

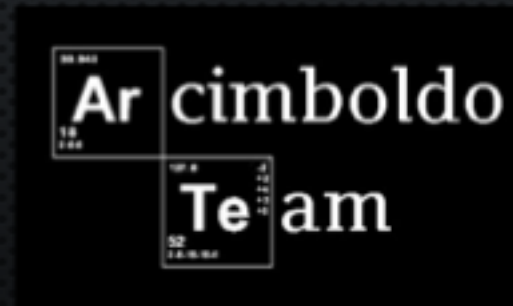
ALEPH: A NETWORK ORIENTED APPROACH FOR THE GENERATION OF FRAGMENT BASED LIBRARIES AND FOR STRUCTURE INTERPRETATION.

PH.D. STUDENT: ANA MEDINA BERNAL, JOSEP TRIVIÑO

PROF. DR. ISABEL USÓN, DR. MASSIMO DOMENICO SAMMITO



Twitter: @InuMassy



ARCIMBOLDO

BORGES

SHREDDER

AMPLE

FRAGON

Fragment based Molecular Replacement

Data quality is crucial:

- Completeness
- Resolution
- Information Content

Merging multiple datasets can help

Significantly **partial models** but very **accurate**

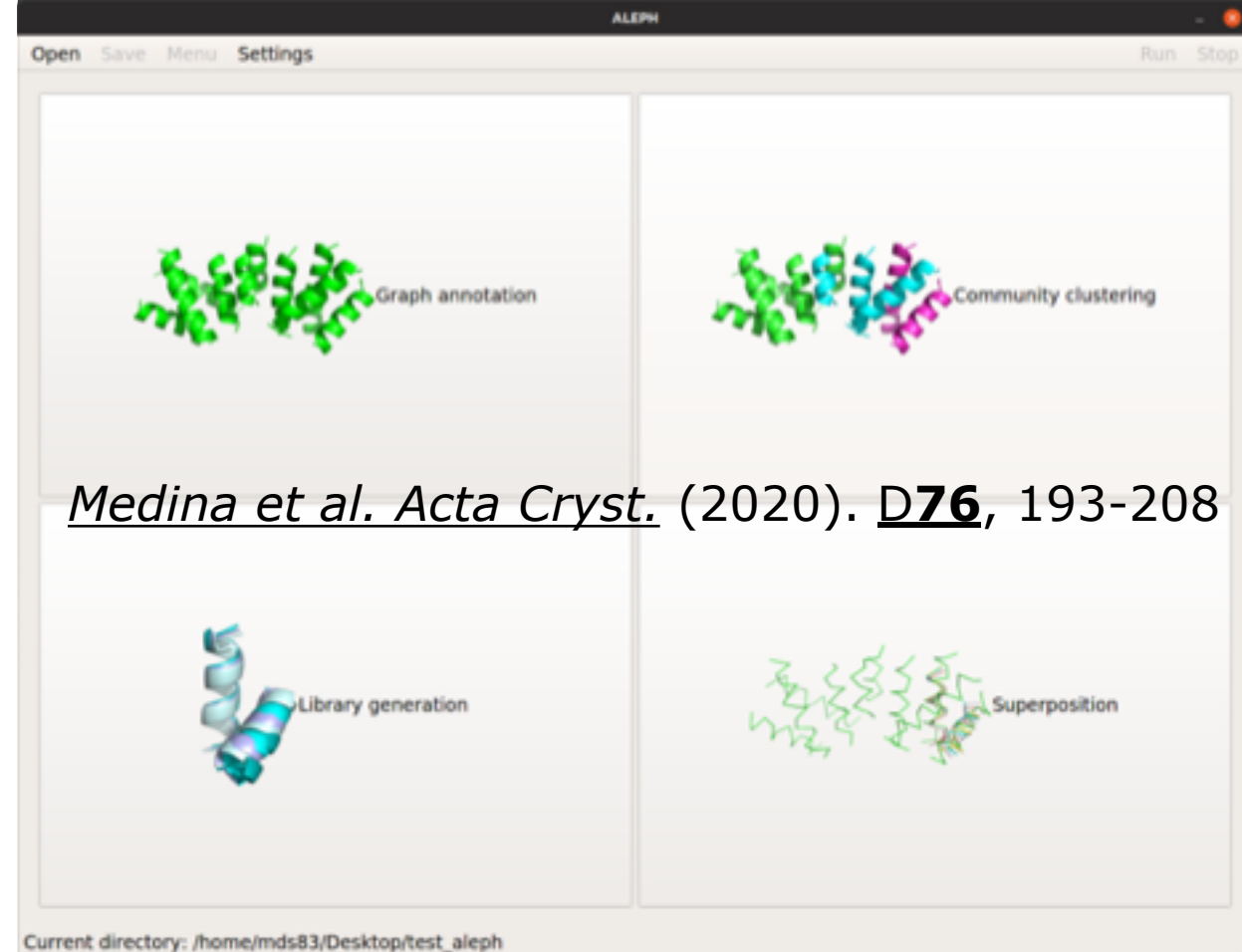
Maximum Likelihood PHASER

Density Modification and Auto-tracing are required to complete the structure

SHELXE

ALEPH to work with folds

- **Accurately analyse small fragments:** annotation of secondary and tertiary structure through geometrical descriptors encoded in networks.
- **Decomposition in local folds:** community clustering to find compact subfolds.
- **Library generation:** fragment based libraries for *ab initio* phasing .
- **Superposition of small, discontinuous fragments:** graph matching.

