Finding Many Stable Molecular Arrangements

Conformational Searching with Genetic Algorithms

Evan Curtin

October 10, 2016

University of Illinois at Urbana-Champaign

Outline

1. Background Information
2. The Genetic Algorithm
3. Conclusion

Background Information

The Problem

Computational methods require knowledge of molecular structure

The Problem

Computational methods require knowledge of molecular structure

$\Rightarrow$ We need to find the lowest energy structure

The Problem

Computational methods require knowledge of molecular structure

$\Rightarrow$ We need to find the lowest energy structure

The potential energy surface (PES) is high dimensional and has many minima

The Problem

Computational methods require knowledge of molecular structure

$\Rightarrow$ We need to find the lowest energy structure

The potential energy surface (PES) is high dimensional and has many minima

$\Rightarrow$ We can't tell for sure if we've found the global minimum

The Problem

Computational methods require knowledge of molecular structure

$\Rightarrow$ We need to find the lowest energy structure

The potential energy surface (PES) is high dimensional and has many minima

$\Rightarrow$ We can't tell for sure if we've found the global minimum

We may need information about one or more low-energy conformations

The Problem

Computational methods require knowledge of molecular structure

$\Rightarrow$ We need to find the lowest energy structure

The potential energy surface (PES) is high dimensional and has many minima

$\Rightarrow$ We can't tell for sure if we've found the global minimum

We may need information about one or more low-energy conformations

$\Rightarrow$ Ok, let's find them all!

Possible Solutions

Many techniques are well established

method	description	implemented, e.g., in
grid-based	based on grids of selected Cartesian or internal coordinates (e.g., grids of different torsional angle values of a molecule)	CAESAR,[9] **Open Babel**,[10] **Confab**,[11] MacroModel,[12] MOE[13]
rule/knowledge-based	use known (e.g., from experiments) structural preferences of compounds	**ALFA**,[14] **CONFECT**,[15] CORINA and ROTATE,[16,17] **COSMOS**,[18,19] OMEGA[20]
population-based metaheuristic	improve candidate solutions in a guided search	**Balloon**,[21] **Cyndi**[22]
distance geometry	based on a matrix with permitted distances between pairs of atoms	**RDKit**[23]
basin-hopping[24]/ minima hopping[25]	based on moves across the PES combined with local relaxation	**ASE**,[26] **GMIN**,[27] **TINKER SCAN**[28]

[a]Names of freely available programs are highlighted in boldface.

Supady, A.; Blum, V.; Baldauf, C. J. Chem. Inf. Model. 2015, 55 (11), 2338–2348.

Possible Solutions

Many techniques are well established

None are perfect

method	description	implemented, e.g., in
grid-based	based on grids of selected Cartesian or internal coordinates (e.g., grids of different torsional angle values of a molecule)	CAESAR,[9] **Open Babel**,[10] **Confab**,[11] MacroModel,[12] MOE[13]
rule/knowledge-based	use known (e.g., from experiments) structural preferences of compounds	**ALFA**,[14] **CONFECT**,[15] CORINA and ROTATE,[16,17] **COSMOS**,[18,19] OMEGA[20]
population-based metaheuristic	improve candidate solutions in a guided search	**Balloon**,[21] **Cyndi**[22]
distance geometry	based on a matrix with permitted distances between pairs of atoms	**RDKit**[23]
basin-hopping[24]/ minima hopping[25]	based on moves across the PES combined with local relaxation	**ASE**,[26] **GMIN**,[27] **TINKER SCAN**[28]

[a]Names of freely available programs are highlighted in boldface.

Supady, A.; Blum, V.; Baldauf, C. J. Chem. Inf. Model. 2015, 55 (11), 2338–2348.

Desirables

- Accurate energies & Structures, *ab initio* or DFT

Desirables

- Accurate energies & Structures, *ab initio* or DFT
- Minimize number of geometry optimizations

Desirables

- Accurate energies & Structures, *ab initio* or DFT
- Minimize number of geometry optimizations
- Find the entire low energy population of conformations

Desirables

- Accurate energies & Structures, *ab initio* or DFT
- Minimize number of geometry optimizations
- Find the entire low energy population of conformations
- Minimal human input

