

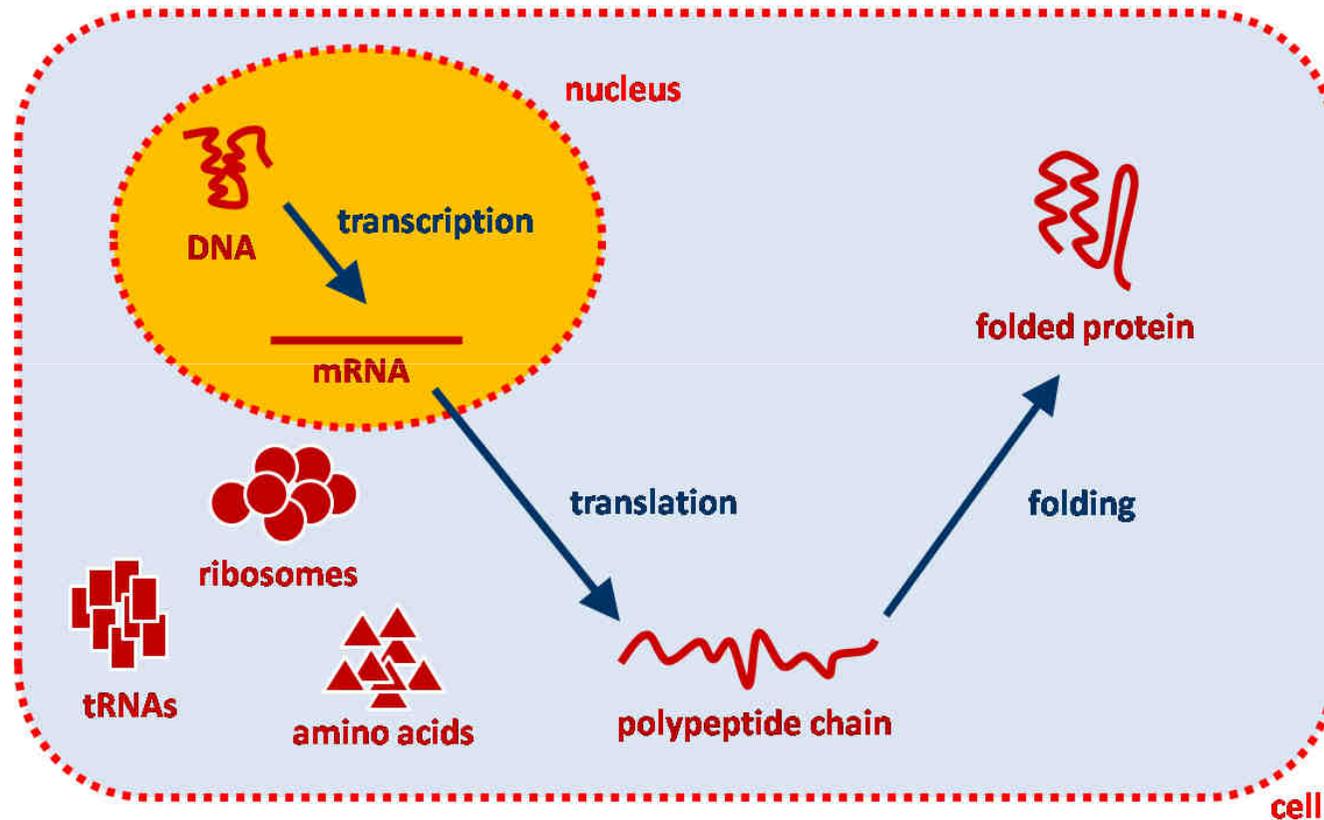
# Detection of non-sequential structural analogs with

