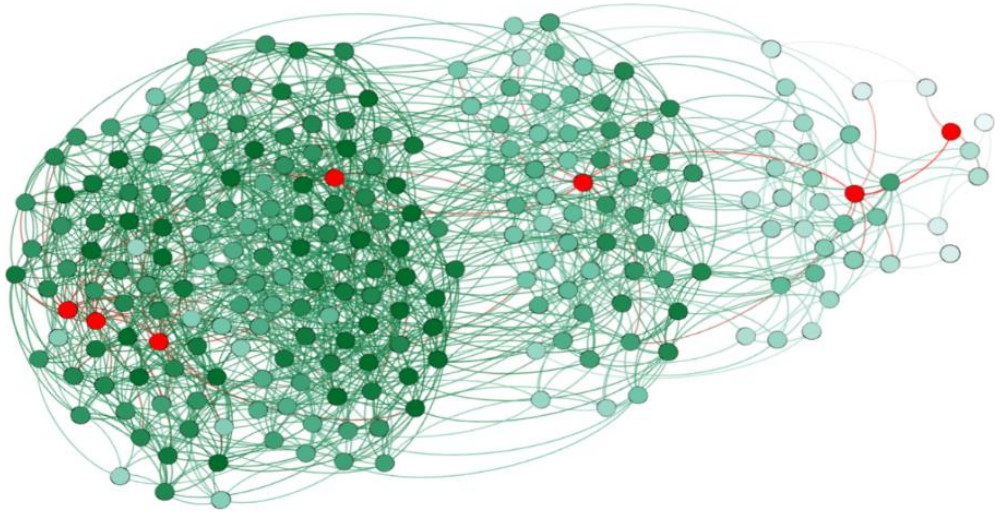
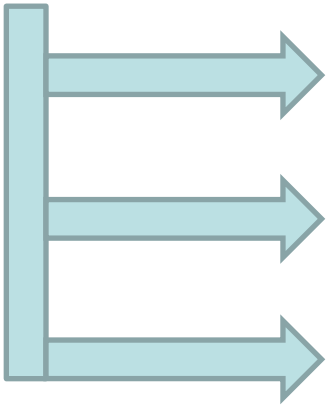
Genetic Algorithm
-
Simulating Materials from first principles



Scope of This Exercise

Electronic Structure Theory is more than pressing buttons



Methodological options

Postprocessing / Analysis of results

Automatisation / Material design

Educational Targets:

- **Identify challenges**
- **Assess solutions**
- **Present results**

Structure Search as a simple-to-implement Example

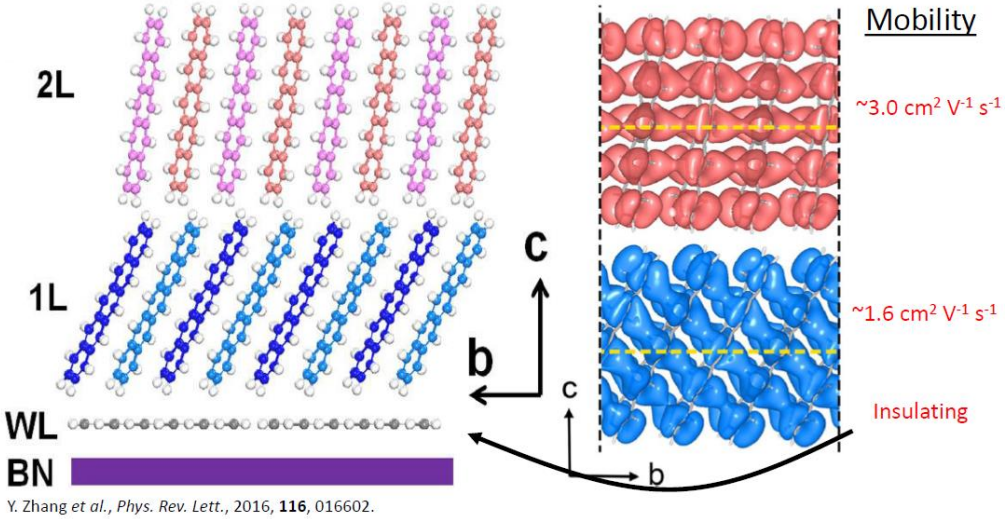
Structure Search Relevance in Practise

Thermal Barrier Coatings

Image by April Gocha http://ceramics.org/ceramic-tech-today/new-ceramic-thermal-barrier-coating-is-long-strong-and-down-to-get-the-friction-on

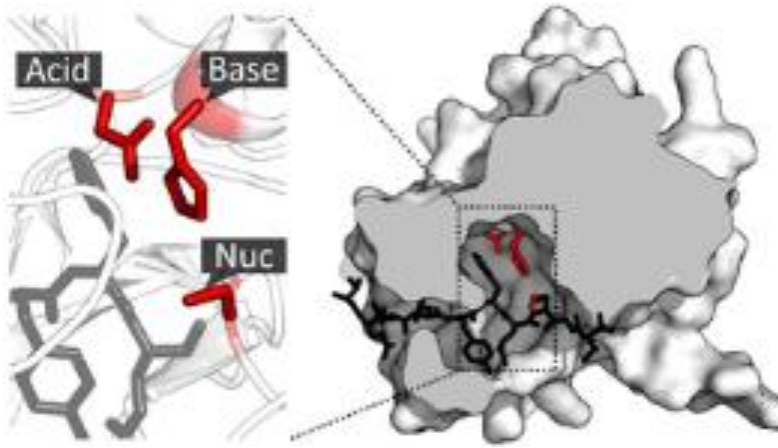
Structure Determines Function

Thalidomid
(German: Contergan)



