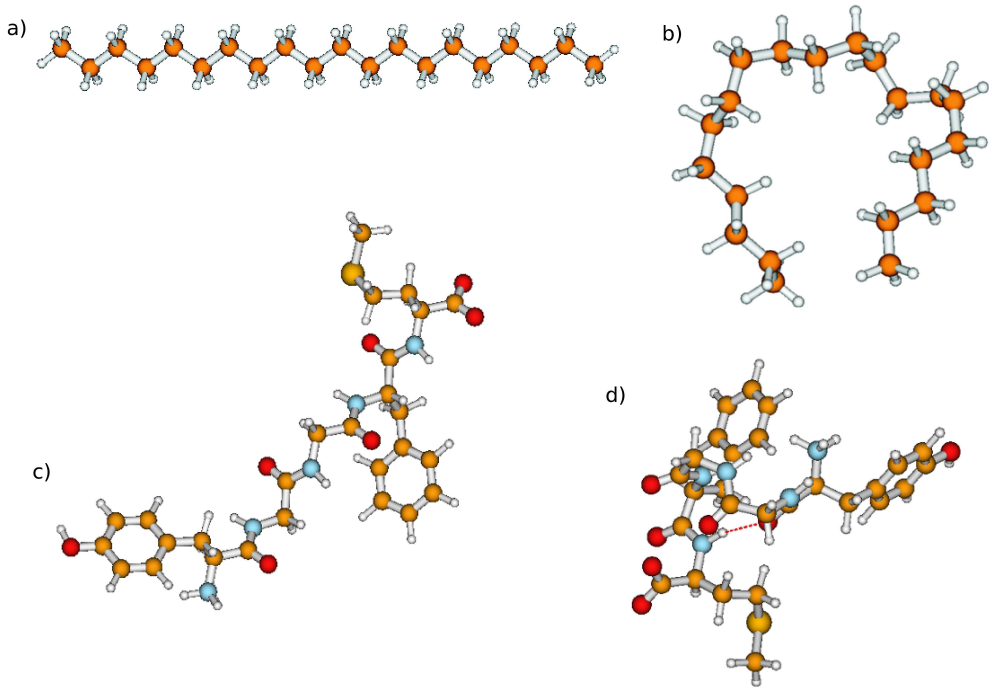
New Neural-Genetic Hybrid Algorithm: Optimization of Molecular Structure

Anna Styrcz, Janusz Mrozek

March 19, 2010

Goal

Medium-sizes molecules global optimization



Global optimization of molecular structure

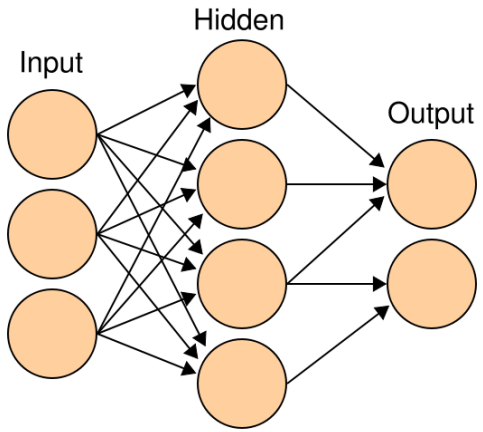
Finding the lowest energy conformation of a big molecule:

- extremely complicated and time-consuming
- number of local minima grows exponentially with the number of atoms
- butane (C_4H_{10}) – 3 conformations
- icosane ($C_{20}H_{42}$) – over 100 000 000 conformations (optimization of 180-dimensional function)

Genetic algorithms

- global search technique
- use techniques inspired by evolutionary biology: inheritance, mutation, selection and crossover
- often can rapidly locate good solutions
- numerous parameters need to be set (mutation rate, crossover rate, selection type, ...)
- different problem – different set of optimal parameters (static or dynamic)

Neural network



- simulates activity of biological neural networks
- modelling relationships between inputs and outputs
- finding patterns
- generalisation

Memetic algorithm

- population – set of potential solutions (conformations)
- chromosomes – coded structural data (bond lengths, angles, torsion angles)
- genetic operations: selection, crossover, mutation
- molecular energy can be obtained from different chemical software (molecular mechanics, semi-empirical, ab-initio, DFT)
- **local optimization** – can take from fractions of a second to many days
- local optimization of all new solutions – most time-consuming part
- natural parallelisation