# **GANGSTA+**

**Aysam Guerler**

**Knapp Lab., Freie Universität Berlin**

# Cellular protein assembly

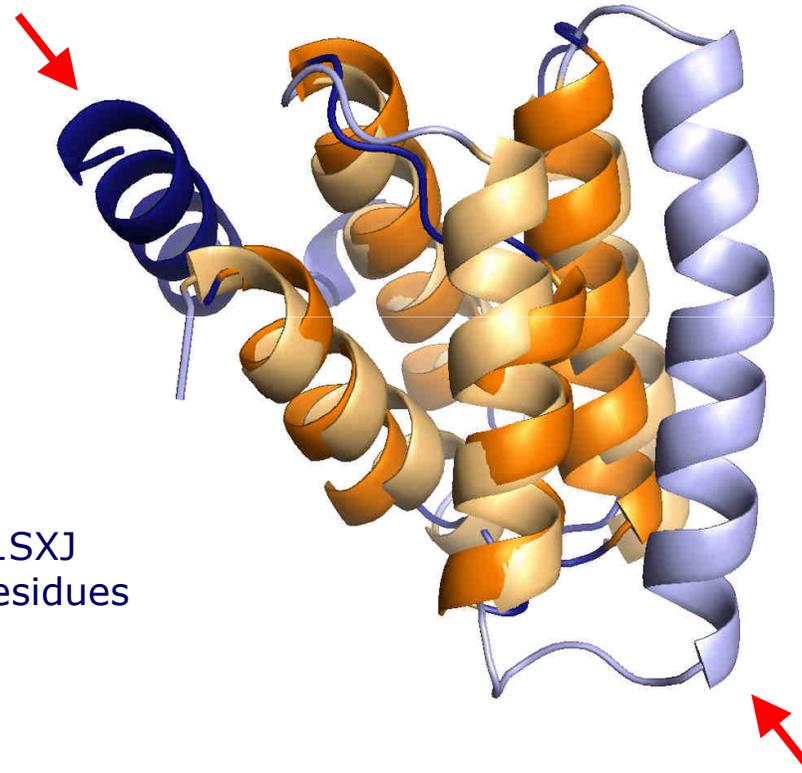


similar sequence → similar structure ↔ similar function

# Sequential similarity

**$\alpha$ -helix of dark blue protein**

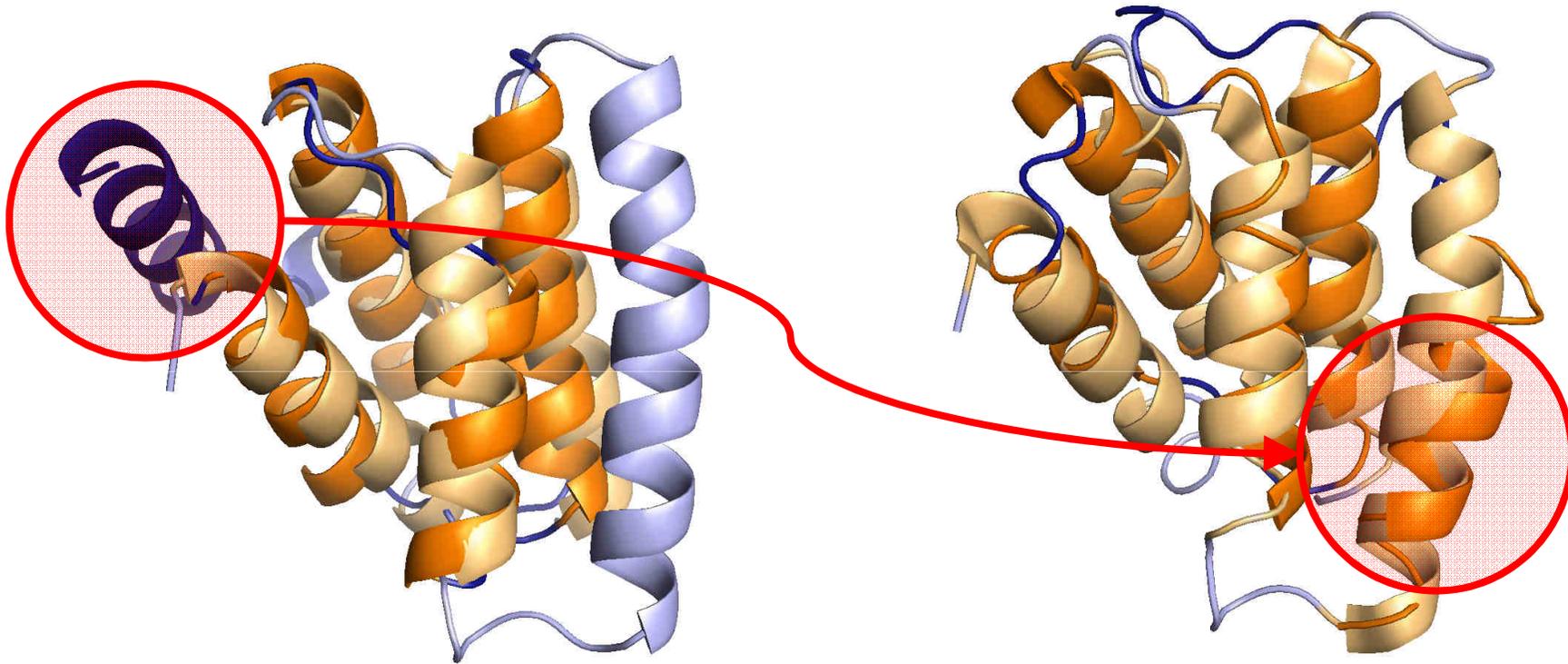
PDB 2ES9 and 1SXJ  
2.5Å RMSD / 57 Residues



**$\alpha$ -helix of light blue protein**

**Is there a more complete structural alignment?**

# Non-sequential similarity

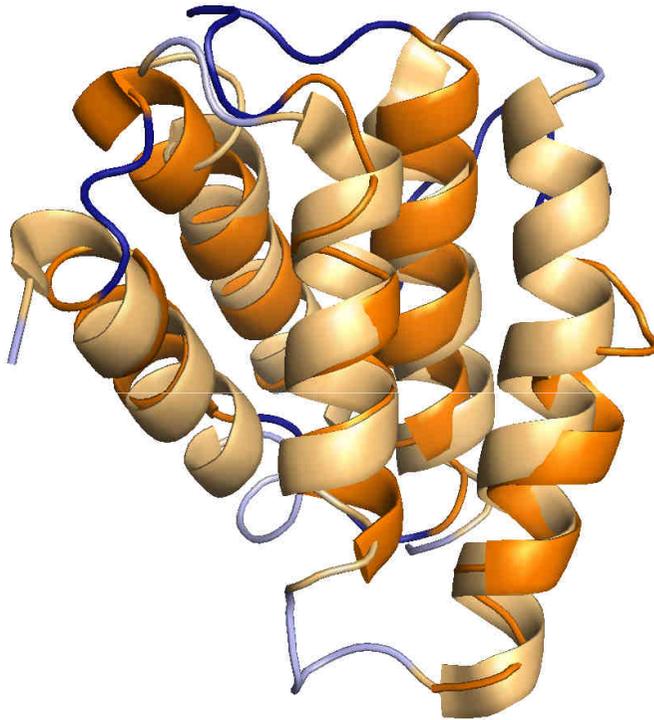


PDB 2ES9 and 1SXJ  
2.5Å RMSD / 57 Residues

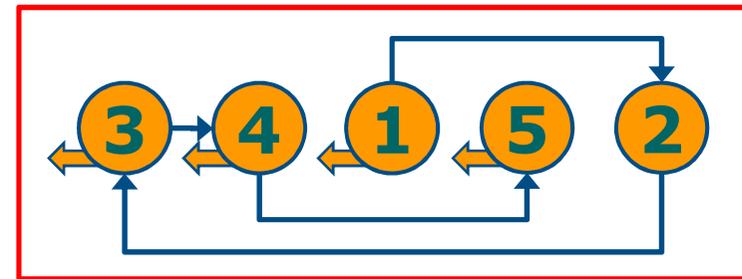
PDB 2ES9 and 1SXJ  
1.8Å RMSD / 69 Residues

