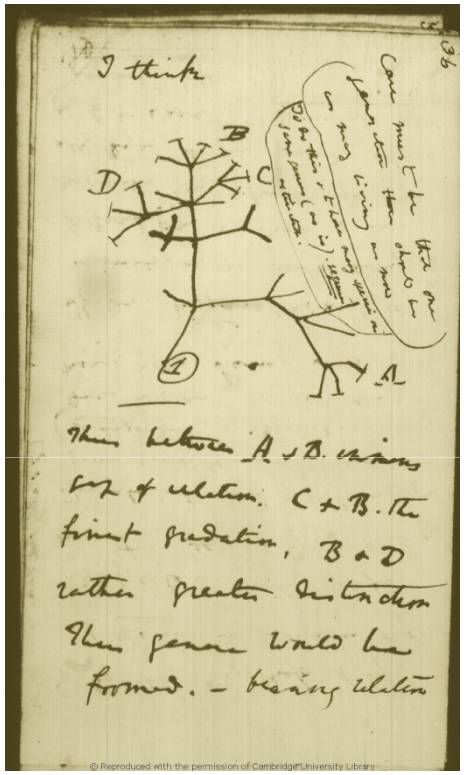
Gisbert Schneider

Beilstein Endowed Chair for Chem- & Bioinformatics

Goethe-University Frankfurt

Transmutation of Species



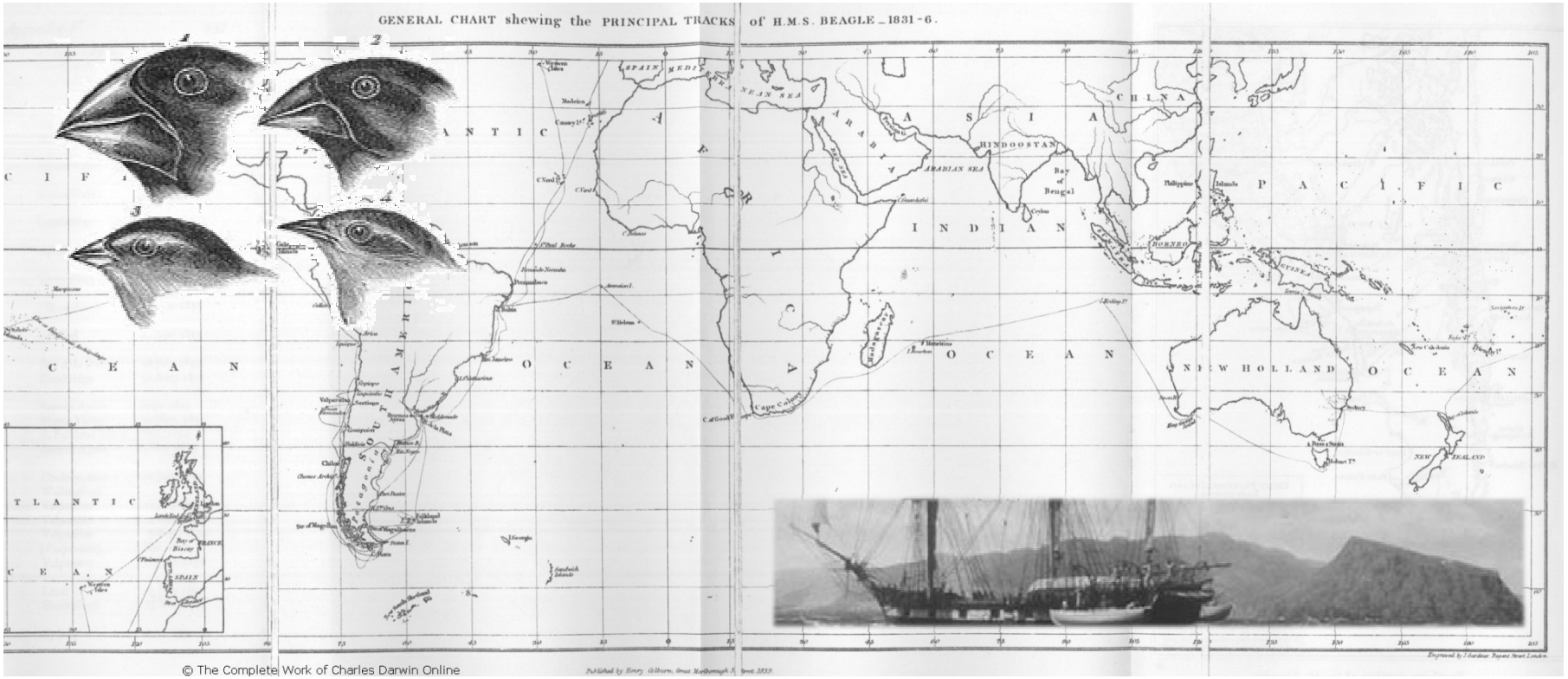


"[..] Thus between A & B immense gap of relation. C & B the finest gradation, B & D rather greater distinction. Thus genera would be formed. — bearing relation"

Charles R. Darwin (*1809)
Notebook B: *Transmutation of Species* (1837-1838), p.36

Source: www.darwin-online.org.uk

Voyages to the (un)known



Darwin and Henslow, letters 1831-1860.
The growth of an idea (N. Barlow, ed. 1967).

Voyages of the H.M.S. *Beagle*

De Novo Design Concepts

Requirements

- **Structure sampling**
- **Structure assessment**
- **Search / Optimization Method**

Implementation

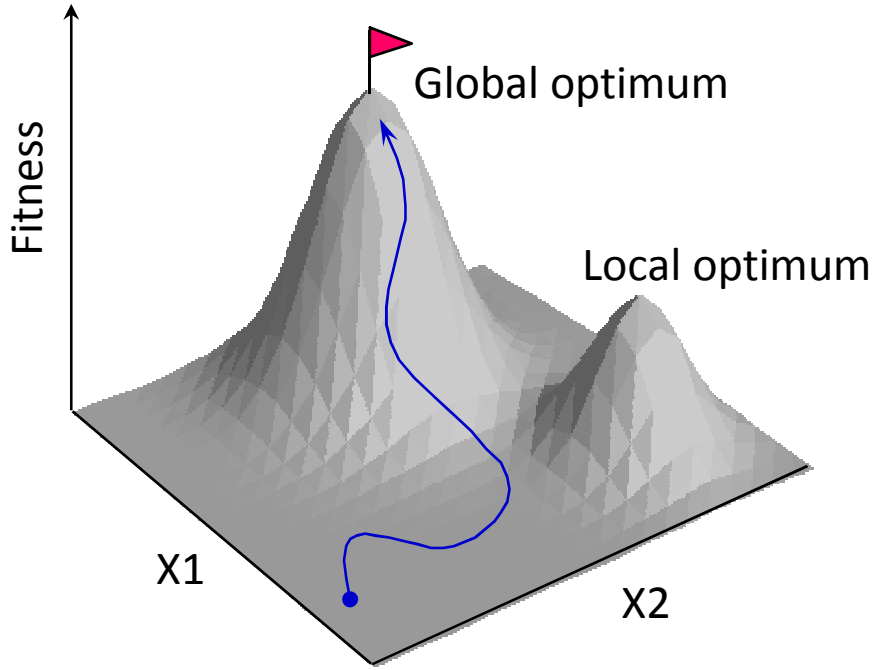
- Grow
- Link
- Lattice
- Stochastic

- Primary constraints (receptor, ligand)
- Secondary constraints

- Depth-first search
- Breadth-first search
- Random search
- Evolutionary Algorithm
- Monte Carlo / Metropolis
- Exhaustive enumeration
- (Free Energy Perturbation)

Schneider & Fechner (2005) Nat. Rev. Drug Discov. 4, 649.

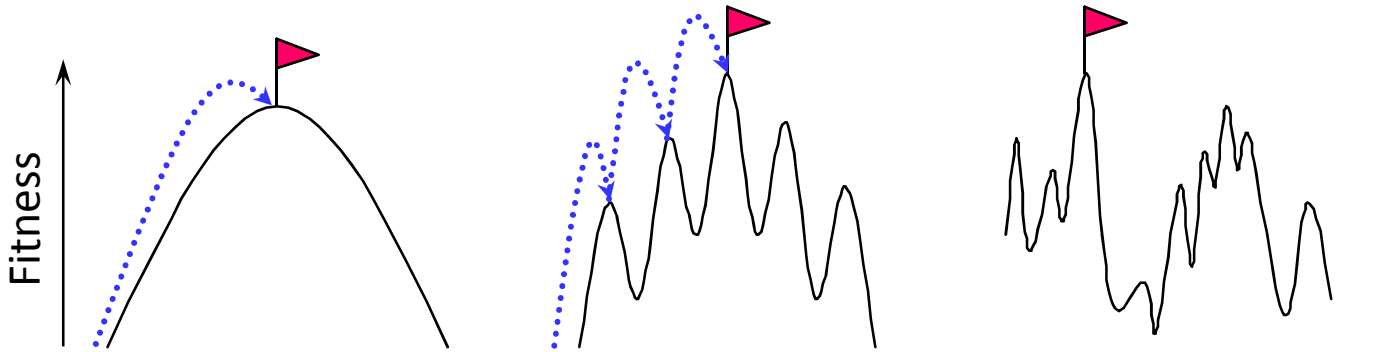
Adaptive Walk in a "Fitness Landscape"



Schneider & Fechner (2005) Nat. Rev. Drug Discov. 4, 649.
Fechner & Schneider (2007) JCIM 47, 656.

23 June 2008, (c) G. Schneider

Types of Fitness Landscapes

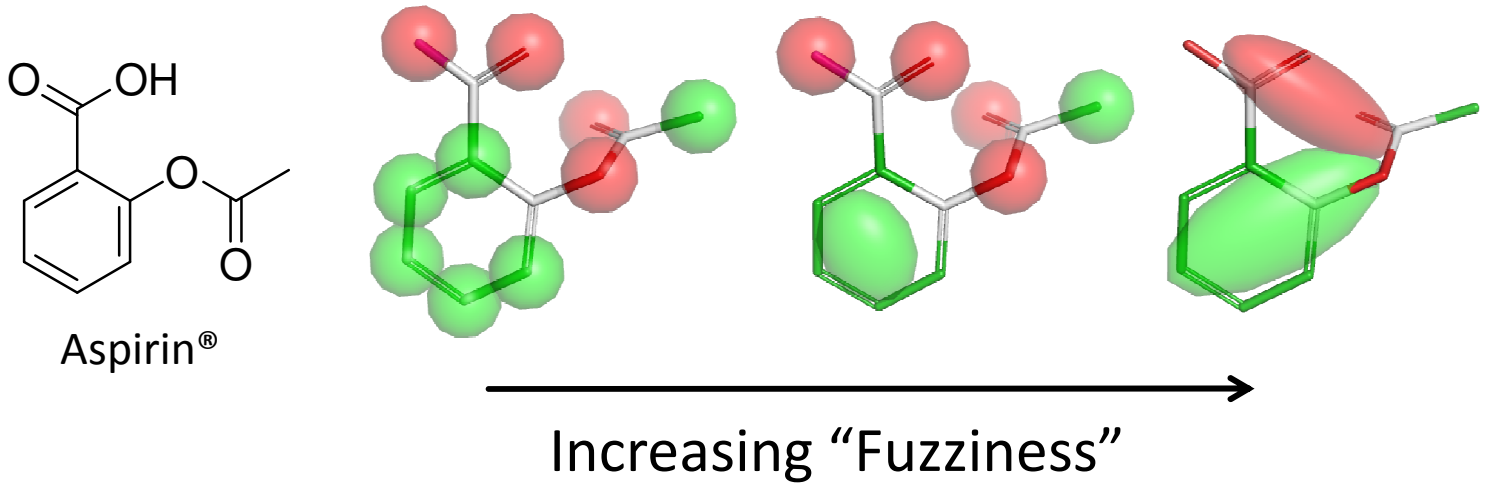


The Dream

The Nightmare

Chemical Similarity Principle!