Neural network

- input – statistical data about population (minimal, average and maximal distance between chromosomes and energies of all solutions)
- output – parameters of evolution (crossover, mutation and selection rate selection type, ...)
- learning set – population statistics in different states of evolution and corresponding optimal genetic parameters
- evaluation of neural network after each generation of memetic algorithm

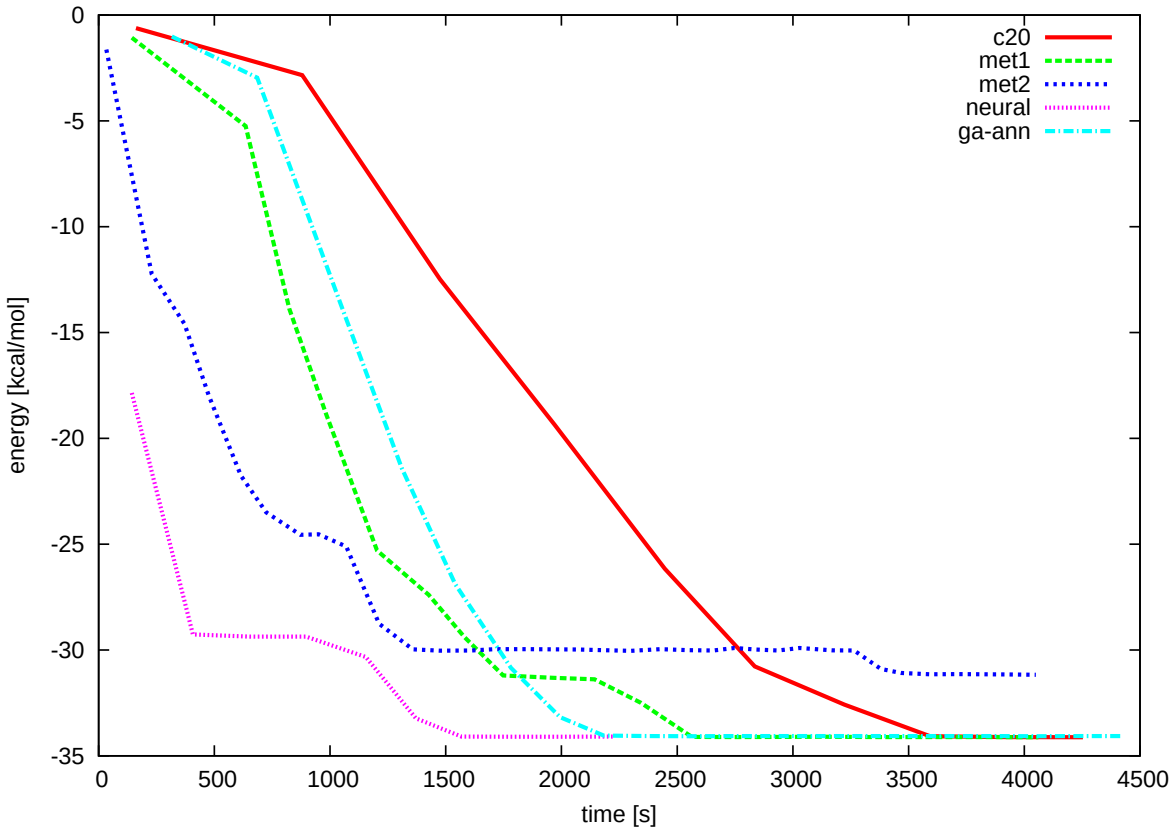
Benchmark

- genetic algorithms
- simulated annealing
- differential evolution
- other genetic-neural hybrids