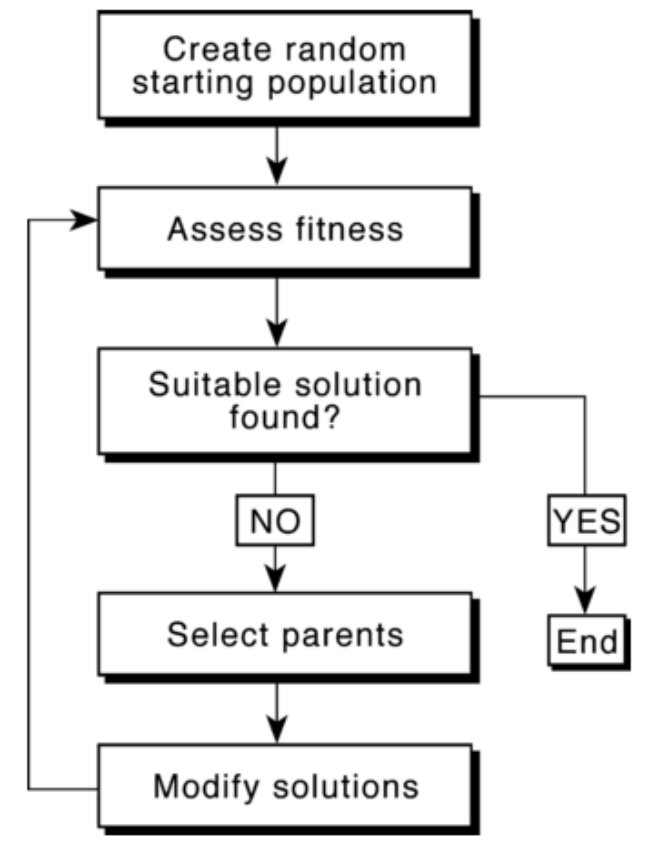
Desirables

- Accurate energies & Structures, *ab initio* or DFT
- Minimize number of geometry optimizations
- Find the entire low energy population of conformations
- Minimal human input
- Parallel-Scalable

The Genetic Algorithm

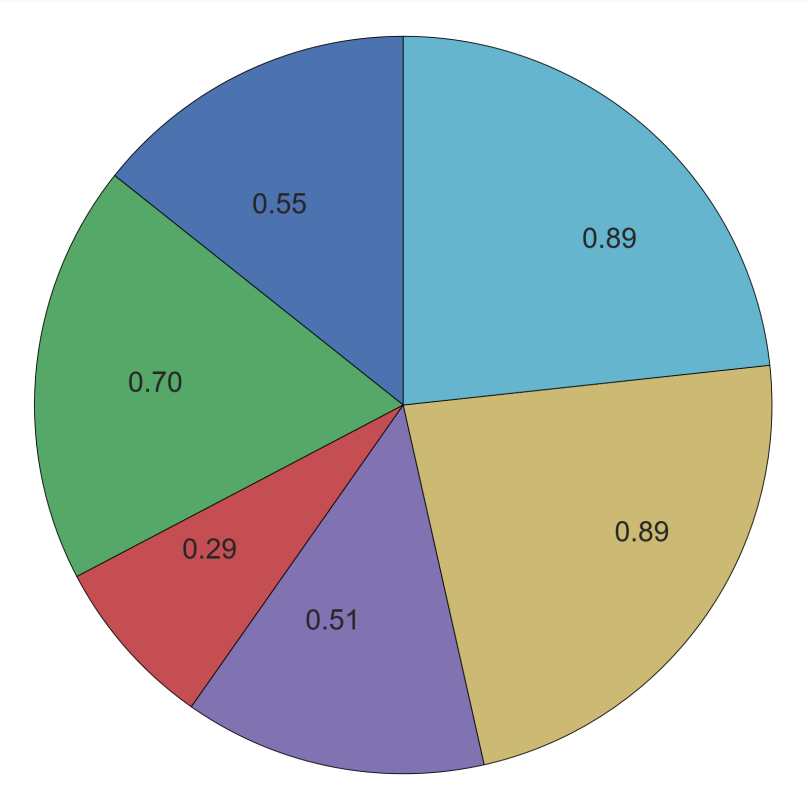
Outline

- Inspired by biological evolution
- Evolve a population over generations
- Survival of the fittest
- Requirements:
  - Represent individuals as vector
  - Fitness function
- $v = \left(x_1 \; y_1 \; z_1 \; x_2 \; y_2 \; z_2 \; \ldots \; x_N \; y_N \; z_N \right)$
- $F = \frac{E_{max} - E}{E_{max} - E_{min}}$



