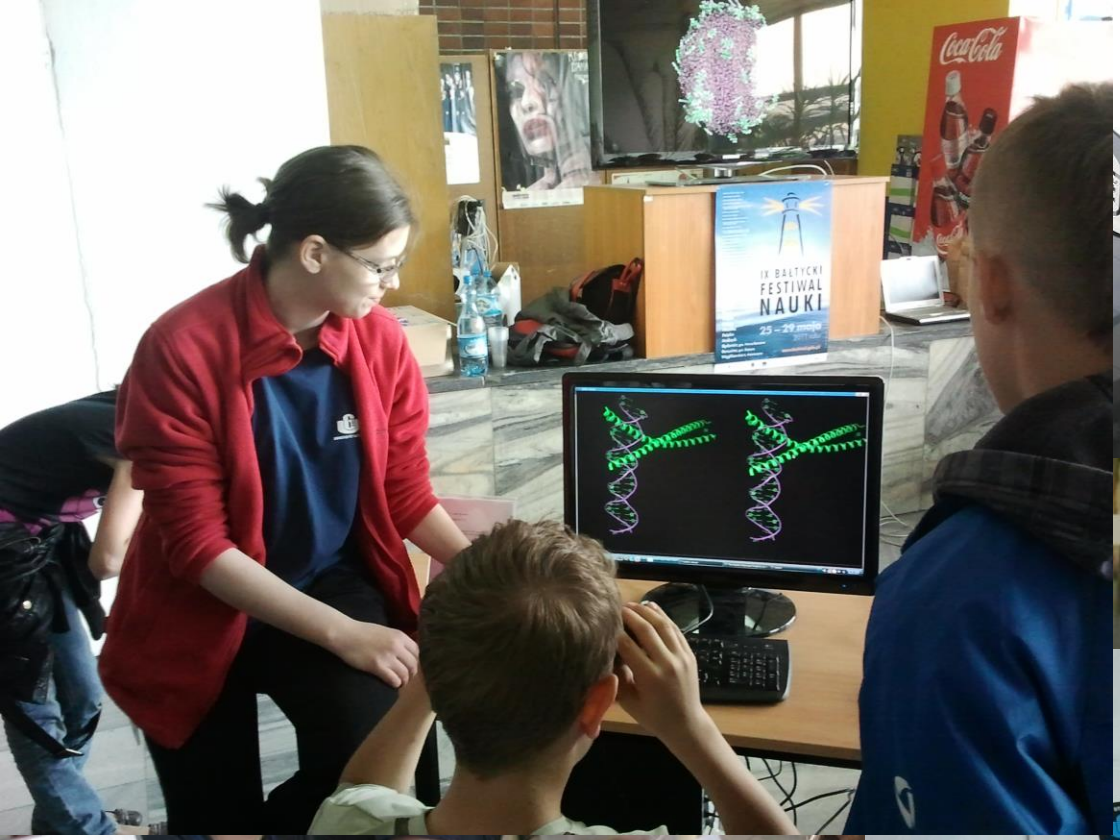
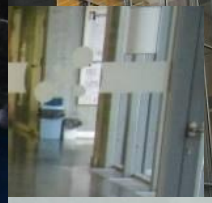
Metody komputerowe przewidywania struktury białek

Magdalena Mozolewska

Instytut Podstaw Informatyki PAN

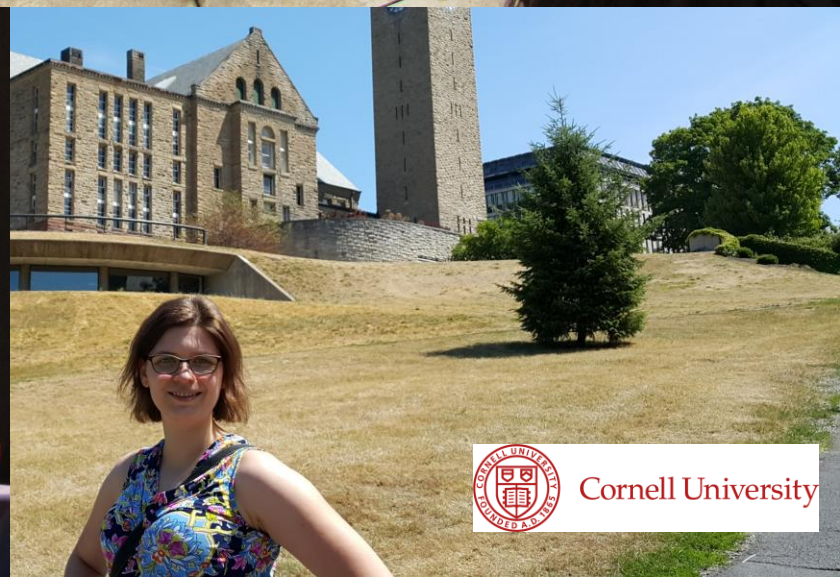

IX BAŁTYCKI
FESTIWAL
NAUKI
Coca-Cola




IX BAŁTYCKI
FESTIWAL
NAUKI


E.Wedel
Mieszanka Wedlowska
Classic

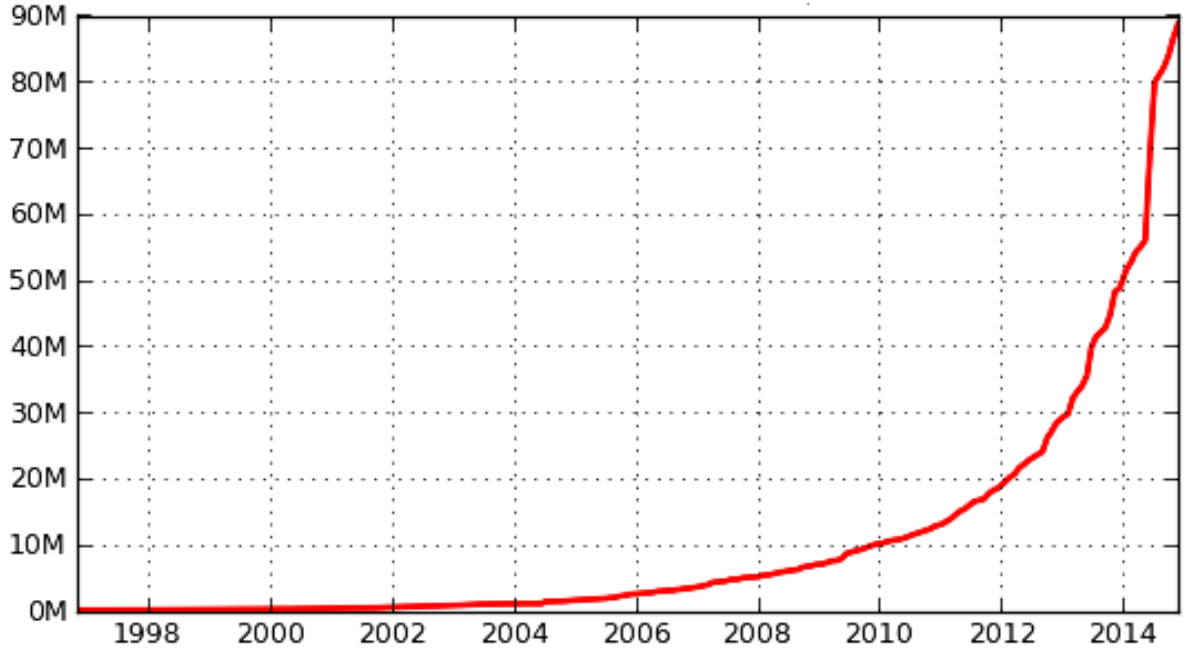

Sciencenter


Cornell University

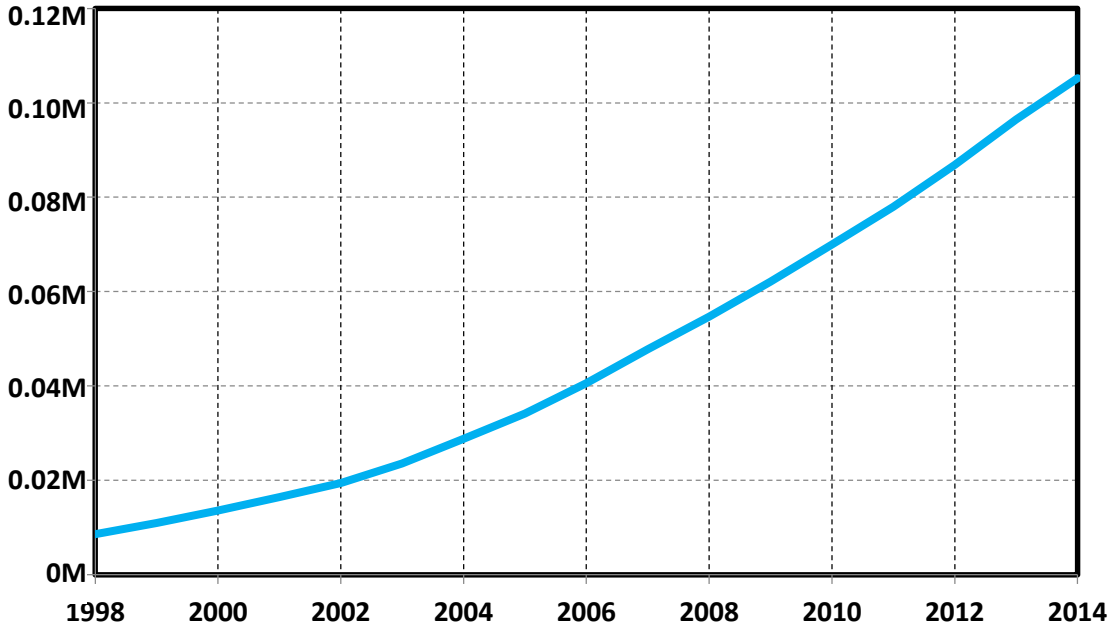
KIAS
KOREA
INSTITUTE FOR
ADVANCED
STUDY

Importance

- In most cases function of the proteins depends strictly on their structure:
  - denatured enzymes lose their catalytic power
  - denatured antibodies can no longer bind antigen
- Single-domain proteins:
  - less than 30% of all proteins in a cell
  - usually easily predicted by template-based methods
- Multi-domain proteins:
  - often each domain has a separate function to perform (e.g. ligand binding, catalytic)
- Intrinsically disordered proteins:
  - often multi-domain proteins connected by flexible linker



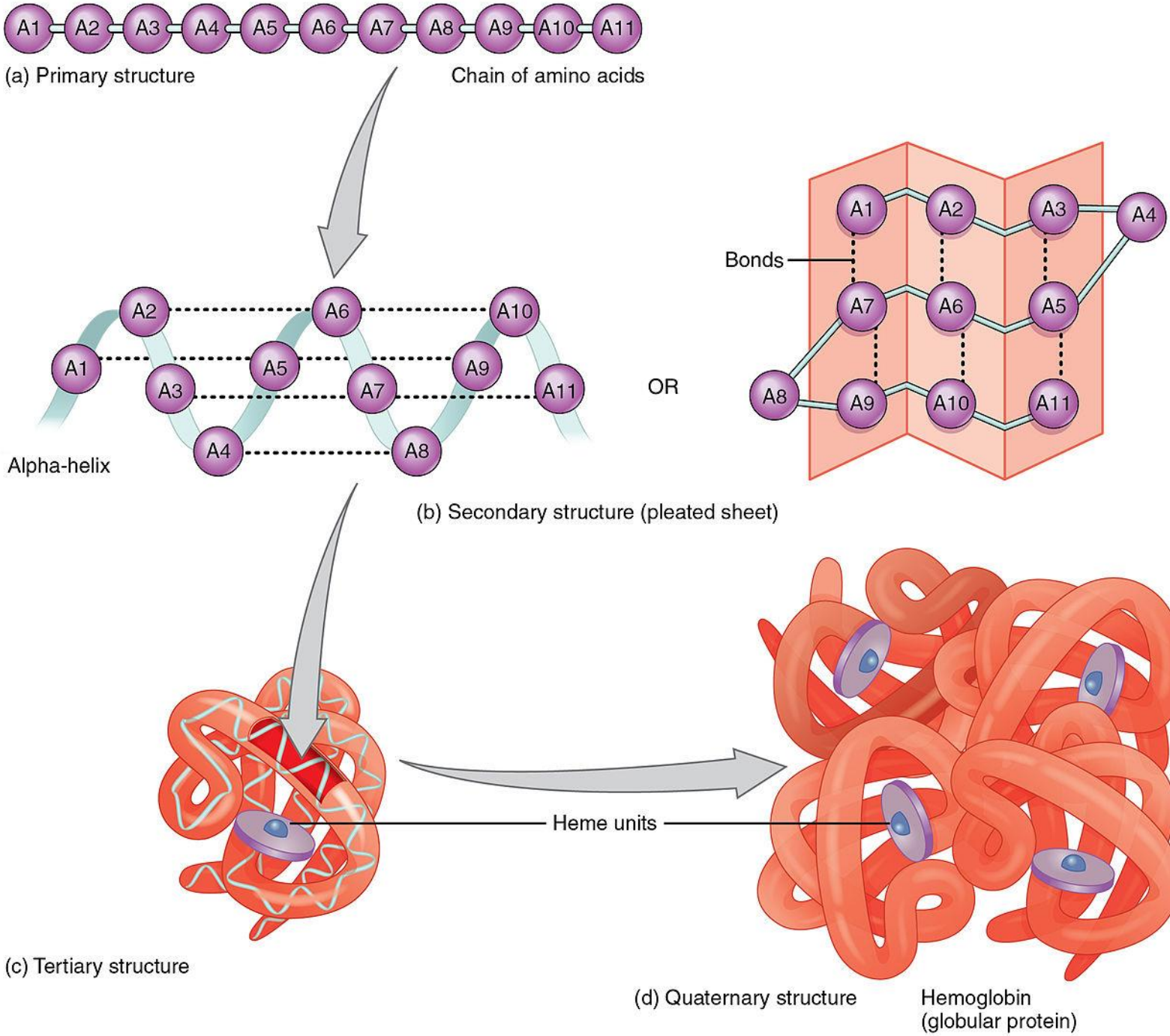
Number of
protein sequences in
UniProtKB database
(November 26th 2014)