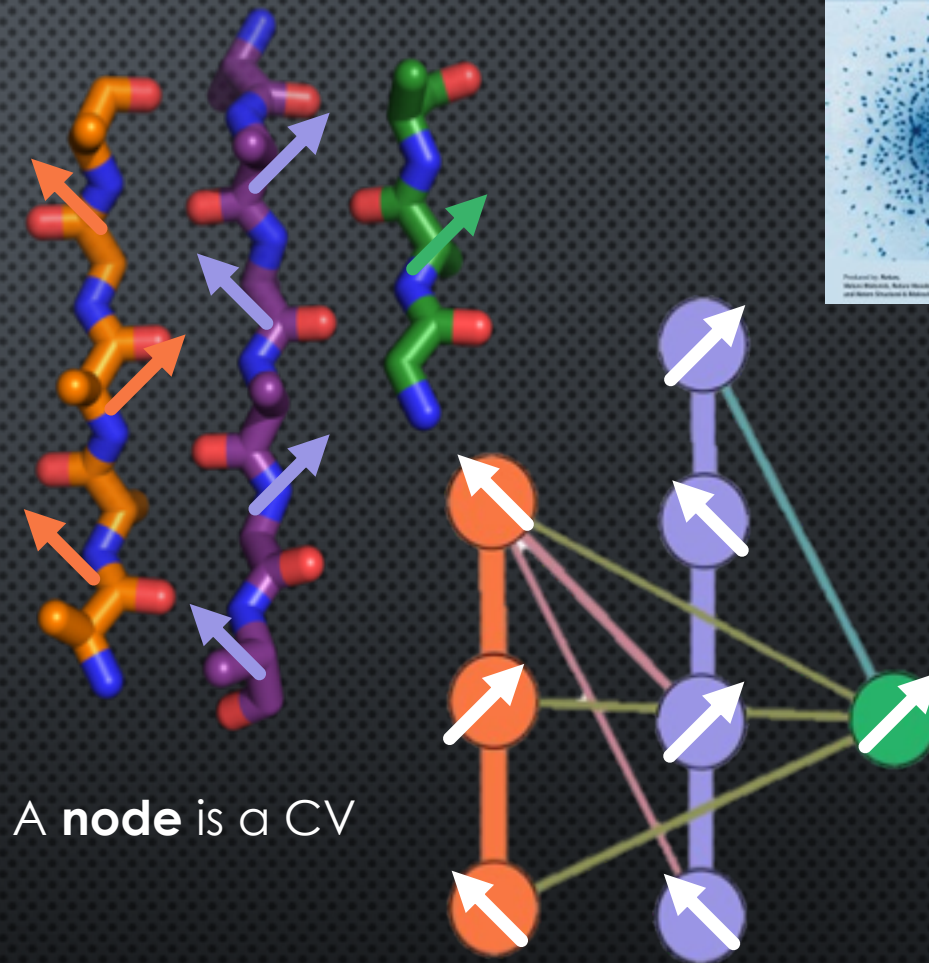
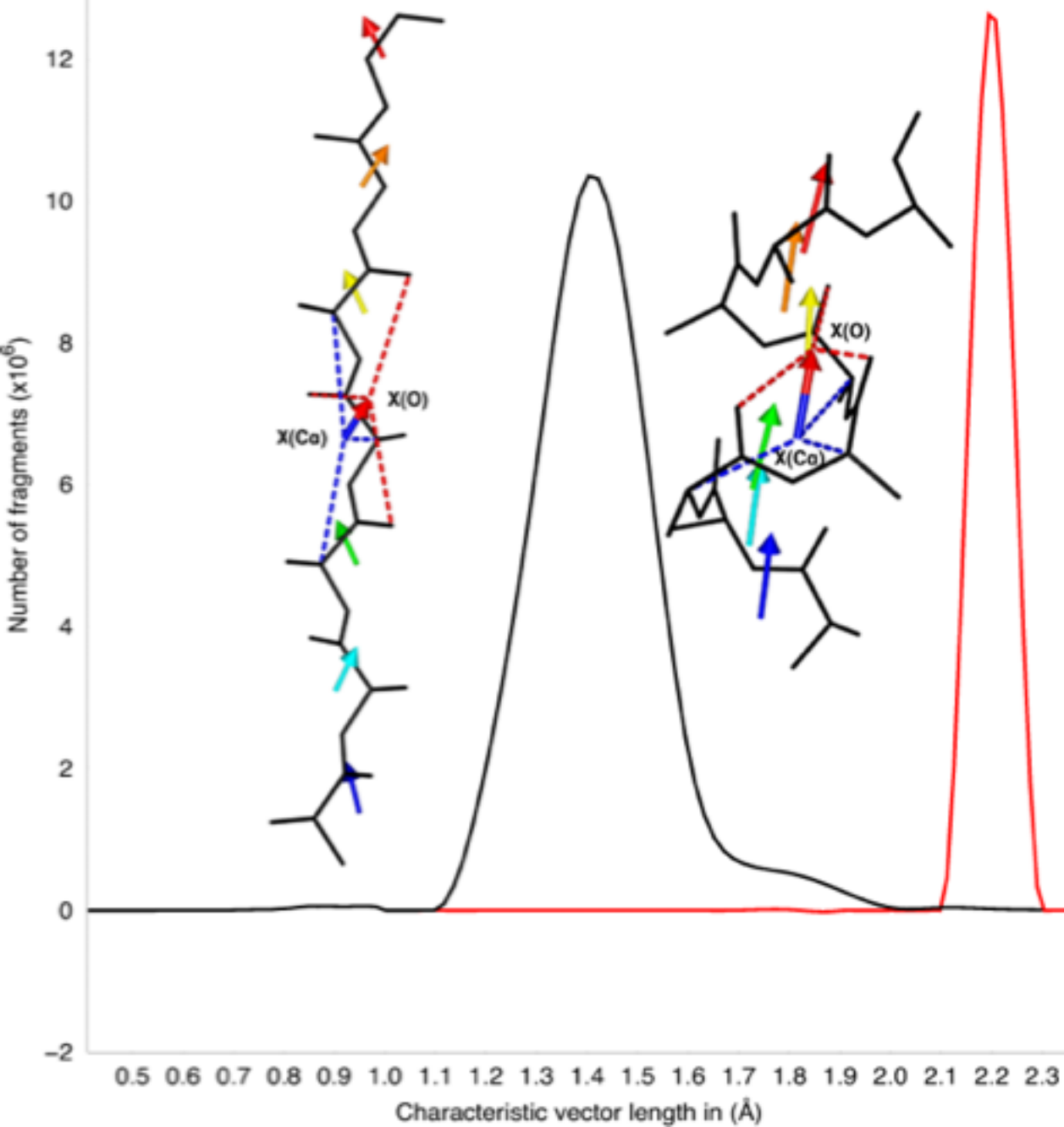


- Python 3 + PySide2 + WebNGL
- Program is available through pip (**pip install aleph**)
- GUI is independent from the main program that can run from command line.



Josep Triviño

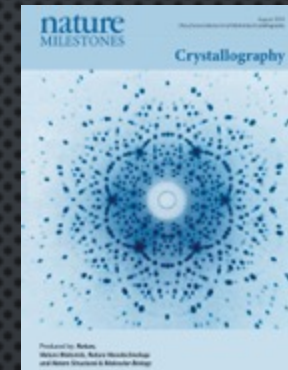
Annotation: characteristic vectors



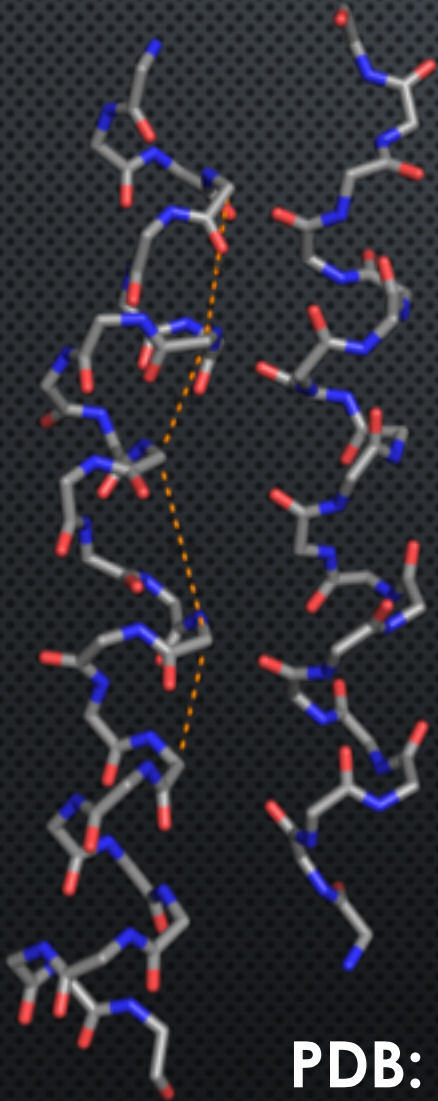
A **node** is a CV

Sammito, M. et al. (2013) Nature Methods 10, 1099-101

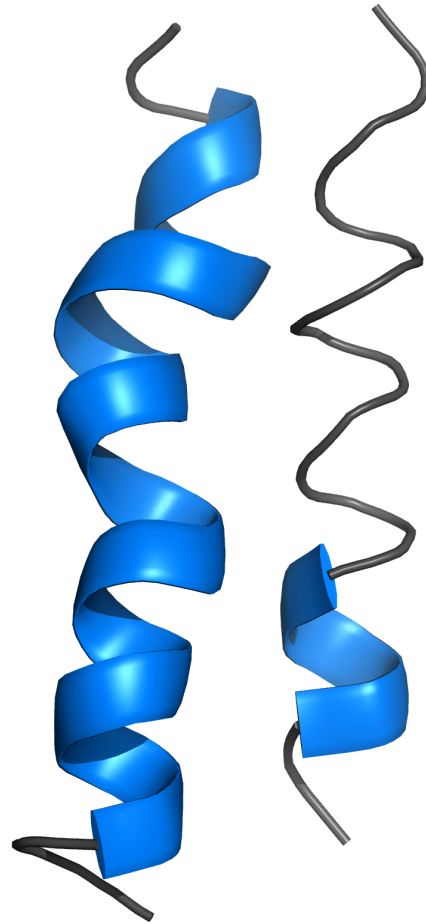
An **edge** is a 3d rel. between two vertices