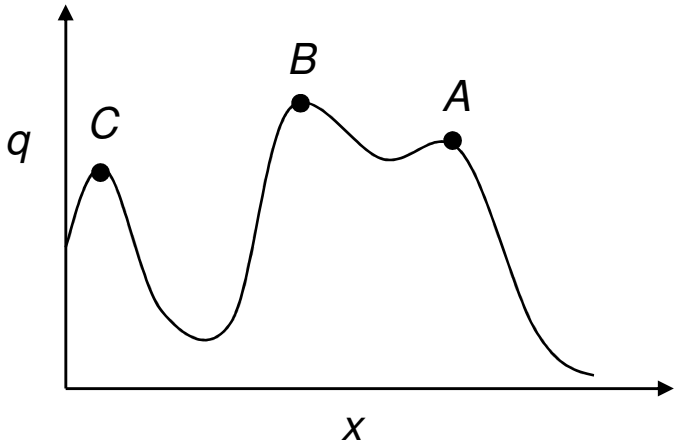
What is a Molecule?

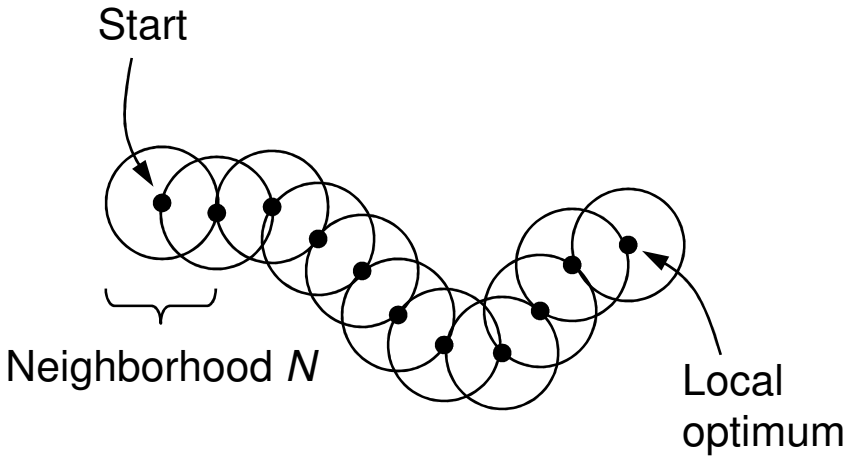


- Pharmacophoric representation on different levels of abstraction

Tanrikulu et al. (2007) ChemBioChem 8, 1932.
Renner & Schneider (2004) J. Med. Chem. 47, 4653.

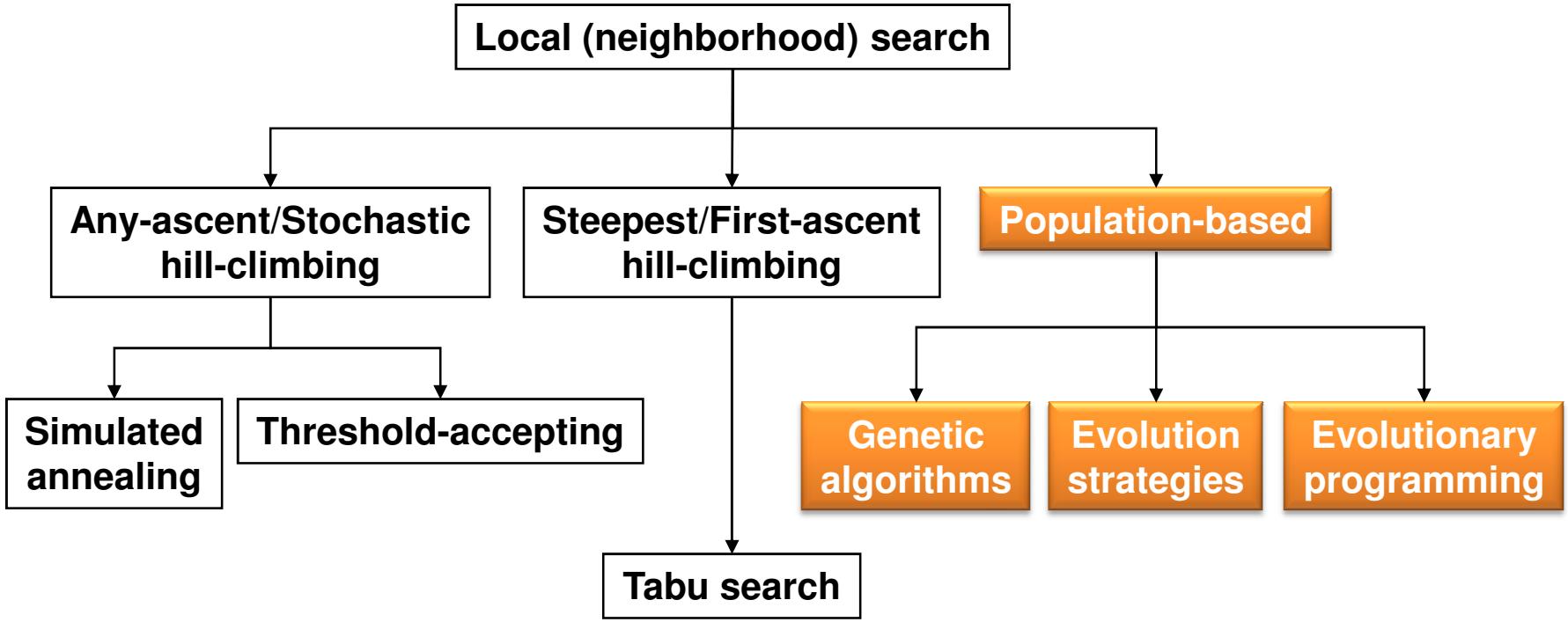
Local Neighborhood Search





- fixed *N*
- variable (“adaptive”) *N*

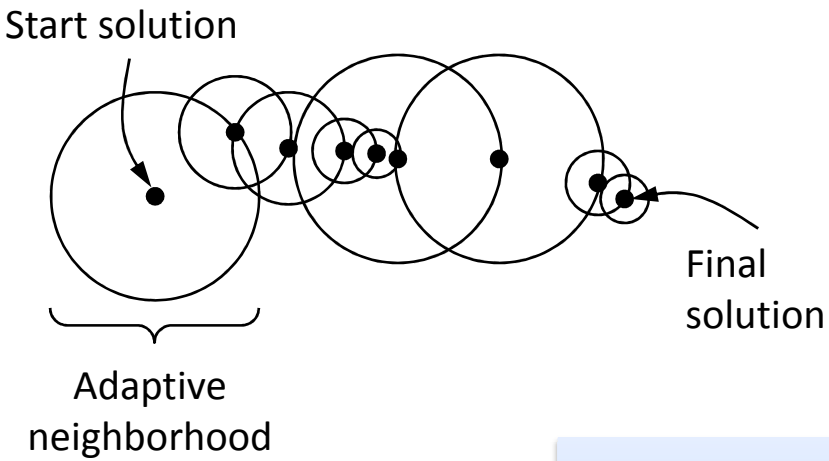
Local Neighborhood Strategies



Elements of Artificial Adaptive Systems (Koza, 1992)

- Structures that undergo **adaptation**
- Initial structures (starting solutions)
- Fitness measure that evaluates the structures
- Operations to modify the structures
- State (**memory**) of the system at each stage
- Method for designating a result
- Method for terminating the process
- Parameters that **control** the process

Why "Adaptive"?



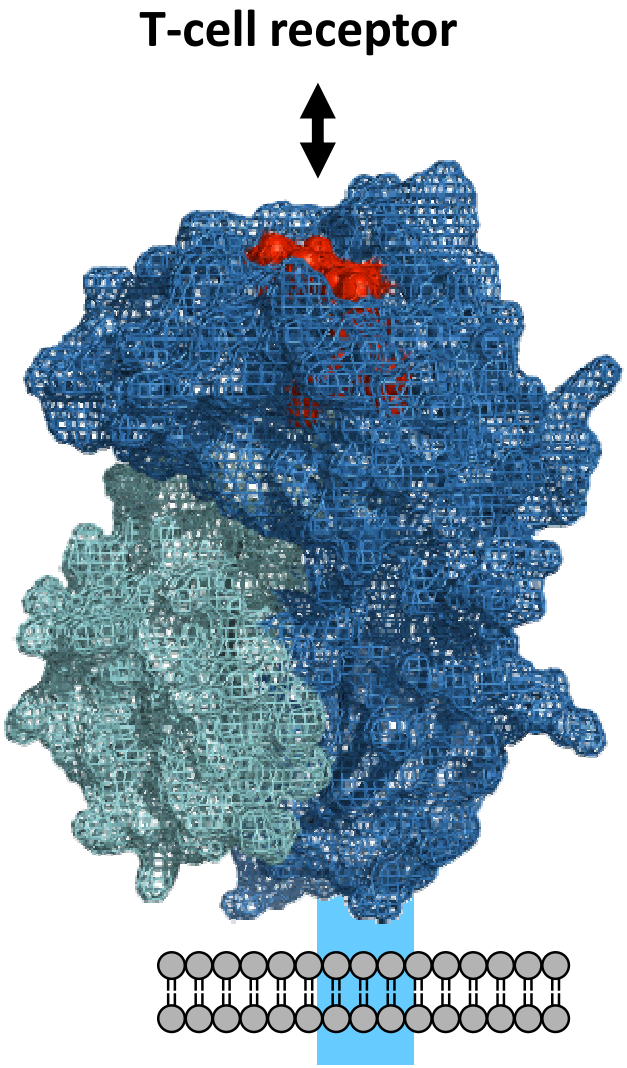
Adaptive optimization copes with …

- non-additive fitness functions
- multiple fitness functions
- very large search spaces
- nonlinear system behavior

Learning from Nature

- **Genetic algorithms**
- **Evolution strategies**
- **Particle Swarm Optimization (PSO)**
- **Ant Colony Optimization (ACO)**

Ant Colony Optimization
A Simple Combinatorial Case: Peptide Design



(PDB: 1HSA, Madden *et al.* 1993)

- Design of novel antigens presented by MHC I (H-2Kb)
- Length: 8 residues (20^8 possibilities)