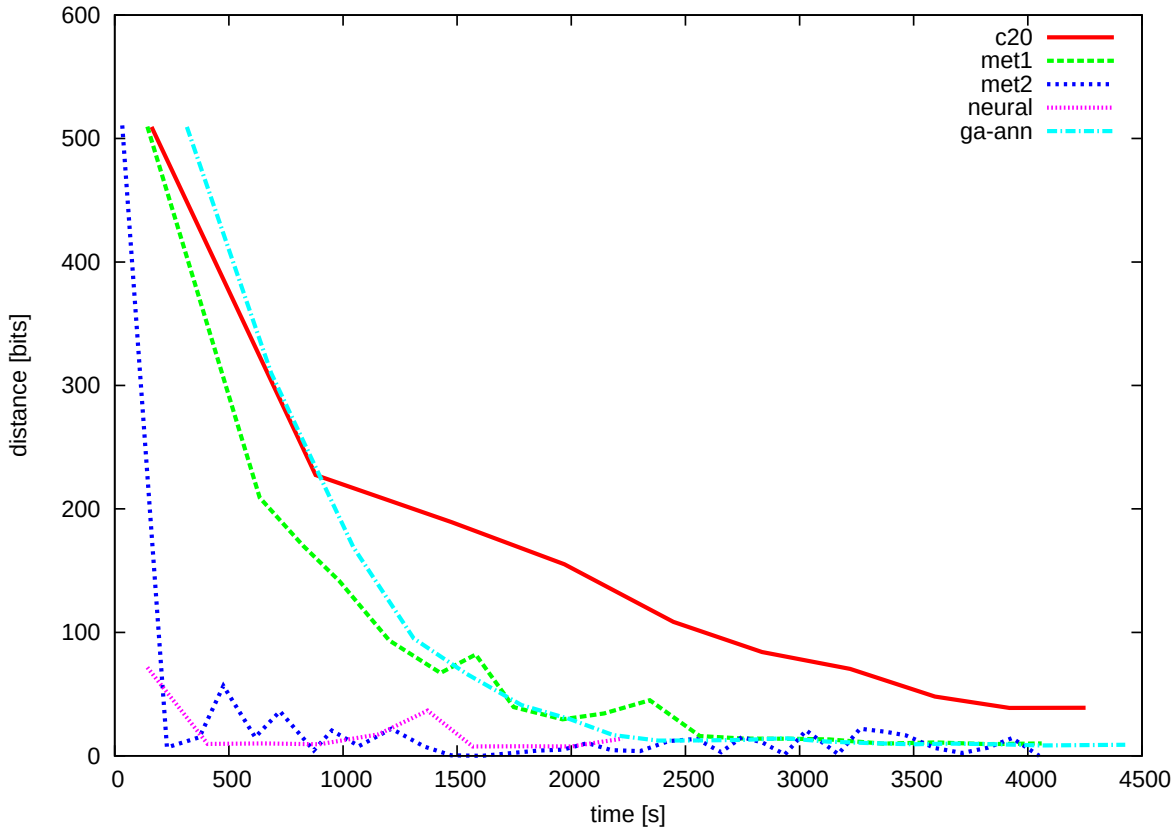
Optimal genetic parameters

parameter	icosane	met-enkephalin
population size	100	50
crossover	singlepoint	multipoint
crossover rate	0.6	1.0
mutation	singlepoint	multipoint
mutation rate	0.3	0.3
selection	roulette	best of 3

Convergence



Influence on population statistics



Results

algorithm	SA+LS	DE+LS	GA+LS
Average	-30.523	-28.783	-33.839
Best	-34.153	-34.157	-34.157

Conclusions

1. Optimal parameters controlling genetic algorithms are strongly dependent on geometry and size of the molecule used for their optimization.
2. Optimal parameters depend on the current state of the evolution – algorithm has to be dynamic.
3. **Neural-genetic hybrid algorithm appears to be the most efficient and universal.**

References

- Pinter, J. D: Global optimization: software, test problems, and applications. In: P. M. Pardalos and H. E. Romeijn (Eds.): Handbook of Global Optimization, Vol. 2, Springer-Verlag, 2002, pp. 515–568.
- Eiben, A. E.—Hinterding, R.—Michalewicz, Z.: Parameter Control in Evolutionary Algorithms. IEEE Transactions on Evolutionary Computation, Vol. 3, 1999, No. 2, pp. 124–141.

Results

algorithm	met1	met2	C_{20}	neural
met	175.759	176.081	187.159	176.303
met-Cl	198.625	198.927	211.000	201.723
met-OH	176.004	176.910	185.207	175.849
met-CH_3	219.395	218.767	227.690	218.762
met-C_3H_7	235.778	234.699	246.763	237.924
met-Ph	312.321	312.758	325.441	311.849
met-Ph-CH_3	312.947	310.904	321.865	310.240