Customizable secondary structure annotation



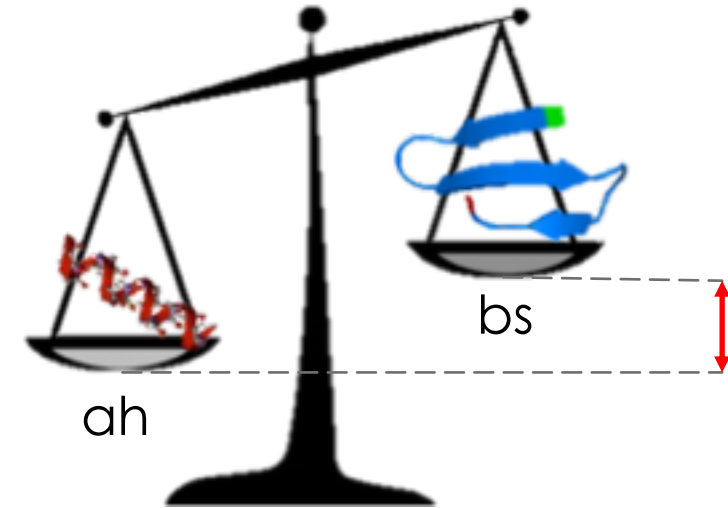
PDB: 2O01



Photosystem I

Range 0.2-0.45

You can control the annotation

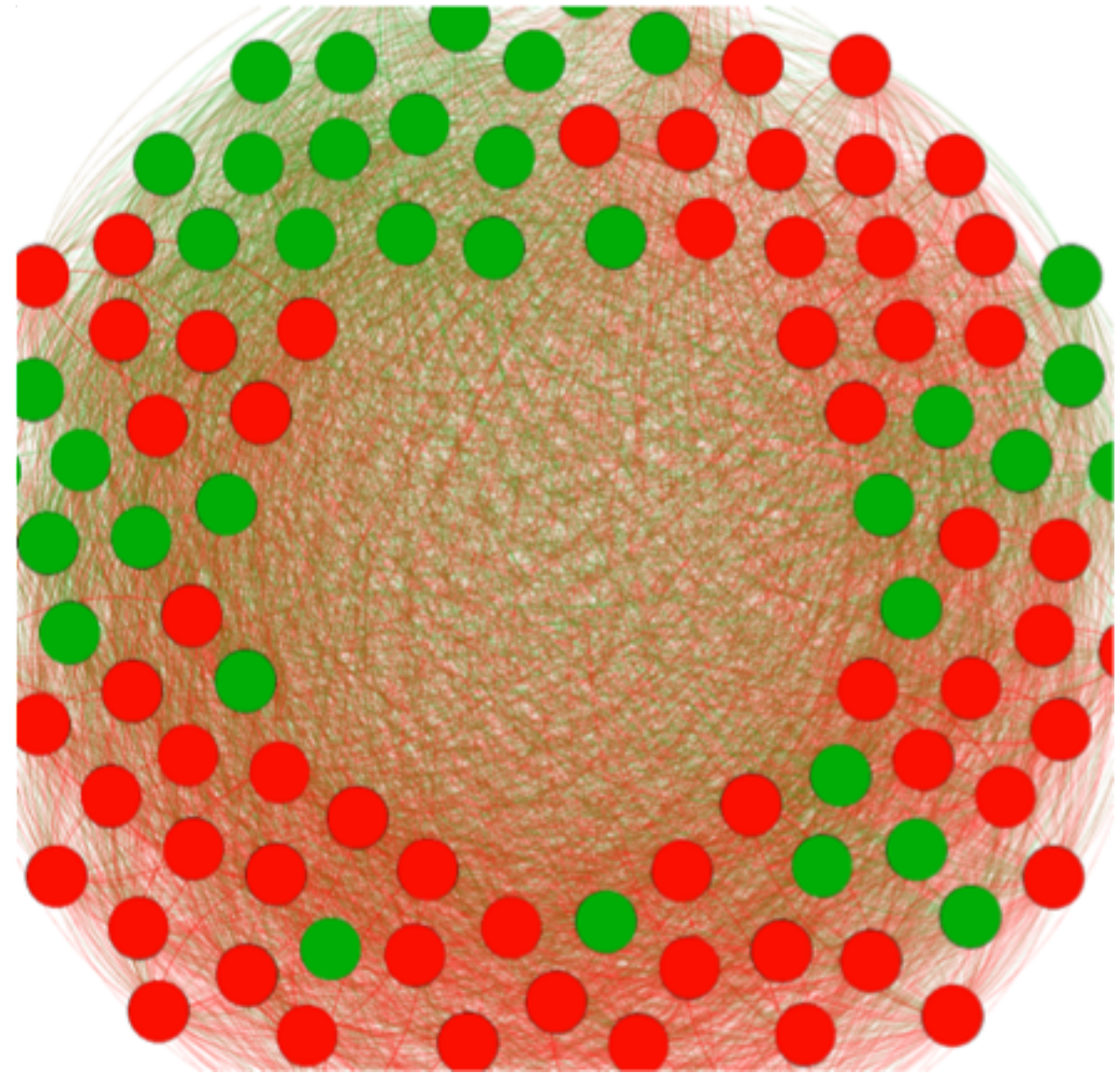
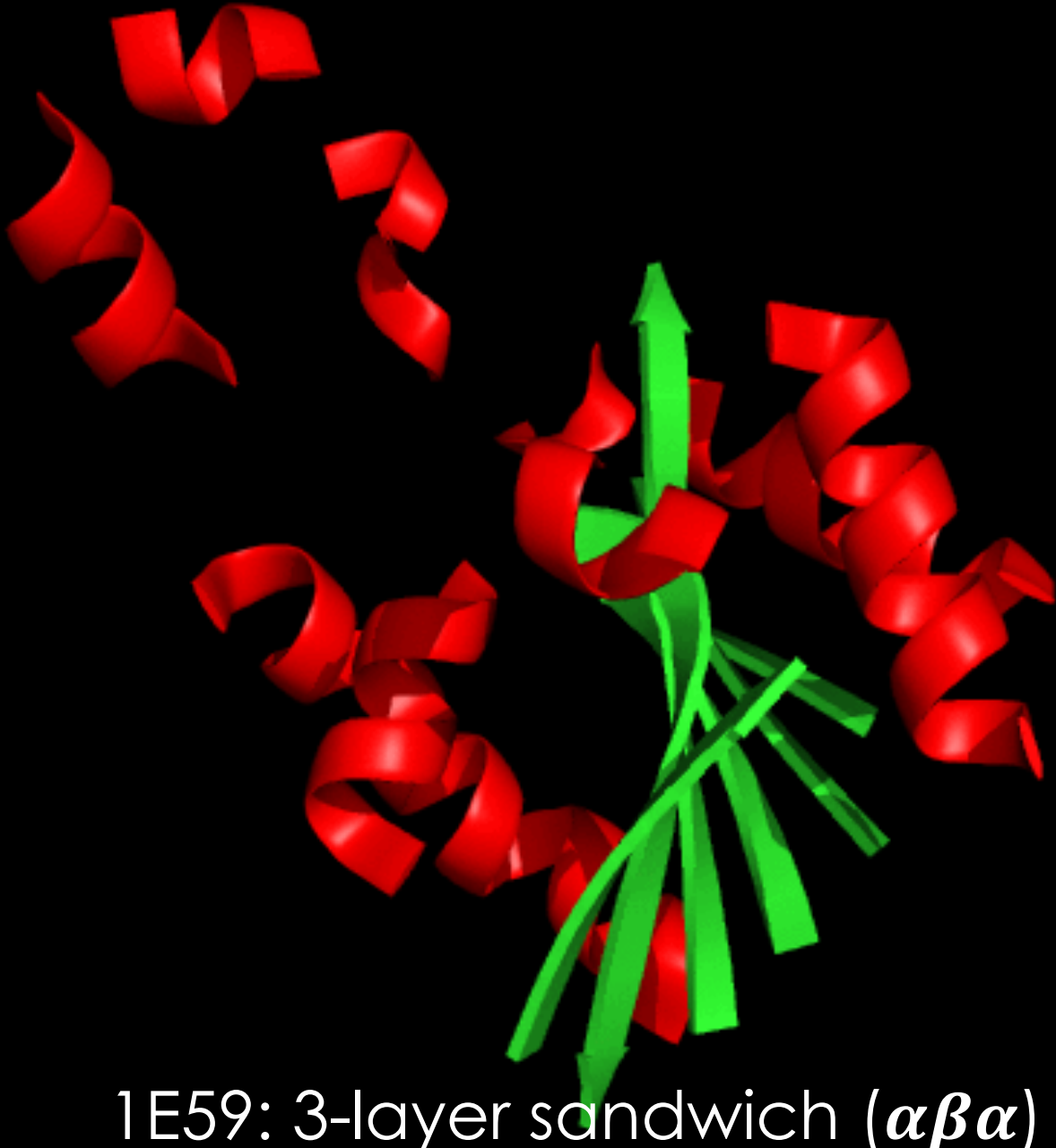


STRICTNESS



Ana Medina

Tertiary structure annotated with graph



Full undirected weighted graph.