  Neural network ensemble + jury

  Artificial ant system

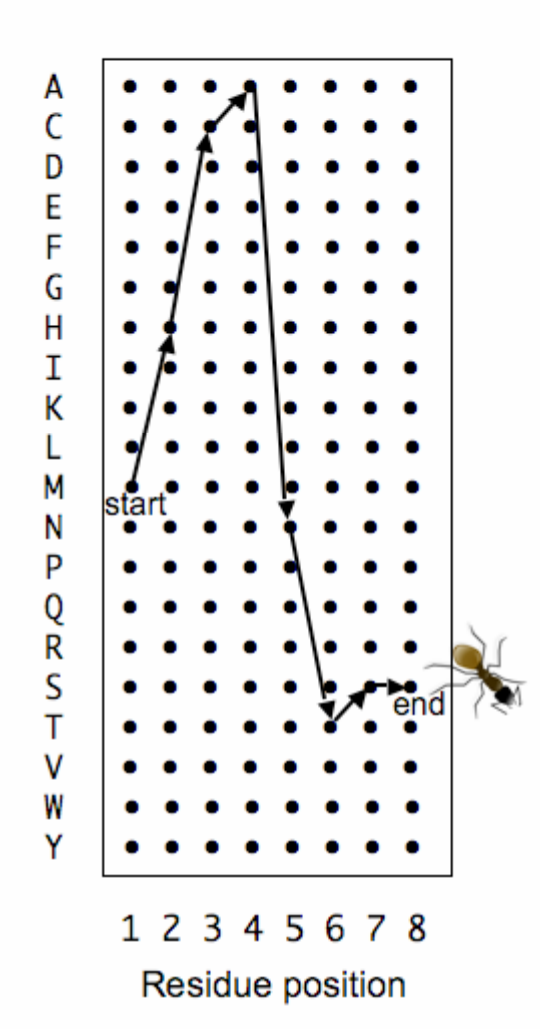
Assay confirms the predictions:

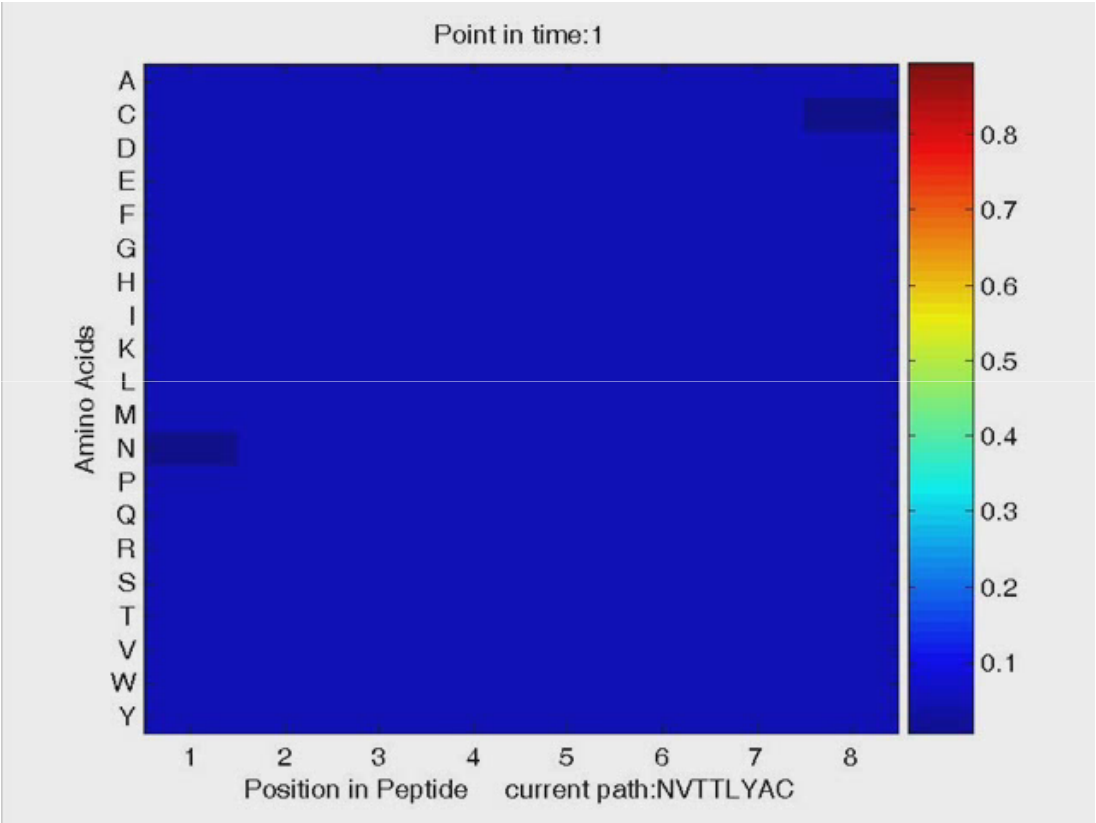
89% correct for binding peptides

95% correct for non-binding peptides

Hiss et al. (2007) PEDS 20, 99.
Schneider et al. (1998) PNAS 95, 12179.

Ant System for Combinatorial Design





Which Search Strategy Should be Applied?

Global & local search control guided by

- Random parameter variation
- Stepwise systematic parameter variation
- Gradients in the fitness landscape ...

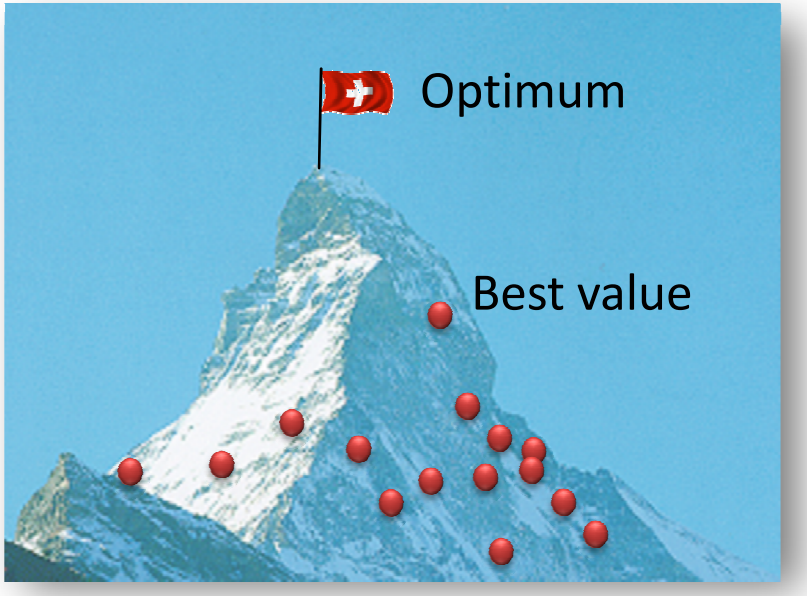
There is no one best method.

Evolutionary algorithms are

- easily implemented
- robust
- able to cope with high-dimensional optimization tasks

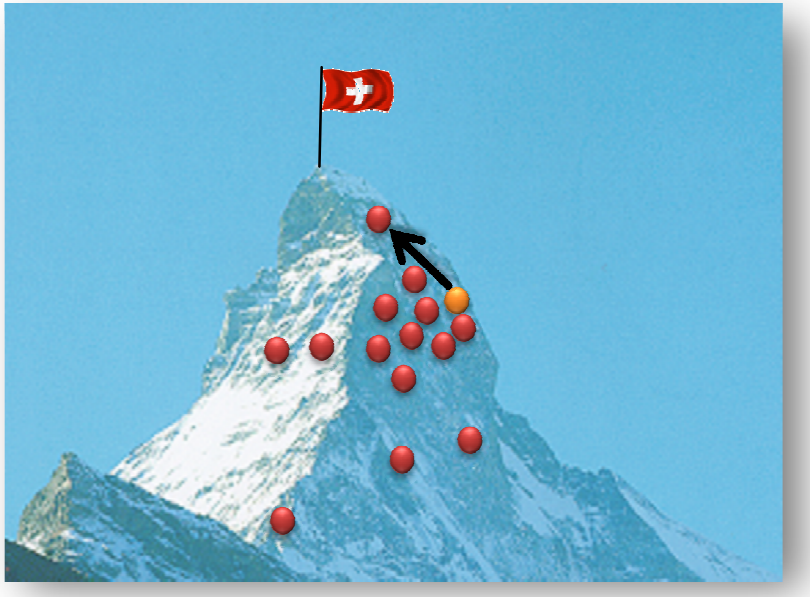
Principle of Evolutionary Strategies

Step 1: Random search



Generate a diverse set of parameter values and hope for a hit

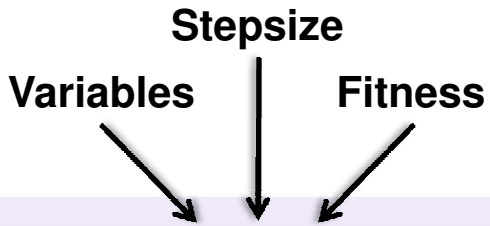
Step 2 and following:
"Local hill climbing"



Generate localized distributions of parameter values and improve steadily

Adaptive neighborhood

Algorithm of the (1,λ) Evolution Strategy (Rechenberg, 1973)



Initialize parent (x^P, s^P, F^P)

For each generation:

Generate λ variants (x^V, s^V, F^V) around the parent:

