PHAISTOS – software for inference of protein structure

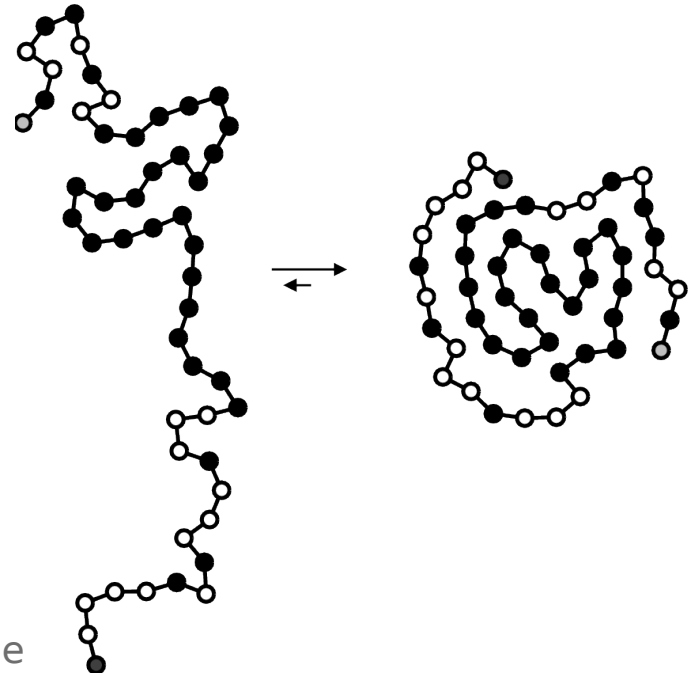
Thomas Hamelryck

Associate professor
Structural Bioinformatics group
Bioinformatics Centre
University of Copenhagen, Denmark

Visiting professor
Department of statistics
School of mathematics
University of Leeds, UK

The structural bioinformatics group

- Bioinformatics Center, KU
  - Two PhD students, one postdoc
- The protein folding problem
  - Biotechnology
    - Enzyme design
  - Medicine
    - Drugs, vaccines
  - New materials
    - Spider silk
- Proteins are linear polymers of amino acids
  - 20 different amino acids
  - Hydrophobic amino acids on the inside
  - Hydrophylic amino acids on the outside
- Sequence encodes a compact 3D shape
  - Protein fold, hydrophobic effect
- Predicting structure from sequence
  - A main open problem in biology and physics

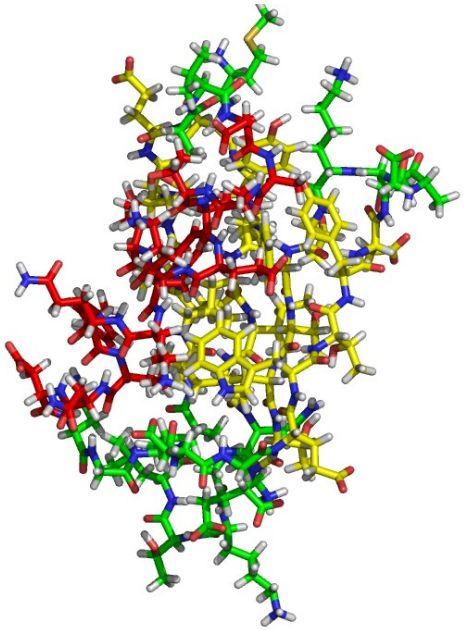


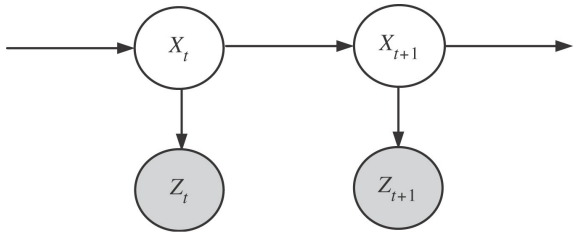
Physics, probabilities, machine learning

