Team meeting 2012

Sune S. Nielsen
PhD student

Computer Science and Communications (CSC)

CSC | COMPUTER SCIENCE AND COMMUNICATIONS RESEARCH UNIT

Outline

Outline

Thesis overview

- Decision-theoretic Fine Tuning of Multi-objective Co-Evolutionary Algorithms
  - Prof Pascal Bouvry, Prof Nikos Vlassis, Dr Grégoire Danoy
  - Started 15. September 2011
  - Part of Evoperf work package 3
- Interdisciplinary research
  - Luxembourg Centre for Systems Biomedicine (LCSB)

Thesis work areas

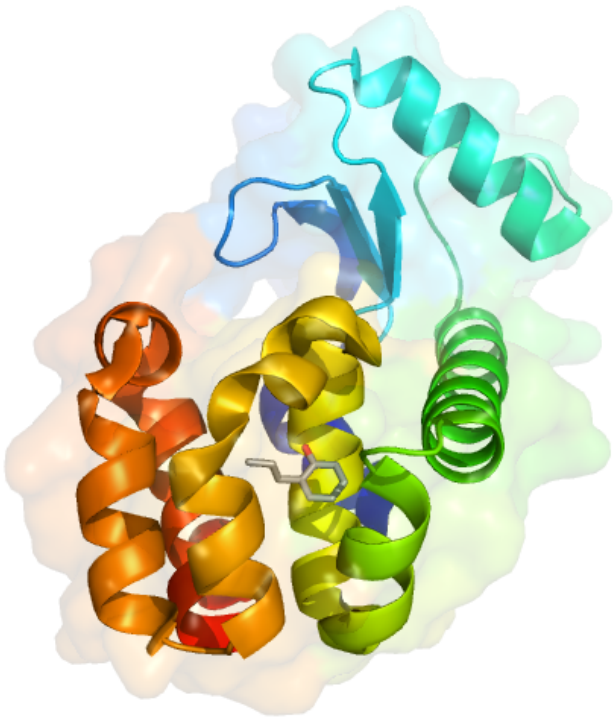
- Co-evolutionary optimisation
  - Fitness evaluation is based on interaction of multiple individuals or sub-populations
- Decision theory
  - Re-inforcement learning
  - Markov-models
  - Fuzzy logic
- Bio-informatics problems
  - RNA assembly (on hold)
  - Protein strucure optimisation and design
- Other
  - Multi-objective benchmark problems
  - VehILux optimisation and model enhancements