**Is there a more complete structural alignment?**

# Non-sequential similarity



PDB 2ES9 and 1SXJ  
1.8Å RMSD / 69 Residues



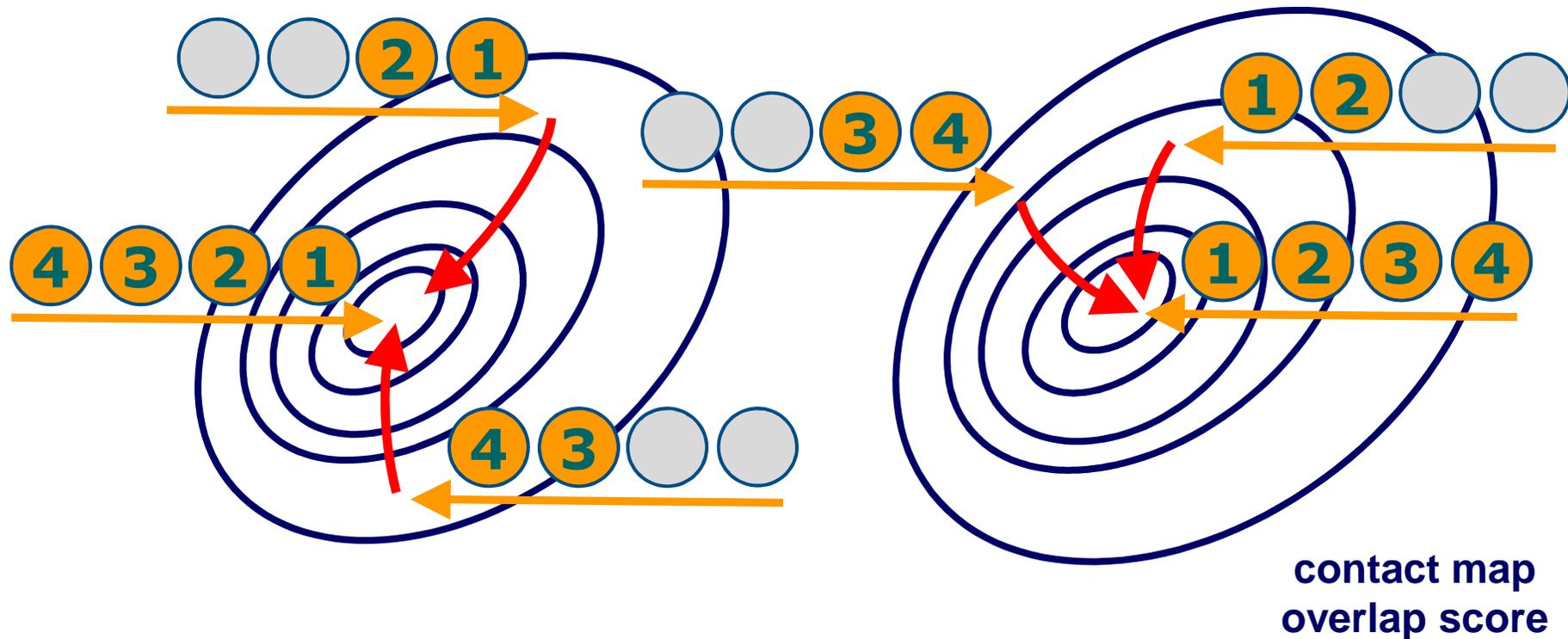
**non-sequential solution,  
illustrated as SSE-assignment map**

**Note: Only helices and sheets are considered  
as Secondary Structure Elements (SSEs)**

**How can we detect this non-sequential solution?**

# Landscape illustration

## 1. Generate all SSE-assignment maps with exactly two assignments

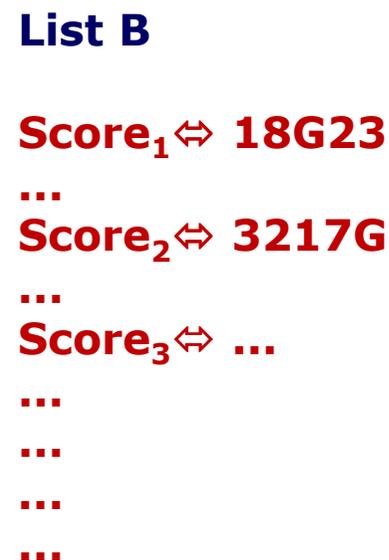


## 2. Merge SSE-assignment maps

Guerler, A., Knapp, E.W. **Novel protein folds and their nonsequential structural analogs**, *Protein Science* 17, 1374-82, 2008

# Optimization of SSE assignments

**Generate more complete  
SSE-assignment maps**



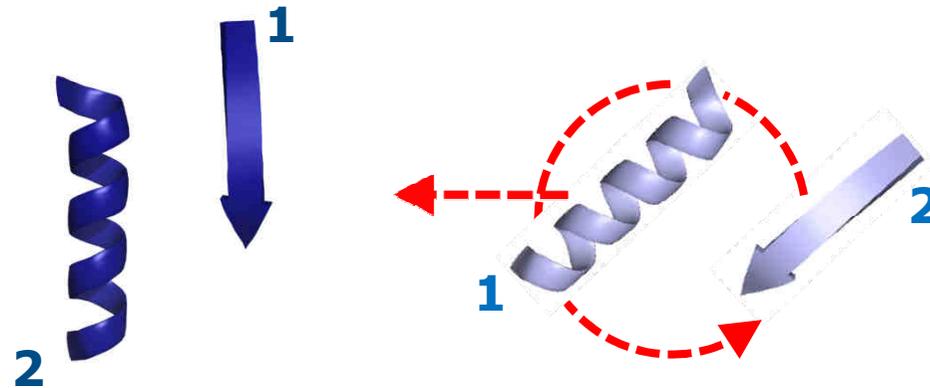
**Save new  
SSE-assignment maps**



**Copy the best entries of List A and List B  
into List A and reiterate the process, where G  
stands for a gap.**

Guerler, A., Knapp, E.W. **Novel protein folds and their nonsequential structural analogs**, **Protein Science** 17, 1374-82, 2008