$s^V = \text{abs}(s^P + G)$

$x^V = x^P + s^V \bullet G$

Calculate fitness F^V

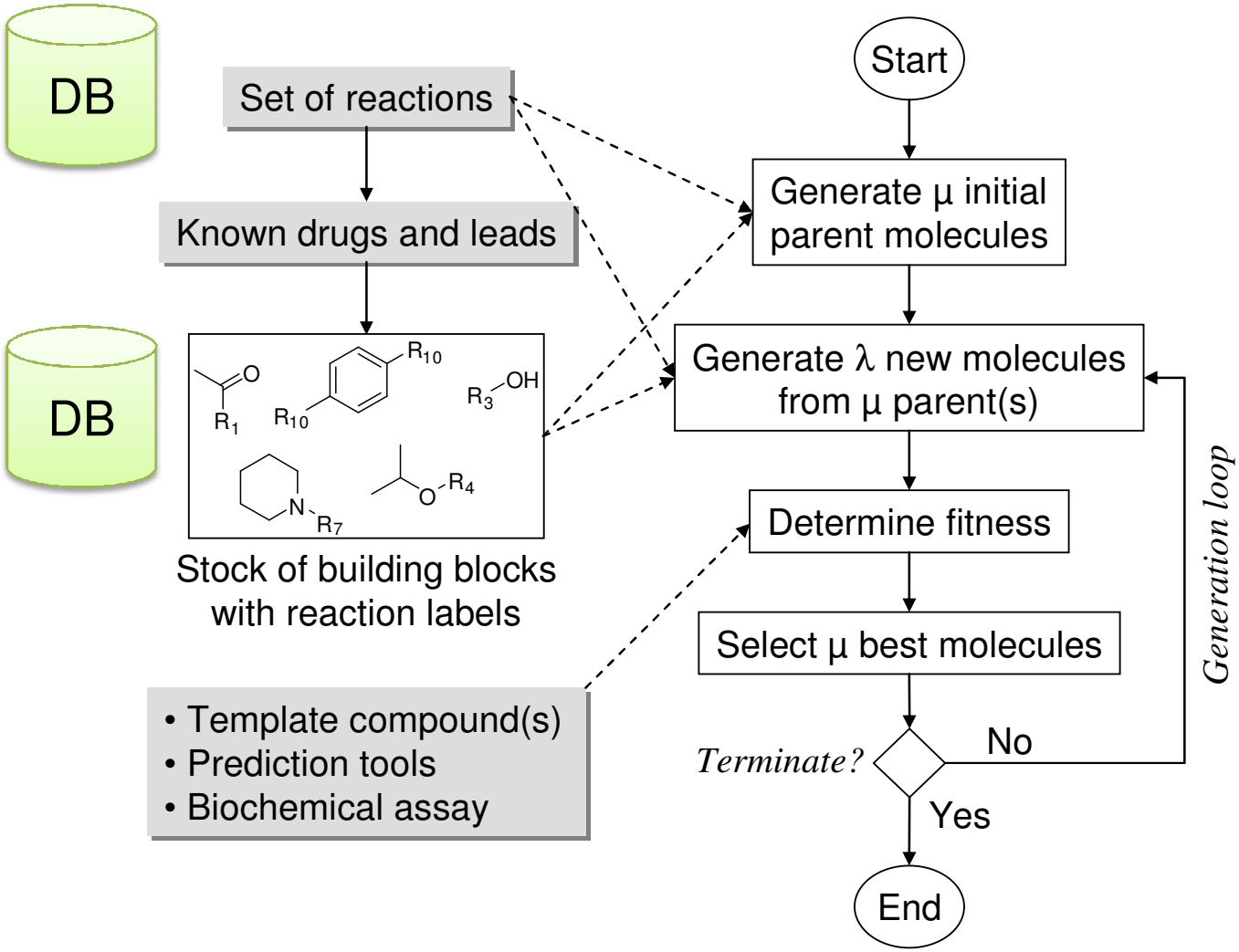
Select best variant according to F^V

Set $(x^P, s^P, F^P) = (x^V, s^V, F^V)^{best}$

$$G(i,j) = \sqrt{-2\ln(i)} \cdot \sin(2\pi j) \quad \text{in } [-\infty,\infty]$$

i,j: pseudo-random numbers in]0,1]

De novo Fragment Assembly



Schneider et al. (2000) Angew. Chem. Int. Ed. 39:4130

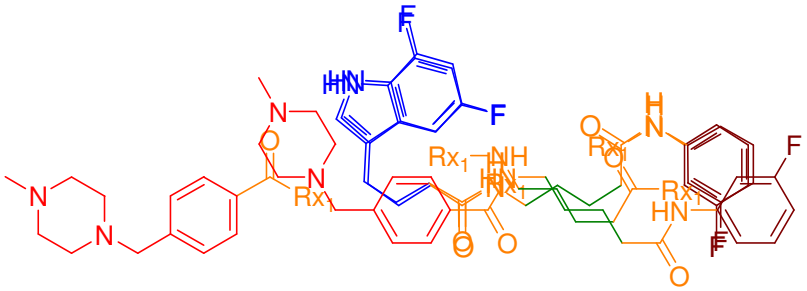
Operators of Evolutionary Optimization

Variation

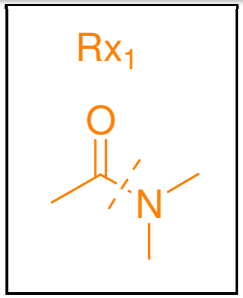
- **Mutation** — Adds new information to a population
- **Crossover** — Exploits information within a population
- **Selection**

Breeding New Molecules

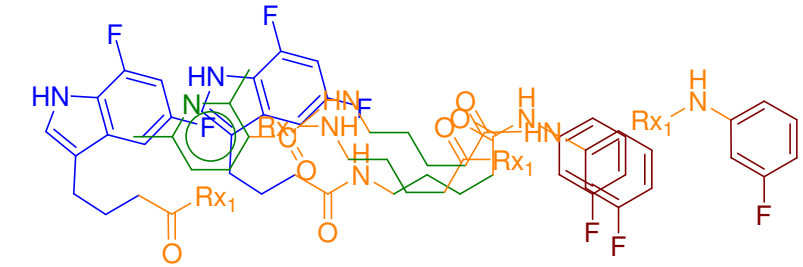
Mutation

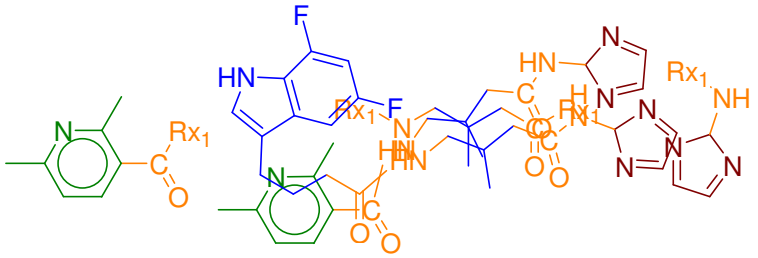


Virtual reaction

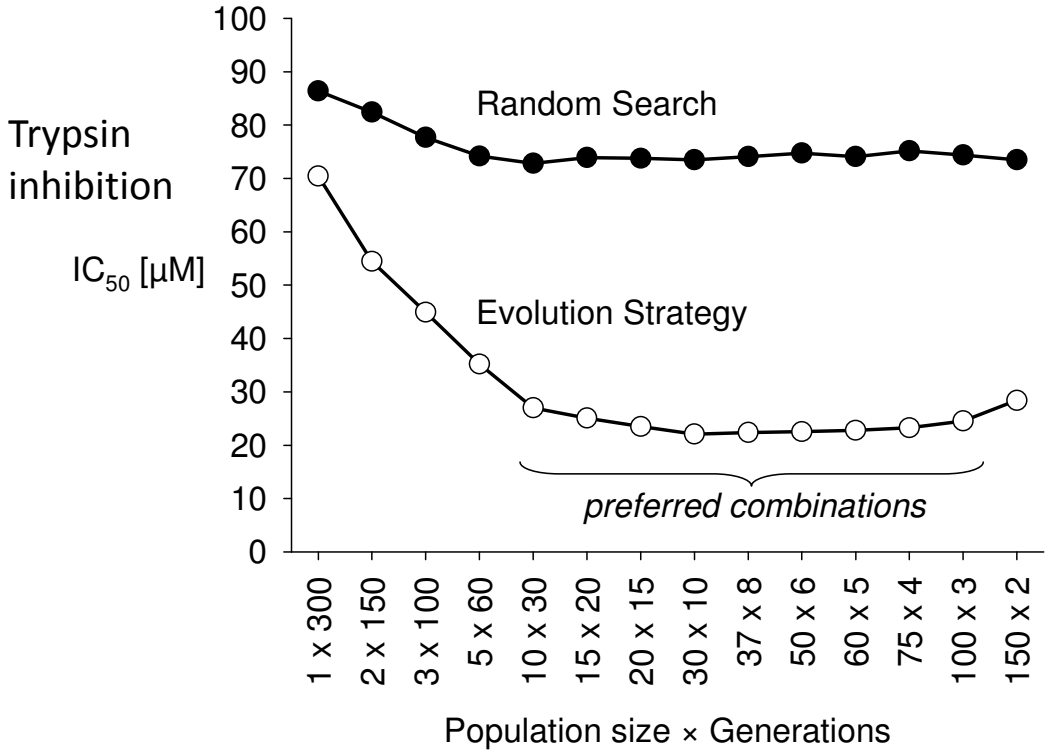


Crossover





How Many Iterations? How Many Compounds?

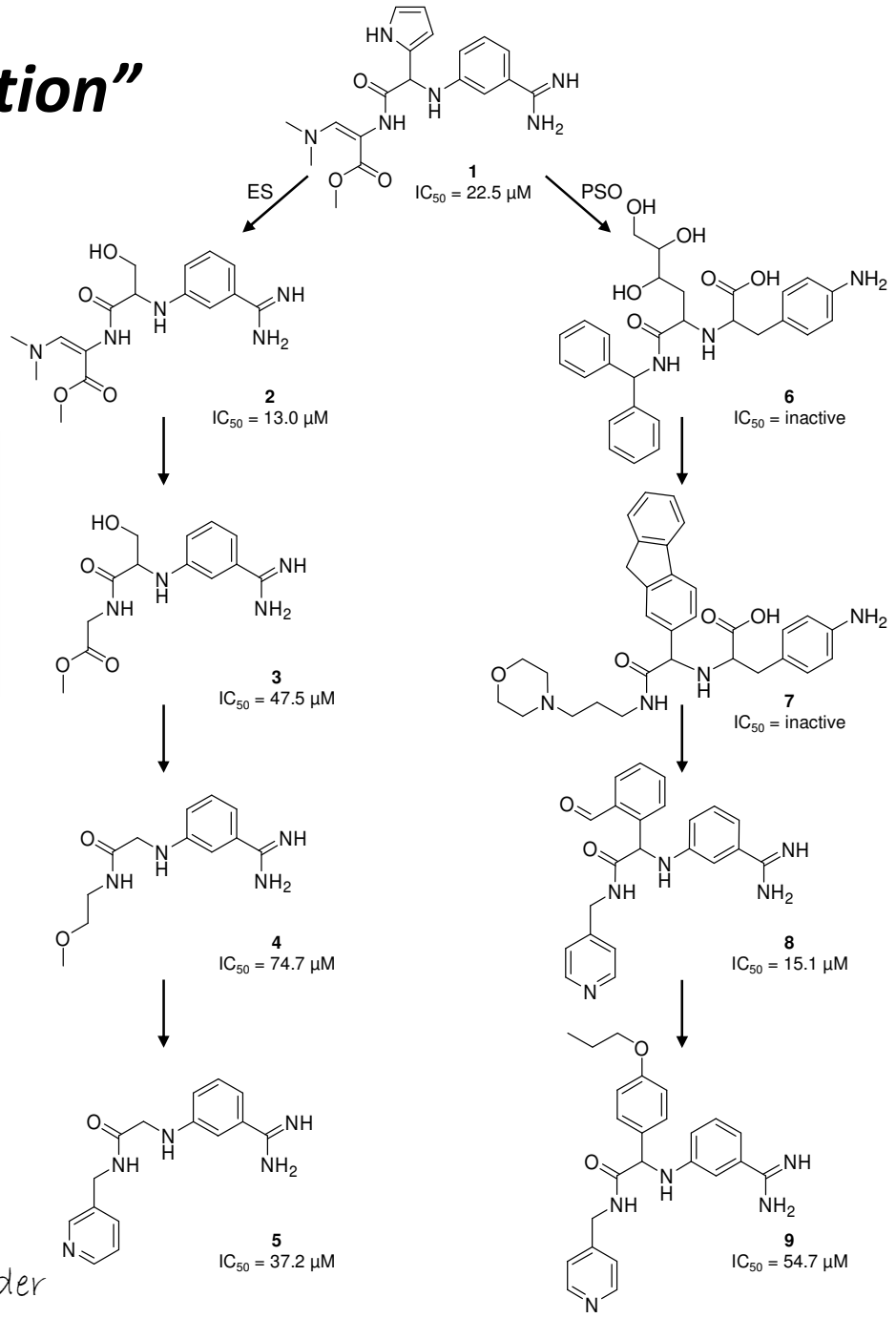


Limited resources:
300 compounds
There is an optimum!