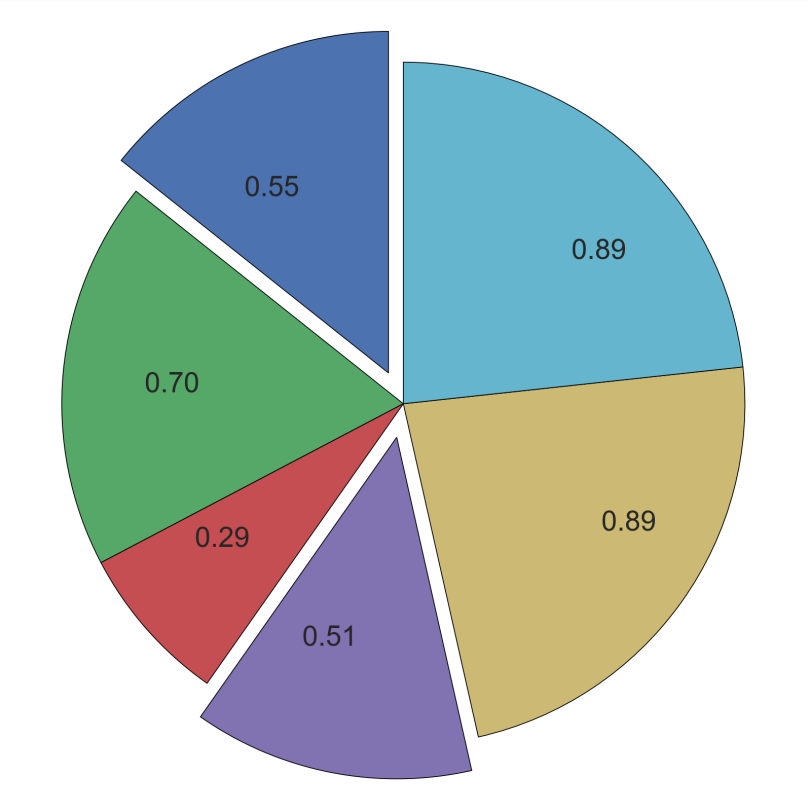
Selecting Parents

- Several methods are common
- Reinforce good characteristics
- Still give losers a chance
- ’Breed’ pairs of winners