Fold template

Given as .pdb
file

Described as
CVs

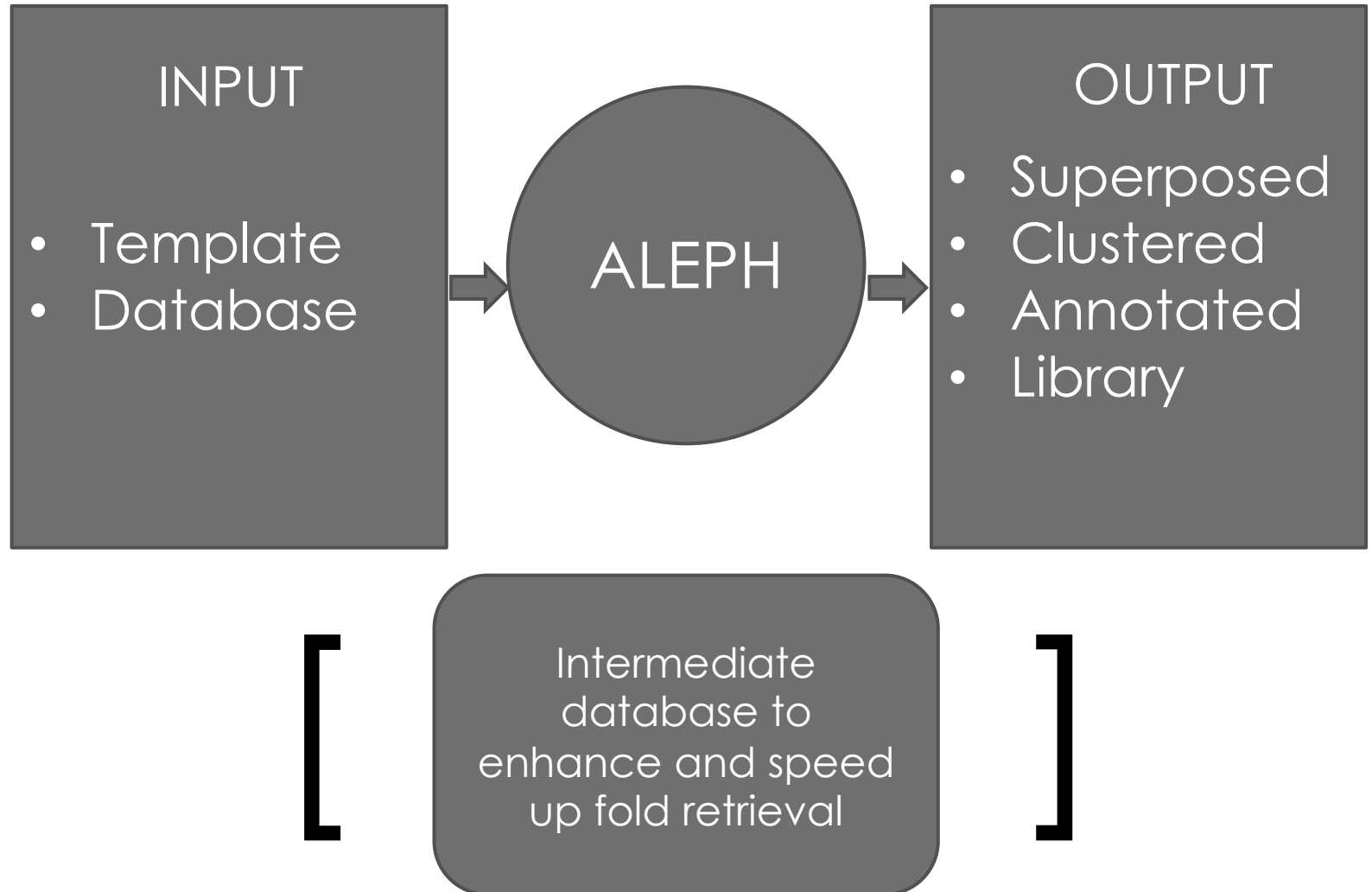
Database
preparation

Downloaded
from RCSB

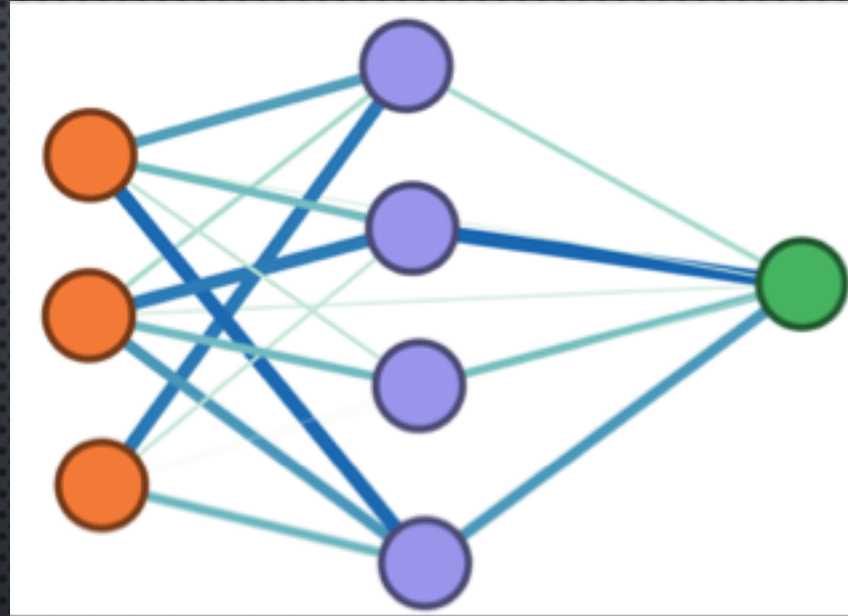
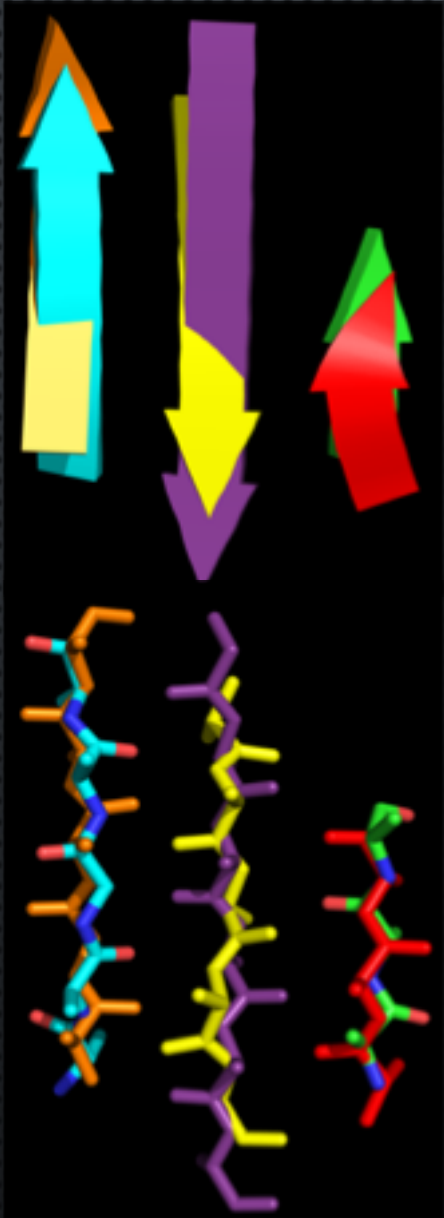
Given as
directory file