Schüller & Schneider (2008) JCIM, in press.

"Molecular Evolution"

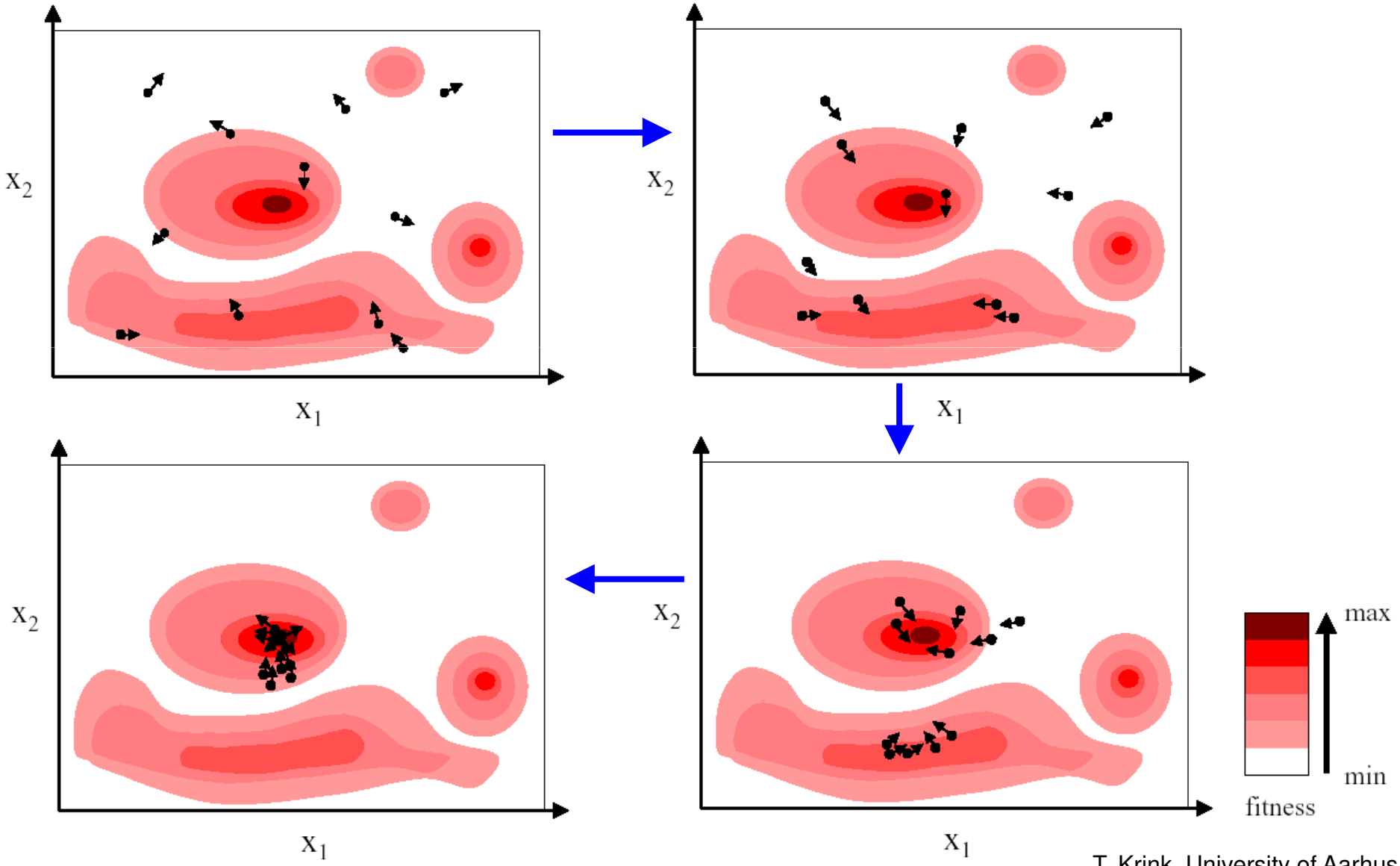
Trajectotries through inactive regions of chemical space are possible



Particle Swarm Optimization (PSO)

- Biological motivated optimization paradigm
- Individuals move through a fitness landscape
- Particles can change their **movement direction** and **velocities**
- **"Social" interactions** between individuals of a population
- **Each** particle knows about its **own** best position
- **All** particles know about the **overall** best position

Particle Swarm Optimization (PSO)



T. Krink, University of Aarhus

PSO: Update Rule 1 (Standard Type)

Individual memory

Social memory

$$\mathrm{v}_\mathrm{i}(t+1) = \mathrm{v}_\mathrm{i}(t) + 2 \cdot r_1 \cdot (p_i - x_i(t)) + 2 \cdot r_2 \cdot (p_b - x_i(t))$$

v: Velocity vector

x: Coordinates of a particle

i: Dimension

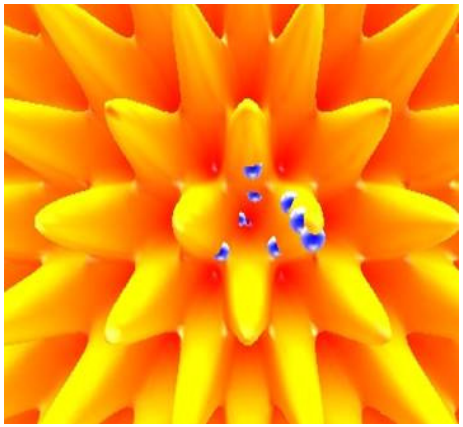
t: Time (epochs)

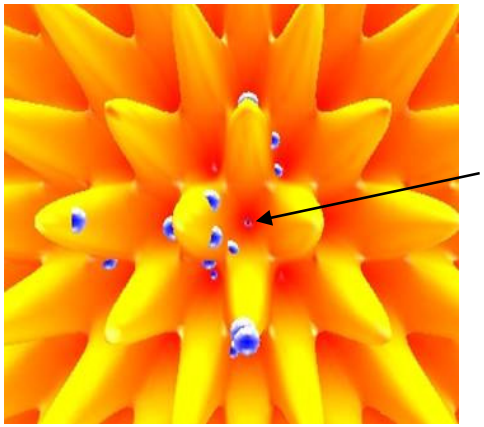
r_1 , r_2: Random numbers in]0,1]

p_i , p_b: Best positions (individual and social)

PSO: Individual and Social Memory

PSO convergence after 100 epochs (Griewangk's function)



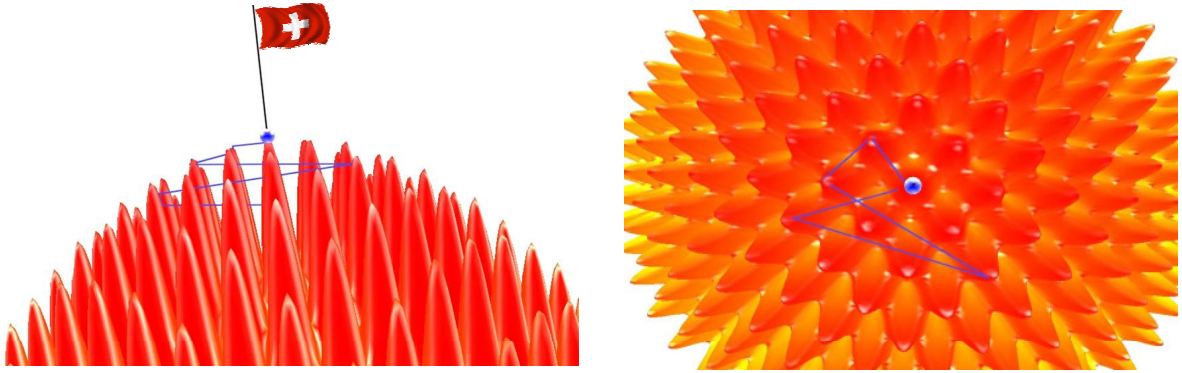


$n_1 = 1$
$n_2 = 2$

Higher weight on the social memory

$n_1 = 2$
$n_2 = 1$

Visualization of Particle Trajectories



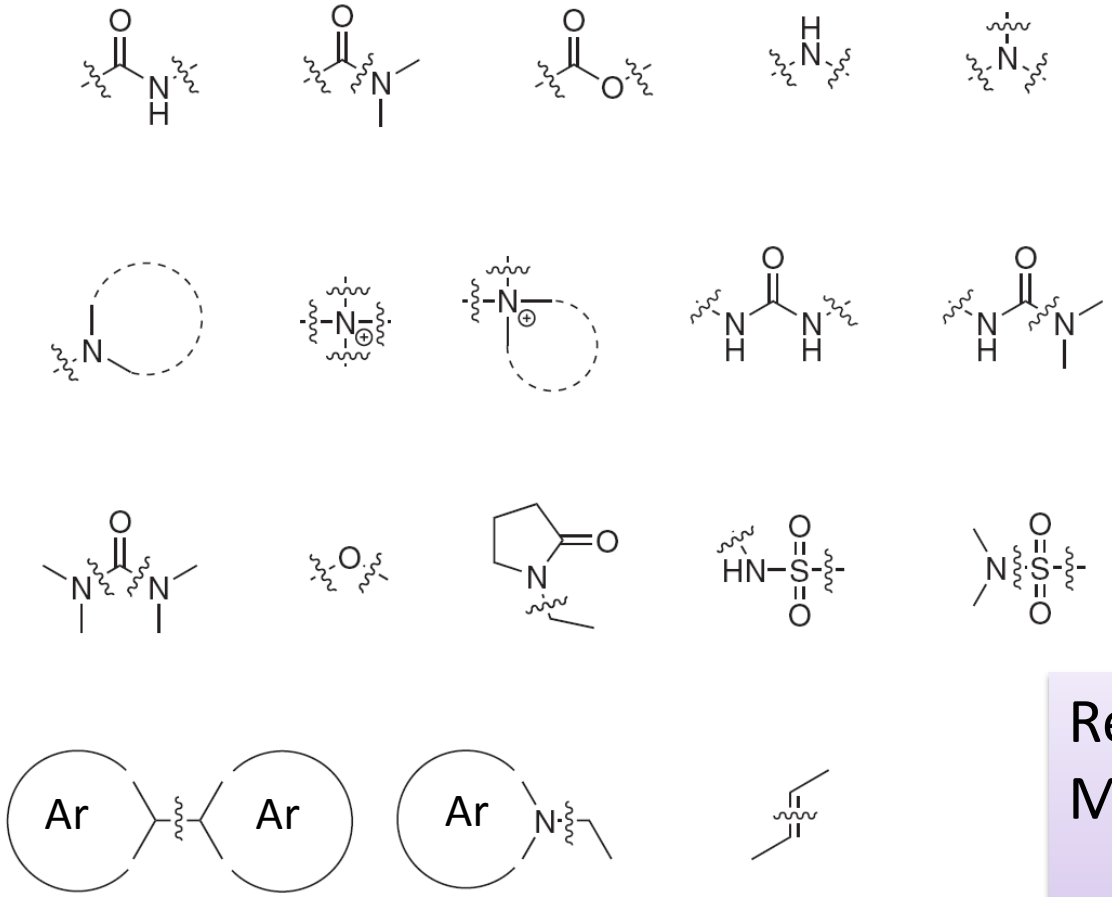
- The particle visited several local optima bevor finding the global optimum