Number of
protein structures in
RCSB database
(November 25th 2014)

http://www.ebi.ac.uk/uniprot/TrEMBLstats

http://www.rcsb.org/pdb/statistics/contentGrowthChart.do?content=total&seqid=100

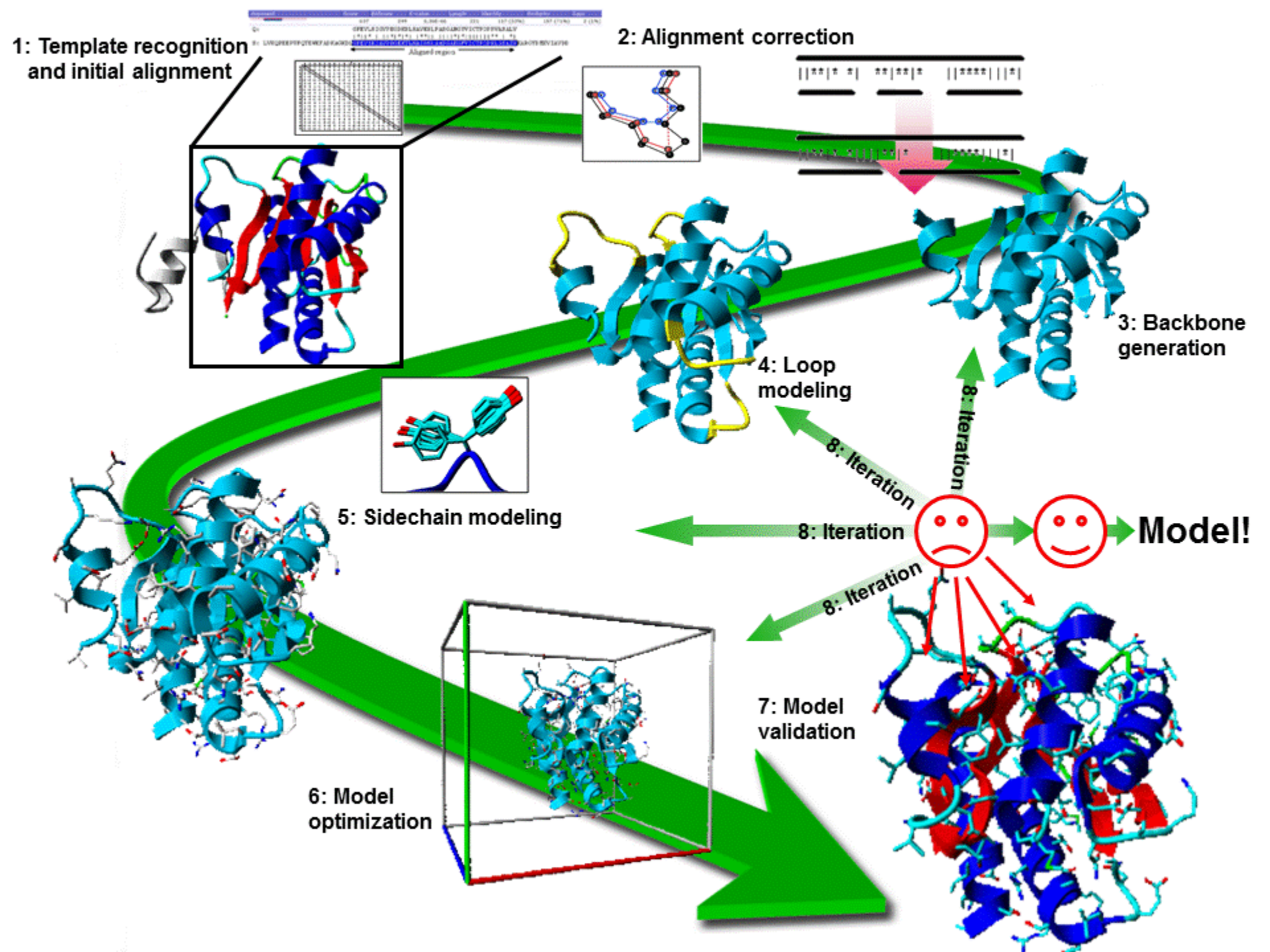

A1
A2
A3
A4
A5
A6
A7
A8
A9
A10
A11
(a) Primary structure
Chain of amino acids
Alpha-helix
Bonds
OR
(b) Secondary structure (pleated sheet)
Heme units
(c) Tertiary structure
(d) Quaternary structure
Hemoglobin
(globular protein)

Predicting protein structure

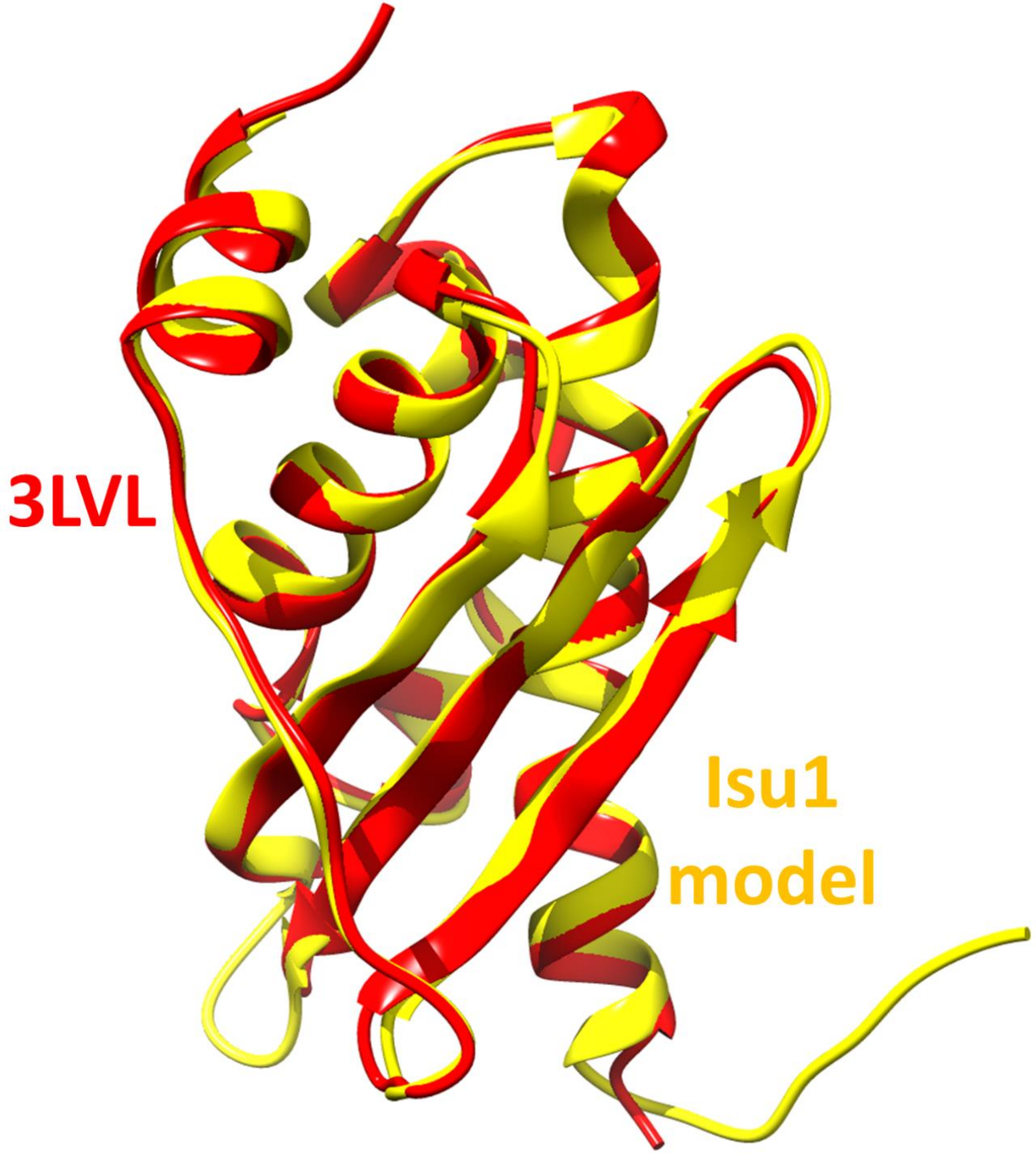
- Homology (template-based) modeling
  - Based on the databases
  - Often servers
- *De novo*
  - Based on physics of interactions
  - E.g. molecular dynamics (MD) or Monte Carlo (MC) simulations
  - E.g. protein folding
- Mixed
  - Tries to combines above approaches

Homology modeling

- Assumption that protein structure is strictly connected to its structure
- High sequence similarity between proteins usually indicates high structural similarity
- Evolutionary information is very useful (e.g. conserved motifs in proteins)

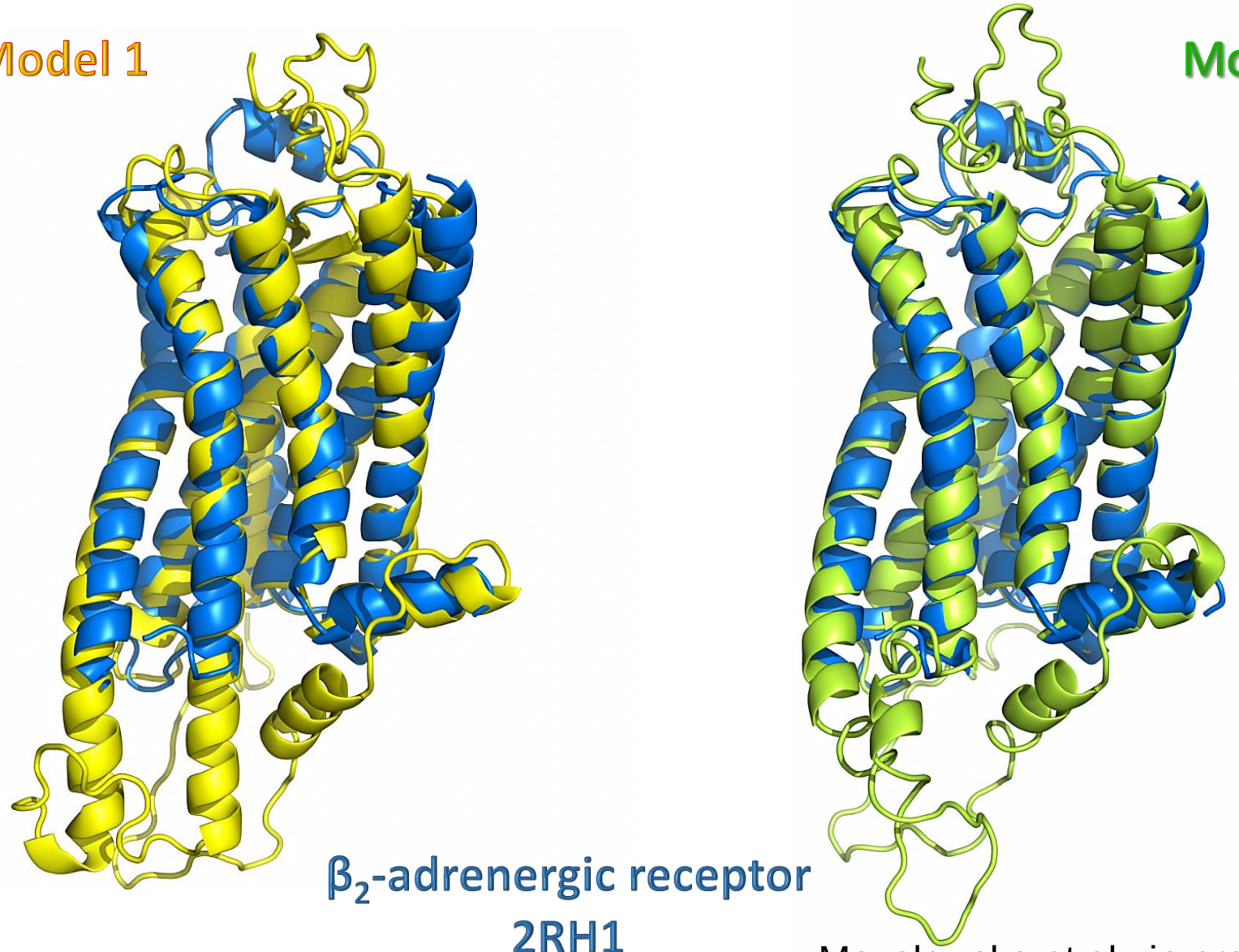

1: Template recognition and initial alignment
2: Alignment correction
3: Backbone generation
4: Loop modeling
5: Sidechain modeling
6: Model optimization
7: Model validation
8: Iteration
8: Iteration
8: Iteration
8: Iteration
Model!