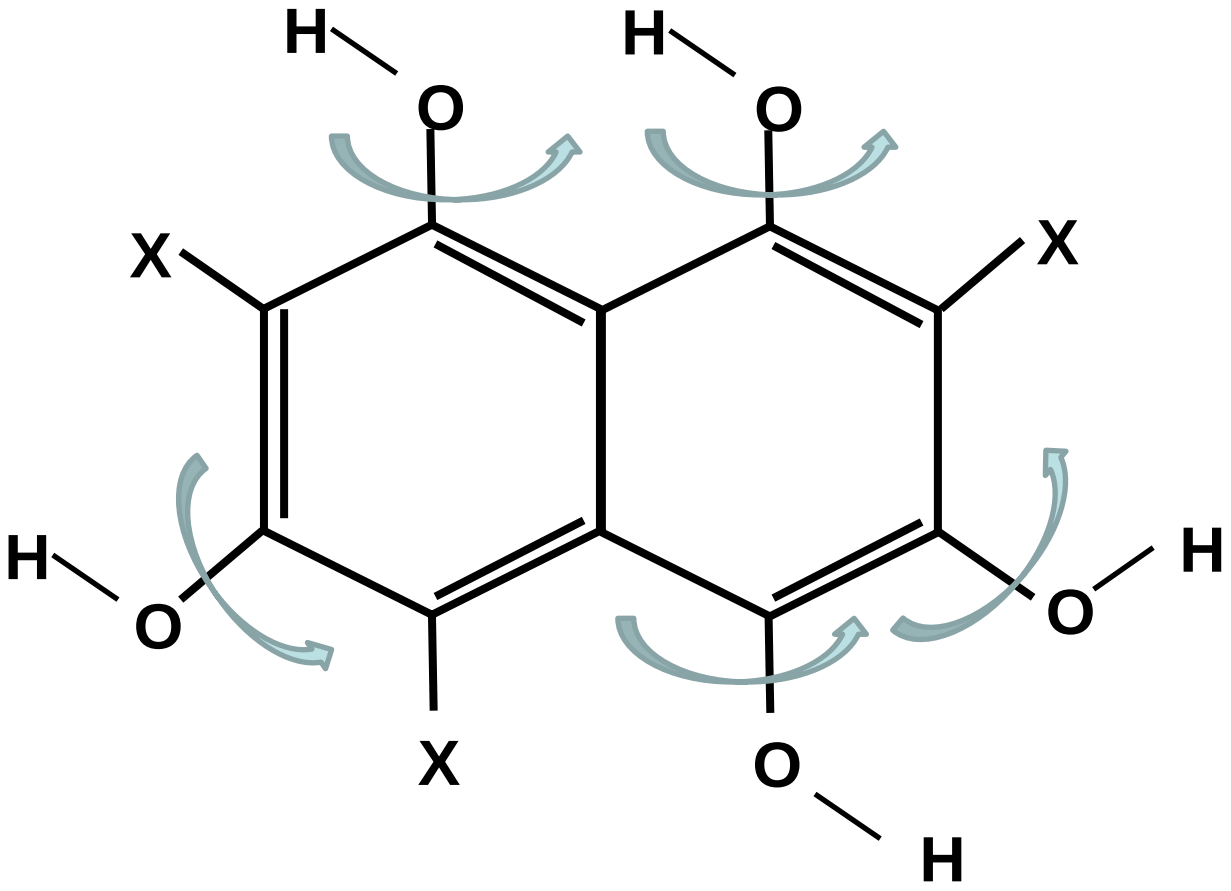
Y. Zhang *et al.*, *Phys. Rev. Lett.*, 2016, **116**, 016602.

Mobilities in Pentacene



Catalytic Triade

Our target:

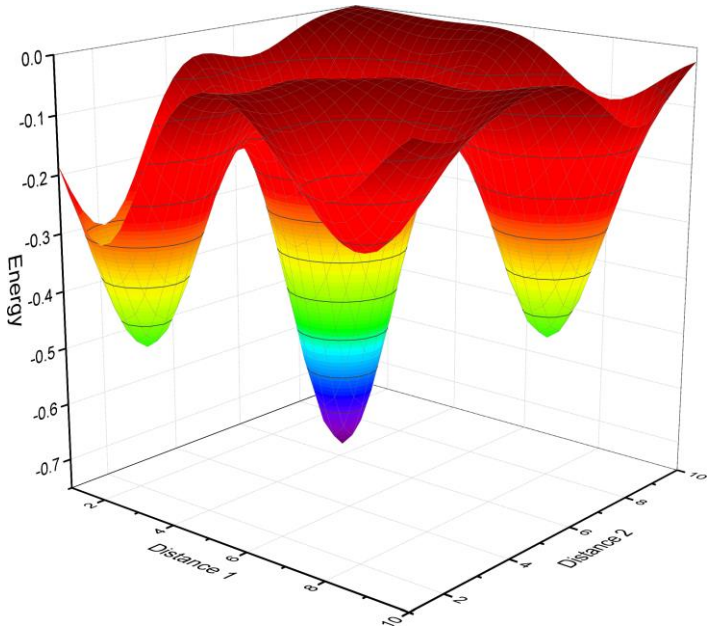


Find the OH-orientations that give the best energy

X choose random mixture of: H, F, Cl, CH3, NO2

Definitions

Potential Energy Surface (PES): result of Born-Oppenheimer approximation. Energy is a function of nuclear coordinates



On the PES:

- Local Minima
- Global Minimum
- Transition States

Attactor Basin: all points leading to the same minimum with a local geometry optimization

