The 7th International Conference
Distributed Computing and Grid-technologies in Science and Education

Russia, Moscow region, Dubna, JINR, LIT
4-9 July 2016

Distributed evolutionary optimization algorithms for peptide structure prediction

S.V. Poluyan
svpoluyan@gmail.com

Dubna State University

Faculty of Natural and Engineering Science

Postgraduate student

N.M. Ershov
ershovnm@gmail.com

Lomonosov Moscow State University

Faculty of Computational Mathematics and Cybernetics

Peptide structure prediction (primary → secondary)

Peptide

- biomolecule
- chain of amino acids linked by peptide bonds
- 20 standard amino acids

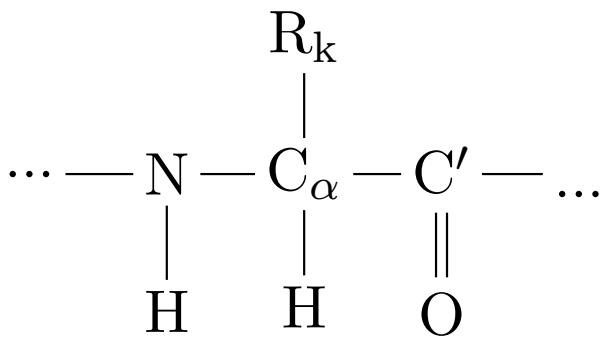
Native conformation from Anfinsen's thermodynamic hypothesis

- Unique three-dimensional structure
- Unique global minimum of free energy

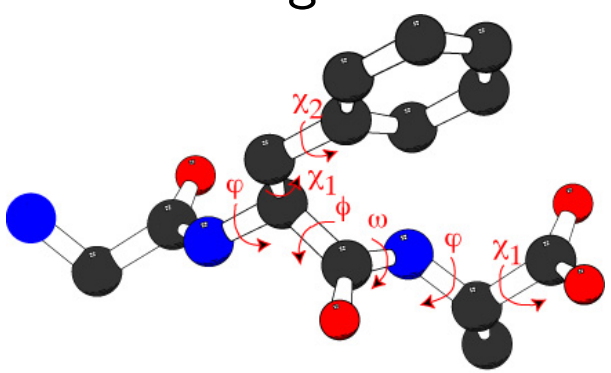
Structures

- Primary: the amino acid sequence
  Example: Alanine-Histidine-Glutamine-· · · -Lysine-Tyrosine
- Secondary: general three-dimensional form of local segments of peptides
  Most common are α-helix and β-hairpin