From sequence
file

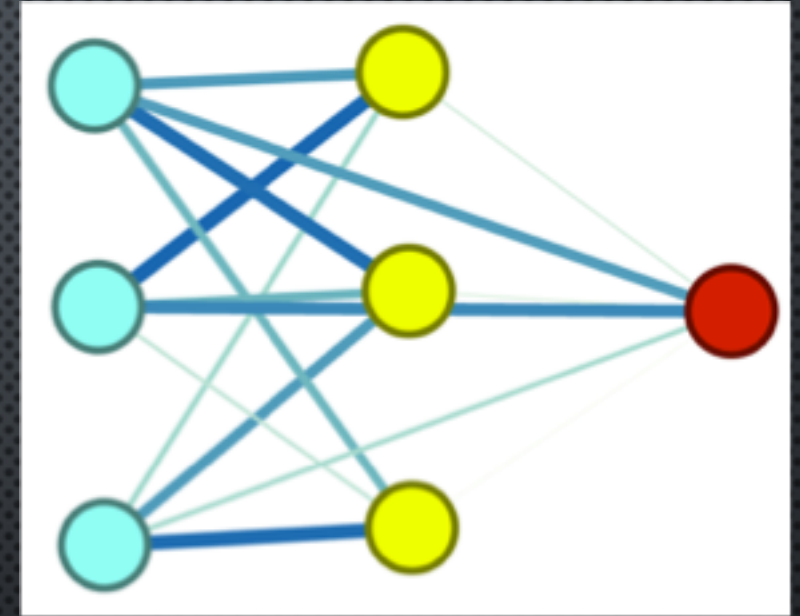
Generating a fragment based library



Comparing structures



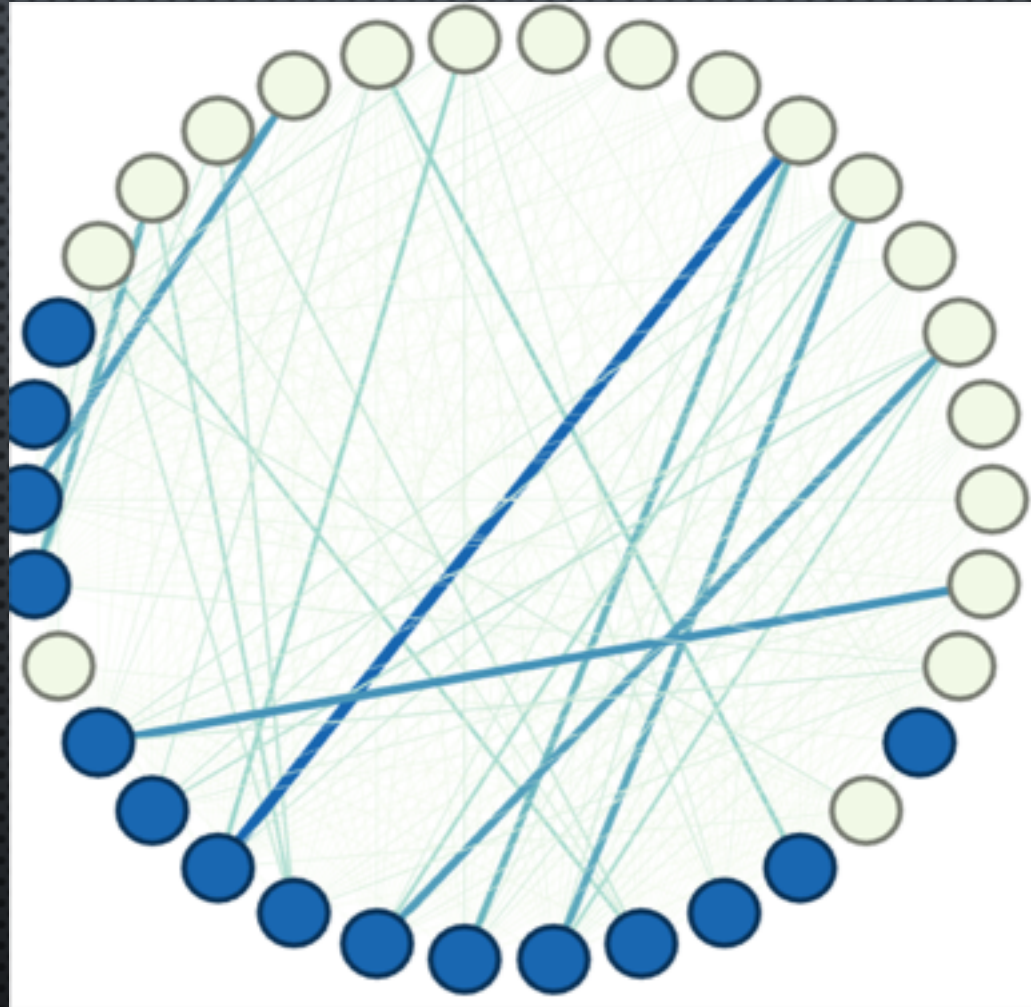
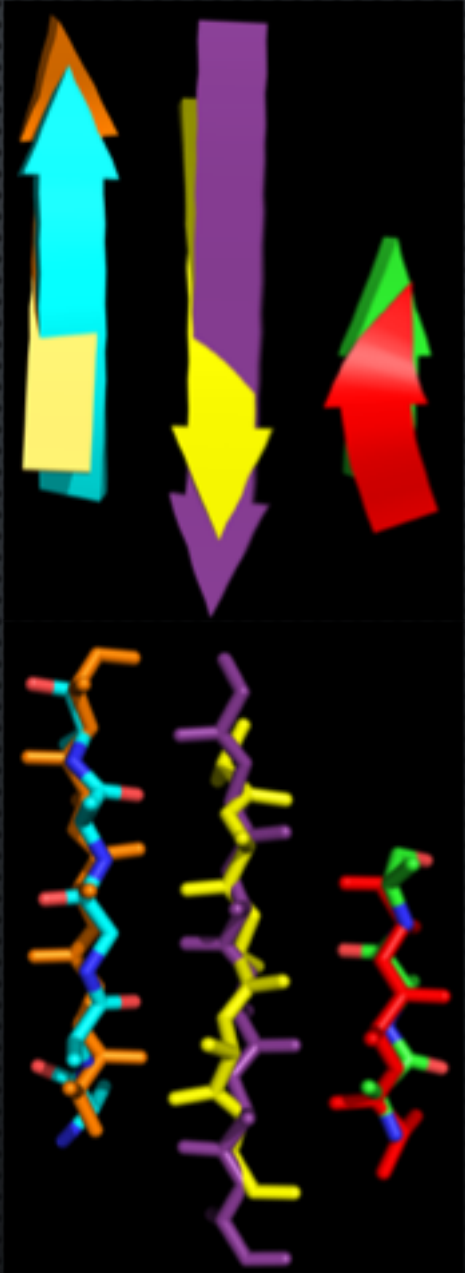
$$G_1 = (V_1; E_1)$$



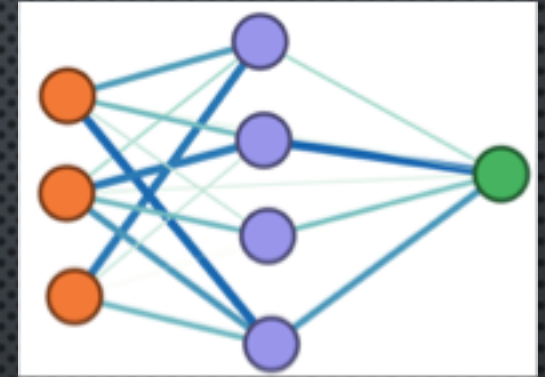
$$G_2 = (V_2; E_2)$$



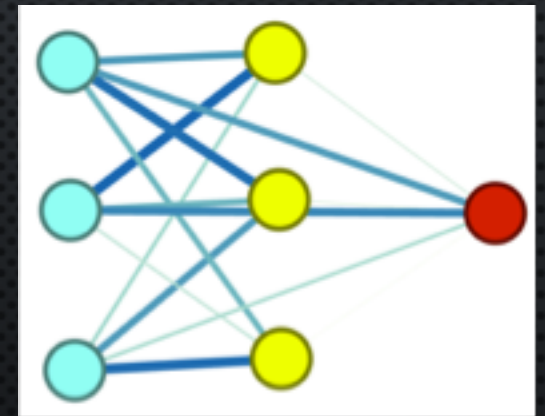
Comparing two structures: edge product graph