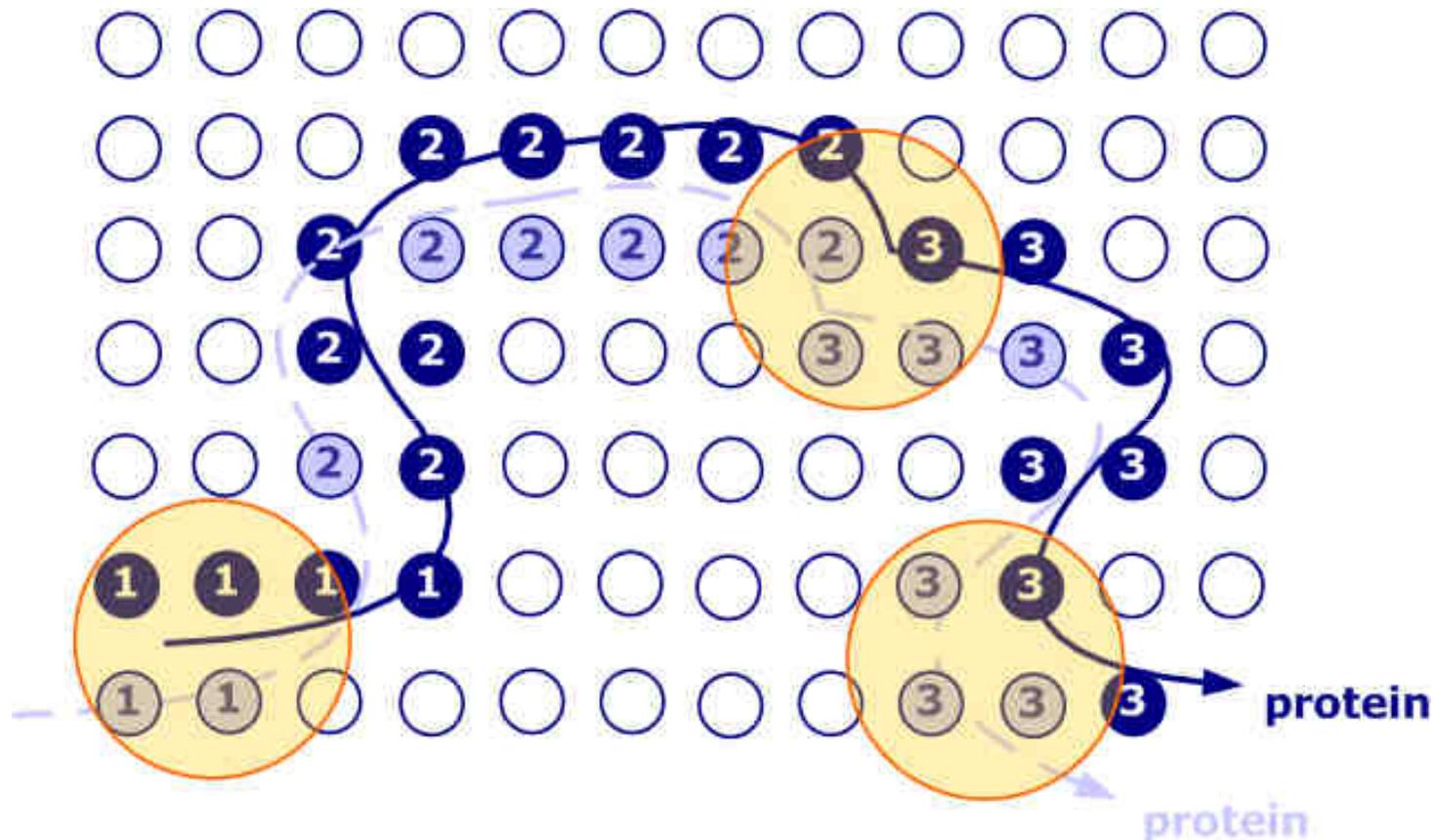
# Transfer to residue-level alignment



**Minimize atomic distances by rigid body transformation.**

Guerler, A., et al., **Selection and flexible optimization of binding modes from conformation ensembles**, **BioSystems** 92, 42-8, 2008

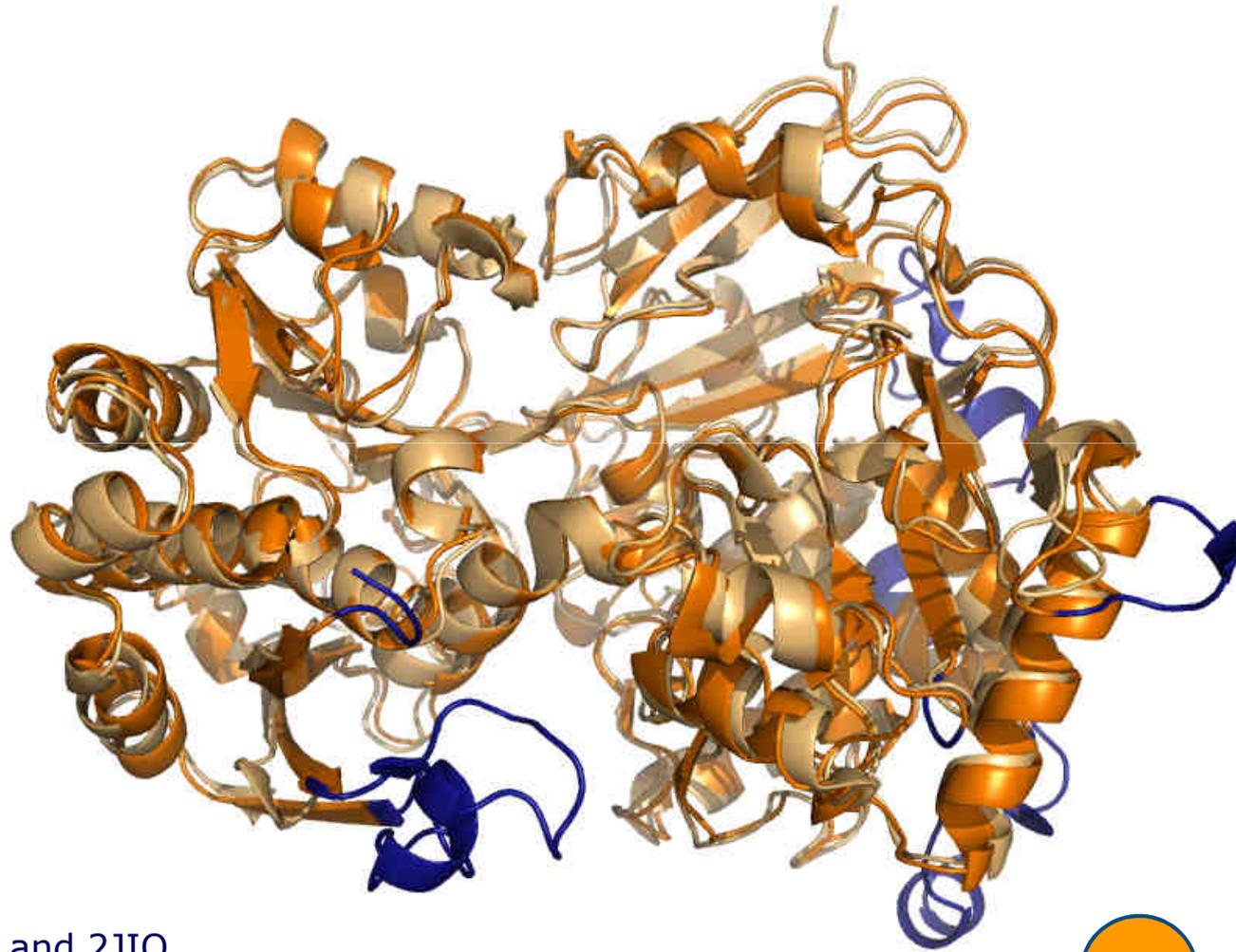
# Grid-based map refinement



## Redetection and optimization of the SSE-assignment maps

Guerler, A., Knapp, E.W. **Novel protein folds and their nonsequential structural analogs**, *Protein Science* **17**, 1374-82, 2008