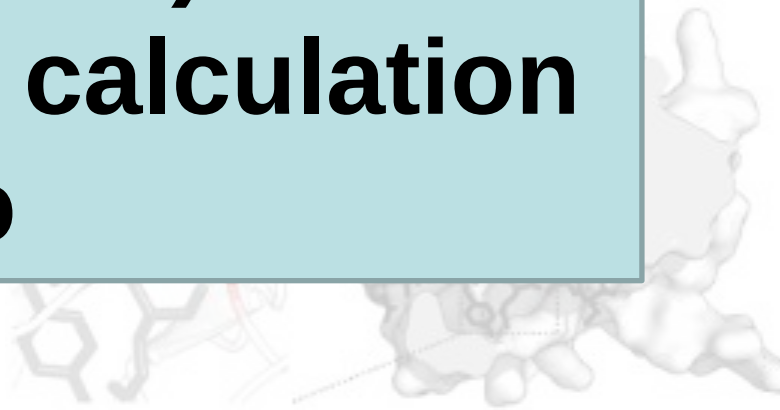
Structure Determines Function

Thalidomid
(German: Contergan)

Global Structure Search (of some kind) is always the first calculation you do

Y. Zhang *et al.*, *Phys. Rev. Lett.*, 2016, **116**, 016602.

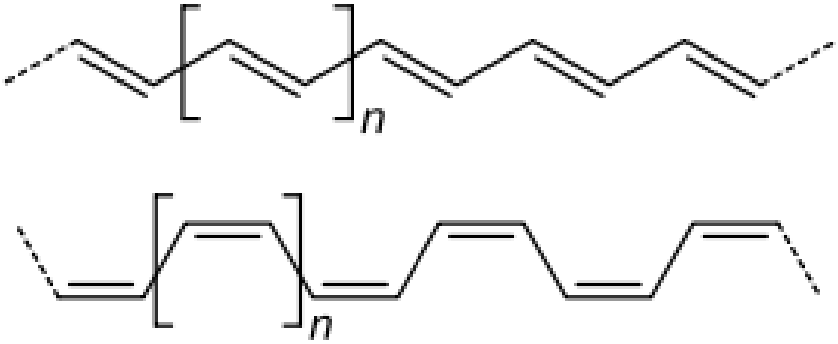
Mobilities in Pentacene



Catalytic Triade

Why Is This a Hard Problem?

Simple Example: Polyacetylene



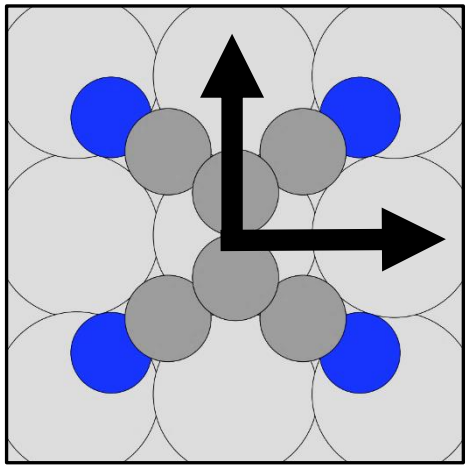
2 options for each double bond:
cis* or *trans

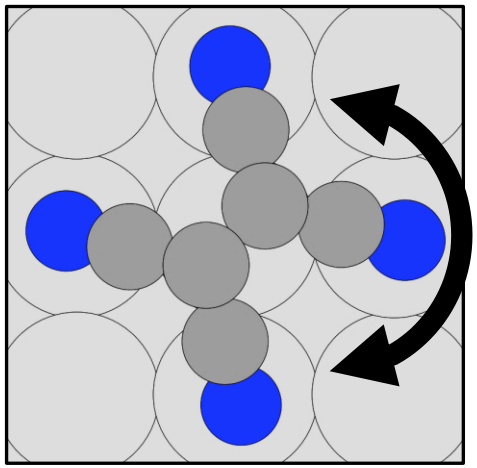
2^n stable geometries

N	# geometries
2	4
4	16
6	64
8	256
10	1024
12	4096
14	16384

„Configurational Explosion“

Non-independent DoF





For each molecule:

Translation x: $\sim$ **10 steps**
Translation y: $\sim$ **10 steps**
Rotation: $\sim$ **10 steps**
3 mol.: $(10 \times 10 \times 10)^3 = 1$ **billion**