Homology modeling



"Molecular modeling of the binding modes of the Iron-sulfur protein to the Jac1 co-chaperone from Saccharomyces cerevisiae by all-atom and coarse-grained approaches ", M.A. Mozolewska, P.K. Krupa, H. A. Scheraga, A. Liwo, published in Proteins: Structure, Function, and Bioinformatics

Homology modeling of OTR

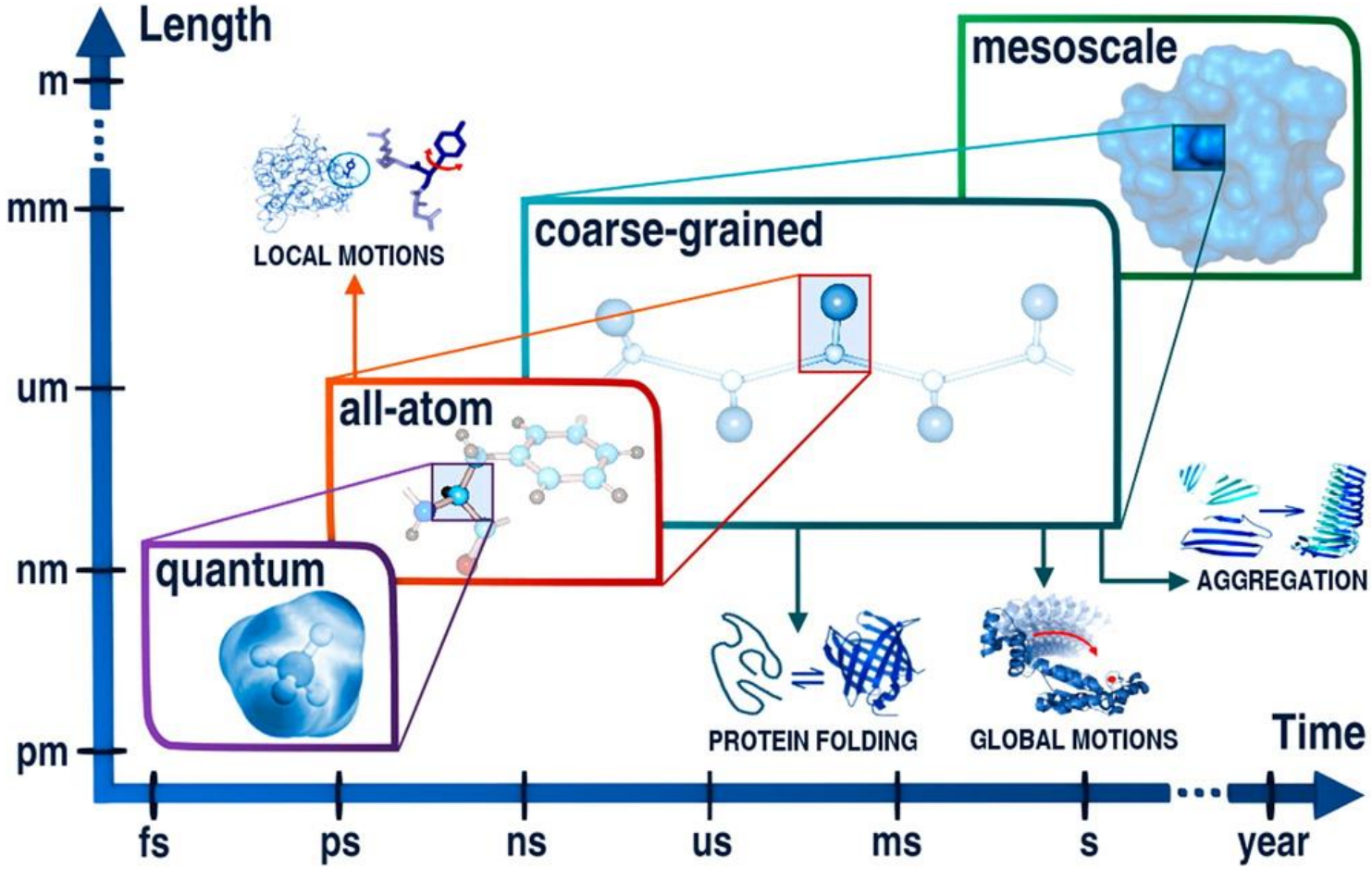


Mozolewska et al., in preparation

De novo methods

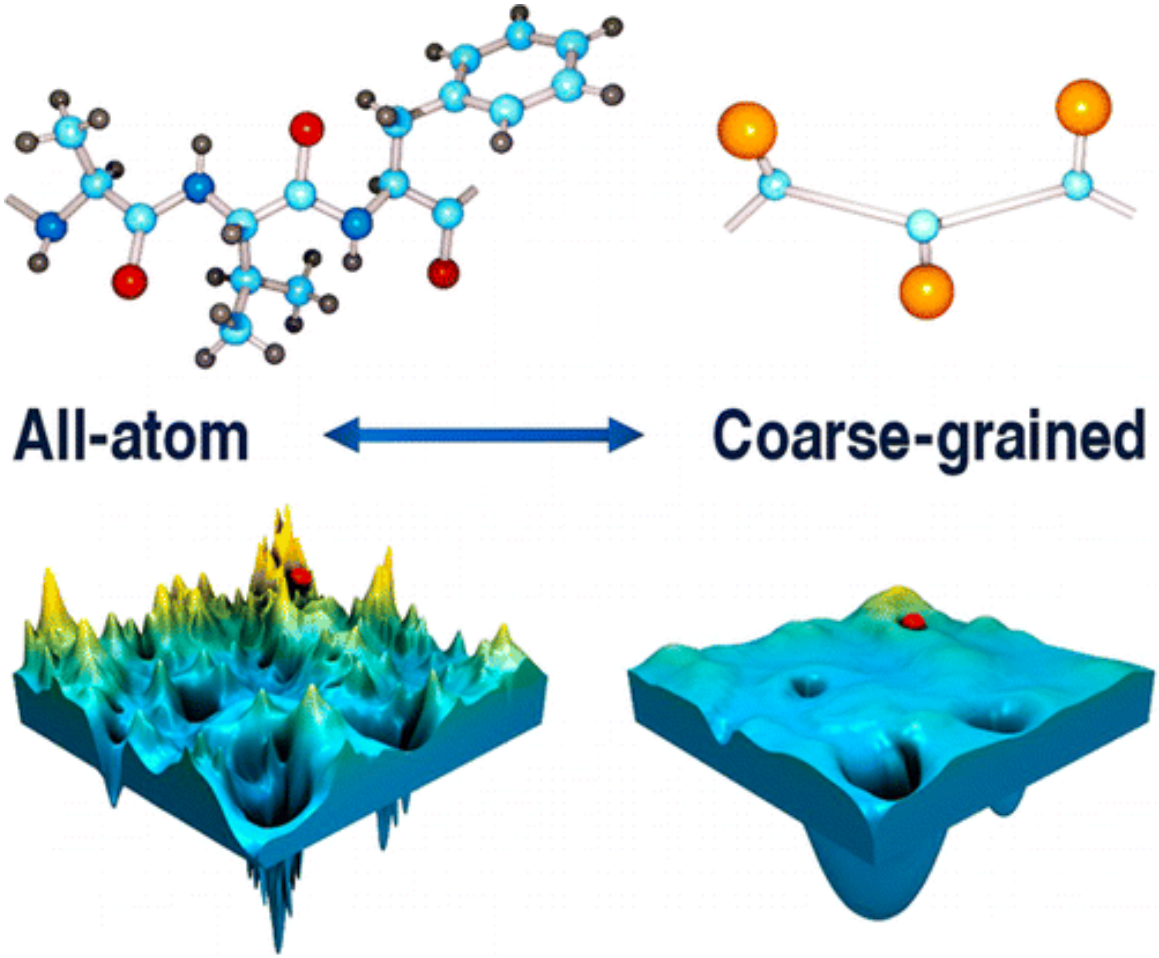
- Based on physics of interactions
- Using molecular dynamics (MD) simulations allows to study protein folding and conformational changes
- Include dynamics of proteins
- Do not rely on databases

Coarse-grained force fields



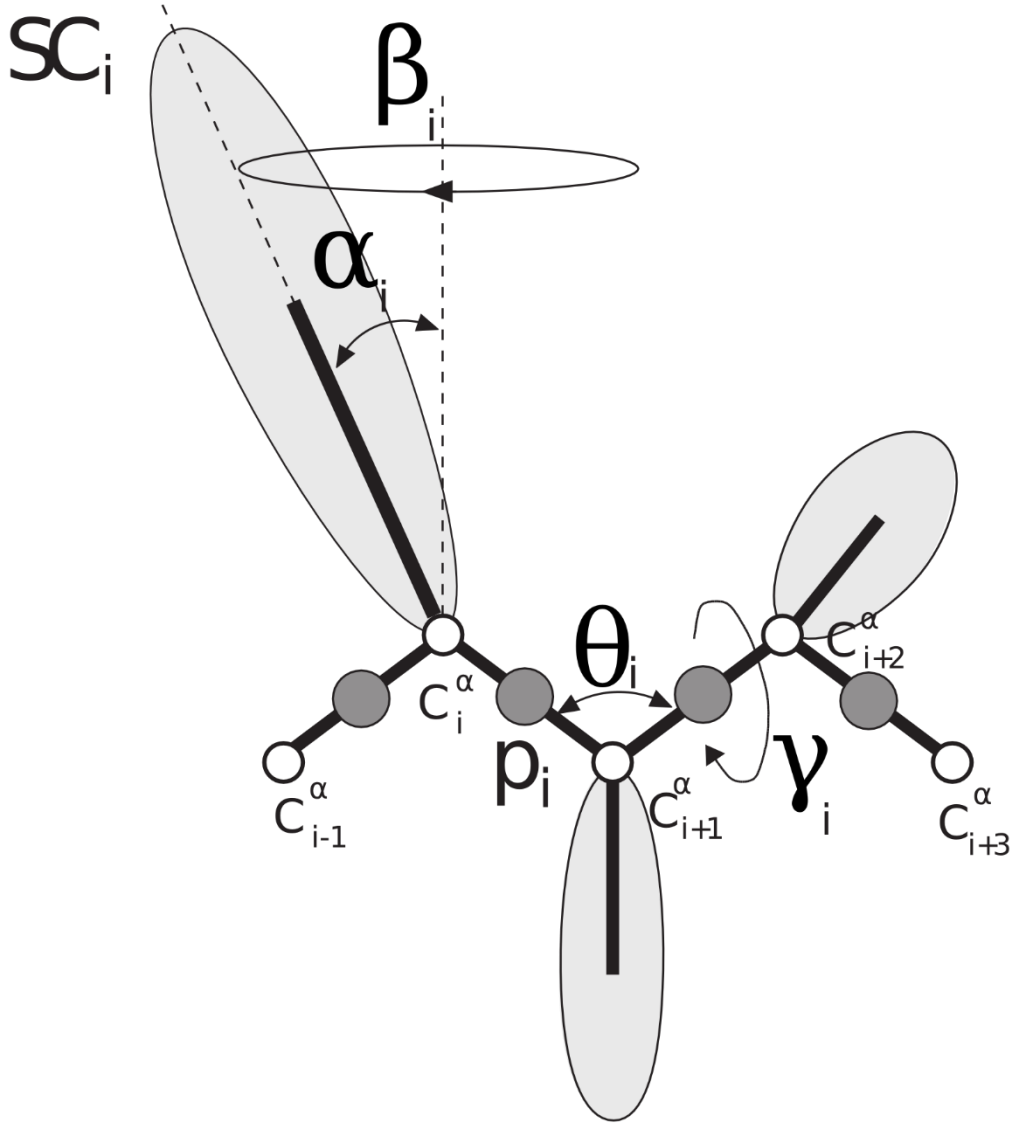
Kmiecik S. et al., *Chem. Rev.*, 2016, 116 (14), pp 7898–7936