# SSE-assignment map completion

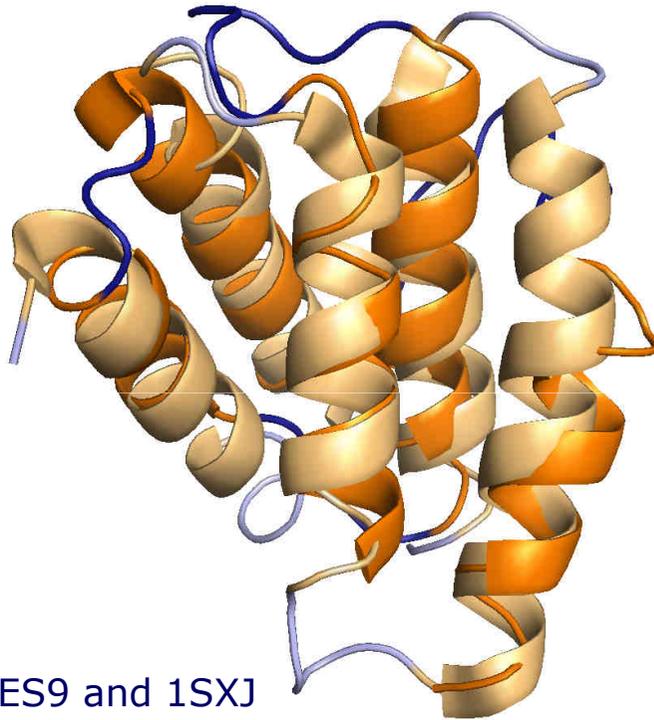


PDB 1OGY and 2JIO  
2.8Å RMSD / 588 Residues

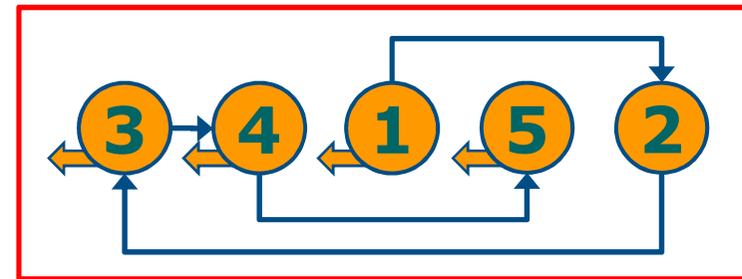
... **15** ... **19** ...

**Note:** Combinatorial optimization has been skipped! → Only two SSEs assigned!

# Strategy summary



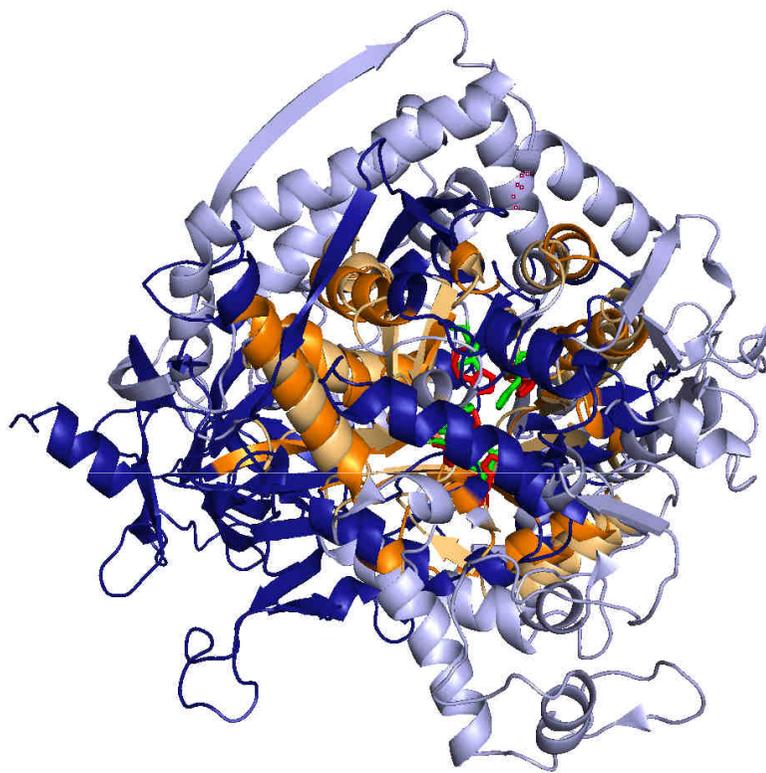
PDB 2ES9 and 1SXJ  
1.8Å RMSD / 69 Residues



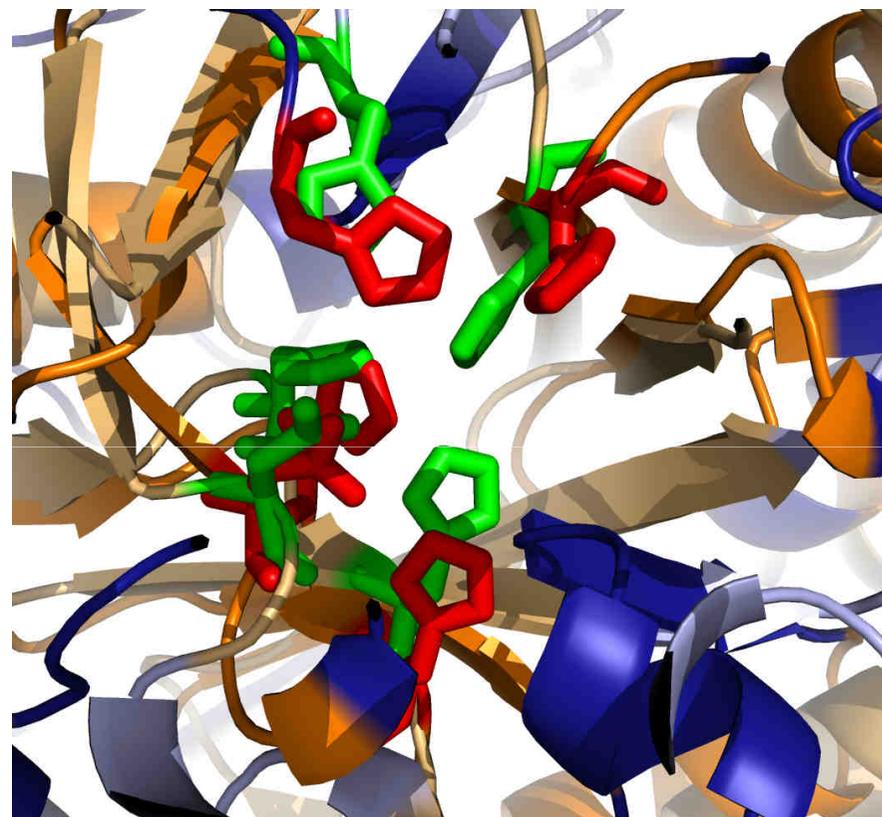
**non-sequential solution,  
illustrated as SSE-assignment map**

- 1. Combinatorial optimization on SSE-level**
- 2. Transfer to residue-level by energy minimization**
- 3. Refinement and optimization in 3D space**