Polypeptide chain



Torsion angles

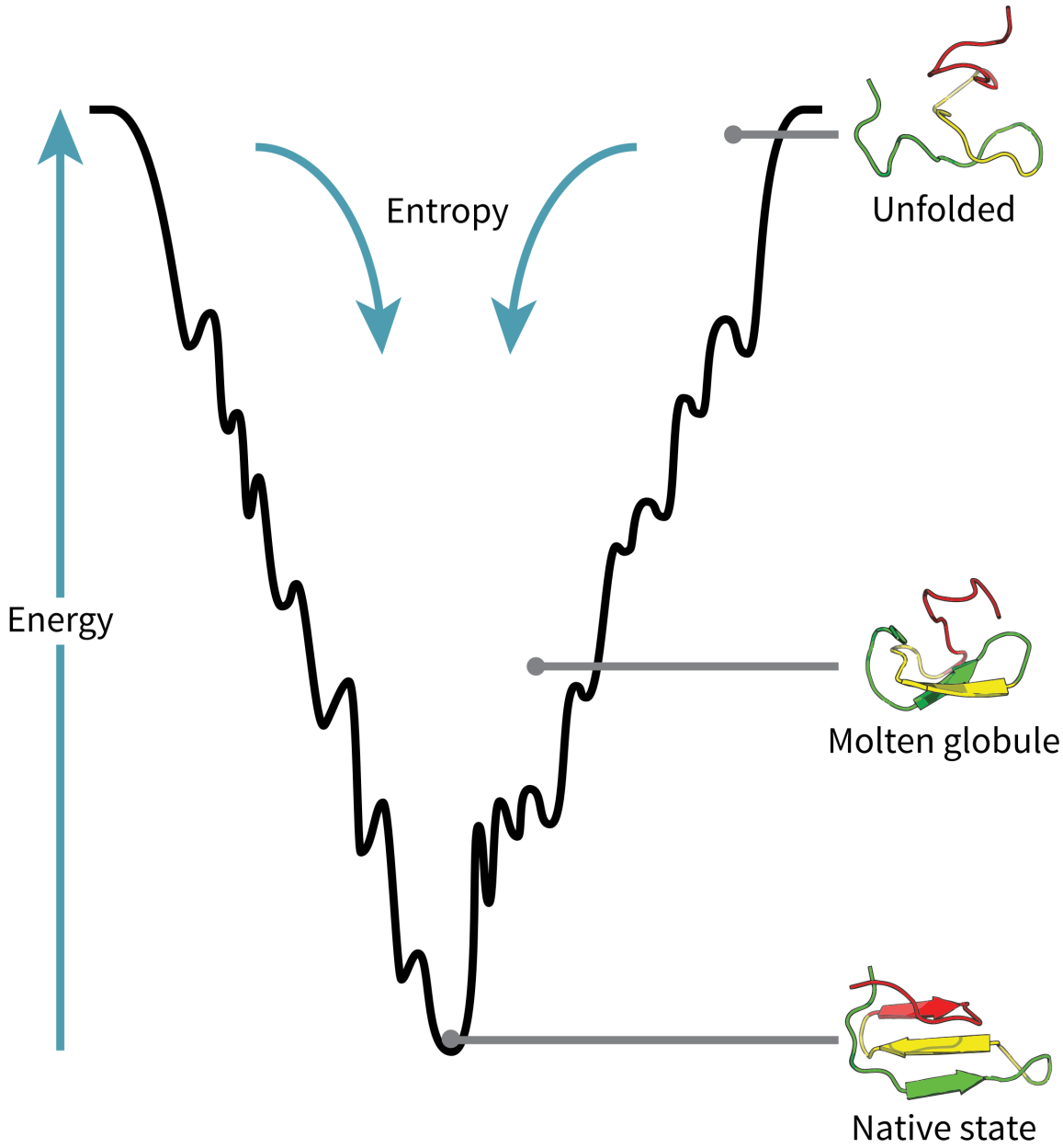


- Side-chain
  $\chi_{1-4} \sim [-\pi, \pi]$
- Main-chain
  $\phi : [-\pi, \pi]$
  $\psi : [-\pi, \pi]$
  $\omega \sim$ planar

Motivation

- Is it possible to predict peptide secondary structure using evolutionary algorithms by global optimization of an all atom potential energy function without any constraints on main-chain/side-chain torsion angles and other approaches (psipred/psi-blast/low-energy fragments/coarse-grained force fields)?

Free energy landscape (concept)



Rosetta all-atom force-field

Rosetta uses a mix of statistical and physical potentials

- Attractive and repulsive forces are modelled with the Lennard-Jones potential
- Lazaridis–Karplus implicit solvation
- Coulombic electrostatic potential with a distance-dependent dielectric

Hydrogen bond terms

- Backbone-backbone hbonds close in primary sequence
- Backbone-backbone hbonds distant in primary sequence
- Sidechain-backbone hydrogen bond energy
- Sidechain-sidechain hydrogen bond energy

Knowledge-based terms

- Ramachandran preferences
- Probability of amino acid at ϕ/ψ
- Omega dihedral in the backbone (harmonic constraint on planarity)
- Internal energy of sidechain rotamers as derived from Dunbrack's statistics

$$Energy = w_1 \cdot term_1 + w_2 \cdot term_2 + \ldots$$

Weights on the score terms are calibrated. Energy score is not kcal/mol.

Selected algorithms

Evolutionary computation

- Stochastic optimization
- Heuristic algorithms
- Evolution of the population
- Similar structures but different mechanisms (operators)
- Easy to parallelize

Evolutionary algorithms & strategies

- JADE: Adaptive Differential Evolution With Optional External Archive
- ESCH: Evolutionary algorithm
- PSO SW: Particle Swarm Optimization with Solis and Wets local search
- SABFO: Self-Adaptive Bacterial Foraging Optimization Algorithm
- CSO: Competitive Swarm Optimizer
- NCGA: Non-uniform Cellular Genetic Algorithm
- CMAES: Covariance Matrix Adaptation Evolution Strategy
- JDE CMAES: Adaptive Differential Evolution with CMAES local search

Simulated annealing for the weight of electrostatic potential.