Coarse-grained force fields



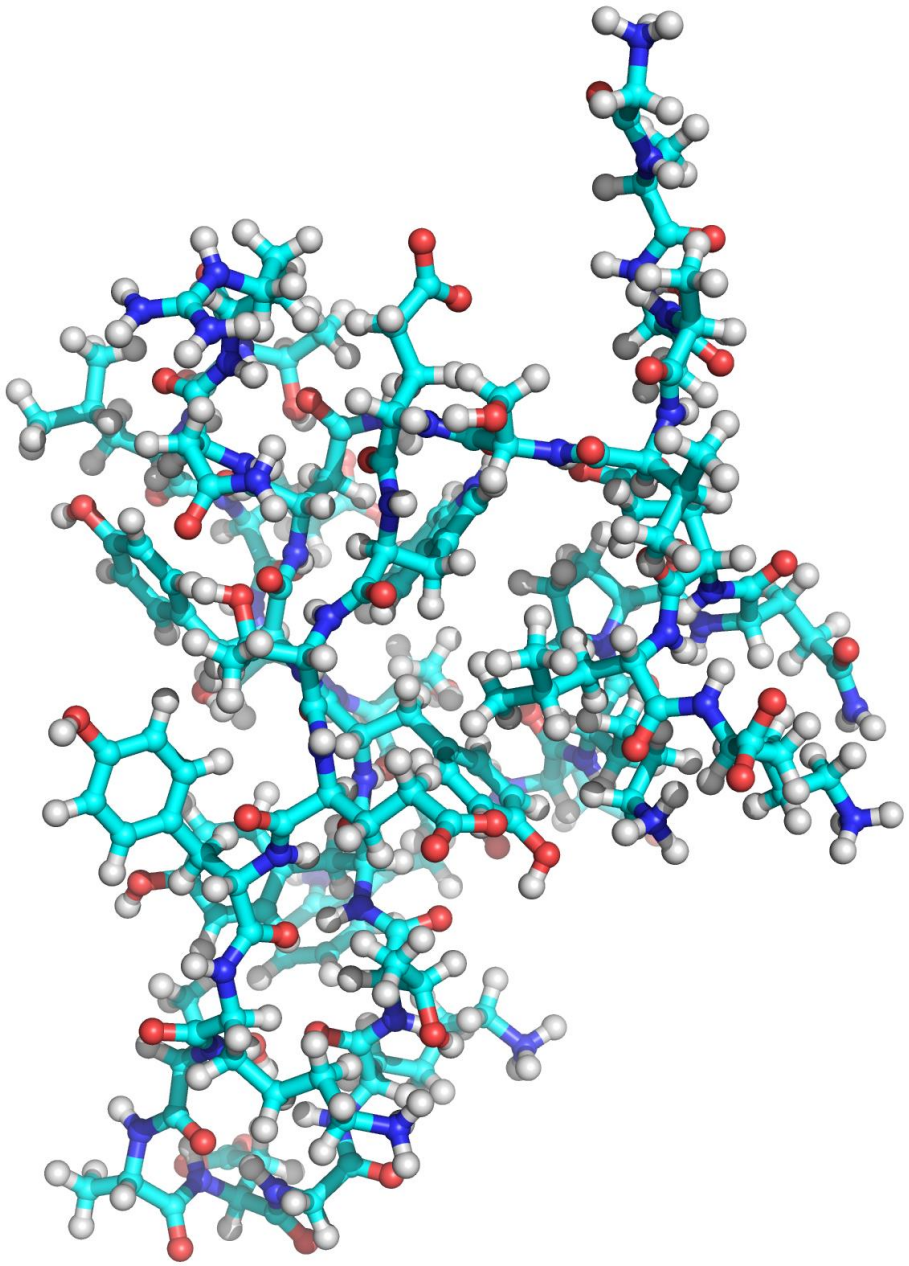
Kmiecik S. et al., *Chem. Rev.*, 2016, 116 (14), pp 7898–7936

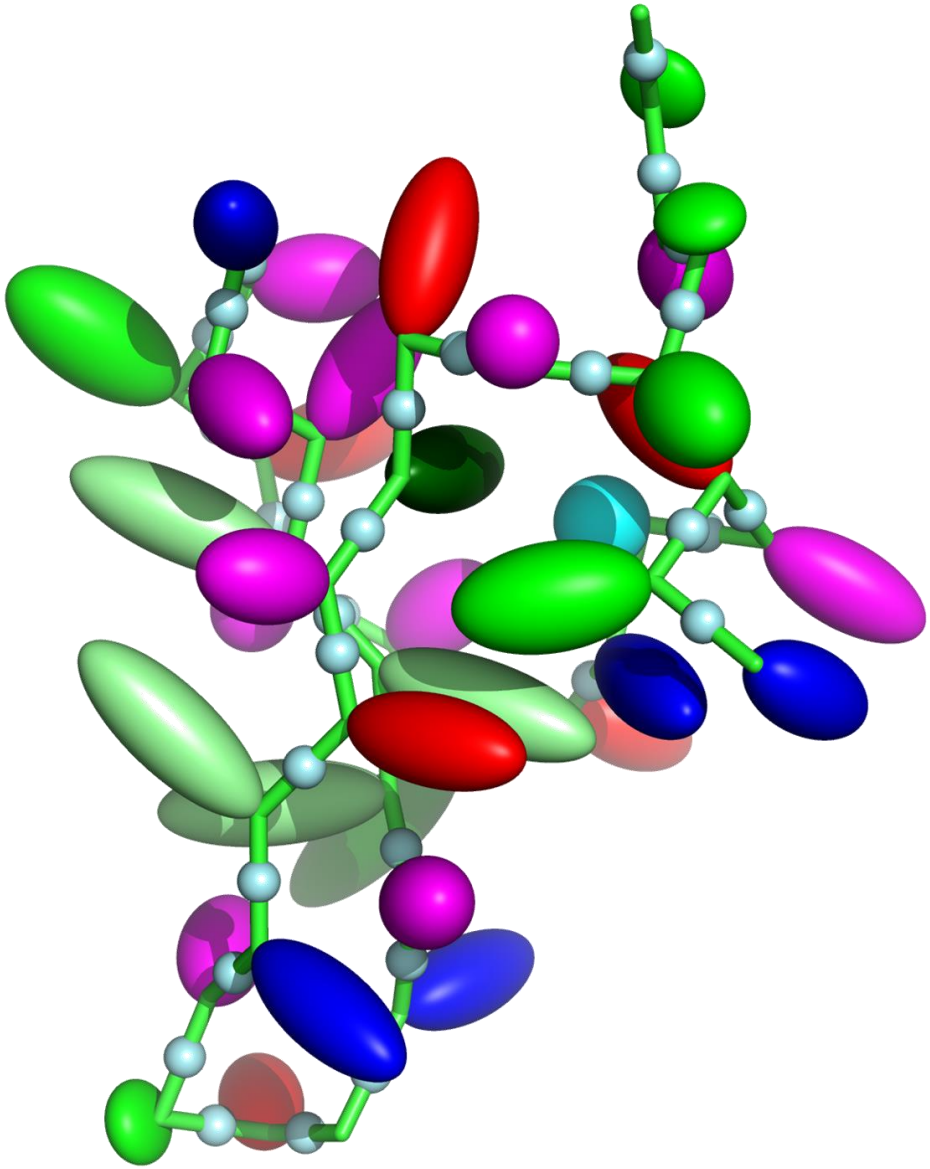
UNited RESidue force field



- Coarse-grained (physics-based) force field for proteins

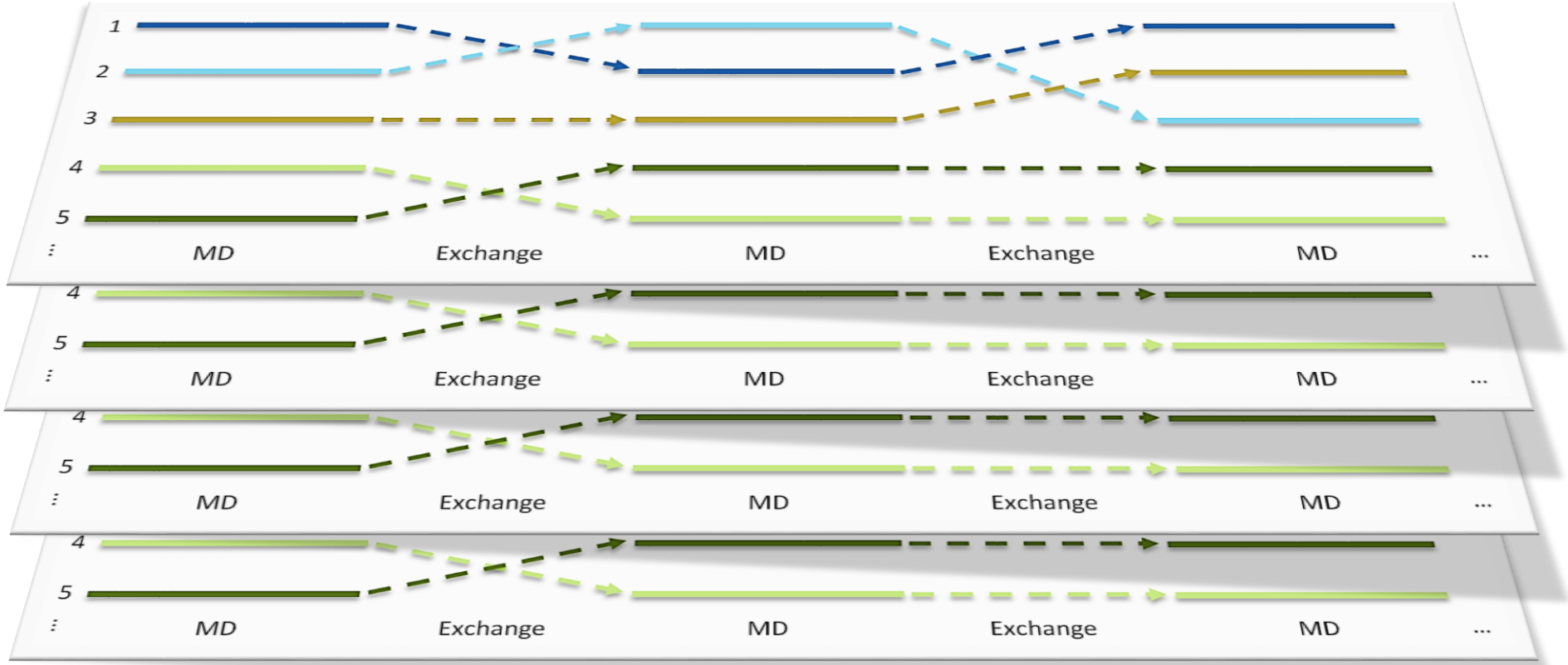
UNRES model of 1E0L





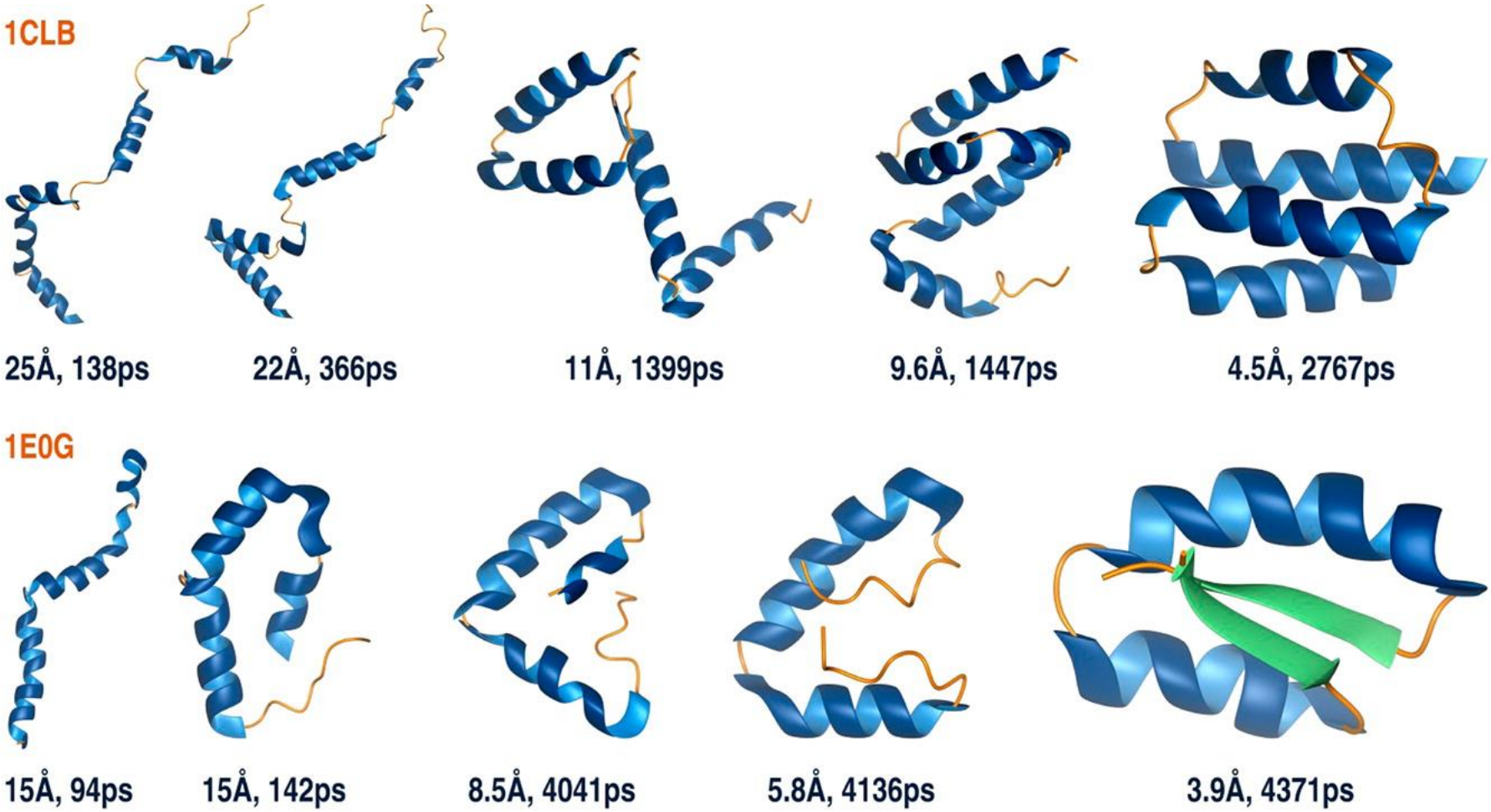
MREMD simulations in UNRES

- Multiplexed Replica Exchange MD simulations
- Up to 75% scalability with 4096 CPUs