PSO Demo

COLIBREE®

Combinatorial Library Breeding

Disassembly Rules

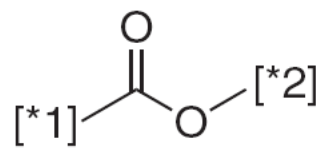


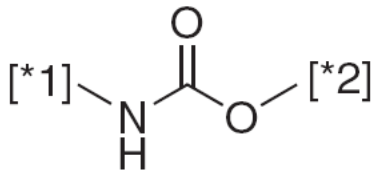
Retain building blocks with MW < 200 Da

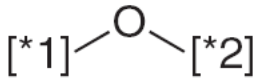
7,184 Building Blocks
avg MW = 141 ± 35 Da

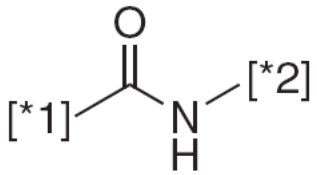
Hartenfeller et al. (2008) Chem. Biol. Drug Des., in press

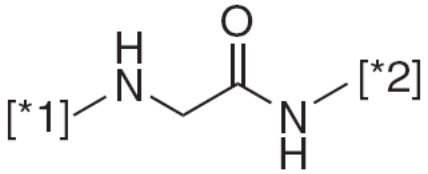
Linker Library

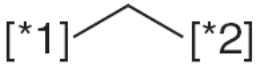


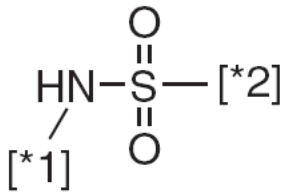


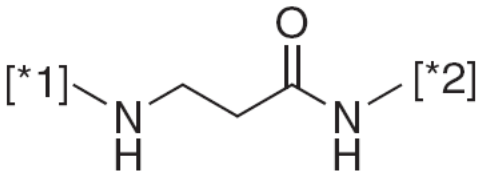


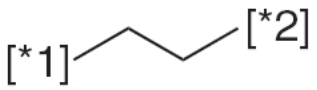


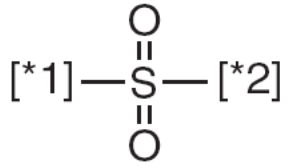


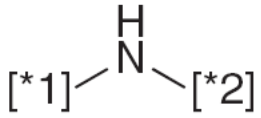


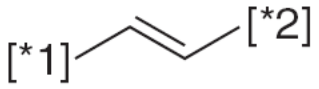




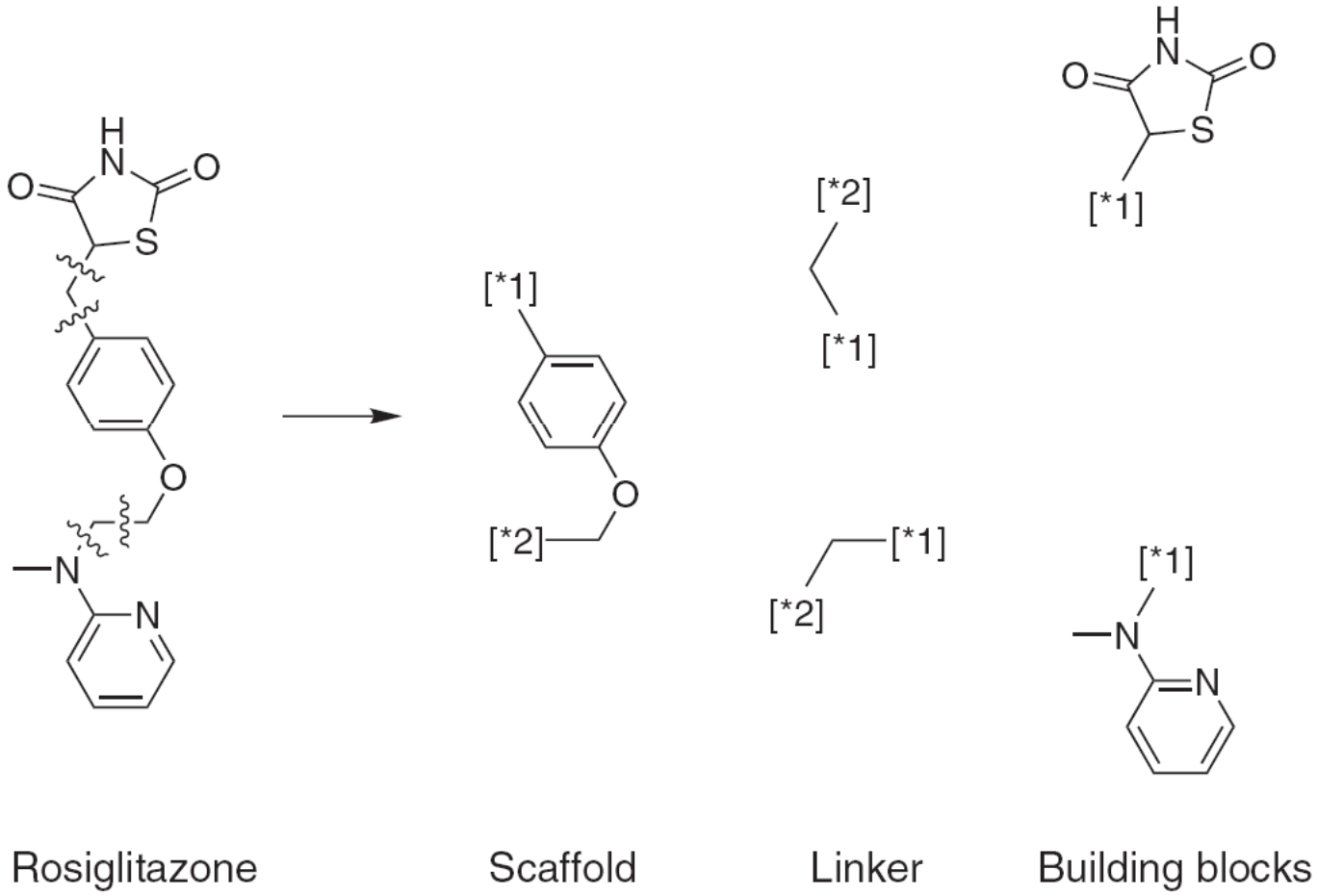




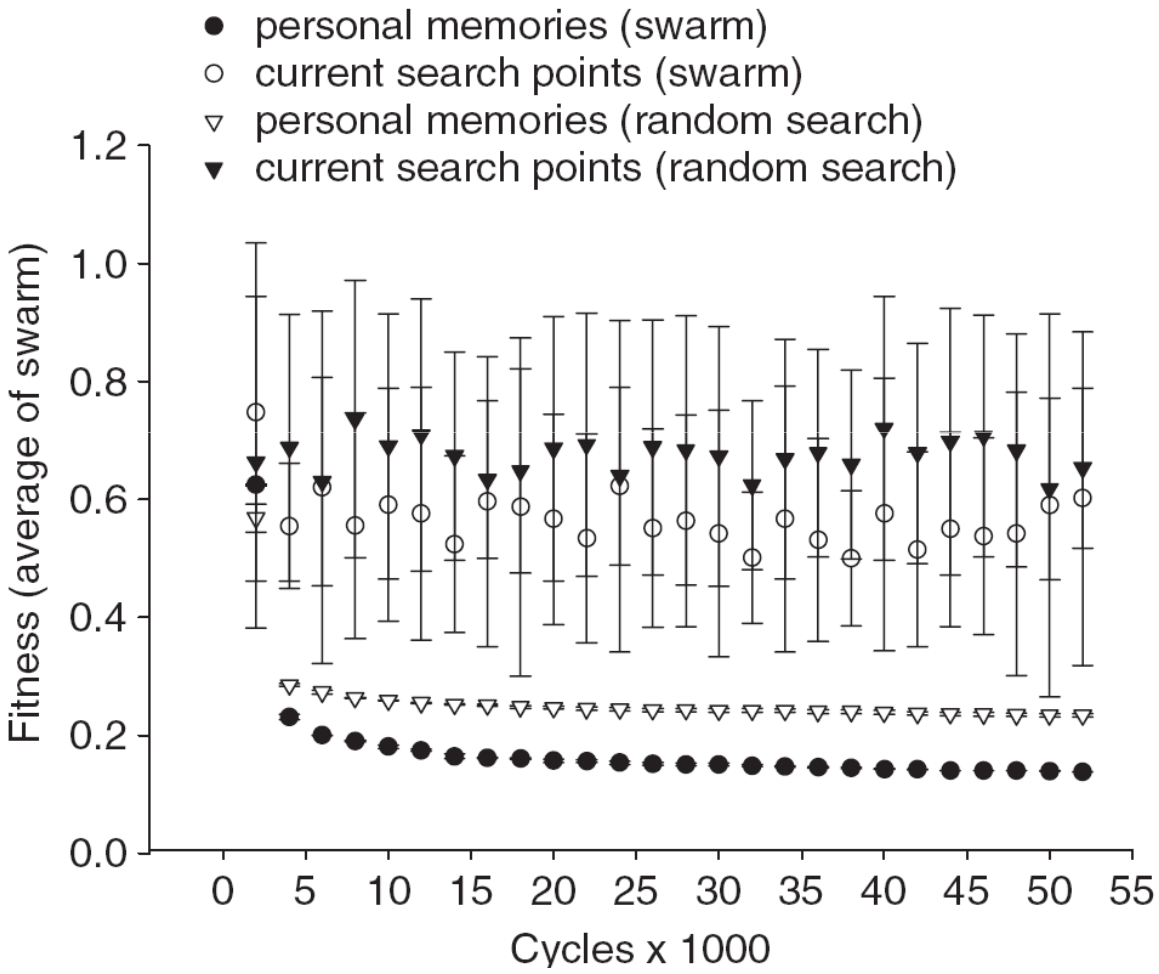




Molecule Dissection



PSO Progress Over Time



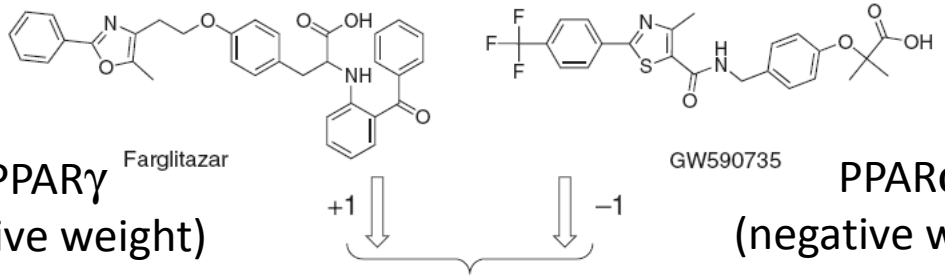
Library size:
40 molecules

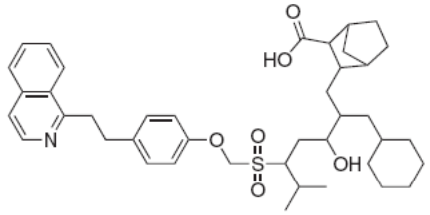
Fitness:
Distance to Reference
(minimization!)