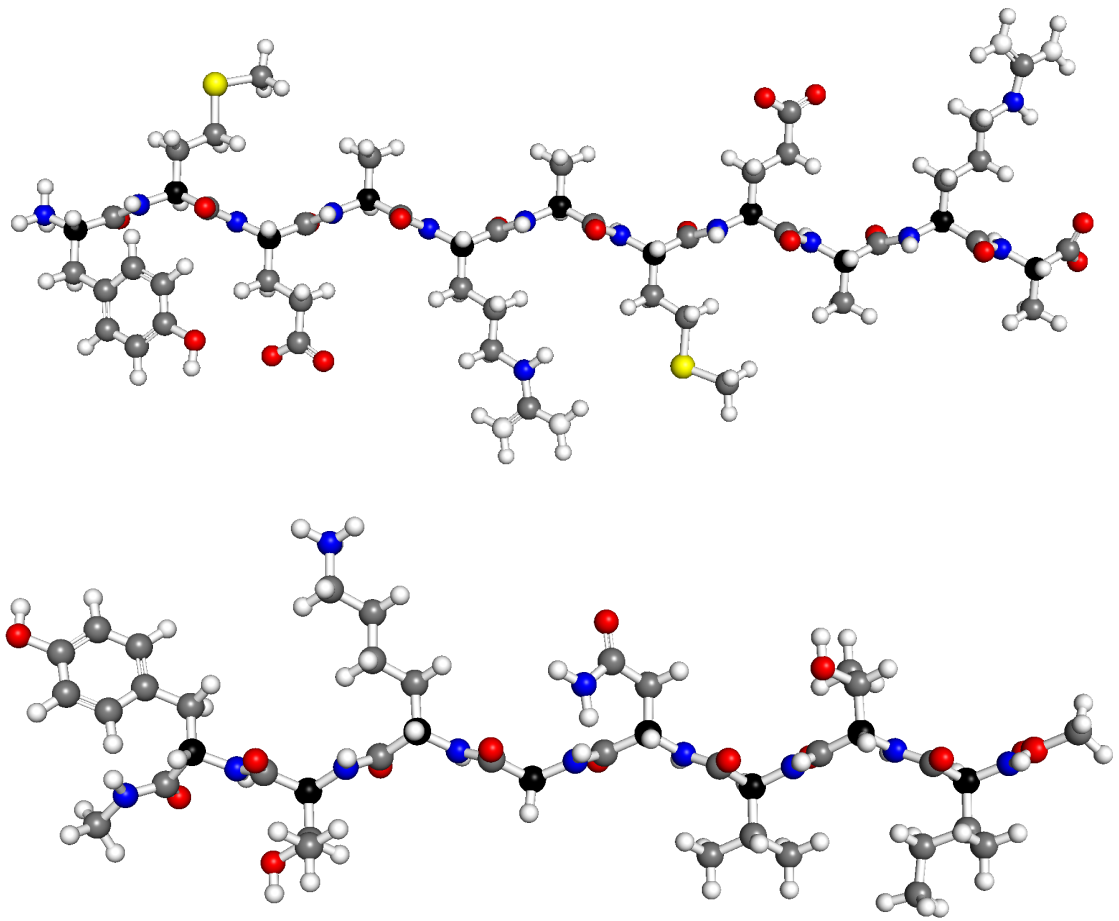
Test set of peptides

α-helices

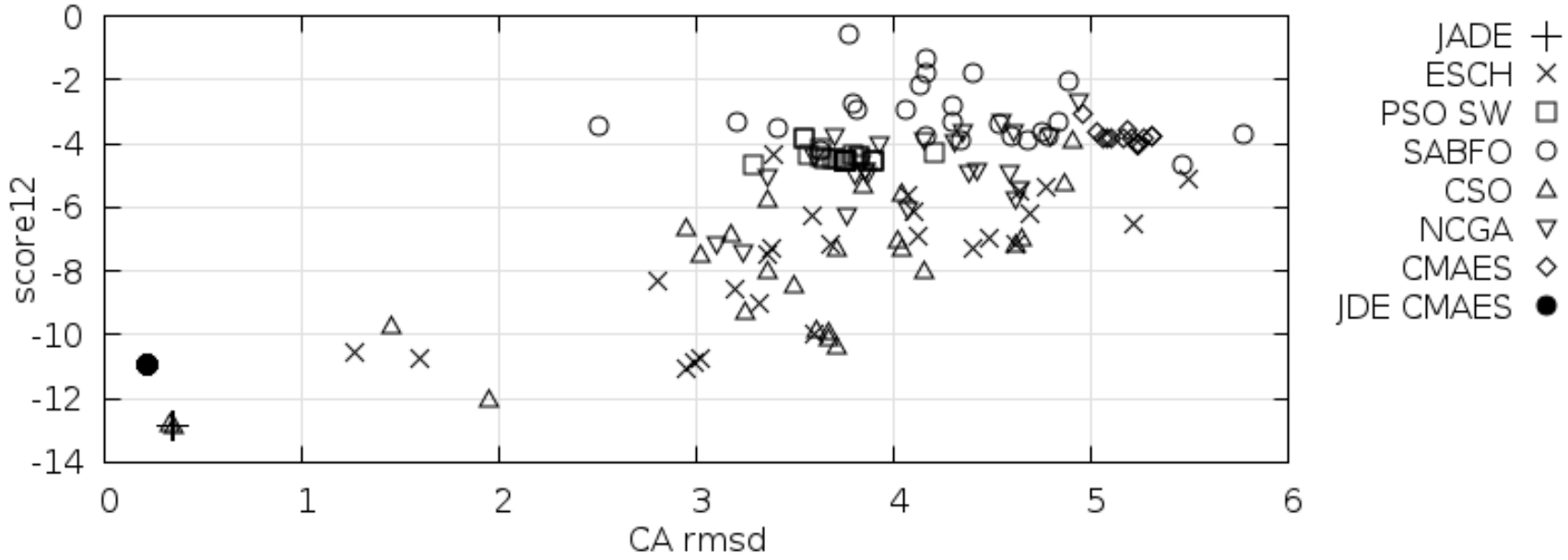
- AAAAAAAAAA
- YMEARAMEARA
- CFSSARLSC

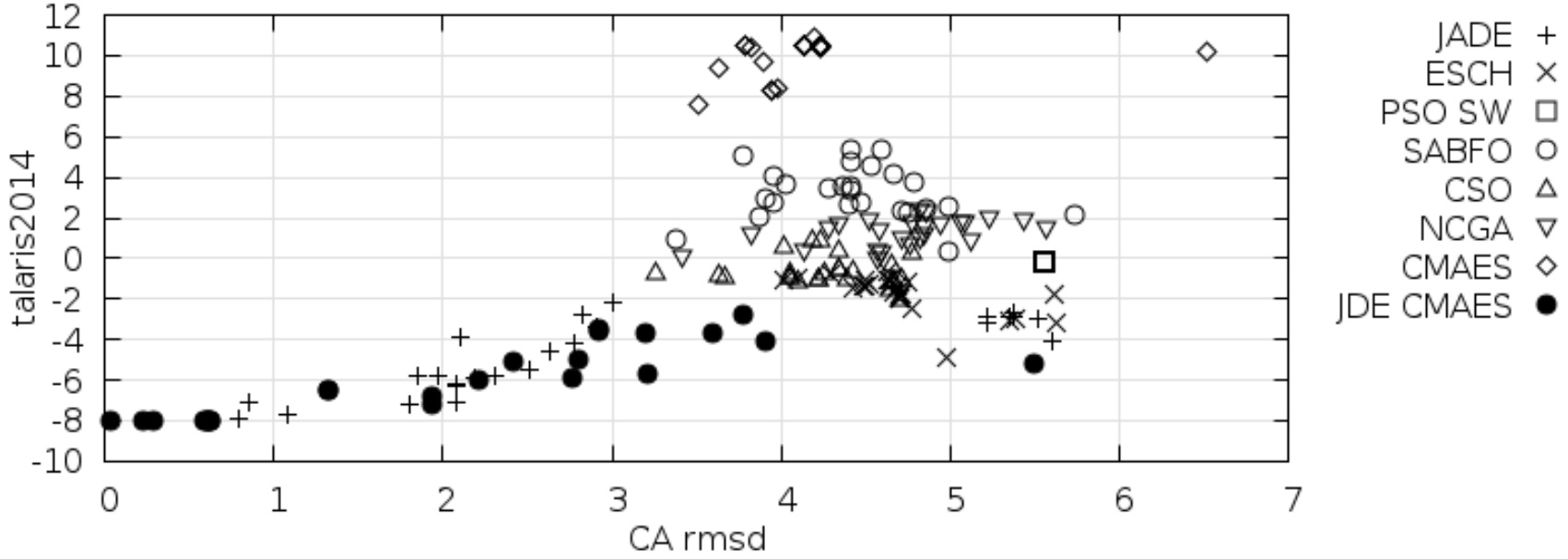
β-hairpins

- VVVVGGVVVV
- Ace-ITVNGKTY-Nme
- GYDPETGTWG

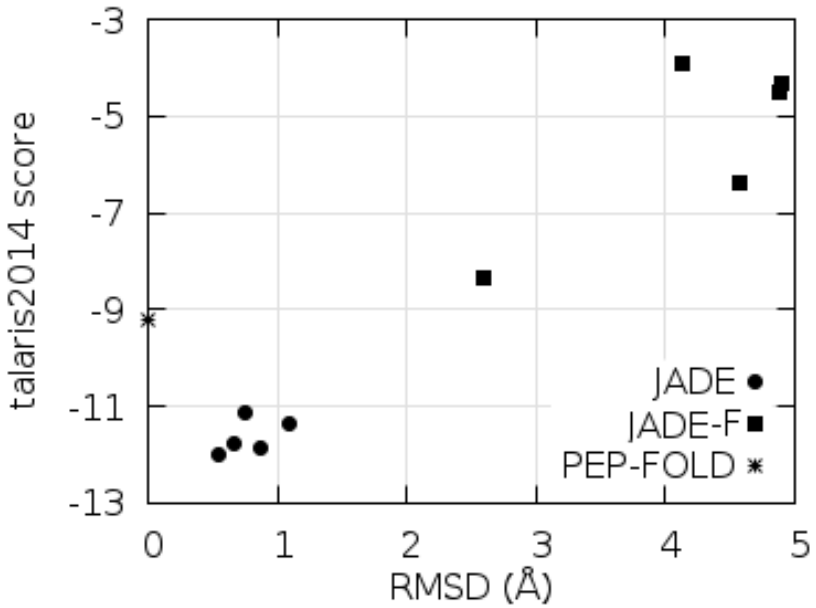


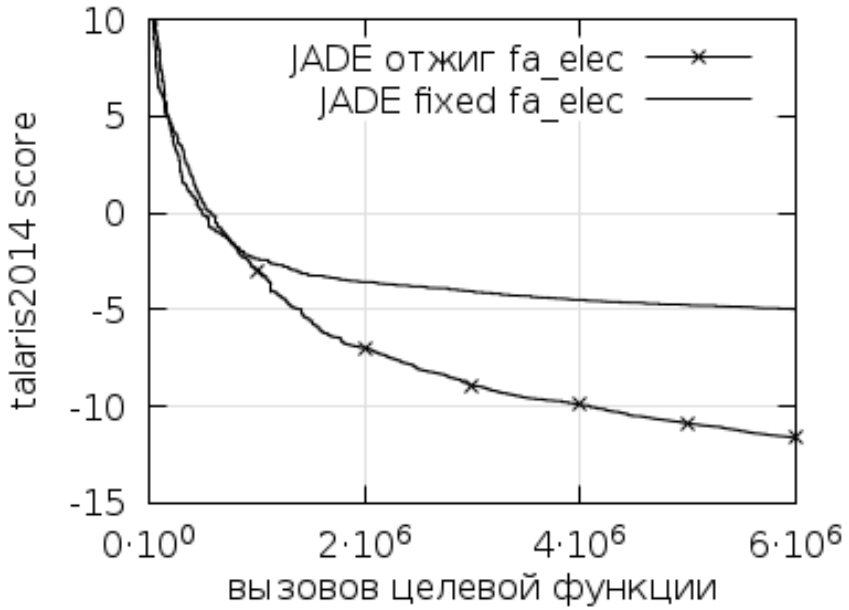