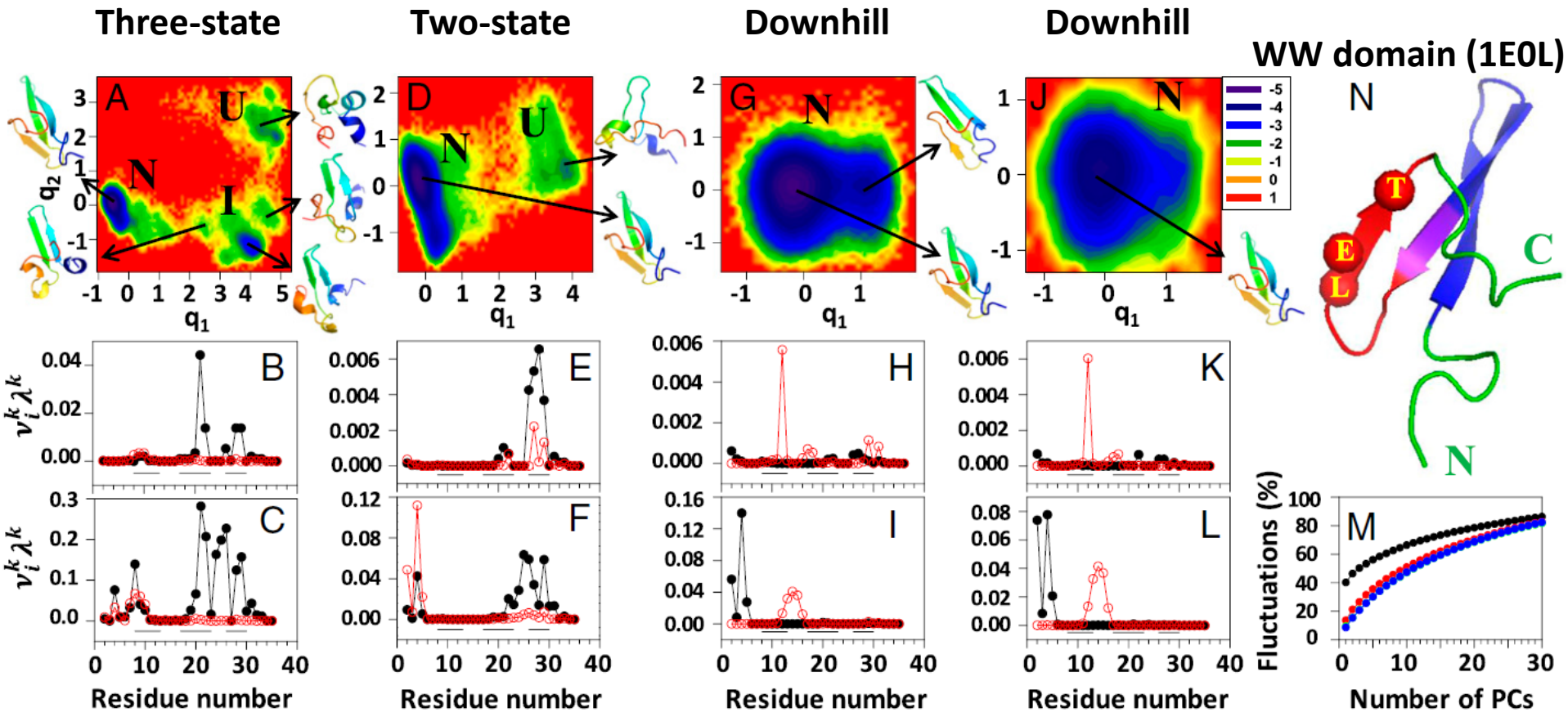
C. Czaplewski, S. Kalinowski, A. Liwo, H.A. Scheraga. Application of multiplexed replica exchange molecular dynamics to the UNRES force field: tests with α and $\alpha+\beta$ proteins. *J. Chem. Theory Comput.*, **2009**, 5, 627-640.

Folding pathways in UNRES



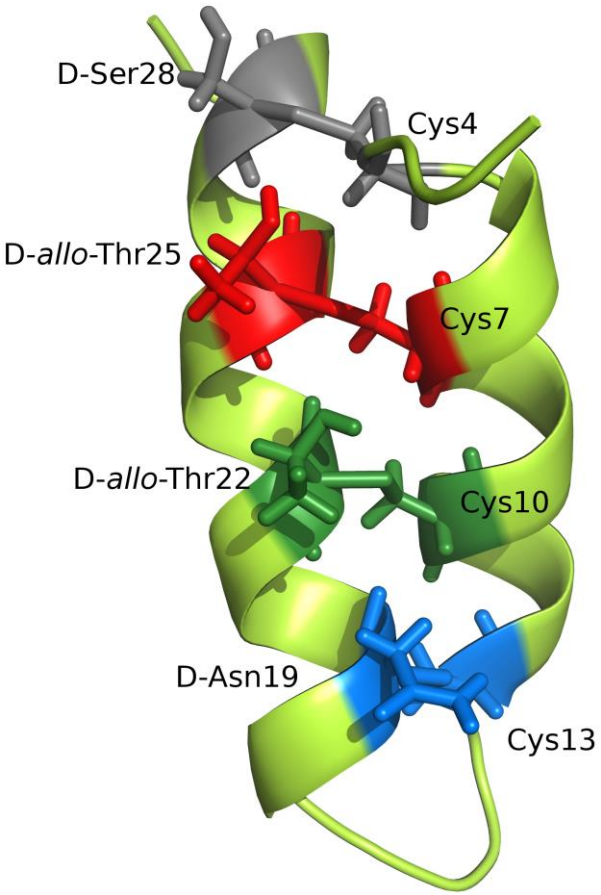
Liwo, A.; Khalili, M.; Scheraga, H. A. Ab initio simulations of protein-folding pathways by molecular dynamics with the united-residue model of polypeptide chains *Proc. Natl. Acad. Sci. U. S. A.* 2005, 102, 2362– 2367

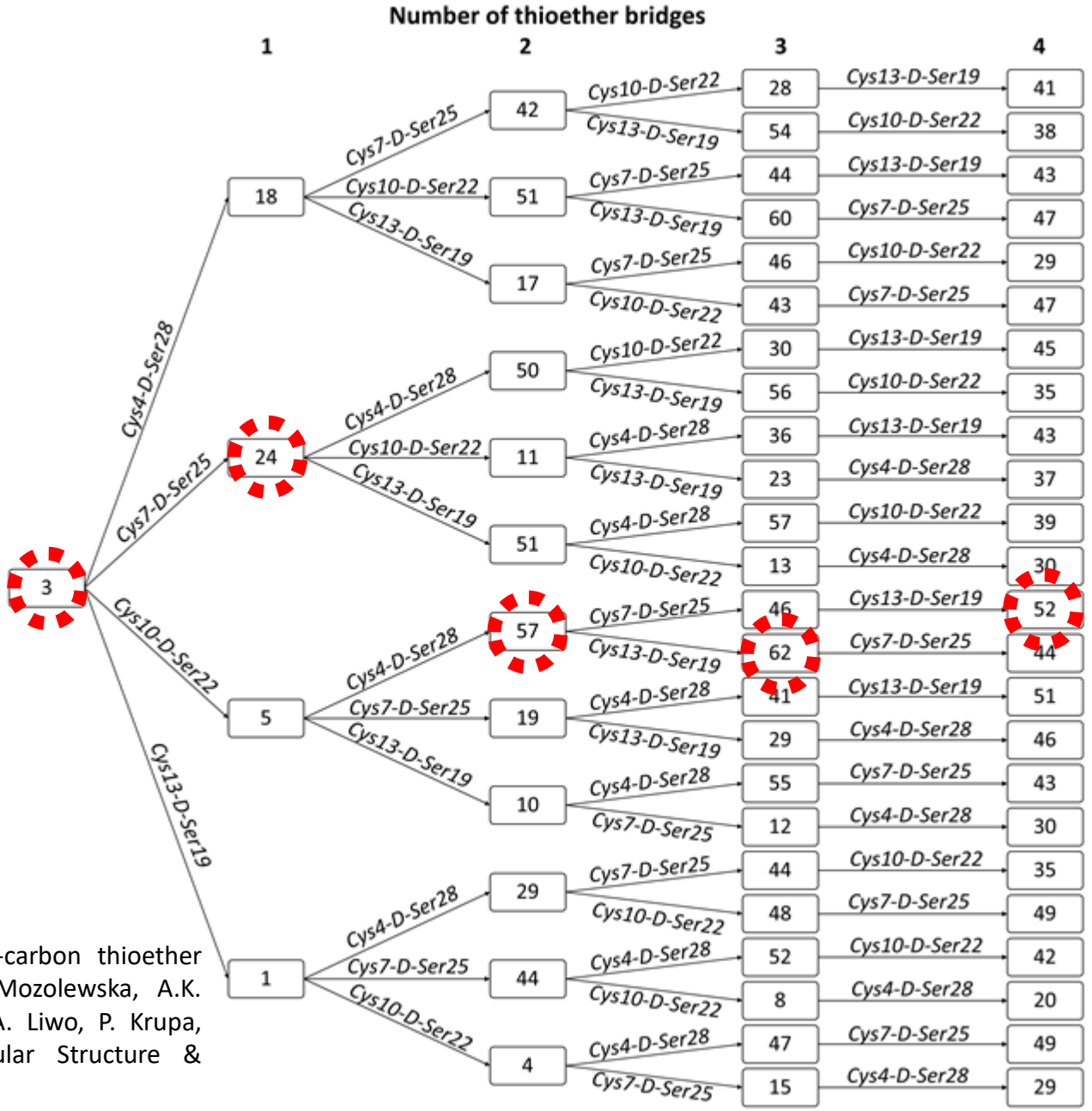
Protein folding studies



"Preventing fibril formation of a protein by selective mutation", Gia G. Maisuradze, Jordi Medina, Khatuna Kachlishvili, Paweł Krupa, Magdalena A. Mozolewska, Pau Martin-Malpartida, Luka Maisuradze, Maria J. Macias, and Harold A. Scheraga*, Proceedings of the National Academy of Sciences (USA), 2015, 112(44):13549-13554.

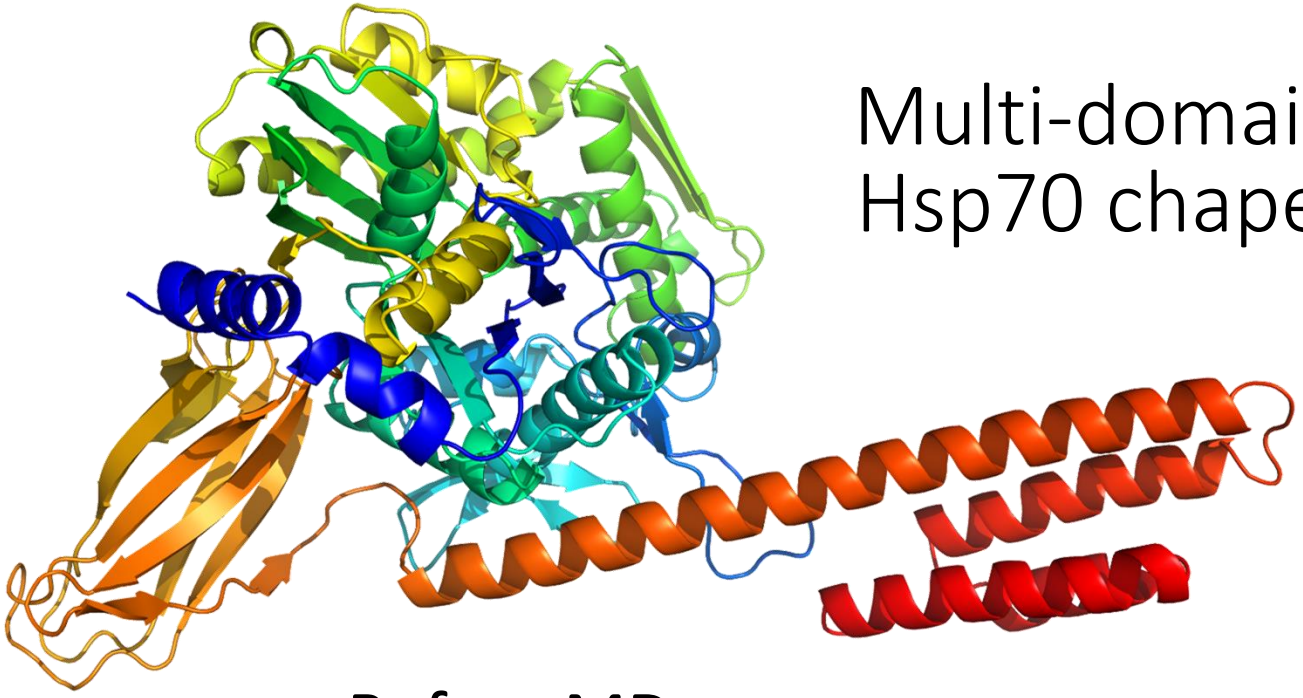
Studies of disulfide-like bonds in UNRES