$$G_p = (E_1 + E_2; E_1 * E_2)$$
$$W(E_1 * E_2) = \{d(e_1, e_2)\}$$

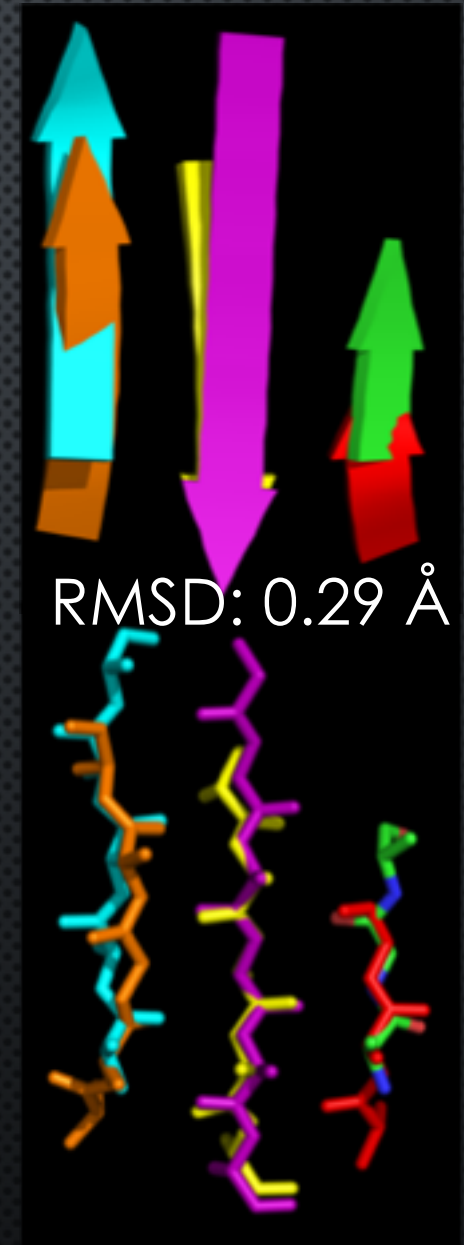
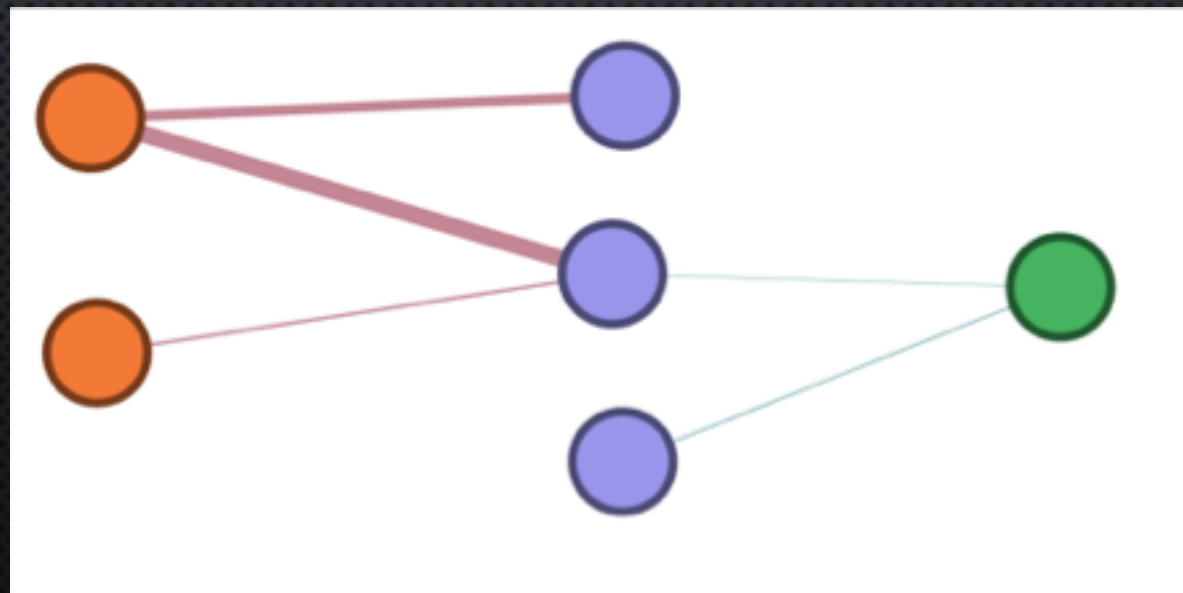
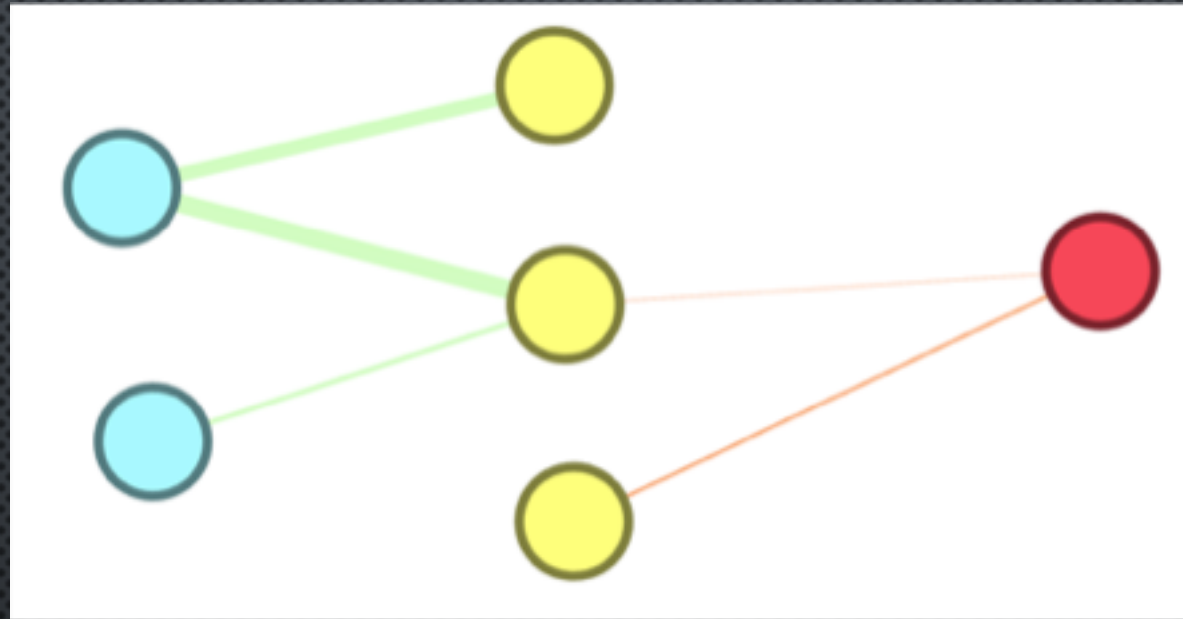