Positive and Negative Design

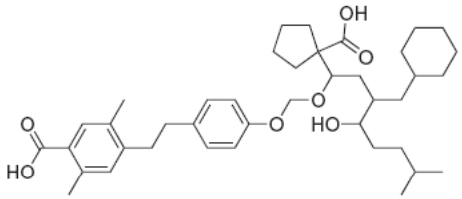
$$F(\mathbf{X}) = \sum_{j=1}^{r} w_j \bullet D_{\text{Euklid}}(\mathbf{X}, \mathbf{Y_j})$$

$$D_{\text{Euklid}}(\mathbf{X}, \mathbf{Y}) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

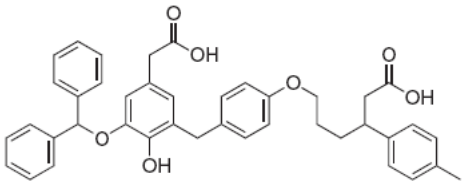




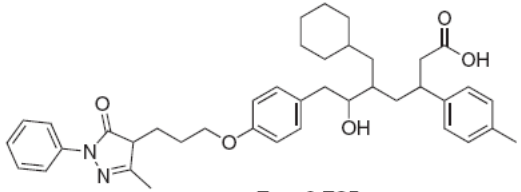
F = –0.728
TC_pos = 0.4805
TC_neg = 0.4217



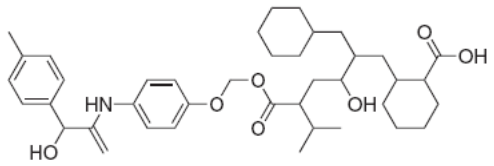
F = –0.727
TC_pos = 0.4923
TC_neg = 0.4225



F = –0.727
TC_pos = 0.5000
TC_neg = 0.3788



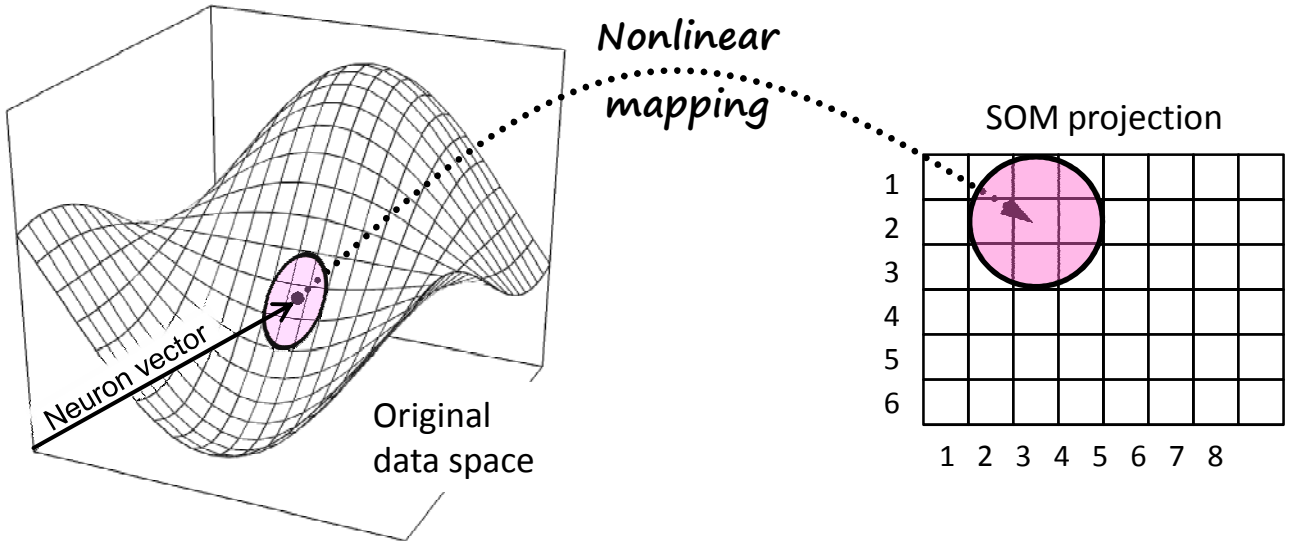
F = –0.725
TC_pos = 0.6912
TC_neg = 0.4875



F = –0.724
TC_pos = 0.5606
TC_neg = 0.3896

Self-Organizing Map (SOM)

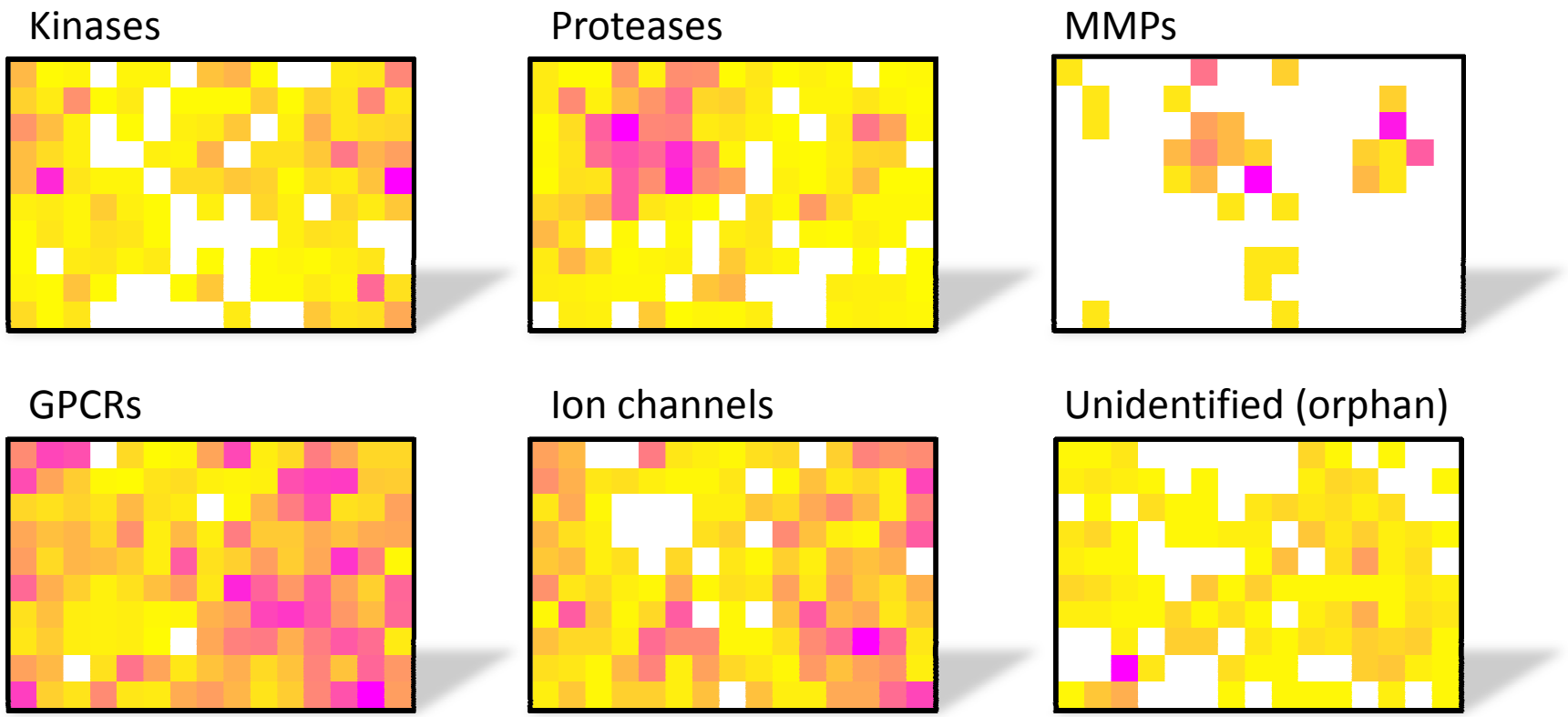
Kohonen , 1982



- Nonlinear projection of high-dimensional space
- Topology-preserving projection
- Data clustering
- Prediction with probability estimation

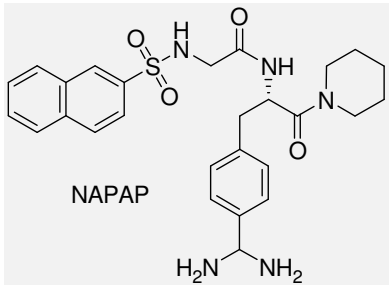
SOMMER Demo

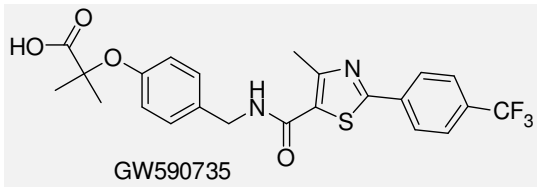
Pharmacophore Road Map of Chemical Space



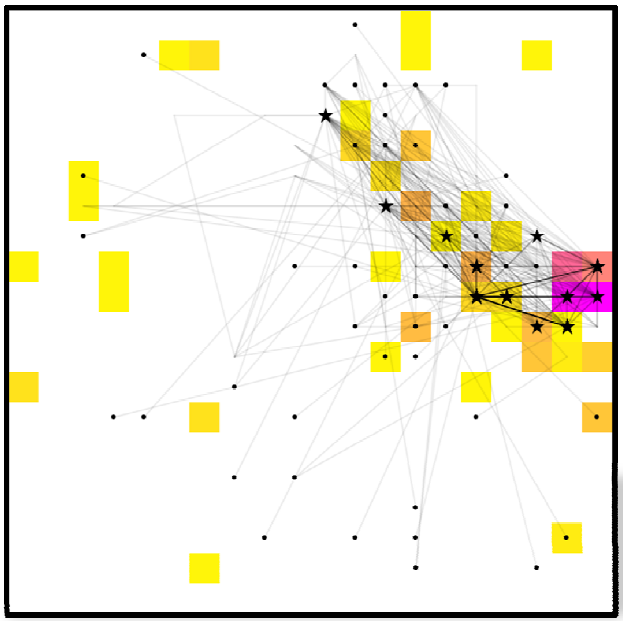
- Topological pharmacophores (CATS descriptor)
- COBRA collection (10,500 drugs and lead compounds)

