Protein Folding from the First Principles on Large-Scale Computer Infrastructures

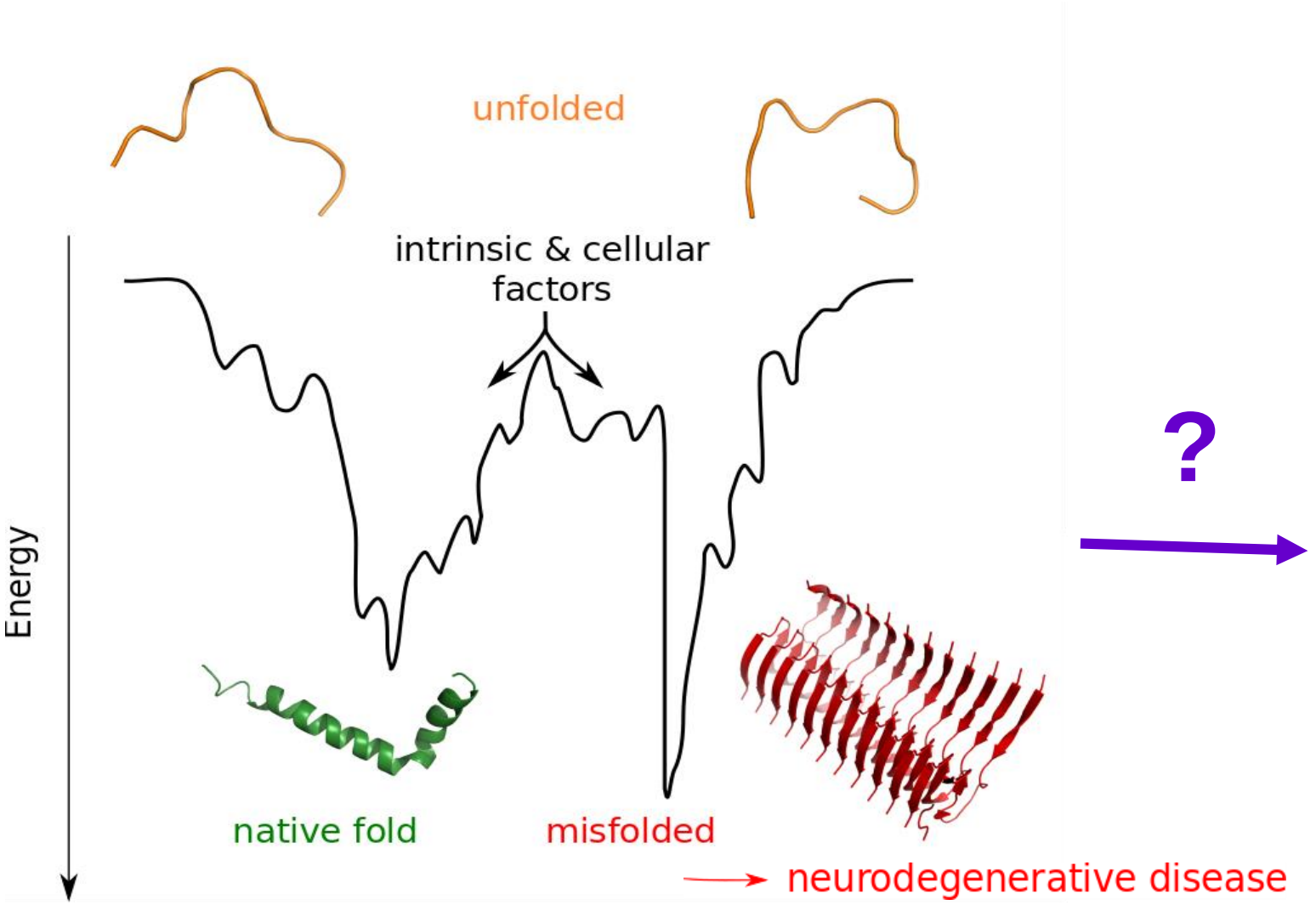
Lubomír Rulíšek

Institute of Organic Chemistry and Biochemistry CAS

Prague, Czech Republic

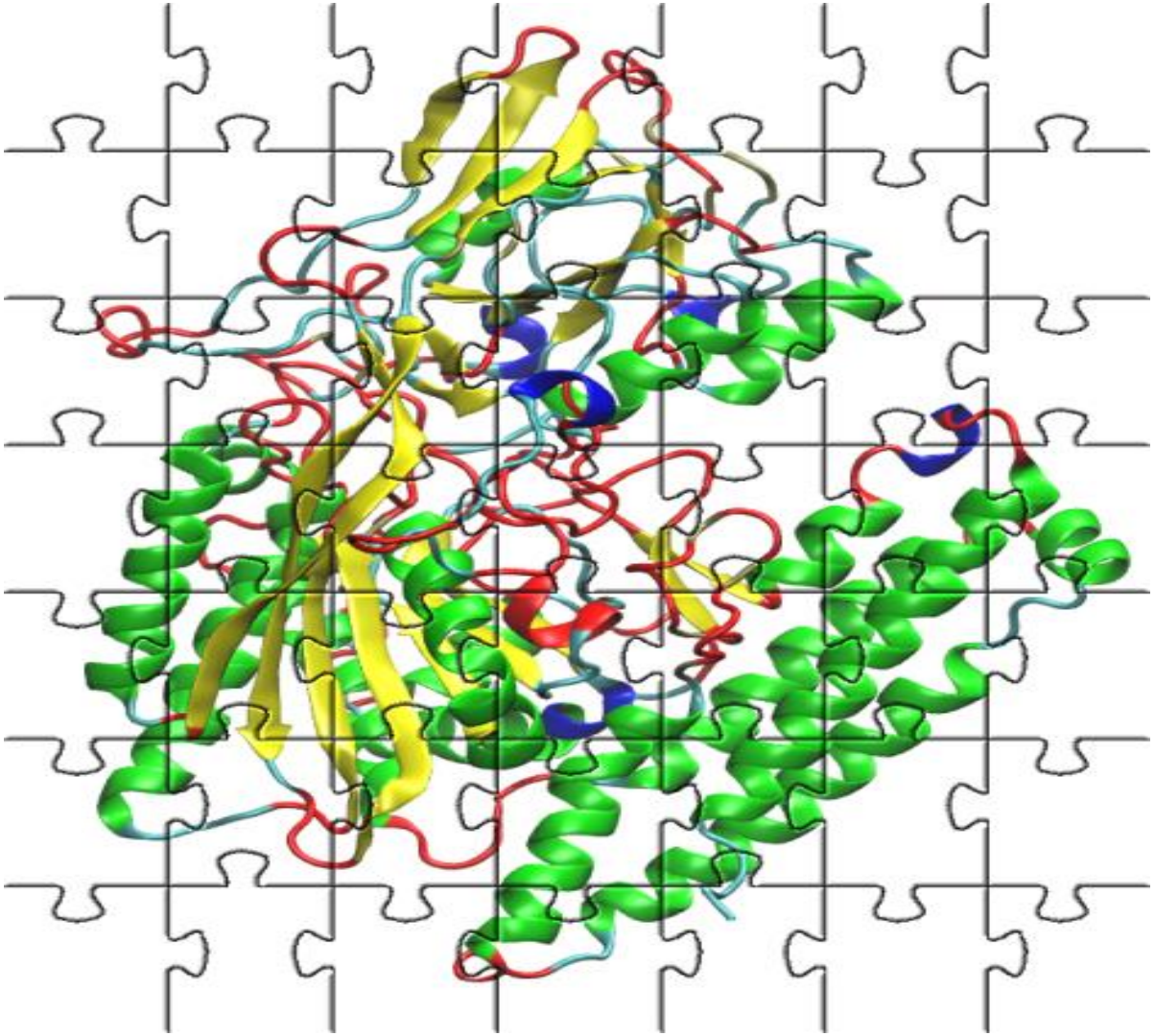
ÚOCHB AV ČR
IOCB PRAGUE

Protein folding as a high-dimensional problem