$$G_1 = (V_1; E_1)$$

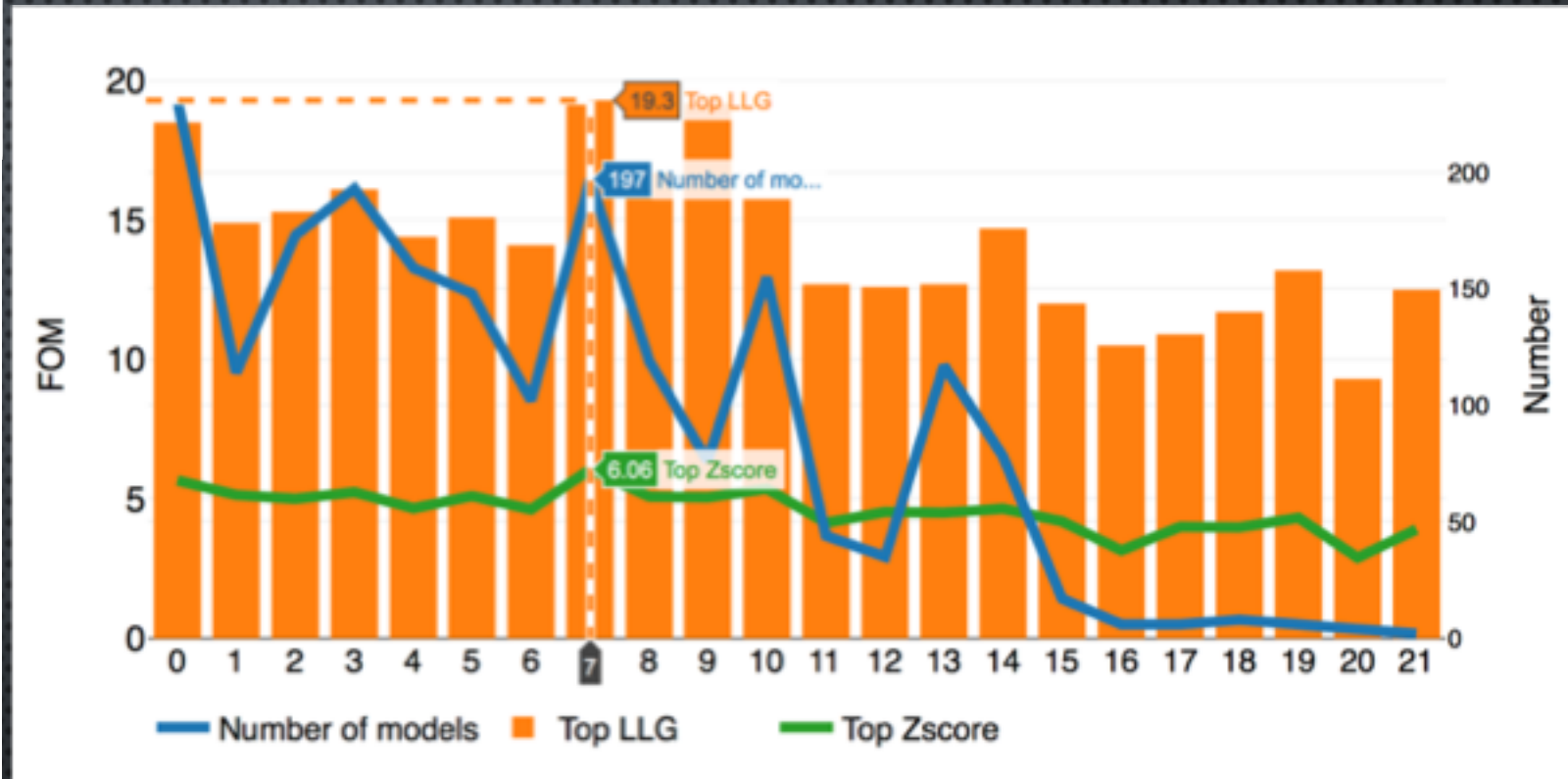
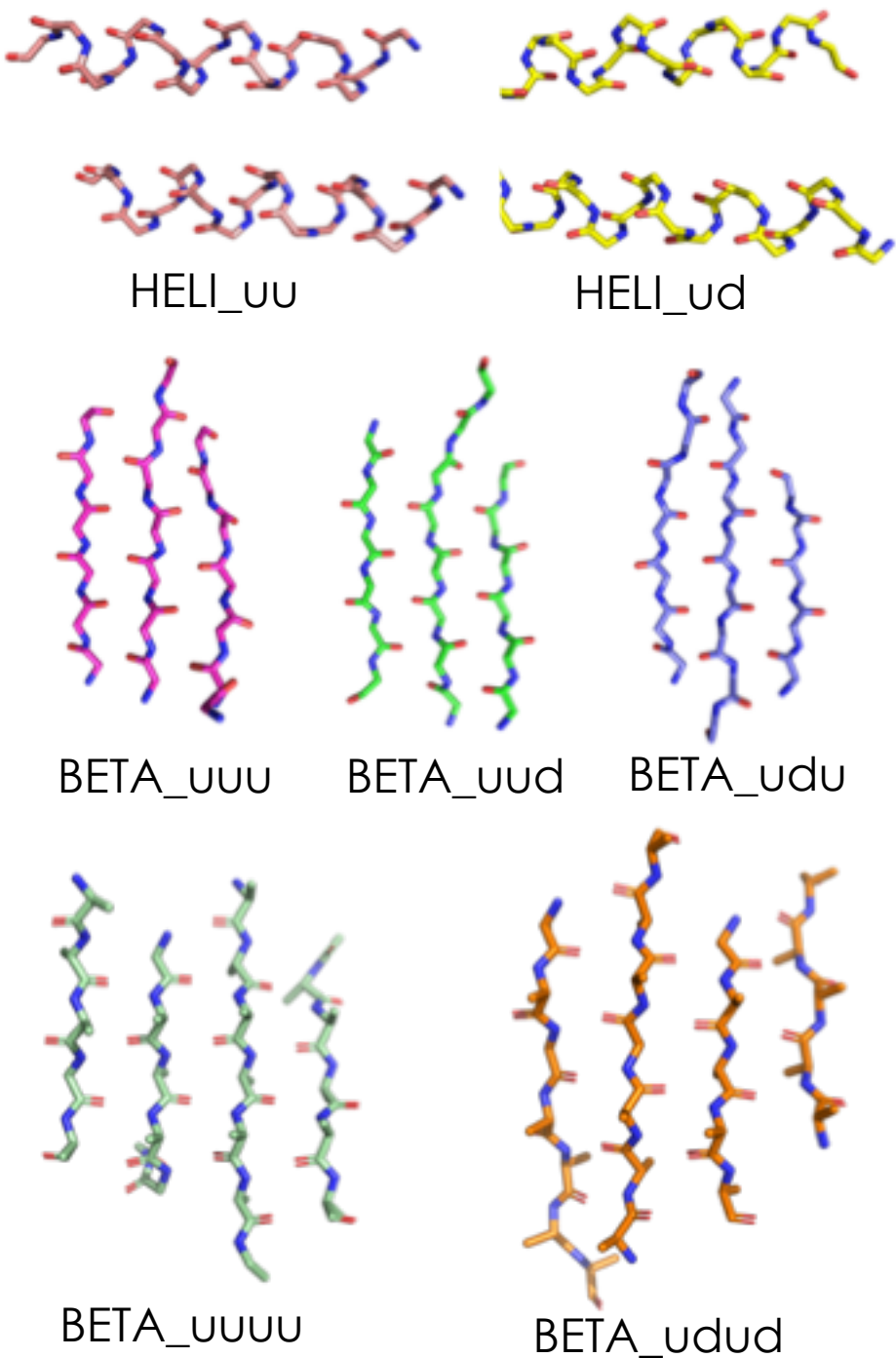


$$G_2 = (V_2; E_2)$$

Comparing two structures: superposition



ARCIMBOLDO_BORGES



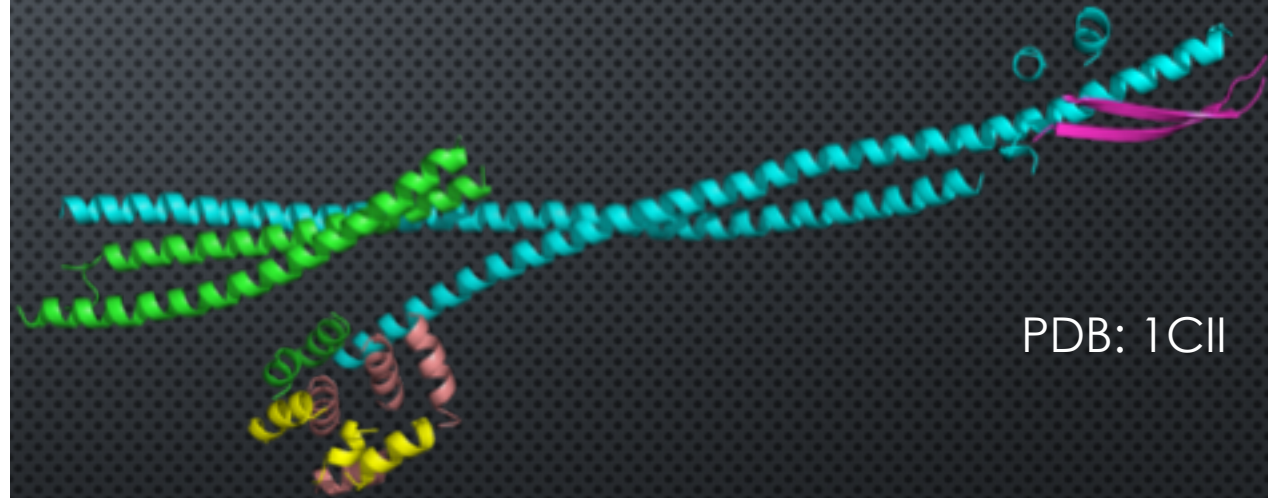
Clustering of the models in the library against the rotation function. The most populated clusters are the ones that usually produce correct solutions.