Land Ho! Exploration of Chemical Space

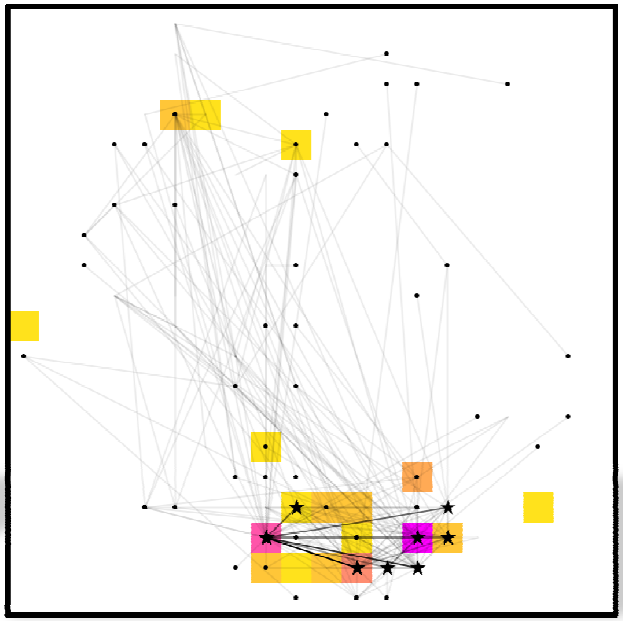




Thrombin inhibitors

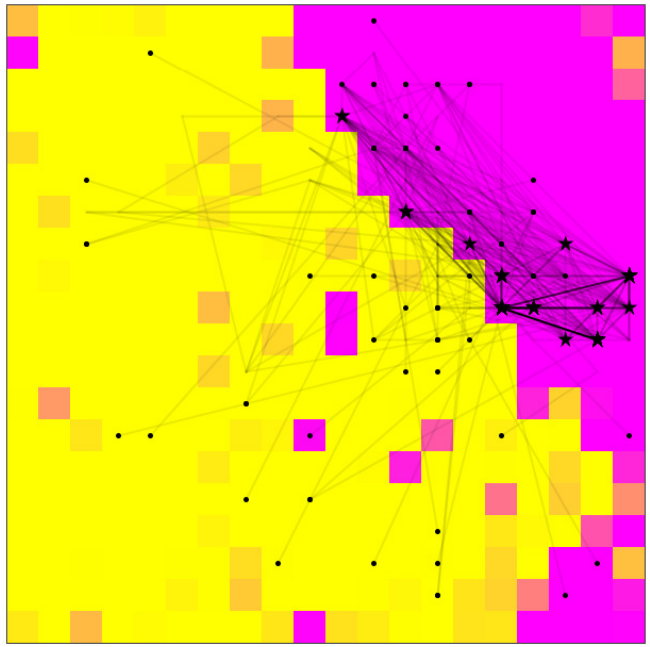


PPARα modulators

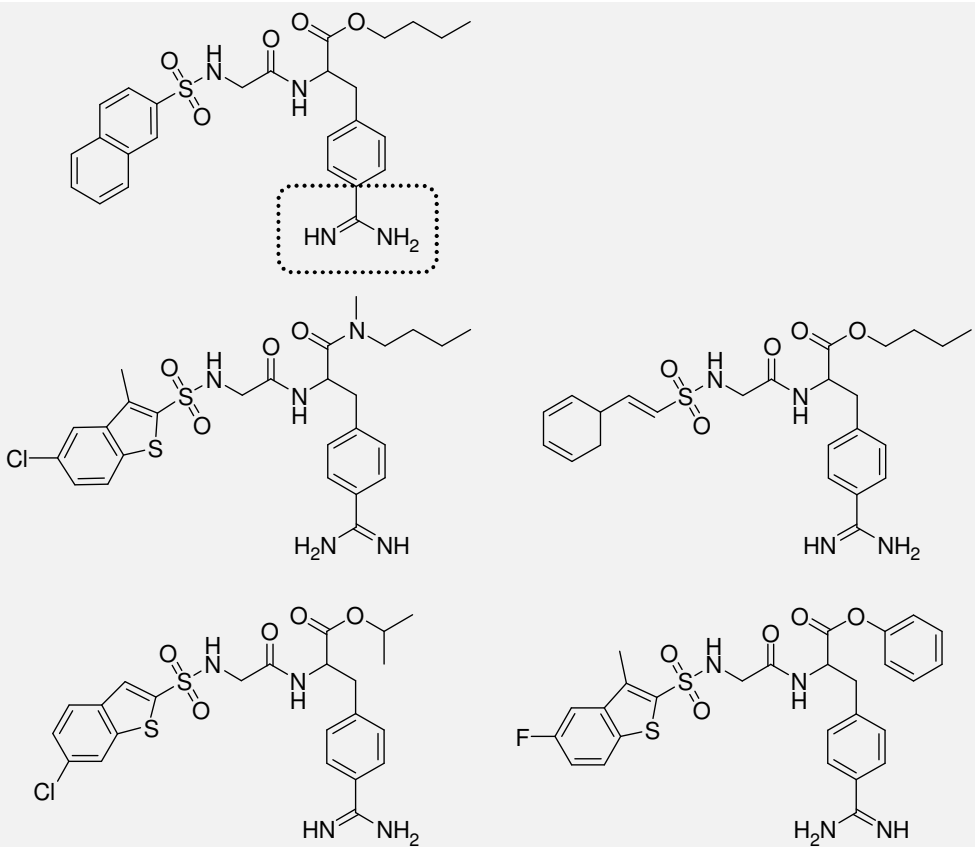


Let's Be Positive!
De Novo Design of Thrombin Inhibitors

CATS descriptor no. 10:
PP0 = ⊕



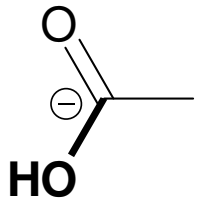
Top-ranking designed compounds

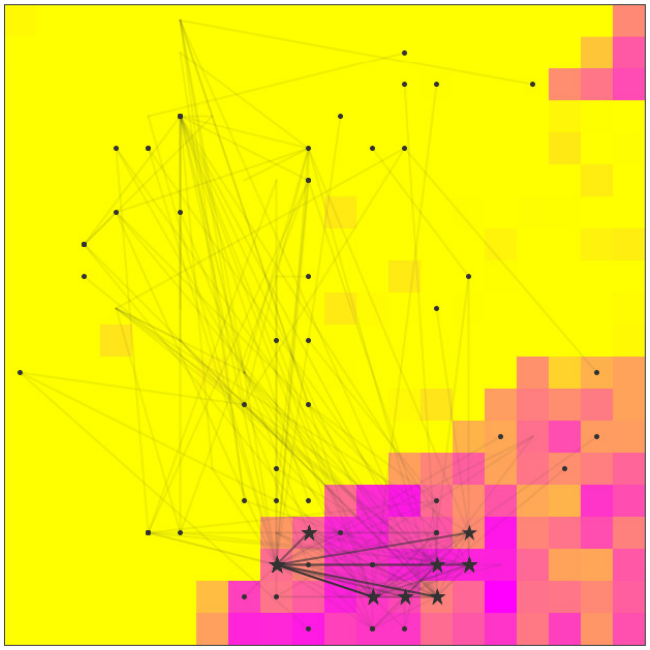


Climbing "Acid hill":
De Novo Design of PPARα Agonists

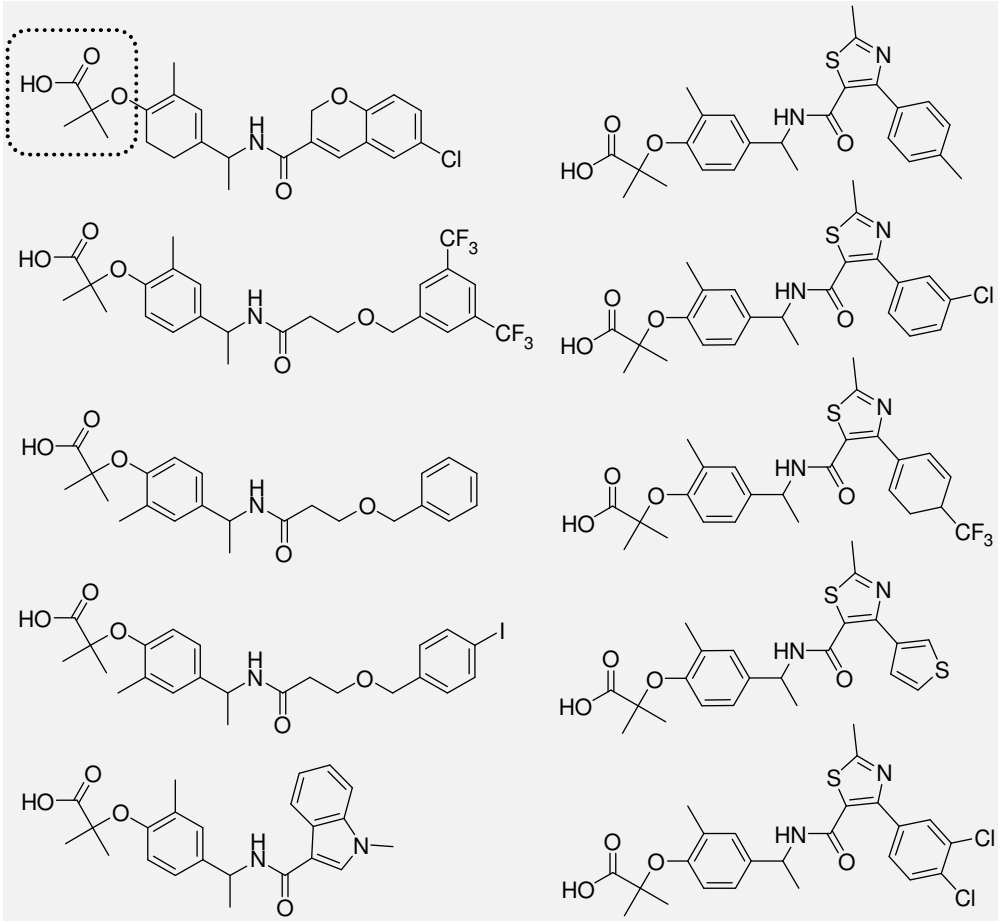
CATS descriptor no. 19:

DN1 =





Top-ranking designed compounds



Conclusions

- Fragment-based design often results in bioactive compounds
- “Novelty” of the designs depends on local convergence
- Adaptive optimization enables “scaffold-hopping”
- Visual inspection / expert knowledge is crucial

Acknowledgement

JOHANN WOLFGANG GOETHE
UNIVERSITÄT
FRANKFURT AM MAIN

Prof. Manfred Schubert-Zsilavecz
Prof. Holger Stark
Prof. Dieter Steinhilber

sanofi aventis
Das Wichtigste ist die Gesundheit

Dr. Karl-Heinz Baringhaus

MERZ

Dr. Tanja Weil

modlab® Team

www.modlab.de