Results of numerical experiments





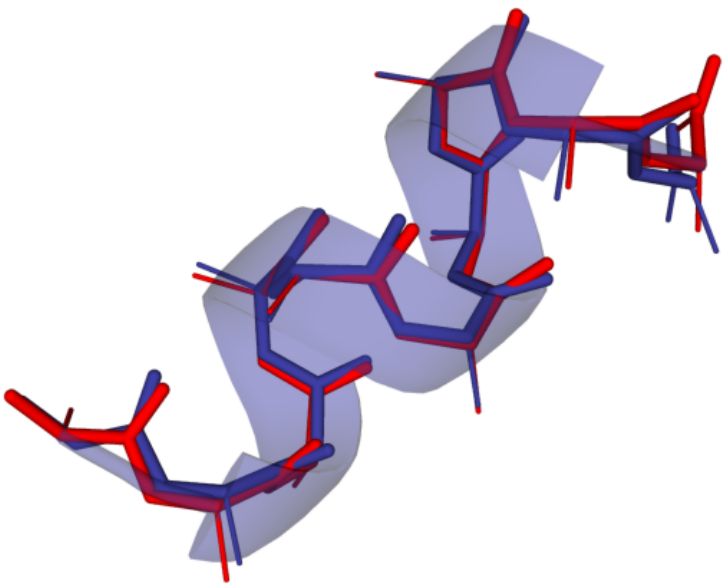
Optimization with simulated annealing



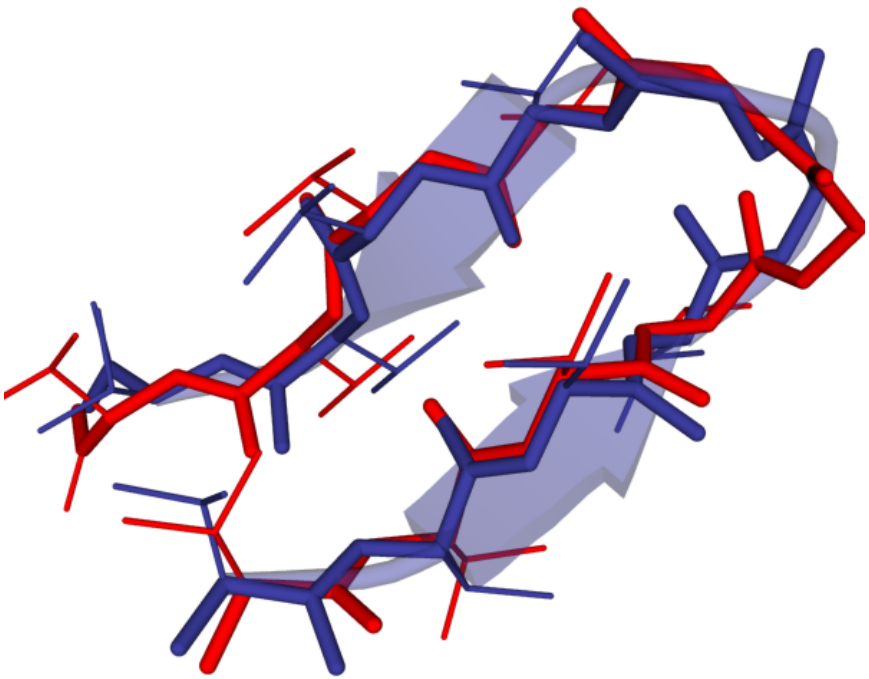


Pep-fold vs. JADE

A10

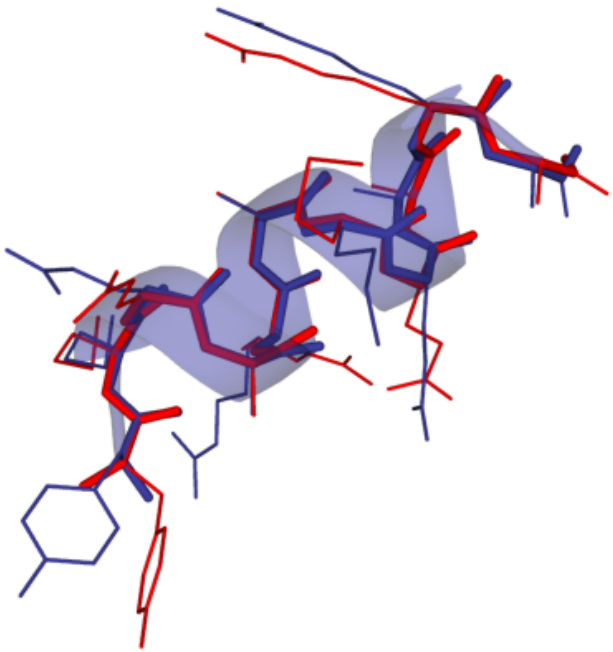


V4GGV4

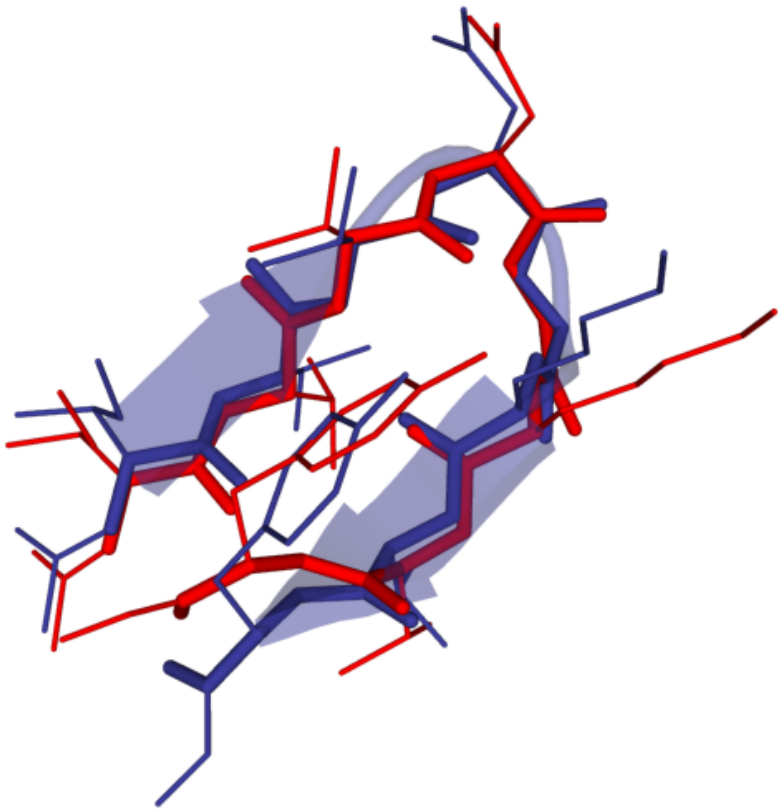


Pep-fold vs. JADE

YMEARAMEARA

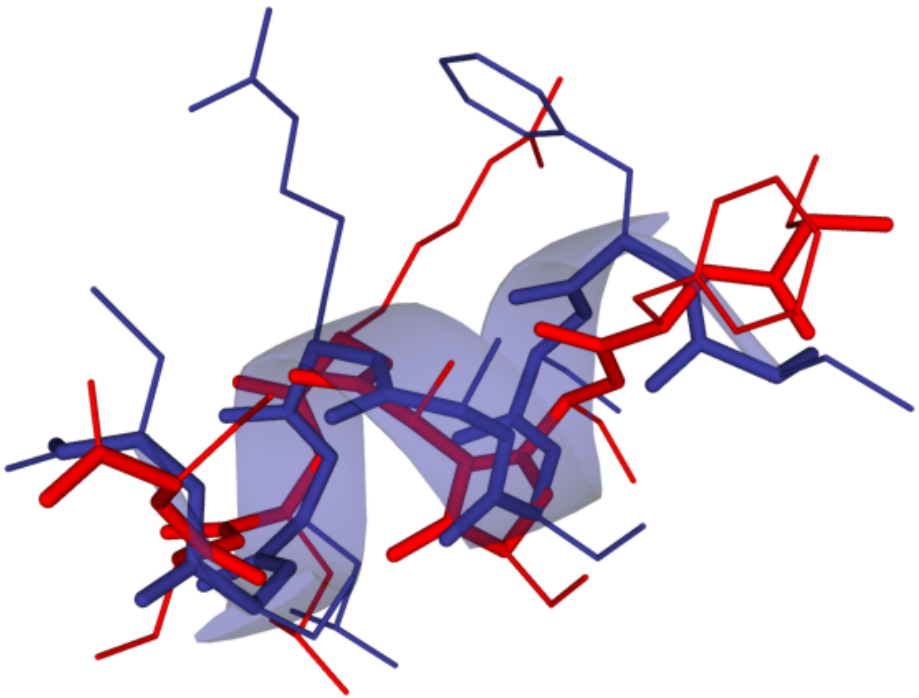