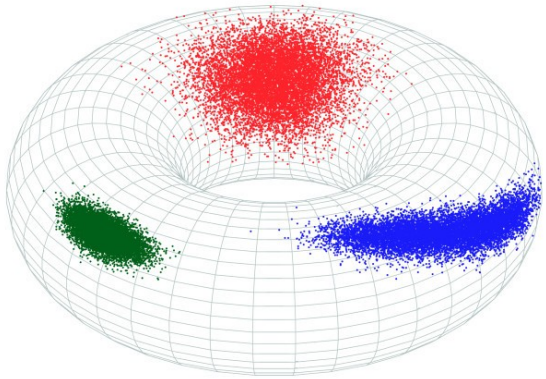
Protein folding problem is a problem in physics
- Fold a protein using physical energy functions

However, it is also a problem in applied statistics
- We extract information from known protein structures using probabilistic models and machine learning
  - Bayesian networks
  - Directional statistics
  - Deep learning
- Combined with physics where needed

Our aims
- Prediction of protein structure from sequence
- Inference of protein structure from sparse data
  - NMR, SAXS, hydrogen exchange,...
- Simulation of protein dynamics
- Protein engineering and design







PHAISTOS

Protein structure prediction, simulation and inference

- Boomsma, W., Frellsen, J., Harder, T., Bottaro, S., Johansson, KE., Tian, P., Stovgaard, K., Andreetta, C., Olsson, S., Valentin, J., Antonov, L., Christensen, A., Borg, M., Jensen, J., Lindorff-Larsen, K., Ferkinghoff-Borg, J., Hamelryck, T. (2013) PHAISTOS: A framework for Markov chain Monte Carlo simulation and inference of protein structure. **J. Comput. Chem.** 34, 1697-705.
- Markov chain Monte Carlo based
  - As opposed to molecular dynamics
- Probabilistic models guide the conformational sampling
  - TORUSDBN, **PNAS** 2008, 2014
- Energy functions
  - OPLS, PROFASI, CHARMM
- Main developer Wouter Boomsma
- Several groups are involved at KU
  - Jan Jensen
  - Kresten Lindorff-Larsen

Freely available from Sourceforge

- Implemented in C++
- http://sourceforge.net/projects/phaistos/

www.phaistos.org

Use case: Prediction of stability

MUMU is a probabilistic model of the local environment of amino acids in protein structures

- Johansson, KE., Hamelryck, T. (2013) A simple probabilistic model of multibody Interactions in proteins. **Proteins** 81, 1340-50.

-Main developer Kristoffer E. Johansson

Can be used to evaluate how well an amino acid "fits" into a given structure

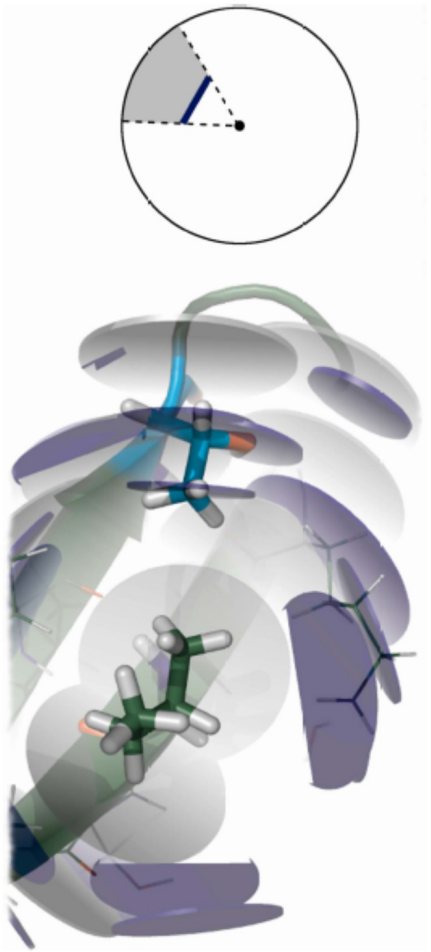
- Amino acid neighbors, volume, backbone angles

$$P(A_i|\mathbf{A}_{\sim i}, \mathbf{X}) = P(A_i|C_i, N_i, V_i, \Phi_i, \Psi_i) = P(A_i|E_i)$$