How much of the folding preference is encoded in the protein sequence?

***Ab initio* investigation of conformational space of peptide fragments**



Starting point

Force field approaches are parametrized to reproduce existing folds ☹

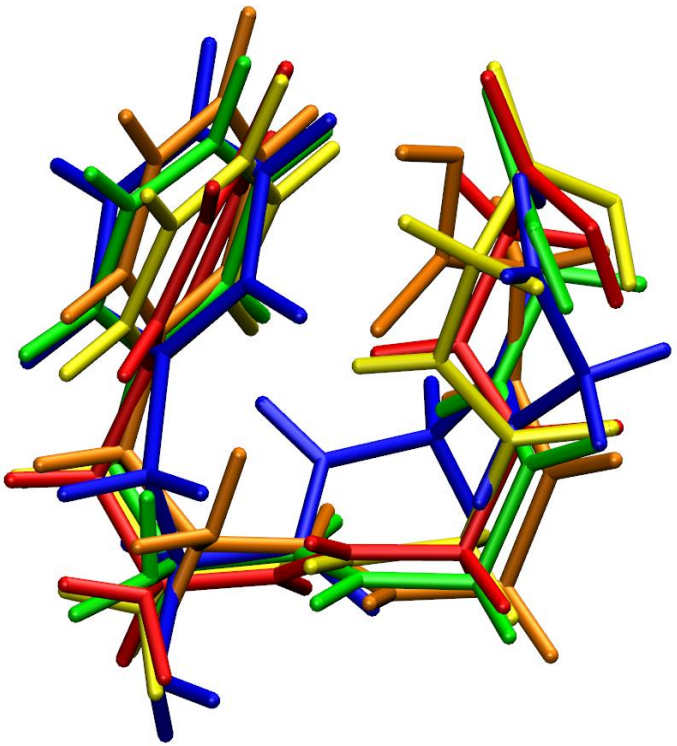
Can we approach the folding *ab initio* without this bias?