Ace-ITVNGKTY-Nme

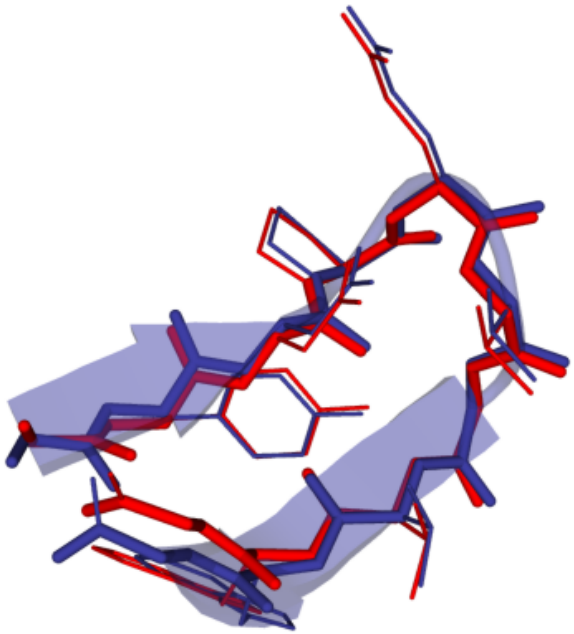


Pep-fold vs. JADE

CFSSARLSC



GYDPETGTWG



Conclusions

Q: Is it possible to predict peptide secondary structure using evolutionary algorithms by global optimization of a potential energy function?
A: Yes, but only for a very short peptides (up to ~ 12 residues) and with scheme