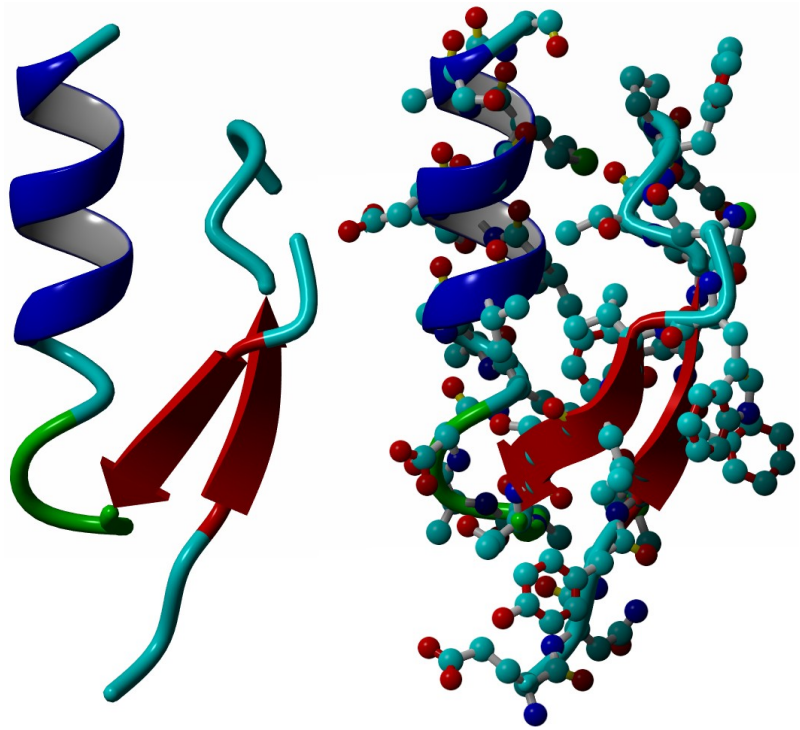
Outline

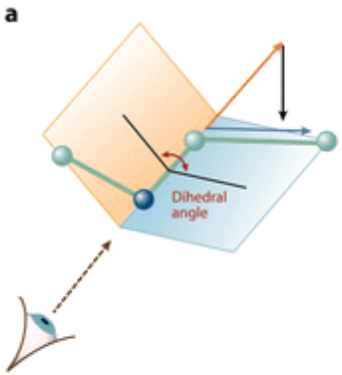
What? Why?

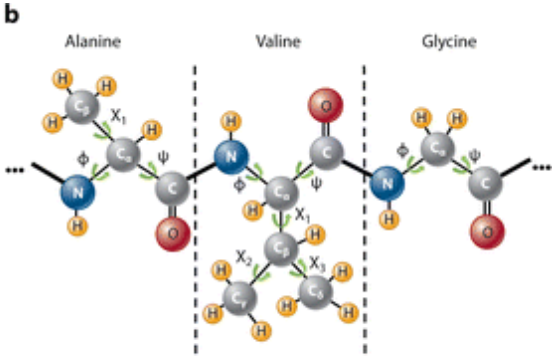
- Cooperate with LCSB and use their theories and knowledge
- Find different proteins with same or similar structure
- Ultimately the results may be used in drug-design etc.

Protein structure









Nucleotides and codons

- Representation of the individual of up to 1000 amino acids
  - The sequnce alone guides folding into the resulting molecular structure, which is hard to predict computationally
- Three nucleotides (A,C,G,T) represent one codon, giving $4^3 = 64$ possbile combinations
- Each codon again represents one amino-acid where multiple codons can correspond to the same amino-acid
  - There are in total 20 standard amino-acids
  - Exception are three possible stop codons