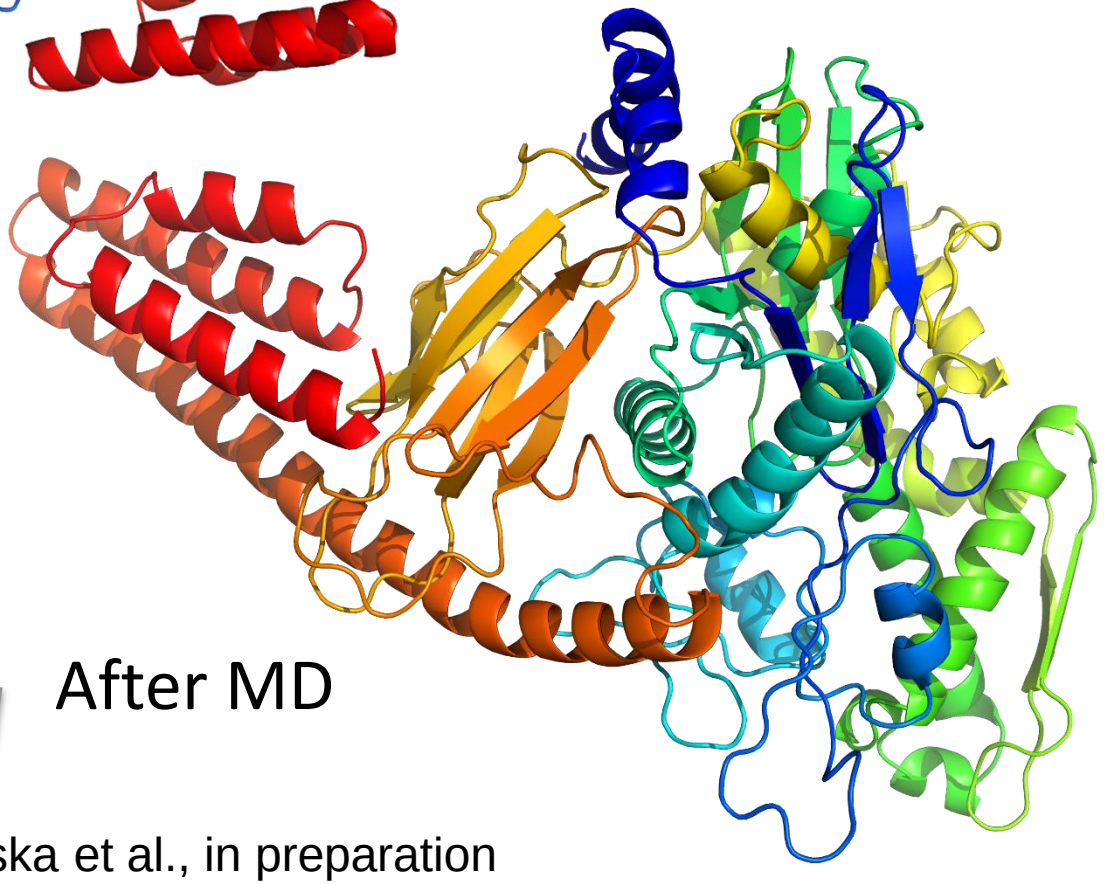


"Influence of the formation of sulfur to α-carbon thioether bridges on folding of thurincin H", M.A. Mozolewska, A.K. Sieradzan*, A. Niadzvedstki, C. Czaplewski, A. Liwo, P. Krupa, published online in Journal of Biomolecular Structure & Dynamics.

Multi-domain dynamics of Hsp70 chaperon in UNRES



Before MD

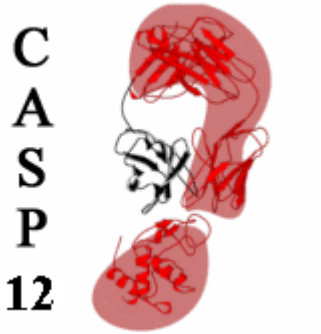


After MD

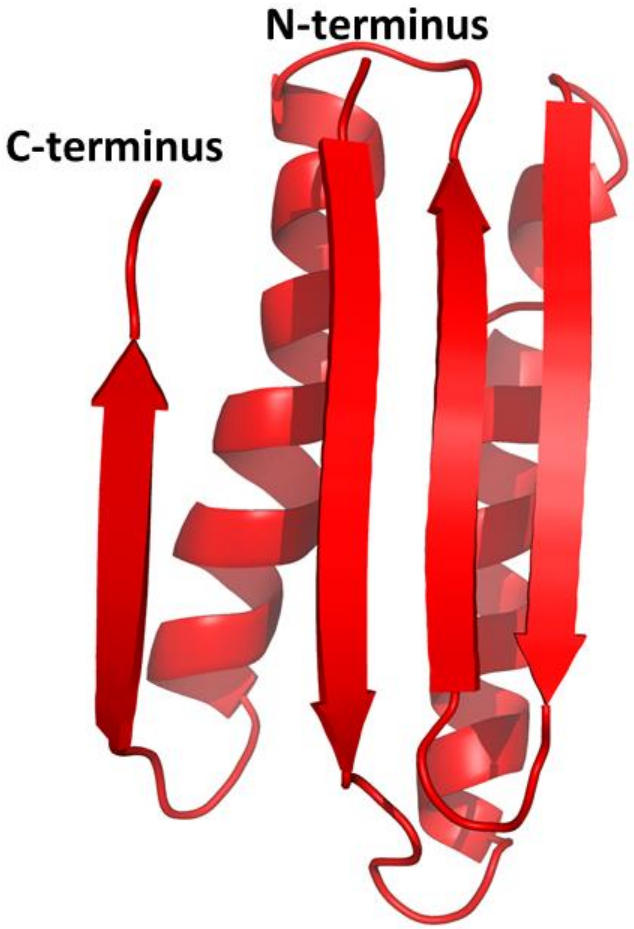
Mozolewska et al., in preparation

De novo protein structure prediction in UNRES

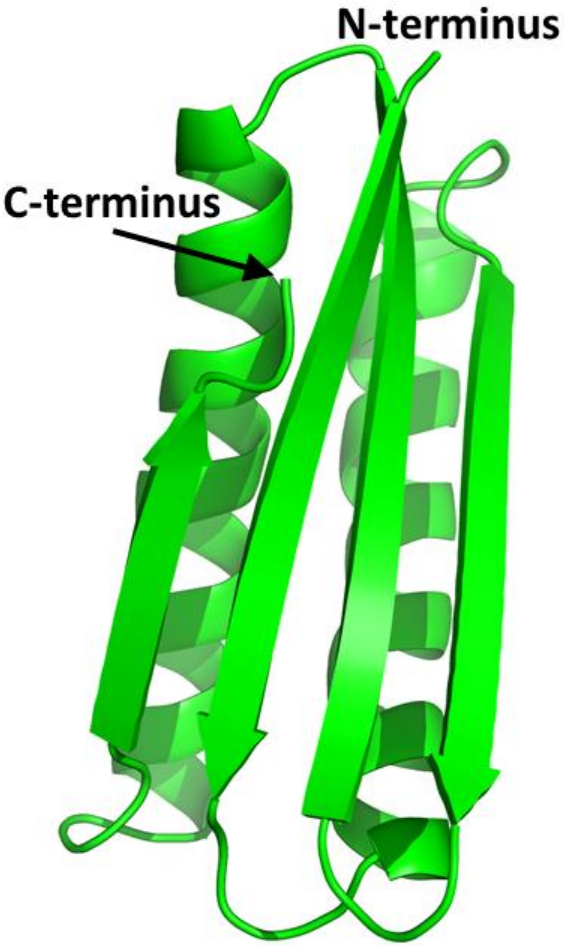
- CASP
  - The Critical Assessment of protein Structure Prediction
  - Bi-annual blind experiment – protein structures are predicted based only on amino-acid sequences
  - In CASP10 (2012) model of target protein T0663 produced by plain UNRES force field was chosen by the CASP assessors as the best one, overperforming 124 other groups and methods
  - After improvements to the force field, in CASP11 (2014) UNRES was able to produce very good structures for several targets with higher resolution than before obtaining e.g. RMSD = 3.8 A for 97-residue target protein T0769.
  - Assessment of current CASP experiment (CASP12) was revealed in late December 2016 – analysis is underway.



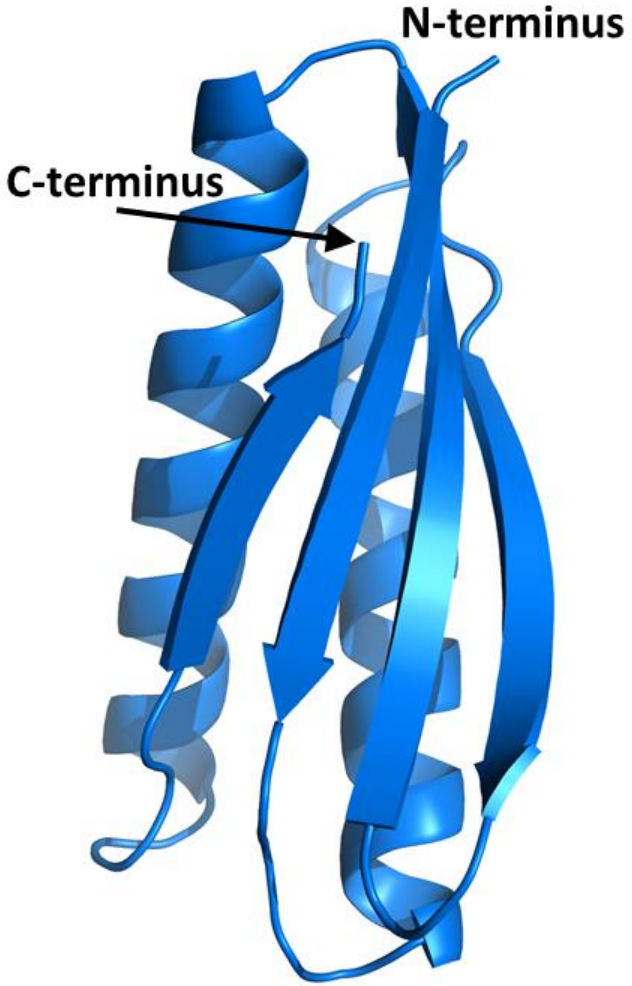
De novo protein structure prediction in UNRES



UNRES model



Native structure



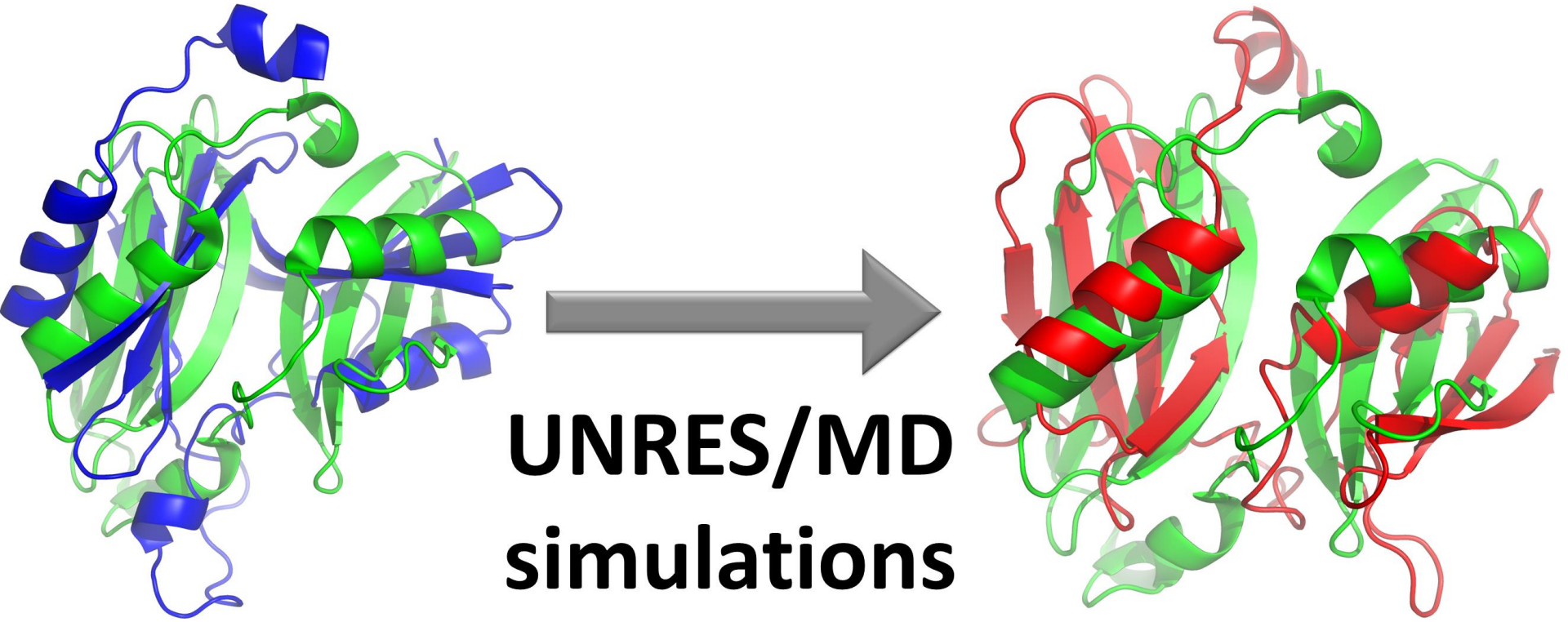
Best TBM model

Paweł Krupa, Magdalena Mozolewska et al., Performance of protein-structure predictions with the physics-based UNRES force field in CASP11, *Bioinformatics* 2016, 32(21), 3270-3278

Mixed methods

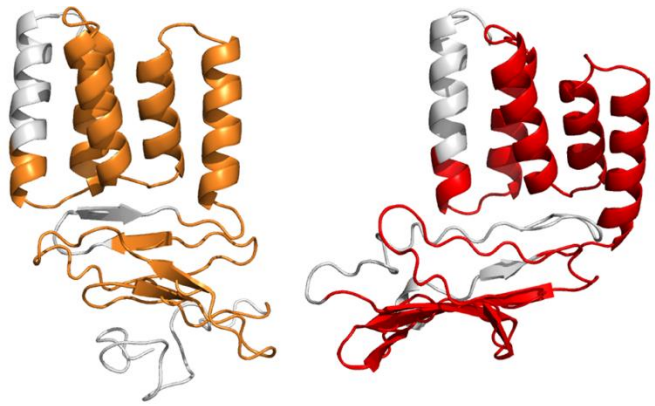
- Combines template-based and *de novo* methods
- Difficult to assess which part of the protein should be predicted by which method
- Excellent peak performance especially for multi-domain proteins