- Very fast, 30.000 amino acids/minute

Web service

- Fast screening of protein mutants



MUMU example

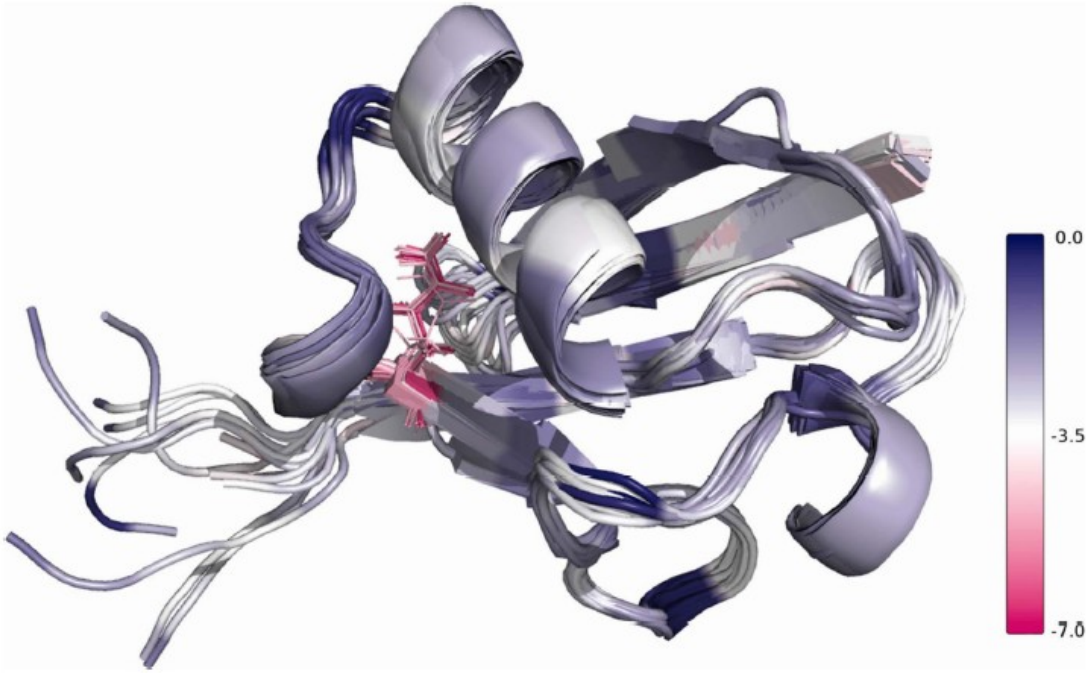


Figure 6

Structural representative ensemble of ubiquitin with residues colored according to the MuMu probability, $\log P(A_i|E_i)$. Log probabilities are indicated in the color bar. Magenta residues have low probability and are thus an unlikely combination of side chain and environment. The side chain of Gln-41 is shown to emphasize that this amino acid has low probability in all structures.

Conclusions & acknowledgements

Phaistos

- Inference of protein structure
- Probabilistic models, physical energy functions
- Use case: protein stability

Acknowledgments

- Wouter Boomsma (TORUSDBN, PHAISTOS, CRISP)
- Jan Valentin (reference ratio *de novo* prediction)
- Sandro Bottaro (CRISP local move)
- Jes Frellsen (MUNINN, reference ratio)
- Kristoffer E. Johansson (MUMU)
- Simon Olsson (TYPHON, ensembles)
- Tim Harder (BASILISK, TYPHON)
- Pengfei Tian (PROFASI implementation)

Collaborators

- Jan Jensen, Kresten Lindorff-Larsen, KU
- Kanti Mardia, John T. Kent, Leeds, UK
- Jesper Ferkinghoff-Borg, DTU, Denmark

http://www.binf.ku.dk
http://www.phaistos.org

Ministry of Science, Innovation and Higher Education