# Binding pocket detection



PDB 1IE7 and 2A3L  
3.2Å RMSD / 144 Residues

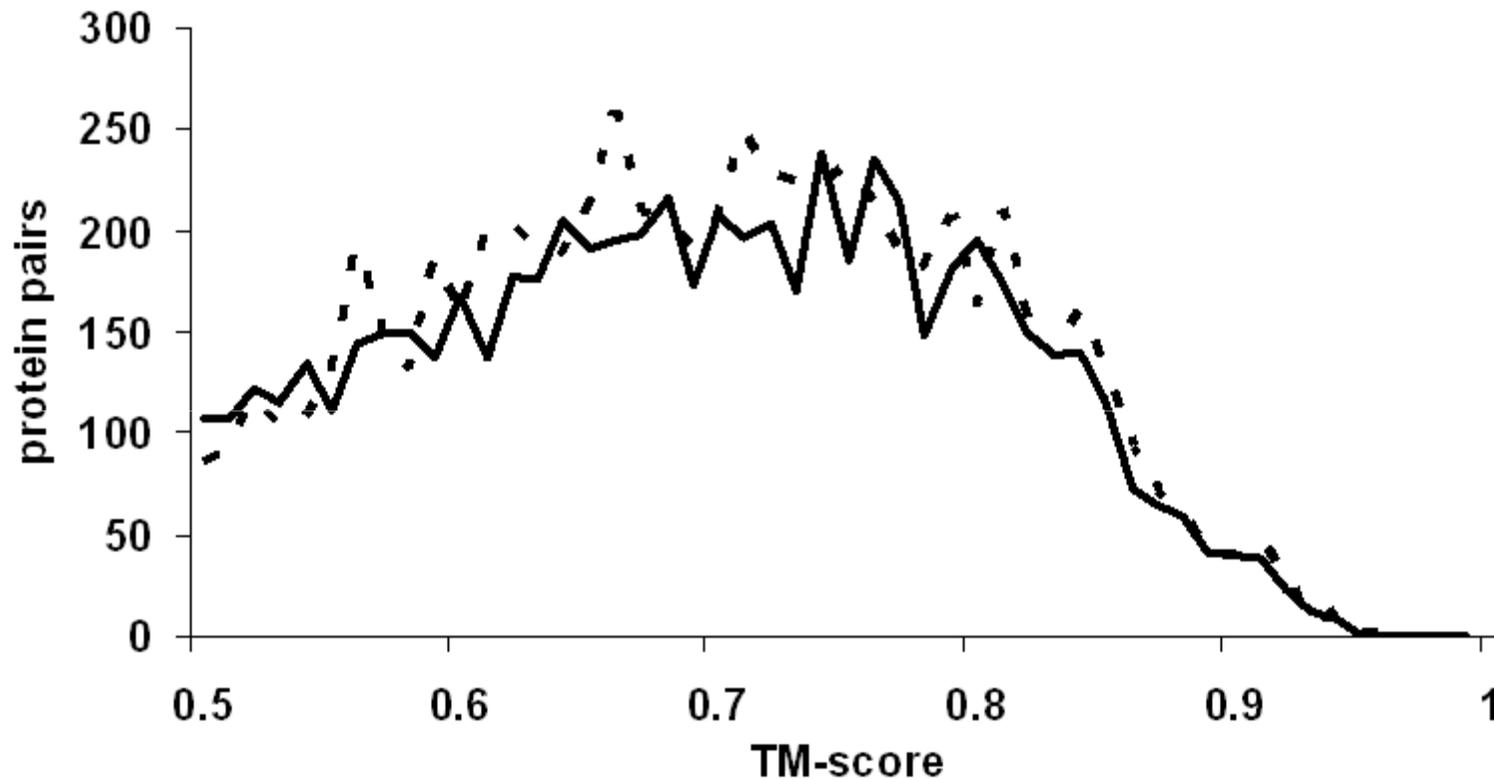


**left: GANGSTA+** structure alignment of urease and adenosine deaminase

**right: enlarged view of active site**

Hasegawa, H., and Holm, L. **Advances and pitfalls of protein structural alignment**,  
**Curr. Opin. in Struct. Biol.** 19, 341-348, 2009

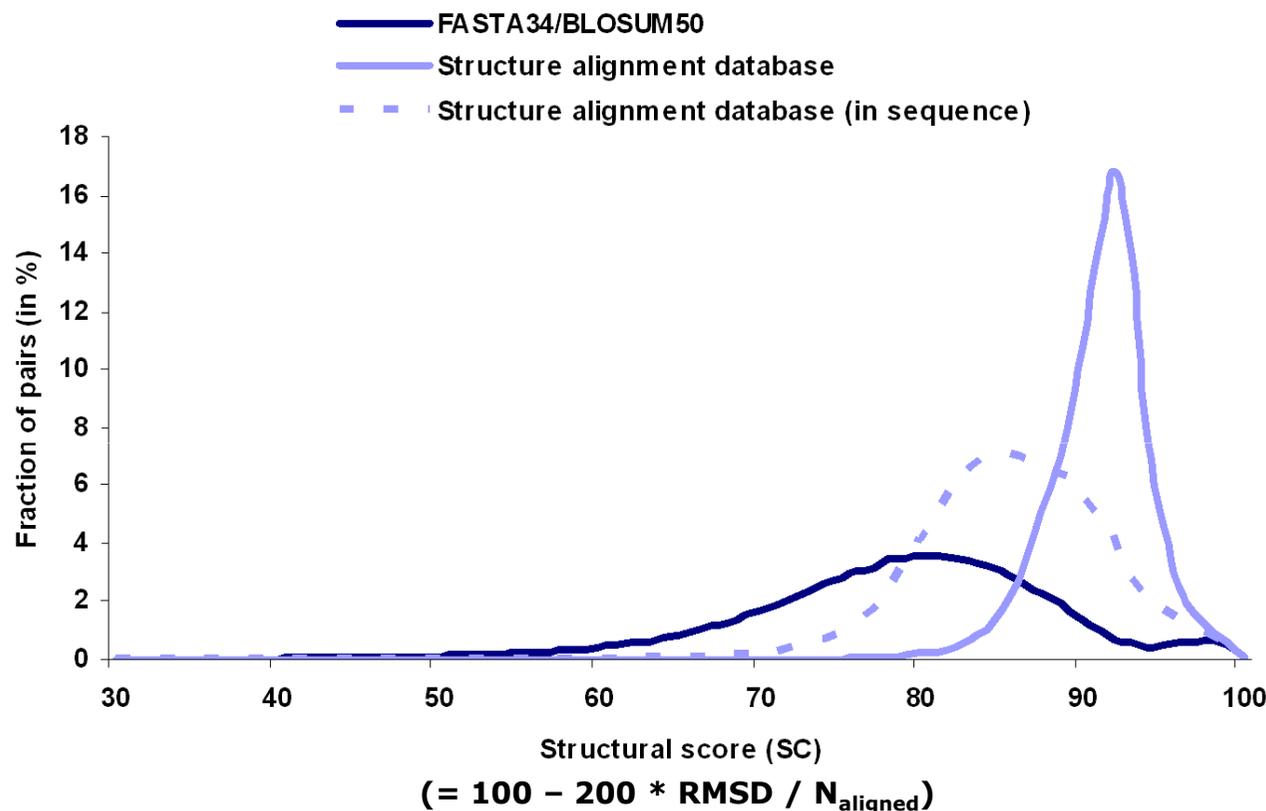
# Sequential analysis



**Sequential structure alignments with  
GANGSTA+ (solid) and TM-align (dashed)**

Guerler, A., Knapp, E.W. **Novel protein folds and their nonsequential structural analogs**, **Protein Science** 17, 1374-82, 2008