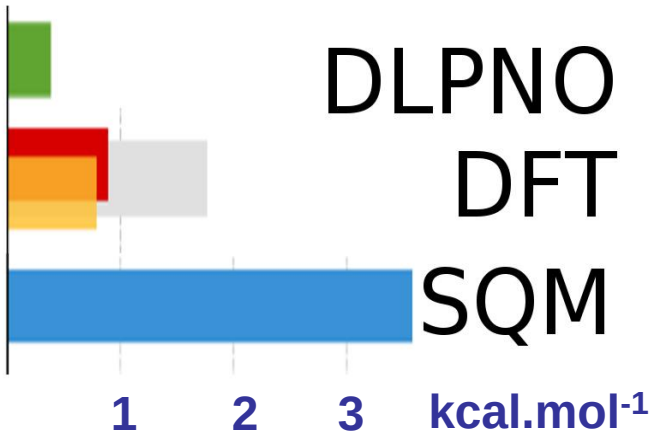
Quantum-chemical approaches demanding
– only smaller peptide fragments can be treated.

Tripeptides: 20 aminoacids → 8000 tripeptides

Selected 4, 5, 6 ... peptides
→ QM geometry optimization of structure ensembles

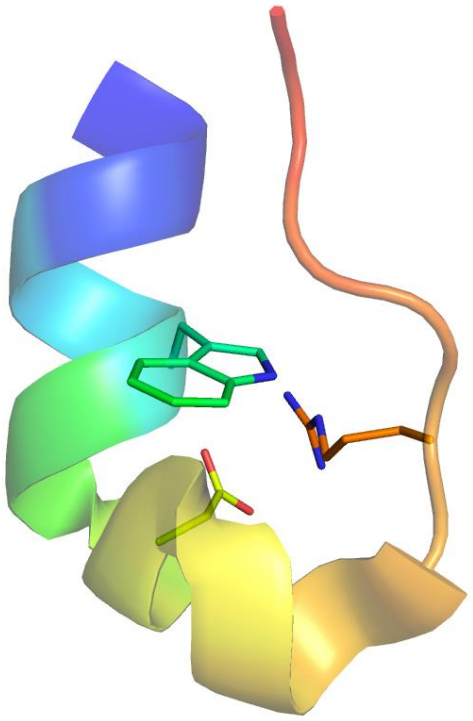
Accuracy of Quantum Chemical Methods



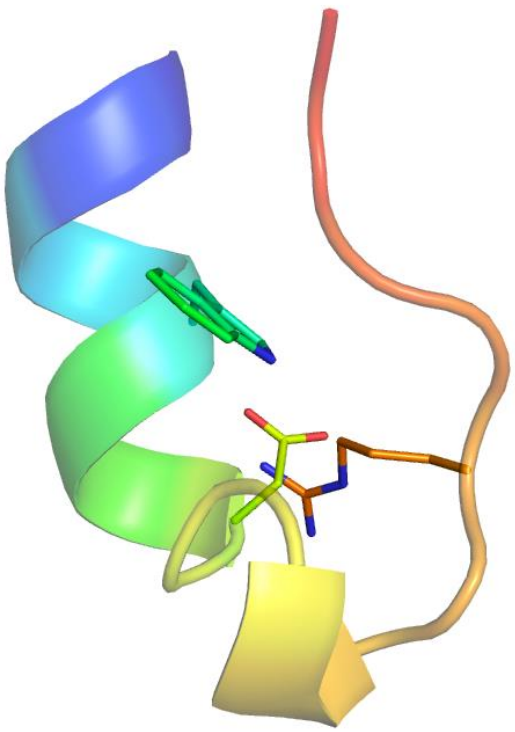


Reduction of dimensionality of the folding problem

- TrpCage miniprotein fold can assembled from tripeptide fragment database



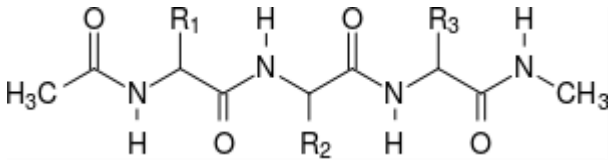
Reference (PDB 2jof)



1 AA from 3pepeptides

Computational resources *estimated*

- G_{conf} (conformer free energy) =
  - $E_{QM(def2\text{-}TZVPD)} + G_{solv(COSMO\text{-}RS)} +$



- <u>For one tripeptide conformer (single CPU):</u>
- QM geometry optimization — 3 - 5 h
- Single points with higher basis set — 2 x 20 min
  - Two calculations for COSMO-RS
- Gas phase optimization & frequency calculation — 4 - 6 h

- CPU time needed for one conformer — up to 12 h
  - Up to 100000 conformers / tripeptide — up to 50000 NCH
  - 8000 tripeptides — up to 400 mil. NCH
- Larger peptide fragments?
- Multiple sampling approaches?