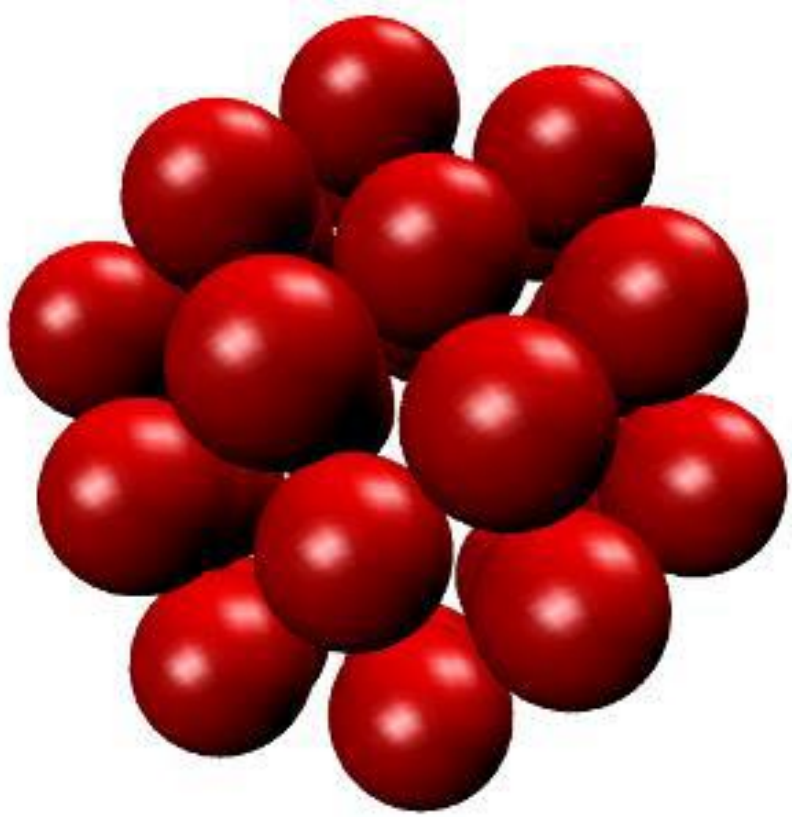
Q: Assuming we can do 1 calc / s,

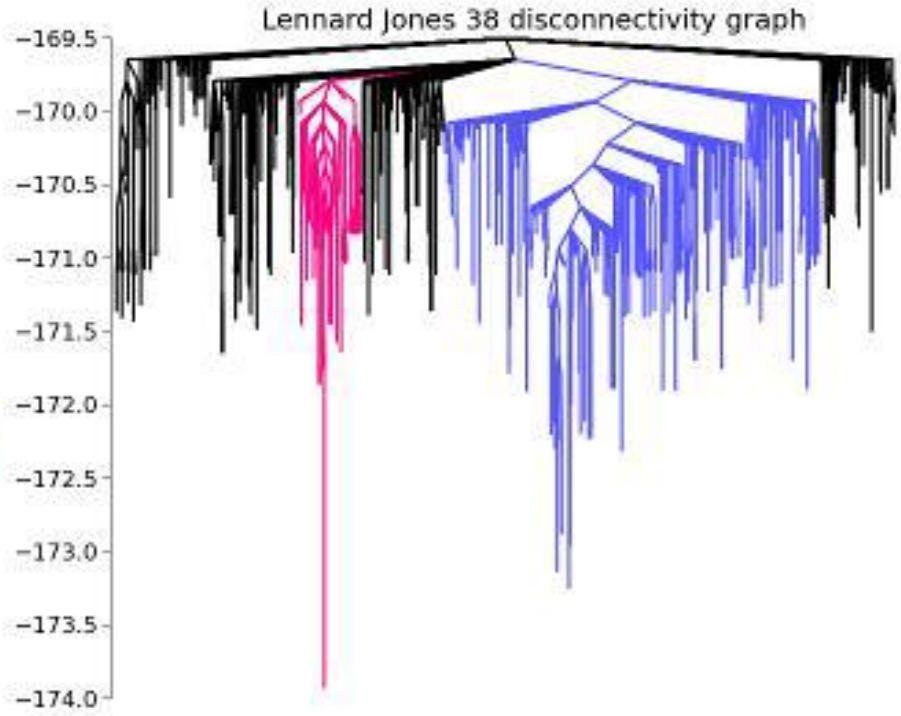
how long would it take?

A: >100 years

Starting points or local minima?

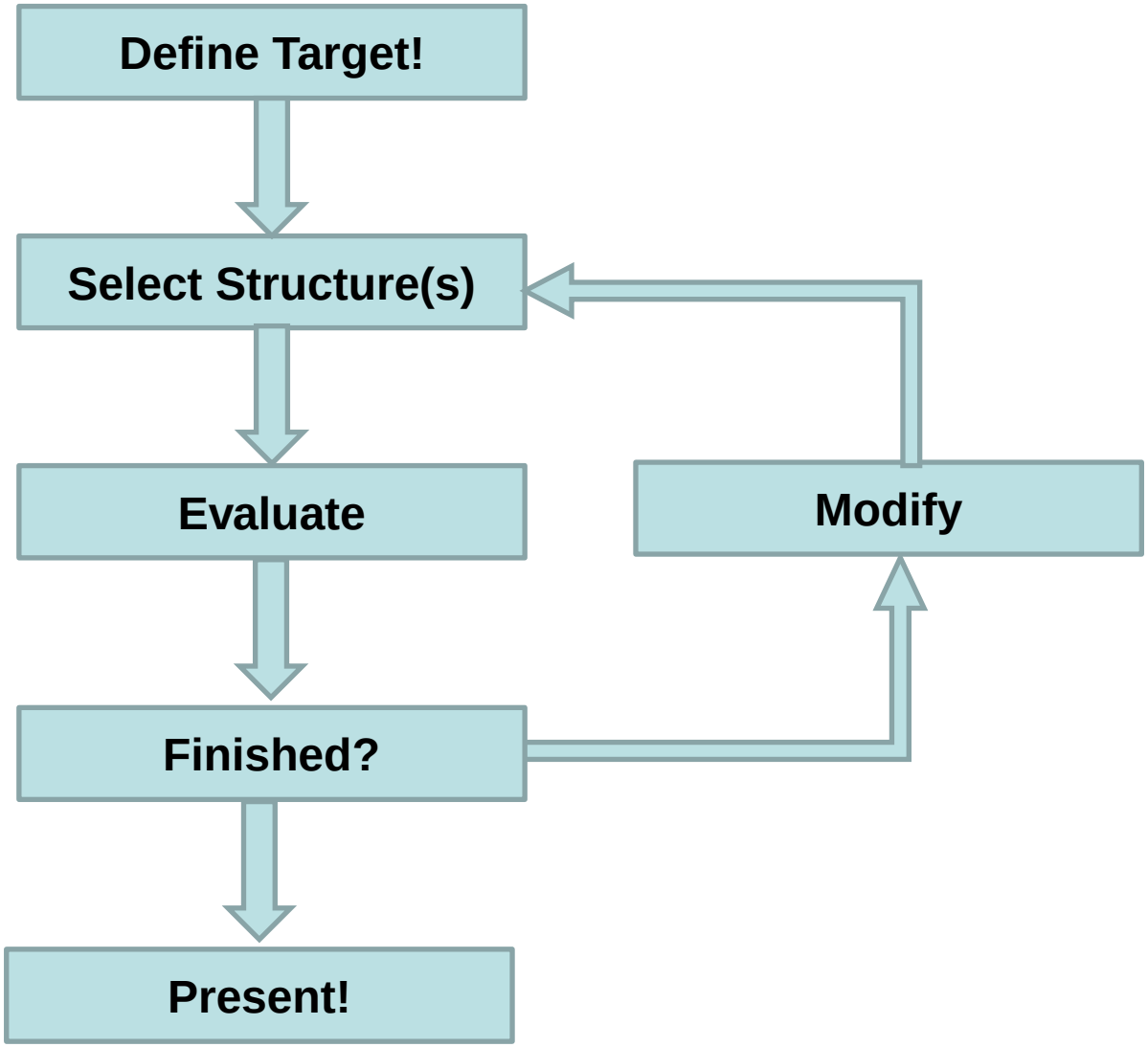
More Complicated Example: Lennard-Jones Cluster