Protein structure prediction in UNRES supplemented by TBM models

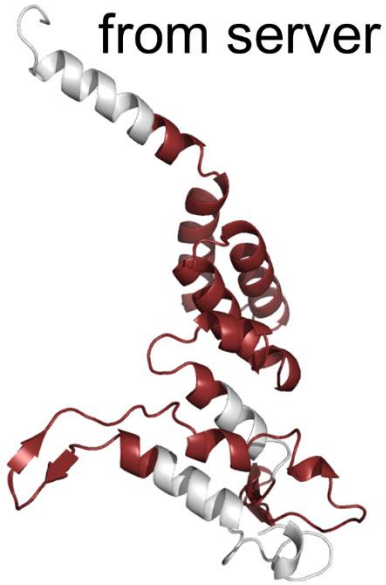


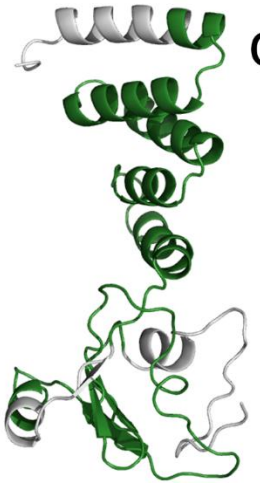
Krupa, P.; Mozolewska, M. A.; Joo, K.; Lee, J.; Czaplewski, C.; Liwo, A. Prediction of Protein Structure by Template-Based Modeling Combined with the UNRES Force Field *J. Chem. Inf. Model.* **2015**, 55, 1271–1281

Use of Restraints from Consensus Fragments from Multiple Server Models as an Approach to Enhance Protein-Structure Prediction Capability of the UNRES Force Field

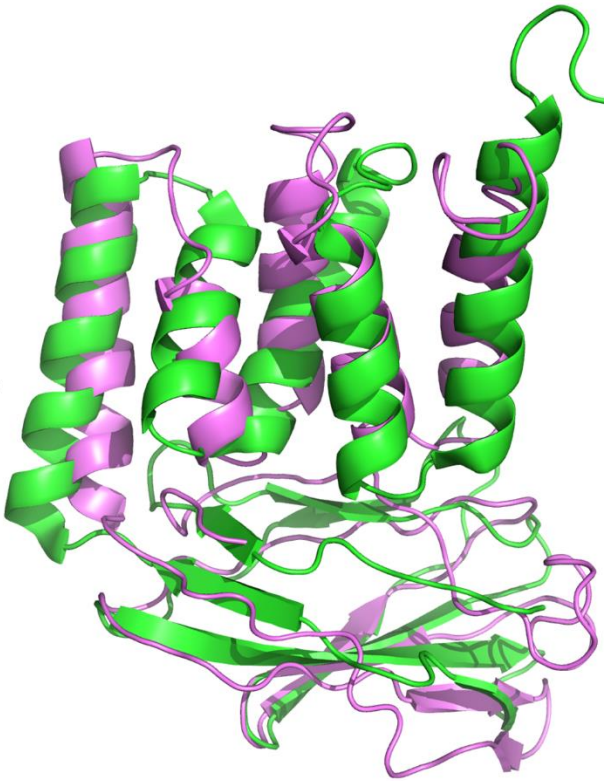


Consensus fragments from server models





UNRES + restraints from consensus fragments



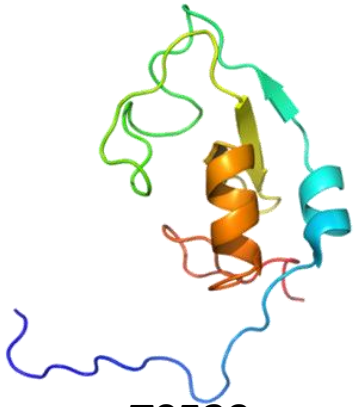
Final model

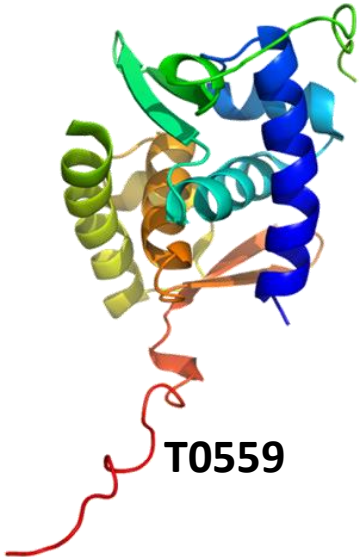
"Use of Restraints from Consensus Fragments of Multiple Server Models to Enhance Protein-Structure Prediction Capability of the UNRES Force Field", M.A. Mozolewska, P. Krupa, B. Zaborowski, A. Liwo*, J. Lee, K. Lee, C. Czaplewski, submitted to Journal of Chemical Information and Modeling.

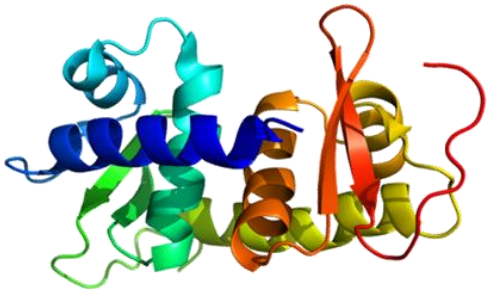
Testing set of proteins

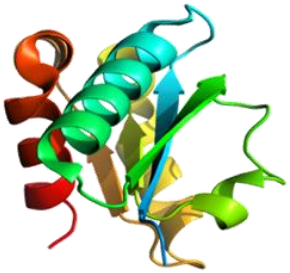
- One-domain


T0538

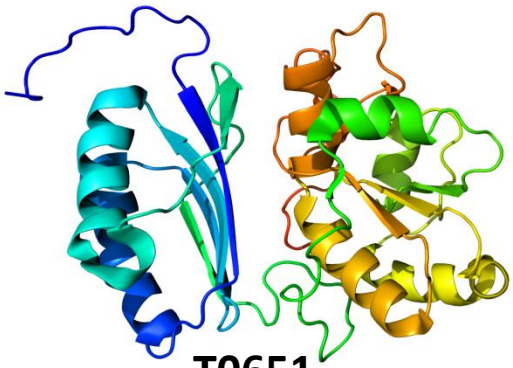

T0539

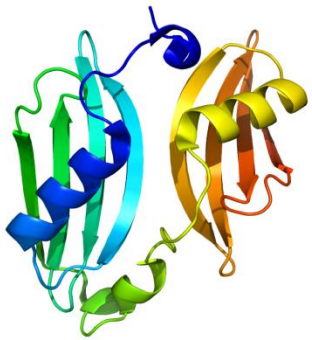

T0559

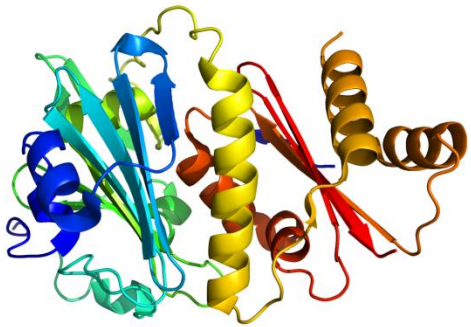

T0560

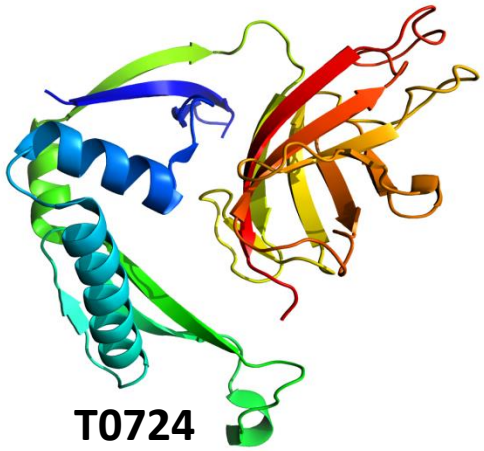

T0580

- Two-domain


T0651

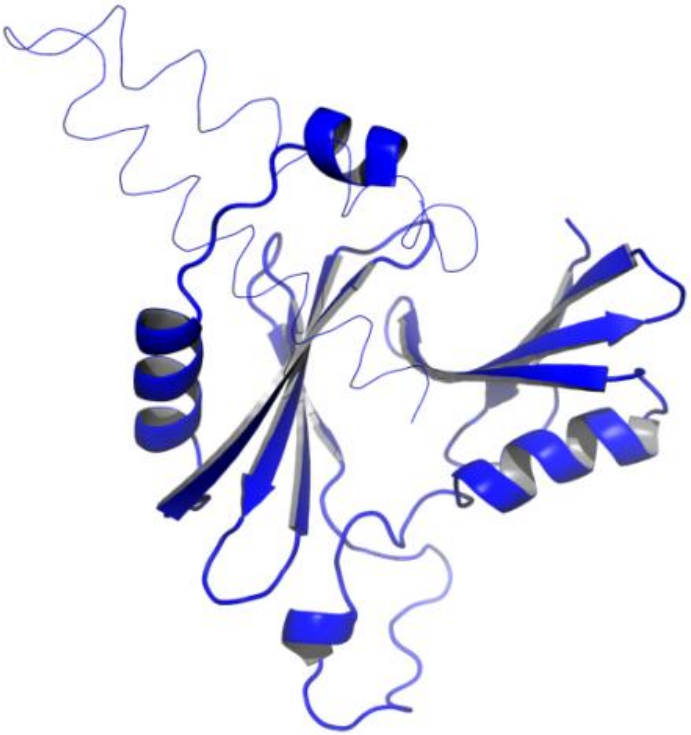

T0663

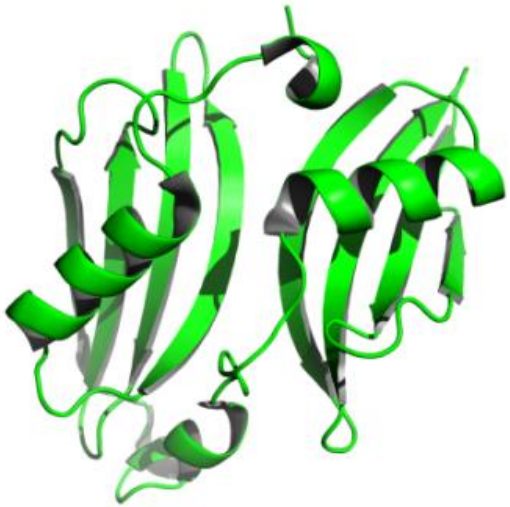

T0717

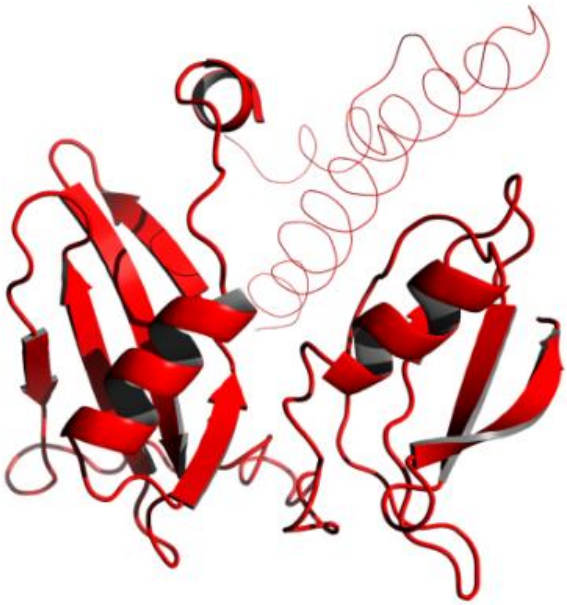

T0724

Prediction of multi-domain proteins in UNRES

T0663







Best TBM model Native structure UNRES model

Krupa, P.; Mozolewska, M. A.; Joo, K.; Lee, J.; Czaplewski, C.; Liwo, A. Prediction of Protein Structure by Template-Based Modeling Combined with the UNRES Force Field *J. Chem. Inf. Model.* **2015**, 55, 1271–1281