Objective functions

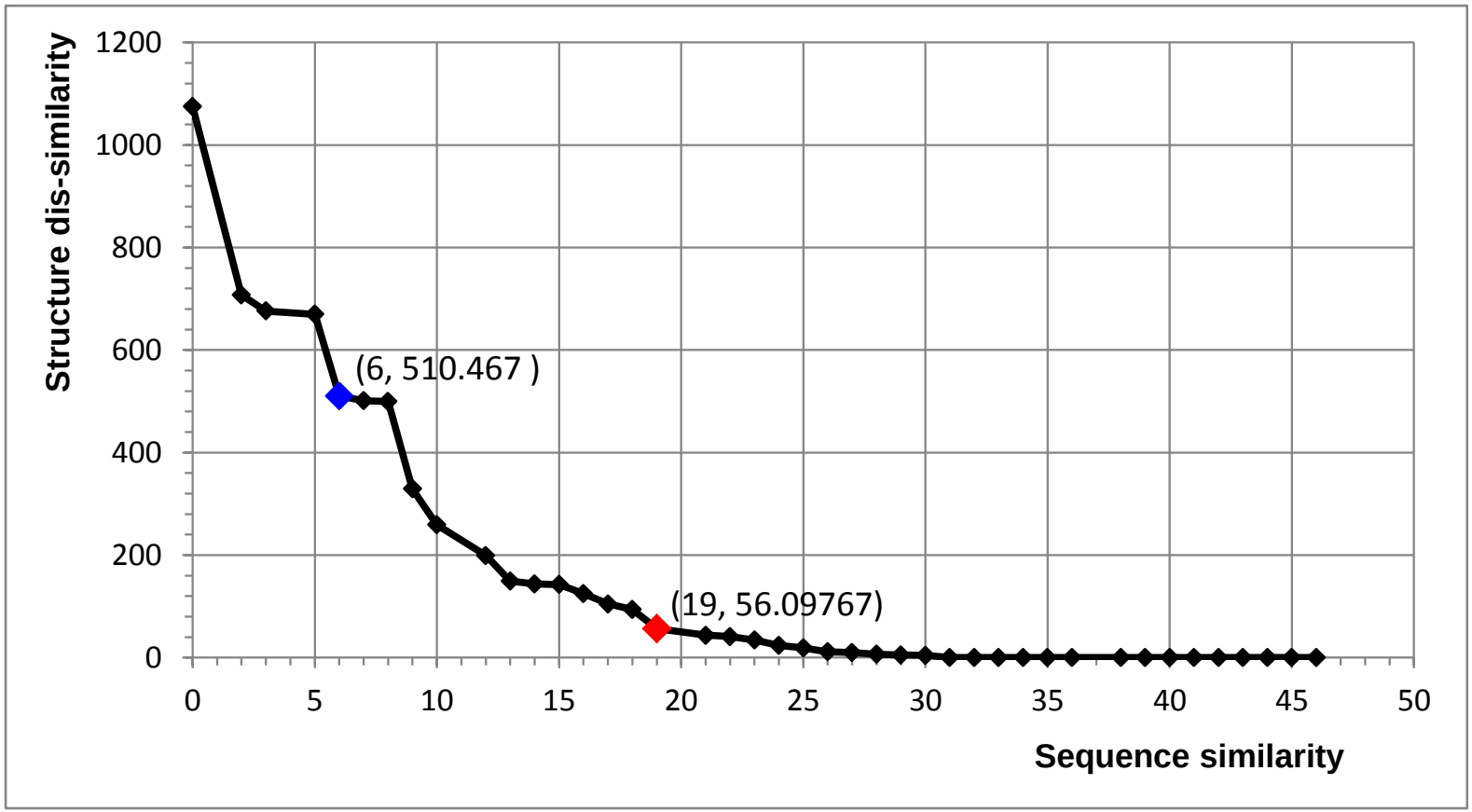
- Structure similarity
  - Local and global alignent(LGA)[Zem03] score
    - Global alignment: root mean square deviation (RMSD) of C_α atoms
    - Local alignment: generates many different local super-positions to detect regions where proteins are similar
- Sequence dis-similarity
  - In bio-informatics this task is often solved using a Dynamic Programming (DP) method [Edd04]
    - Here, a positive score is assigned to each matching codon and penalties to gaps and mismatches.
  - In preliminary experiments a simple count of the number of same codons is used.
- Energy estimate
  - Aggregated energy functions, $E_{vanderWaals}$, $E_{electrostatic}$, E_{bond} etc...