for changing electrostatic term during optimal structure search.

Q: What algorithms (operators) are more preferable for such tasks?
A: From our experiments, it appears that the Differential Evolution (DE) shows better results. It's most likely that using JADE or hybrid evolutionary algorithms with DE will show good results.
The most productive mutation scheme:
$v_i = x_i + F_i(x^p_{best} - x_i) + F_i(x_a - \tilde{x}_b)$.

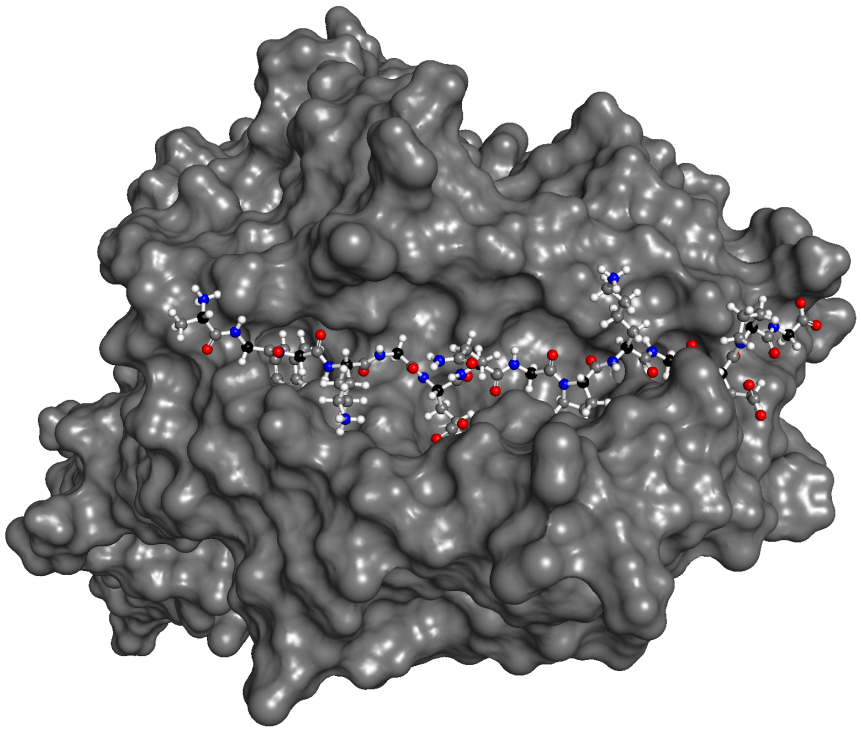
Crucial problem for all algorithms is population degeneration.

Implementation features

- All algorithms are implemented using C++ with MPI/OpenMP
- Benefits of using one node (24 threads) of HybriLIT cluster (JINR) – The largest speed-up is up to 20 times

Future Work

- Multiobjective optimization – bonded/non-bonded energy terms or others terms presented in Rosetta
- Other algorithms: Asynchronous differential evolution with adaptive correlation matrix
- Ligand–Protein docking: evolutionary algorithms show good results
- Peptide–Protein docking: ?



Thank you!