THE UNIVERSITY
OF BIRMINGHAM

Applications of Evolutionary Computation in the Pharmaceutical Industry: *The Cercia Perspective*

Dr. Thorsten Schnier, Cercia

the centre of excellence
cercia
for research in computational
intelligence and applications

Overview

I. Cercia and Nature Inspired Computation

II. Evolutionary Computation

III. Applications of Evolutionary Computation in the Pharmaceutical Industry

THE UNIVERSITY OF BIRMINGHAM

CERCIA

Centre of **E**xcellence for **R**esearch in **C**omputational **I**ntelligence and **A**pplications

- Technology transfer centre in the School of Computer Science, University of Birmingham
- Focuses on *applications* of Nature Inspired methods to practical *business problems*

Natural Computation Group at University of Birmingham

- 50 Researchers including staff and students
- World leading group with links to industrial and academic partners in UK, Europe, Asia, Australia, US

the centre of excellence cercia for research in computational intelligence and applications

Advantage West Midlands

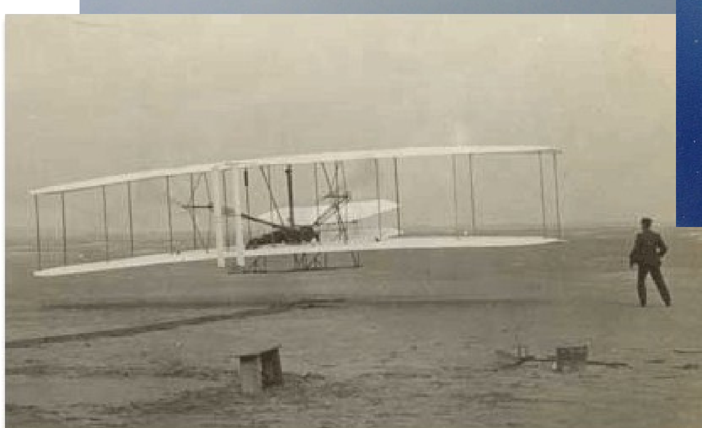
Why Natural Computation?

- **Flexible**: Applicable to different problems
- **Robust**: Can deal with noise and uncertainty
- **Adaptive**: Can deal with dynamic environments through self-adaptation
- **Autonomous**: Function without human intervention
- **Decentralised**: Without a central authority

THE UNIVERSITY OF BIRMINGHAM

Nature Inspired Computation

Inspired, but not a direct copy



the centre of excellence for research in computational intelligence and applications

II. Evolutionary Computation

THE UNIVERSITY
OF BIRMINGHAM

Darwinian Evolution

- Individuals within species are variable
- Some of the variations are passed on to offspring
- In every generation, more offspring are produced than can survive
- The survival and reproduction of individuals are not random: The individuals who survive and go on to reproduce, or who reproduce the most, are those with the most favourable variations. They are naturally selected.

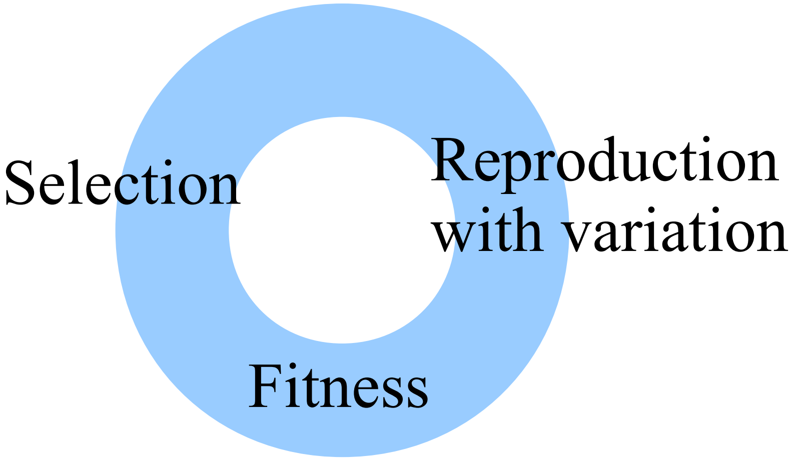
On the Origin of Species by Means of Natural Selection
(Darwin 1859)

the centre of excellence
cercia
for research in computational intelligence and applications