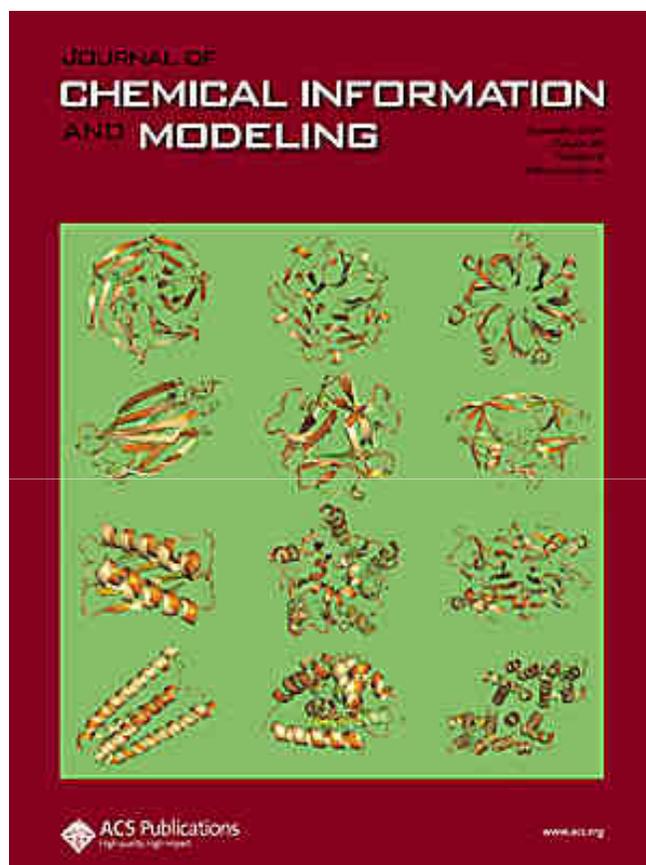
# Database of structural alignments



## Distribution of the structural similarity for the 100 most similar proteins for each ASTRAL40 entry

Guerler, A., et al. **Evaluation of sequence alignments of distantly related sequence pairs with respect to structural similarity**, *Genome Informatics* 18, 183-92, 2007

# Symmetric protein structures



Protein family by SCOP	n
ferredoxin-like	68
immunoglobulin-like $\beta$ -sandwich	31
beta-trefoil	23
four-helical up-and-down bundle	16
DNA clamp	14
7-bladed beta-propeller	13
TIM beta/alpha-barrel	12
gamma-crystalline-like	10

**376 of 8738 protein structures with at minimum 80% of self-aligned residues and less than 4 Å RMSD**

Guerler, A., et al. **Symmetric structures in the protein database**, **Journal of Chemical Information and Modeling** 49, 2147 – 2151, 2009

# Symmetry of frequent proteins



Degree of symmetry for all structures (**light colors**) and for the 20 most frequently appearing structures (**dark**) in the ASTRAL10 (**blue**) and the ASTRAL40 (**orange**) data set of protein structures

Guerler, A., et al. **Symmetric structures in the protein database**, *Journal of Chemical Information and Modeling* 49, 2147 – 2151, 2009

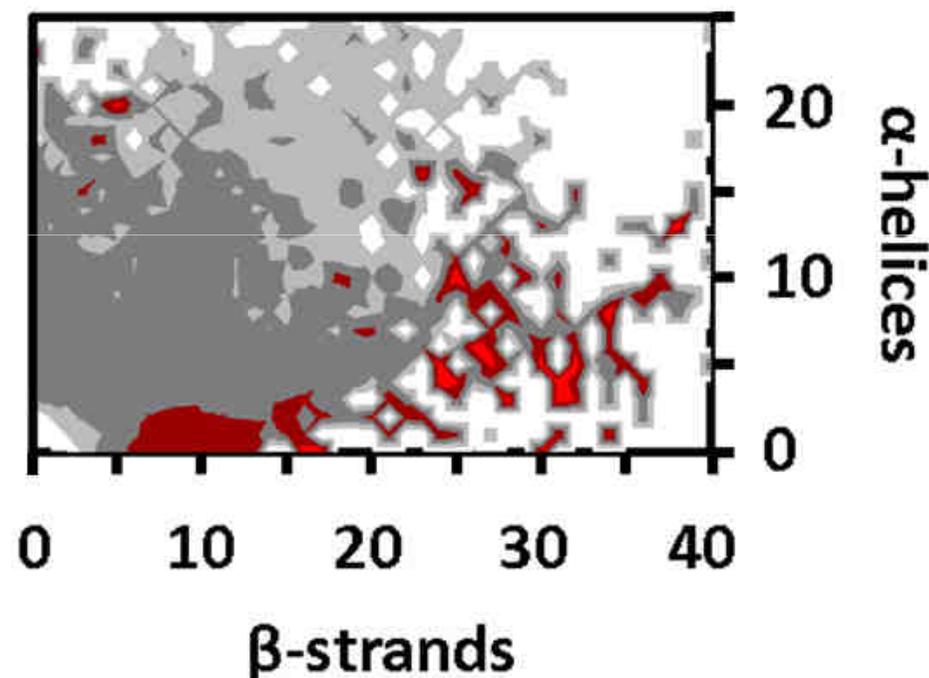
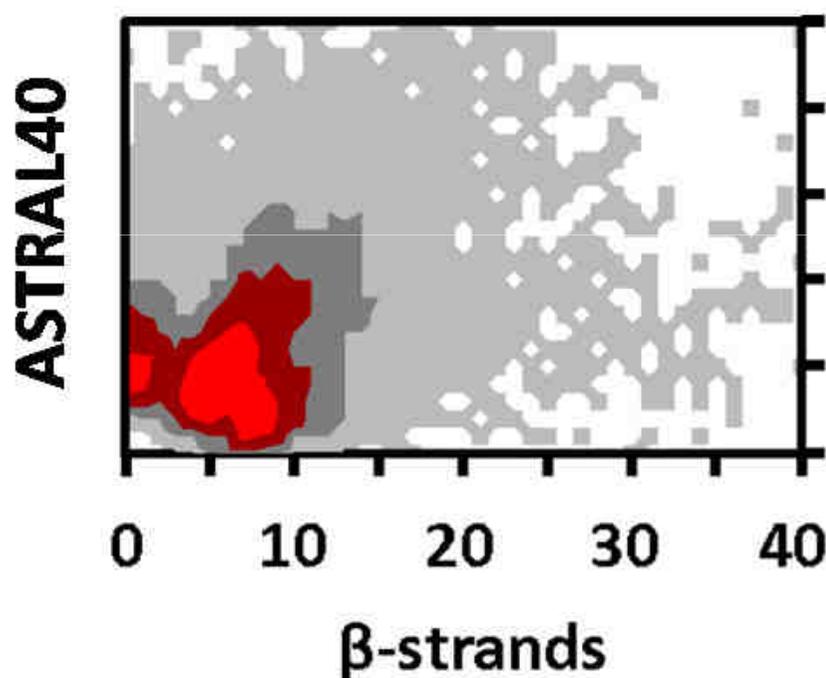
# Symmetric protein structures