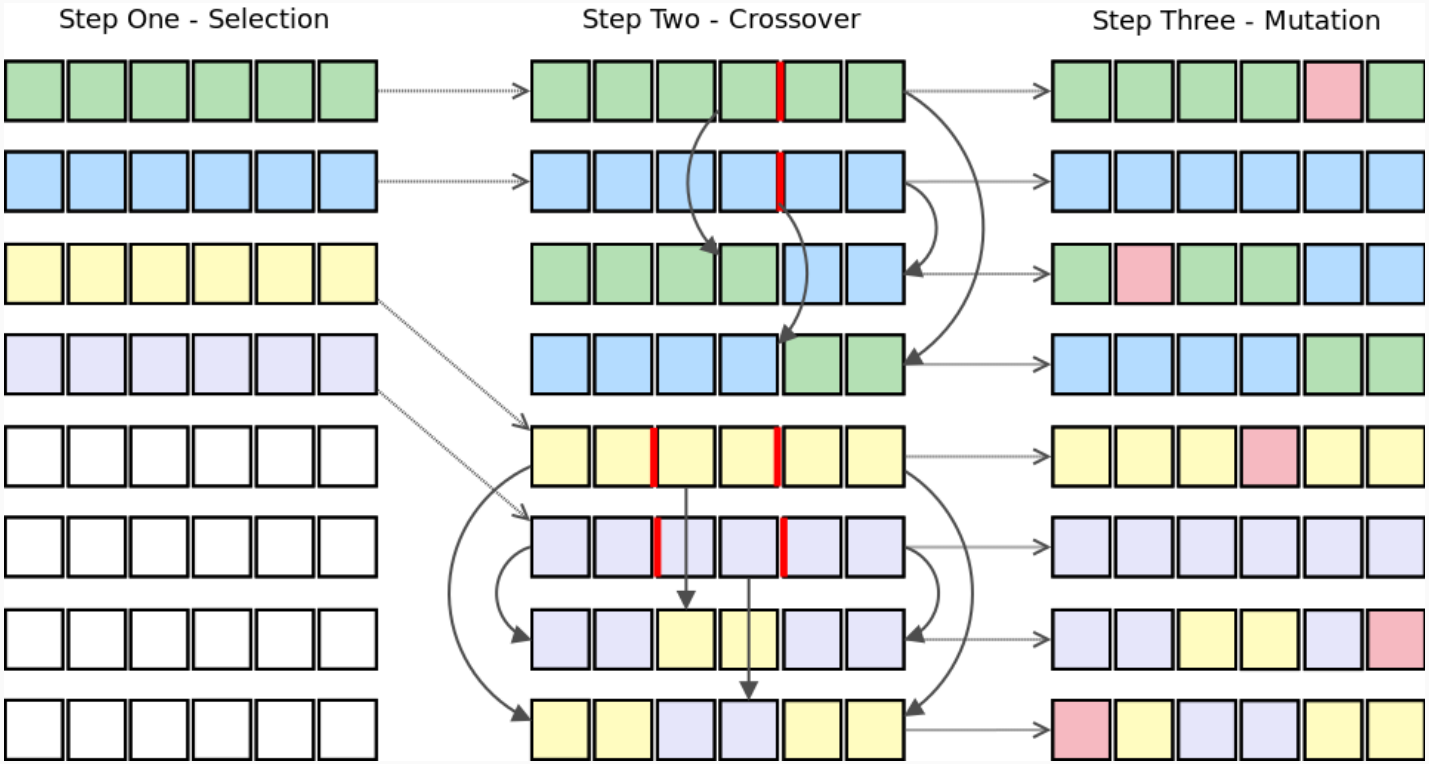


Selecting Parents

- Several methods are common
- Reinforce good characteristics
- Still give losers a chance
- ’Breed’ pairs of winners

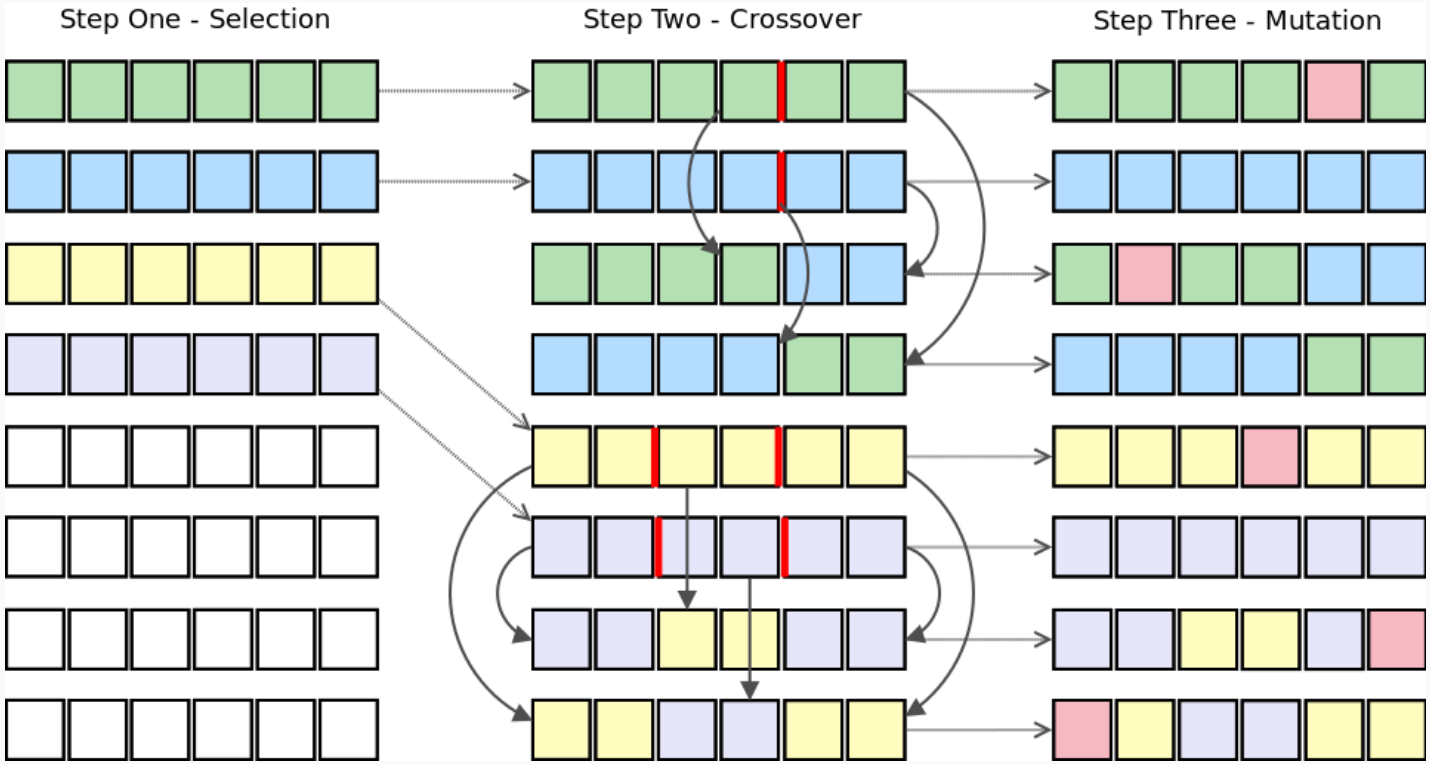


The Next Generation



http://www.turingfinance.com/computational-finance-updates-ieee-world-congress-computational-intelligence

The Next Generation



Crossover distinguishes this from Monte Carlo

http://www.turingfinance.com/computational-finance-updates-ieee-world-congress-computational-intelligence

Conclusion

Questions?

Backup slides

a