Evolutionary Algorithms

Four Main Elements

- Group of Individuals - *Population*
- Reproduction with Variation - *Genetic Operators*
- Reproductive Success - *Fitness*
- Survival of the Fittest – *Selection*



Advantages of EC, #1

Advantage #1: Evolutionary Computation is an automated search process

- Trial and Error: Recipe for next trial
- Huge range of applications (design, optimization, prediction, exploration, ...)
- Ranges from 10s of trials to 100s Millions trials
- Trivially parallelized for computer clusters

Advantages of EC, #2

Advantage #2: EC is a knowledge-lean process

Only needs very little knowledge:

- Fitness needs comparison only
- You don't need to understand the problem
- It can be non-linear, non-static, complex,...
- In fact: the harder it is. the better !

But you can (and should) integrate any knowledge you have.

Advantages of EC, #3

Advantage #3: EC produces better results

Conventional processes often

- Require linearization
- Make invalid assumptions
- Ignore potential solutions because they cannot understand them

Evolutionary Computation

- Does not have these problems
- Can also explore far more potential solutions than a human could

Advantages of EC, #4

Advantage #4: Flexibility, Simplicity, Robustness

- Conventional approaches are brittle
  - Change anything, and you may have to start from scratch
- EC is flexible
  - Small change in fitness often sufficient
- EC is simple
  - Very fast to set up
- EC is robust
  - Copes with faulty/noisy/missing fitness evaluations

Advantages of EC, #5

Advantage #5: EC can deliver explanations

Genetic programming derives rules, functions, programs (LISP trees)

- Readably by the user
- No 'black-box'
- Can be implemented in control systems, classifiers,...

Advantages of EC, #6

Advantage #6: Multiobjective Optimization

Real-world optimization is (nearly) always multi-objective

EC can deliver many solutions

- Exploring all trade-offs
- No need for prioritization
- User can explore and chose after the run

Can also deal with hard and soft constraints

Disadvantages of EC

There are disadvantages, too:

May take a long time

- Conventional (direct) solutions are always faster
- Good heuristics may be faster

No guarantees

- Finds good solutions, but not always the best
- May sometimes not find a solution at all

THE UNIVERSITY OF BIRMINGHAM

III. EC Applications in the Pharmaceutical Industry

the centre of excellence
cercia
for research in computational intelligence and applications

Prediction

- Task: predict some numerical property of a molecule based on its structure
  - E.g. solubility in water, bio-availability, QSAR, …
- Genetic programming approach:
  - Population: LISP program trees
    - Input: structure description, or pre-computed values; Output: prediction (real value)
  - Fitness: predictive performance
- EA advantage: Good extrapolation, Interpretable results [Lan2003, Nico2002]

Docking

Task: Given two molecules, predict the interaction – find the lowest-energy complex

- Evolutionary approach:
  - Population: vectors encoding position, orientation, conformation
  - Fitness: energy, as computed by a given model
- EA in docking software: AutoDock, LGA, DockEA, GOLD
- EA advantage: Allows docking of complex molecules, including flexible ligands [Budin2001, Morris2000]

Library Design

Task: from large number of molecules or fragments, select n-dimensional subset

  - Optimized for diversity, or specifically for a target
  - Secondary objectives (cost, availability, similarity to known target,…), constraints (molecular weight, patents)

- Evolutionary Approach:
  - Population of libraries (e.g. using integer strings)
  - Multi-objective fitness, constraints influence selection
- EA advantage: no restriction on objectives, can work with extremely large search spaces
  [Sheridan2000, Gillet2002, Jamois2003]

THE UNIVERSITY OF BIRMINGHAM

De-Novo Design

Task: create a new substance (e.g. ligand)

- From fragments in library

- Evolutionary Approach:
  - Population of candidate drugs: acyclic graphs of fragments
    - Alternative: SMILES notation
  - Fitness: Through docking evaluation with known target
    - Alternatives: QSAR predictors, or in-vitro experiments
- EA Advantage: efficient exploration -> novel analogs, better inhibitors [Dougouet1999, Pegg2001]
- EAs used: for library selection, for fitness function, and for ligand design !
- Next step: evolve combinations of ligands…

the centre of excellence cercia for research in computational intelligence and applications