absolut number of proteins

**$\geq 80$**  - 40 - 20 - 1

average degree of symmetry

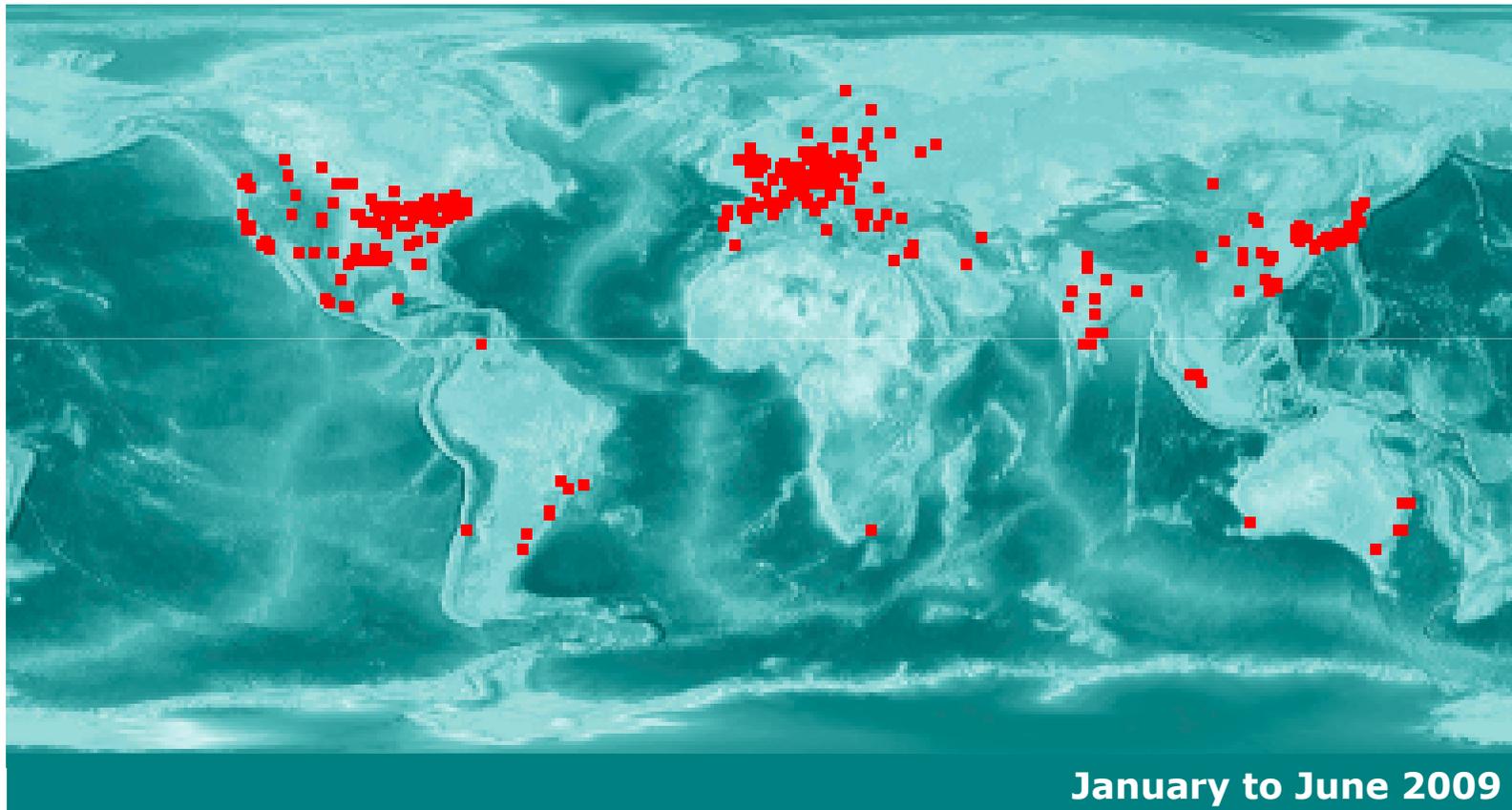
**$\geq 80\%$**  - 60% - 40% - 20%



Guerler, A., et al. **Symmetric structures in the protein database**, **Journal of Chemical Information and Modeling** 49, 2147 – 2151, 2009

# Tracking of user queries

Available at <http://agknapp.chemie.fu-berlin.de/gplus>



**~120 unique users per month**

# Summary

- **Efficient** method for non-sequential structure alignment
- **Validation** with other methods
- **Application** on several million protein structure pairs
- **Analysis** of rotational symmetries in protein structures
- **Highly available** and frequently used

# Thanks to the IRTG and in particular to **Prof. E.-W. Knapp** and to you for your attention

- (1) **Schmidt-Gönner, Guerler et al.** Detection of circular permuted protein structures  
*Proteins*, 2009, *in review*
- (2) **Guerler et al.** Strategies of non-sequential structure alignments  
*Genome Informatics*, 22, 21 – 29, 2009
- (3) **Guerler, Wang, Knapp** Symmetric structures in the protein database  
*Journal of Chemical Information and Modeling*, 49, 2147 – 2151, 2009
- (4) **Guerler, Knapp et al.** Sampling geometries of protein-protein complexes  
*Genome Informatics* 20, 260 – 269, 2008
- (5) **Guerler, Knapp** Novel folds and their non-sequential structural analogs?  
*Protein Science* 17, 1374-1382, 2008
- (6) **Guerler\* et al.** Selection and flexible optimization of binding modes from conformational ensembles *Elsevier BioSystems* 92, 42-48, 2008
- (7) **Bauer, et al., Guerler et al.** Superimpose: A 3D structural superposition server  
*Nucleic Acids Research*, W47-W54, 2008
- (8) **Guerler, Knapp** Evaluation of sequence alignments of distantly related sequence pairs with respect to structural similarity *Genome Informatics* 18, 183-191, 2007