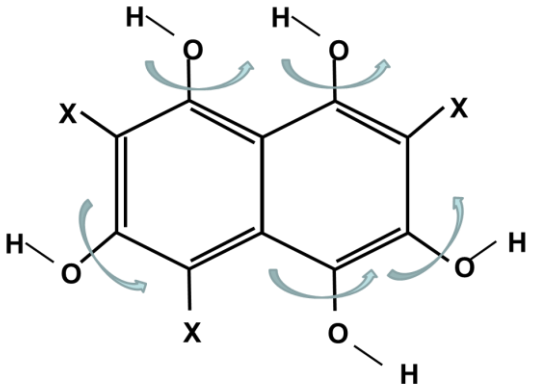




Structure and Energy Disconnectivity Graph for a 38 atom Lennard-Jones Cluster. Taken from https://pele-python.github.io/pele/

General Workflow





General Workflow

Define Target!

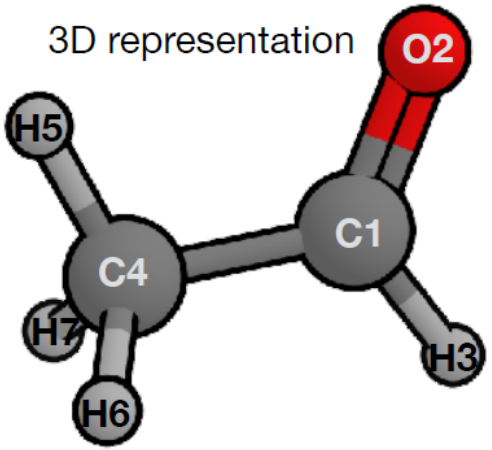
Select Structure(s)

Initialization:

- Random?
- All oriented the same?
- Mixture?

Representation:

Cartesian or Internal Coordinates?
Reflect symmetries?



atom name	atom connect	distance	angle connect	angle	torsion connect	torsion angle
C1	0	0.00	0	0.0	0	0.0
O2	1	1.20	0	0.0	0	0.0
H3	1	1.10	2	120.0	0	0.0
C4	1	1.50	2	120.0	3	180.0
H5	4	1.10	1	110.0	2	0.0
H6	4	1.10	1	110.0	2	120.0
H7	4	1.10	1	110.0	2	-120.0

Coordinate Choice

Cartesian Coordinates:

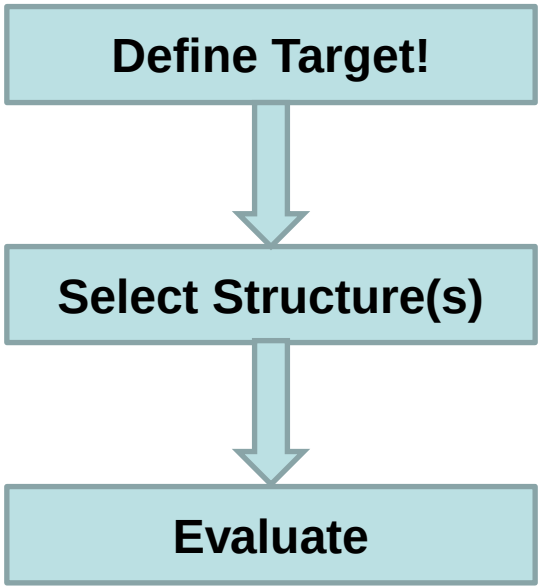
- Easy to implement
- Straightforward to connect to external code
- Reasonable choice for non-covalent bonds
- Perturbations easily create bogus structure

versus

Internal Coordinates:

- Requires (some) more thought
- Smoother potential energy surface
- Easier to modify (esp. for genetic algorithm)
- Freezing of DOFs more straigthforward

General Workflow



Select Level Of Theory:

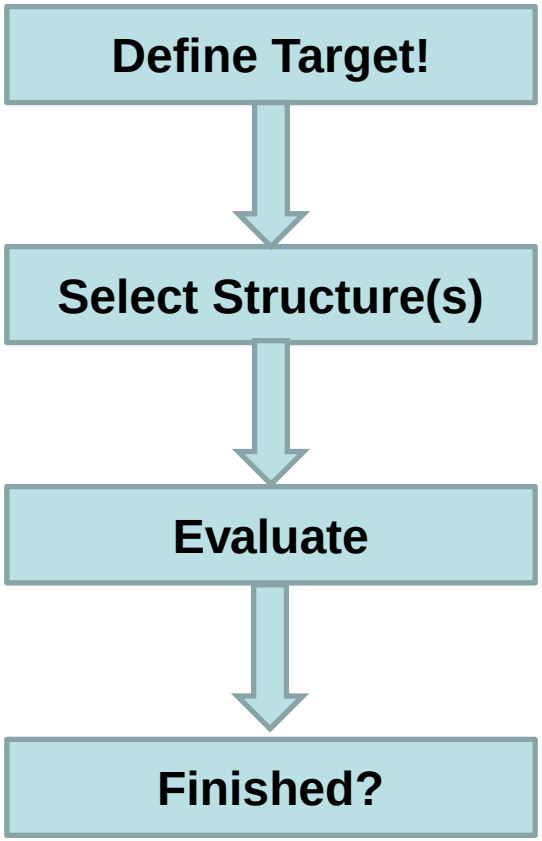
- Connect to Electronic Structure Code (PySCF, FHI-aims, Gaussian, VASP)

→ Low Level Theory will be fast
(Basis set? Method?)

Obtain energy

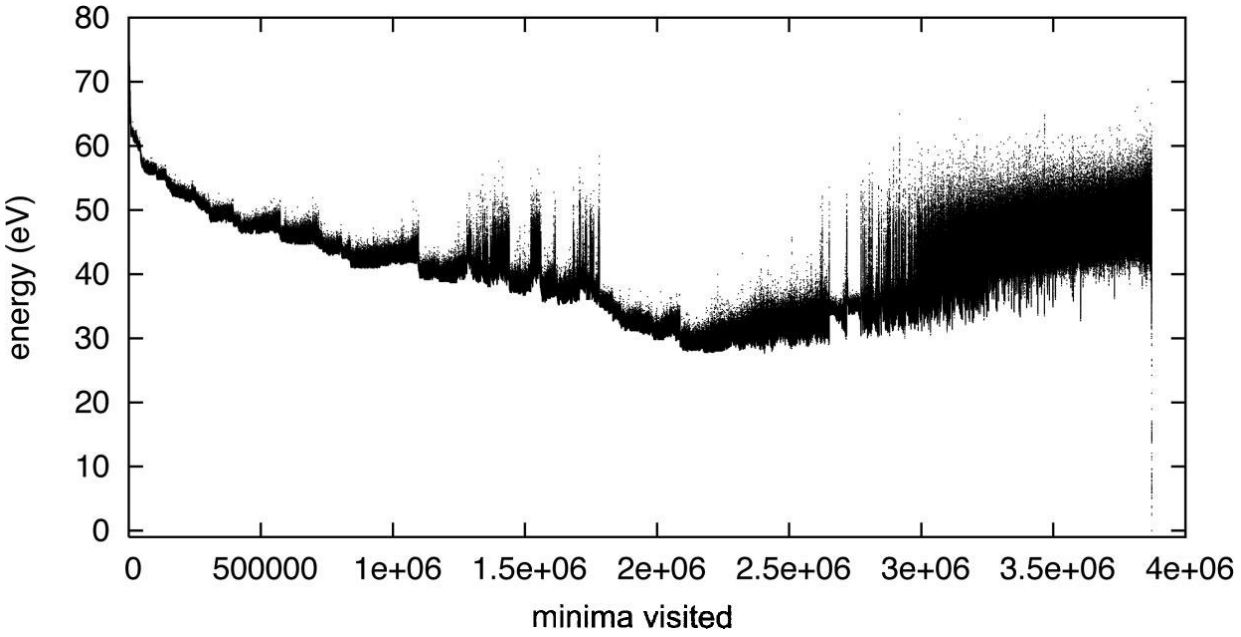
- Just single point
- Local Geometry Optimization
- Think about thresholds (energy, forces)