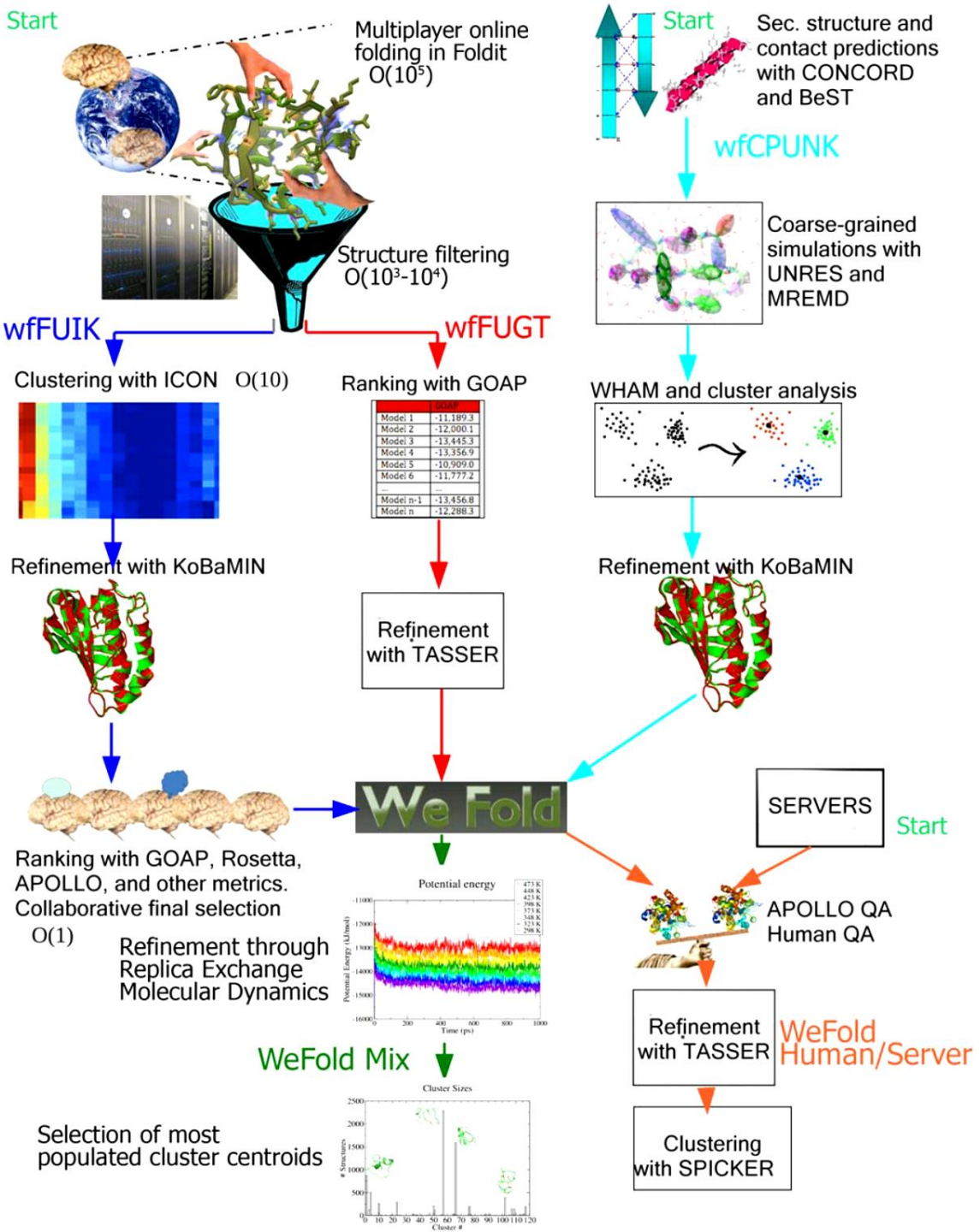
„Lessons from application of the UNRES force field to predictions of structures of CASP10 targets", Yi He, Magdalena A. Mozolewska, Paweł Krupa, Adam K. Sieradzan, Tomasz K. Wirecki, Adam Liwo, Khatuna Kachlishvili, Shalom Rackovsky, Dawid Jagieła, Rafał Ślusarz, Cezary R. Czaplewski, Stanisław Ołdziej, Harold A. Scheraga*; Proceedings of the National Academy of Sciences (USA), 2013, 110(37), 14936-14941.

Collaborative initiatives

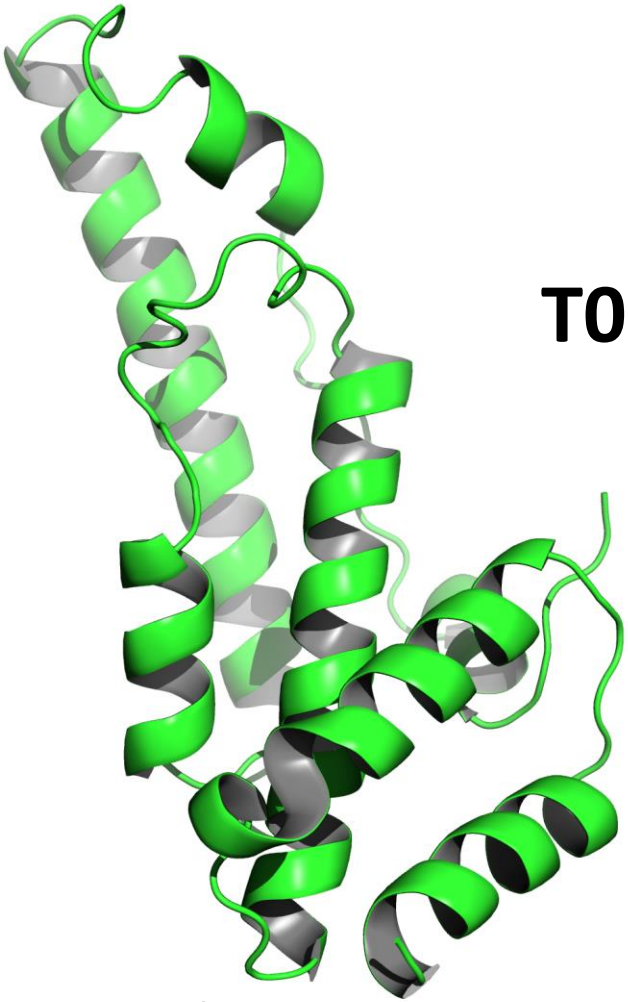
- WeFold – co-opetitive (cooperative competition) effort of 21 groups to combine various methods for protein structure prediction to achieve more reliable results

S.N. Crivelli[1,2], D. Bhattacharya[3], L. Blake[2,4], L. Bortot[5], G. Chopra[6,7], R. Faccioli[8], E. Faraggi[9,10], Sambit Ghosh[11], Soma Ghosh[11], Y. He[12], L. Heo[13], C. Keasar[14], M. Khan[15], F. Khatib[15], G. Khoury[16], C. Kieslich[17,18], P. Krupa[19], G.R. Lee[13], J. Li[3], A. Lipska[19], A. Liwo[19], L. McGuffin[20], M. Mirdita[21], **M.A. Mozolewska[19]**, S. Ovchinnikov[22], A. Shah[15], A.K. Sieradzan[19], J. Smadbeck[16], P. Tamamis[17,18], N. Trieber[15], B. Wallner[23], T. Wirecki[19], Y. Yin[12], Y. Zhang[24], J. Bacardit[25], M. Baranowski[19], N. Chapman[26], S. Cooper[27], A. Defelicibus[8], J. Flatten[26], R. Ganzynkowicz[19], A. Giełdoń[19], B. Koepnick[22], Z. Popović[26], M. Ślusarz[19], R. Ślusarz[19], B. Zaborowski[19], D. Baker[22], J. Cheng[3], C. Czaplewski[19], A. Delbem[8], C.A. Floudas[17,18], A. Kloczkowski[10], S. Ołdziej[19], M. Levitt[28], H.A. Scheraga[12], C. Seok[13], J. Söeding[21], S. Vishveshwara[11]*

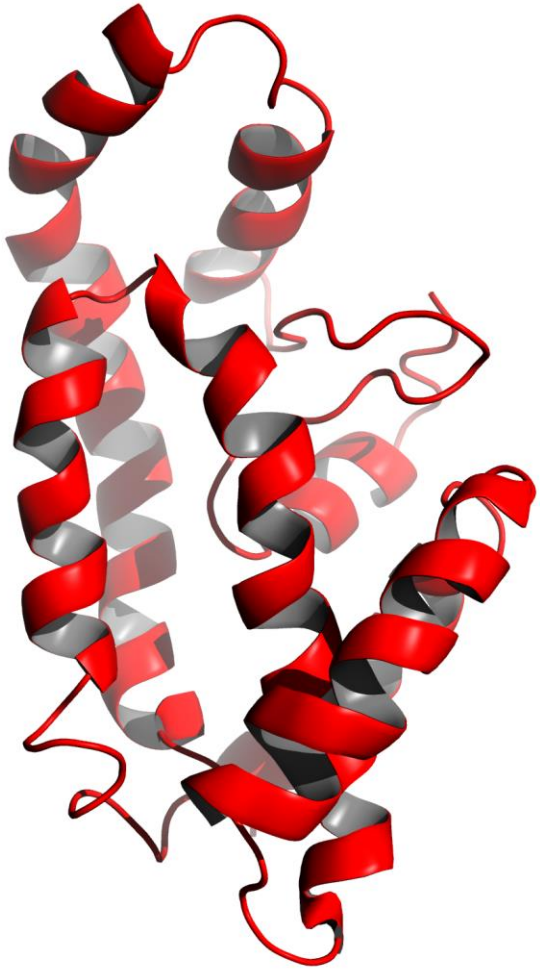
WeFold scheme of pipelines



WeFold collaboration



T0740

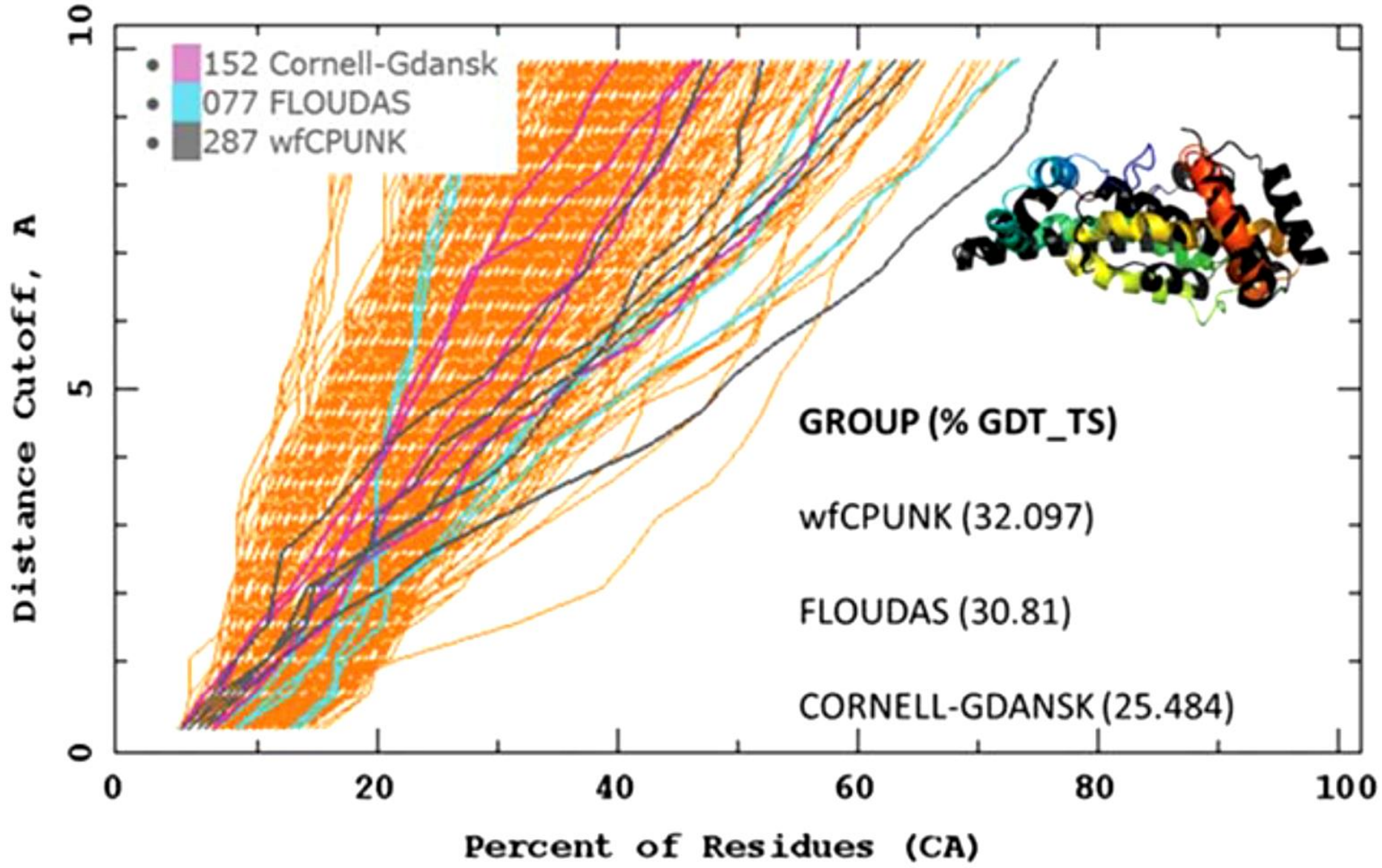


Native structure

UNRES (wfCPUNK) predicted structure

„WeFold: A Coopetition for Protein Structure Prediction”, George A. Khoury, Adam Liwo, Firas Khatib, Hongyi Zhou, Gaurav Chopra, Jaume Bacardit, Leandro O. Bortot, Rodrigo A. Faccioli, Xin Deng, Yi He, Pawel Krupa, Jilong Li, Magdalena A. Mozolewska, Adam K. Sieradzan, James Smadbeck, Tomasz Wirecki, Seth Cooper, Jeff Flatten, Kefan Xu, David Baker, Jianlin Cheng, Alexandre C. B. Delbem, Christodoulos A. Floudas, Chen Keasar, Michael Levitt, Zoran Popović, Harold A. Scheraga, Jeffrey Skolnick, Silvia N. Crivelli* and Foldit Players; Proteins: Structure, Function, and Bioinformatics, 2014, 98(9):1850–1868.

Performance of UNRES in WeFold



http://www.predictioncenter.org/casp10/gdtplot.cgi?target=T0740-D1&models=all&group=290

Acknowledgments

Adam Liwo

Cezary Czaplewski

Paweł Krupa

Harold A. Scheraga

Adam Sieradzan

Funding

UMO-2015/17/N/ST4/03935 from the Polish National Science Centre

NATIONAL SCIENCE CENTRE
POLAND