Current and future process items

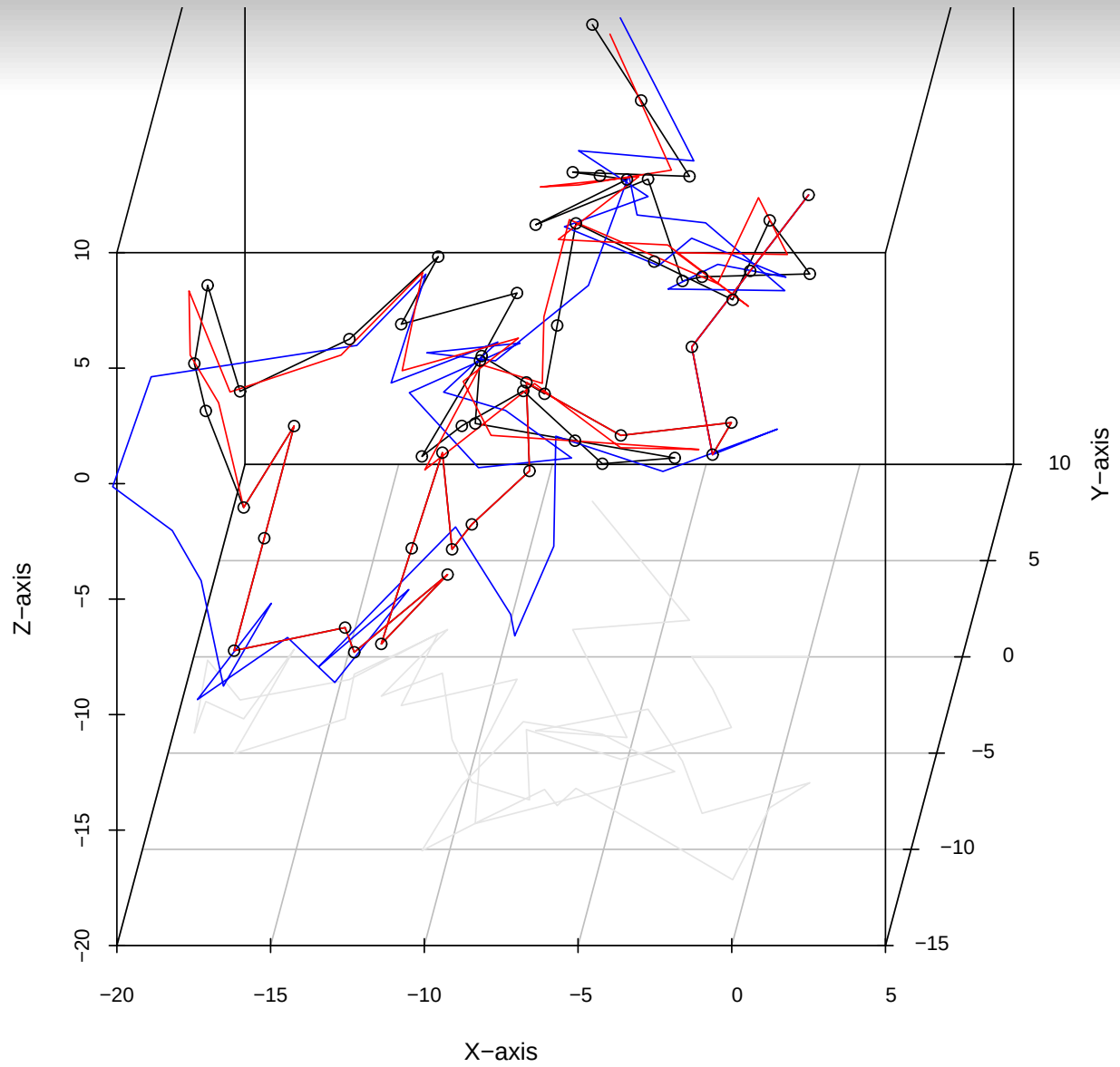
- Optimisation algorithm generates individual sequences
- Sequnce similarity relatively cheap
- Remaining objective evaluations in different levels of precision
  - Before and after costly relaxation (folding) of structure
  - Energy computation on backbone alone or with atoms and side-chains added
- Fast optimisation of prospect individuals to then undergo heavier molecular dynamics simulation
  - Skip evaluation of bad individuals
  - Skip evaluation of similar individuals already evaluated thoroughly
  - Split evaluation in parts of the protein and make optimisation co-evolutionary
- Further constraints to increase likelihood of obtaining similar sructure

Pareto front

- The below pareto front of best solutions was found during a preliminary experiment on the protein Nisin53



Generated backbones



Outline

Discussion and questions

Outline

Bibliography I

Sean R. Eddy.
What is dynamic programming?
Nat Biotech, 22(7):909–910, July 2004.

Adam Zemla.
Lga: a method for finding 3d similarities in protein structures.
Nucleic Acids Research, 31(13):3370–3374, 2003.