Other Uses, Part 1

- Genetic Marker Acquisition
  - Select a subset of markers, multi-objective fitness [Hubley2004]
- Reverse Engineering of Regulatory Networks
  - EA evolves candidate networks [Repsilber2002, Eriksson2004]
- Structure from Diffraction Data
  - EA evolves candidate models [Harris2004]

Other Uses, Part 2

- Classification (e.g. two-class cancer data)
  - Gene subset selection and classification [Deb2003]
- Optimization of Enzymatic Synthesis
  - Parameters: concentration, enzyme activities, temperature, pressure
  - Goal: maximum concentration and selectivity
  - Evaluation by experiment [Hoh2002]

THE UNIVERSITY
OF BIRMINGHAM

Summary

- Evolutionary Algorithms are "population based trial and error" algorithms
- Very powerful, flexible,with a wide range of applications
- Application focus in pharma R&D currently is in Evolutionary Chemistry, but there are other promising areas

the centre of excellence
cercia
for research in computational intelligence and applications

THE UNIVERSITY OF BIRMINGHAM

References (1)

- Budin2001: Nicolas Budin, Nicolas Majeux and Amedeo Caflish: *Fragment-Based Flexible Ligand Docking be Evolutionary Optimization*, Biol Chem. Vol 382, pp 1365-1372, Sept. 2001
- Deb2003: Kalyanmoy Deb and A. Raji Reddy: *Reliable classification of two-class cancer data using evolutionary algorithms*, BioSystems 72 (2003) 111-129
- Douguet1999: Dominique Douguet, Etienne Thoreau and Gerard Grassy: *A Genetic Algorithm for the automated generation of small organic molecules: Drug design using an evolutionary algorithm, Journal of Computer Aided Molecular Design 14 (2000) 449-466*
- Eriksson2004: R. Eriksson and B. Olsson: *Adapting genetic regulatory models by genetic programming,* BioSystem 76 (2004) 217-227
- Gillet2002: Valier J. Gillet et al: *Designing Focused Libraries using MoSELECT*, Journal of Molecular Graphics and Modelling 20 (2002), 491-498
- Harris2004: Kenneth D.M. Harris, Roy L. Johnston and Scott Habershon: Applications of Evolutionary Computation in Structure Determination from Diffration Data, in *Applicaions of Evolutionary Computation in Chemistry*, Springer Verlag, Heidelberg (2004) pp. 55-94
- Hoh2002: Christoph Hoh, Gregor Dudziak and Andreas Liese: *Optimization of the enzymatic synthesis of O-Glycan Core 2 structure by use of a genetic algorithm*, Bioorganic and medicinal chemistry letters 12 (2002) 1031-1034

the centre of excellence cercia for research in computational intelligence and applications

References (2)

- Hubley2004: Robert M. Hubley *MAGMA: Mutiobjective Analyzer for Genetic Marker Acquisition,* http://snp-magma.sourceforge.net
- Jamois2003: Eric A. Jamois, Chien T. Lin and Marvin Waldman: *Design of focused and restrained subsets from extremely large virtual libraries,* Journal of Molecular Graphics and Modelling 22 (2003) 141-149
- Langdon2003: W.B. Langdon, S.J. Barrett and B.F. Buxton: *Predicting Biochemical Interactions – Human P450 2D6 Enzyme Inhibition* CEC 2003, pp 807-812
- Morris2000: Protein-Ligand Docking, in: *Evolutionary Algorithms in Molecular Design* Wiley-VCH, Weinheim, 2000
- Nicolotti2002: Orazio Nicolotti et al: *Multiobjective Optimization in Quantitative Structure-Activity Relationships: Delivering Accurate and Interpretable QSARs* J.Med.Chem 2002, 45, 5069-5080
- Pegg2001: Scott C.-H. Pegg, Jose J. Haresco and Irwin D. Kuntz: *A Genetic Algorithm for structure-based de novo design*, Journal of Computer Aided Molecular Design 15(2001): 911-933
- Repsilber2002: Dirk Repsilber and Hans Liljenstroem and Siv G.E. Andersson: *Reverse Engineering of regulatory networks: Simulation studies on a genetic algorithm approach for ranking hypotheses,* BioSystems 66 (2002) 31-41
- Sheridan2000: Robert P. Sheridan, Sonia G. SanFeliciano and Simon K. Kearsley: *Designing Targeted Libraries with Genetic Algorithms*, Journal of Molecular Graphics and Modelling 18 (2002), 320-334