General Workflow

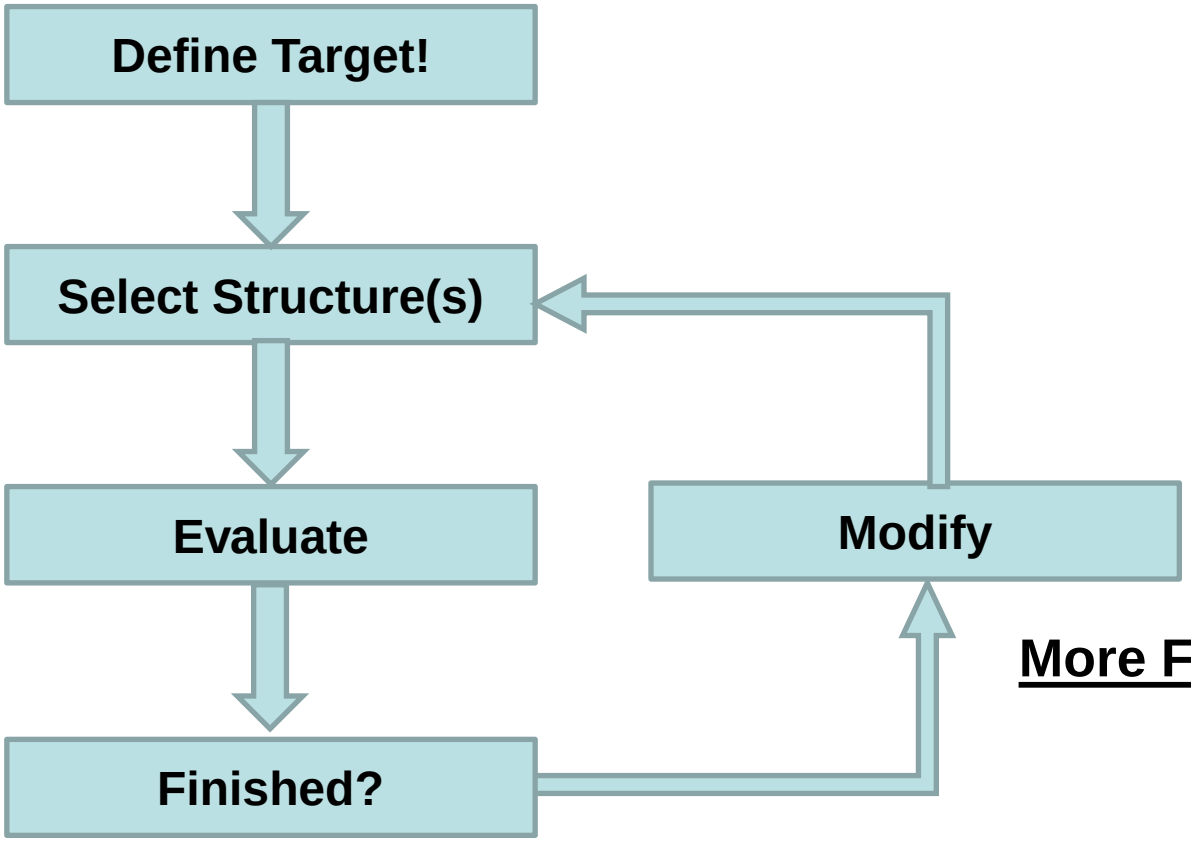


Convergence

- Run for fixed number of steps
- Run maximum time
- Stop if no better structure is found long enough

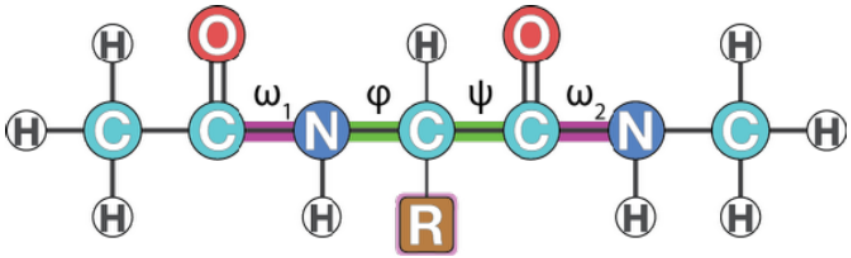


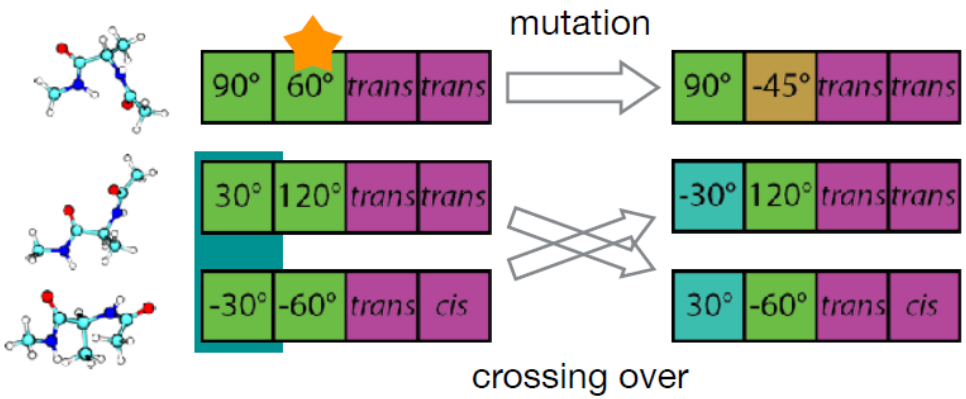
General Workflow



More Fun: Genetic Algorithm

Genetic Algorithm

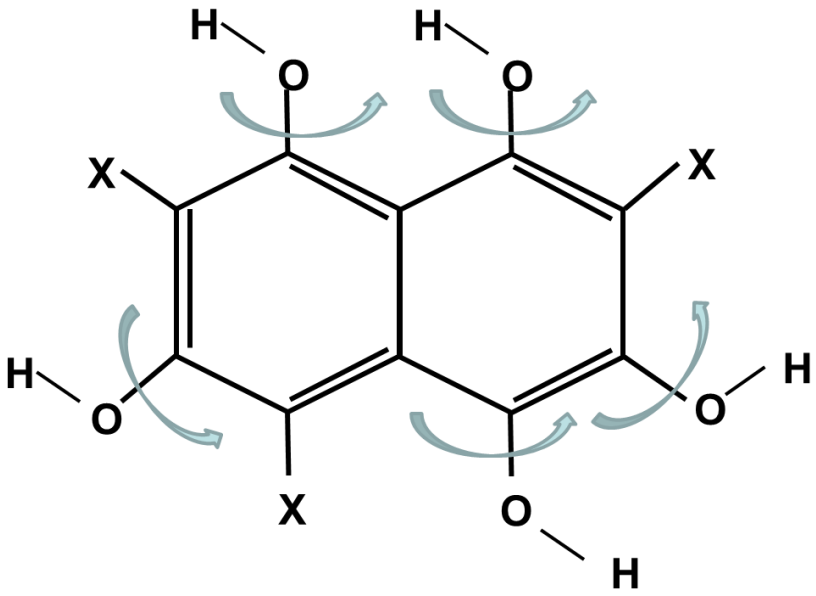




Work by A. Supady

Select sensible „genes“:

- **Internal coordinates, including connection to different atoms**

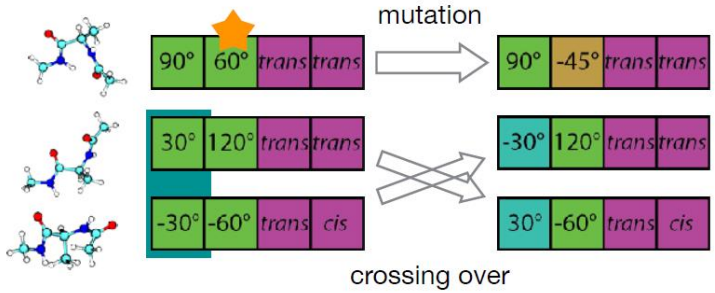


Select sensible „genes“:

- **Here: Only C-C-O-H torsions**
- **Single points: Use 6-fold discretization: 0,60,120,180,240,300 °**
- **Geometry optimzation: Darwin or Lamark**

Genetic Algorithm

Select (pairs of) parents to mate and form offspring



Work by A. Supady

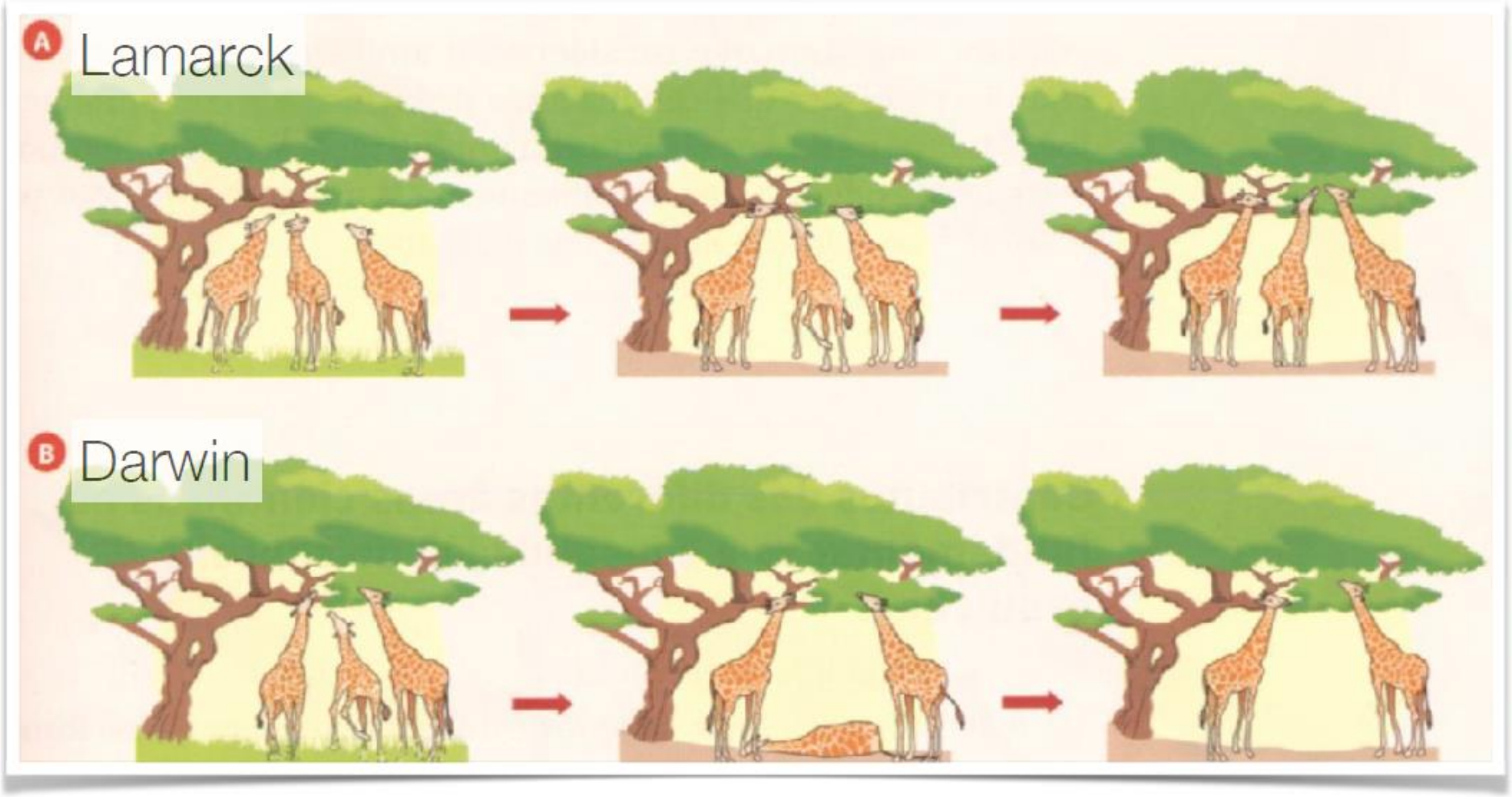
Mating:

- Mutation: Change gene in sensible range
- Crossover:
  - Set of genes from M1
  - ***complimentary*** set of genes from M2

Selections:

- Roulette Wheel: full population, weigh probably to mate by fitness
- Inverse Roulette Wheel: full population, less fit members are more likely to mate
- Tournament: randomly select subset, fittest mate

→ **Remember to keep some parents in the pool**


A
Lamarck
B
Darwin

Present Results

- **Chosen Molecule**
- **Level of Theory**
- **Other choices (population size, mutation rate, ..)**
- **Outcome**
  - **Best Geometry**
  - **Energy spread of explored geometries**
  - **Success rate**
  - **Time to find global minimum**

Target Excersice

- **Write a global optimization program**
  - **Select a system (cluster or molecule)**
  - **Make methodological choices and justify them**
  - **Explore parameters, report impact**
  - **Report results from the search. What could be relevant questions?**

Note: The performance will not be graded!

Questions:
Contact me (o.hofmann@tugraz.at)