The analysis of the library as a whole, reveals new patterns that are impossible to extract with individual molecular replacement jobs.

Structure decomposition

- Cluster secondary structure elements into compact rigid groups.
- The new PHASER **gyre/gimble** refinement locally refine rigid groups at the correct position.
McCoy, A. J., et al (2018) Acta Cryst D74, 279-289
- **Modularity** score: maximize the intra-group connections while minimizing the inter-group ones.
Clauset et al. Phys Rev E 70, 066111 (2004)
- **Homogeneity** score: help to distribute the number of elements uniformly among groups.

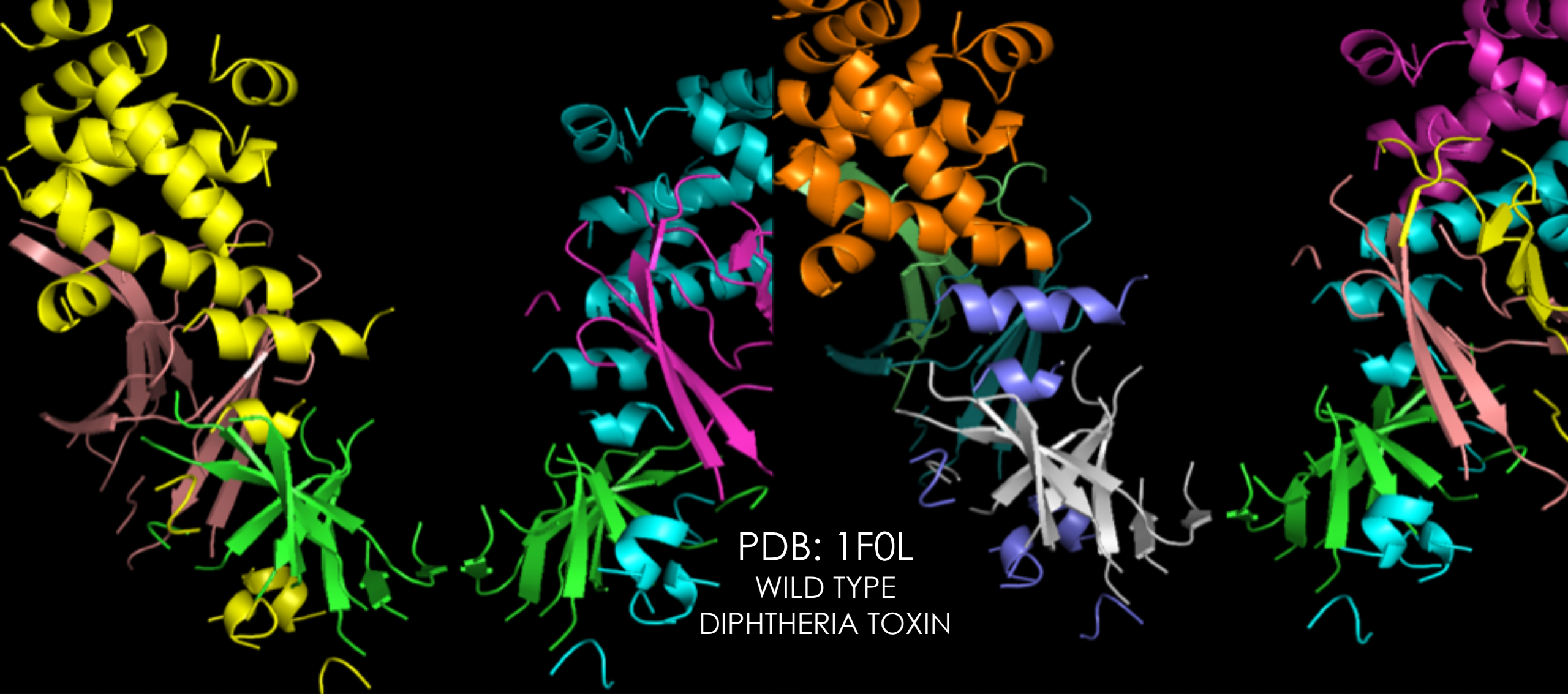
Modularity Score (q) is the **fraction** of the edges that fall **within** the given **groups minus** the expected **fraction** if edges were distributed at **random**



PDB: 1CII

Homogeneity Score

Penalize clustering resulting in groups with different cardinalities.



Homogeneity ON:
Decomposition enforces
uniform cluster sizes

Homogeneity OFF:
Only modularity is
optimized: compactness



PDB: 1YGS
SMAD4 TUMOR
SUPPRESSOR C-
TERMINAL DOMAIN



PACK BETA SHEET ON:
Beta Sheet are favored to
form independent clusters

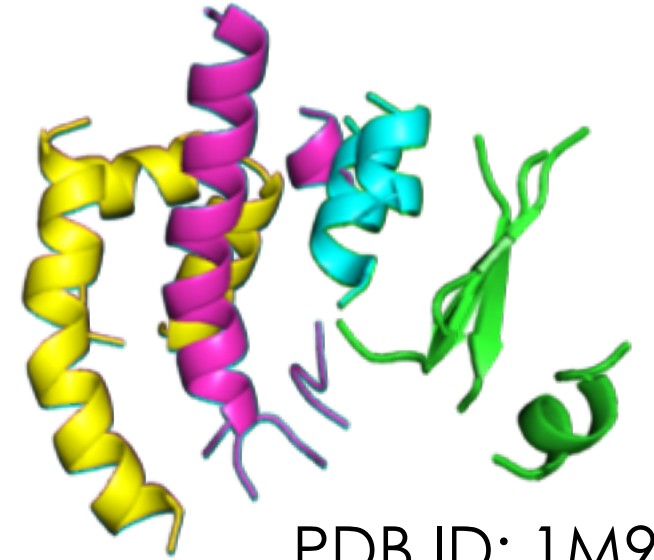
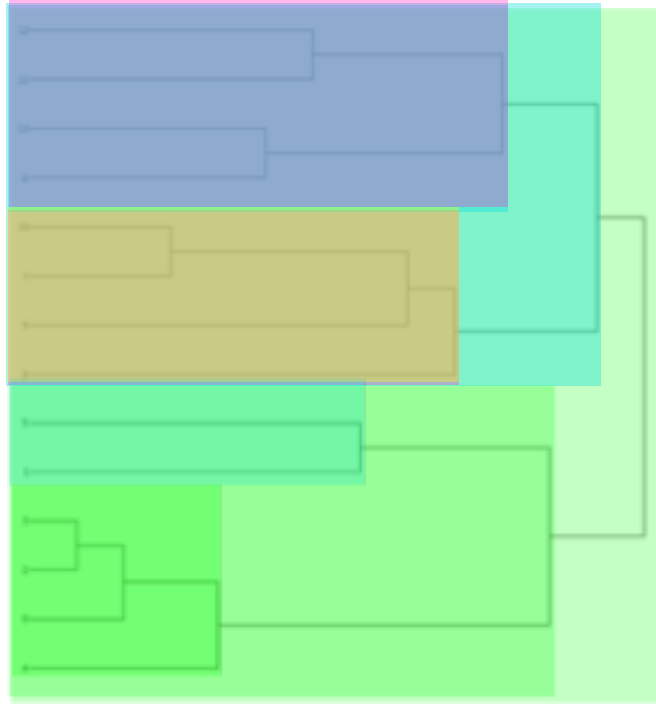


PDB: 1YGS
DEHYDROGENASE/
REDUCTASE

ARCIMBOLDO_SHREDDER cuts models from templates thus annotated, which are also used with PHASER Gyre/Gimble functions for rigid group refinement.

Hierarchical decomposition

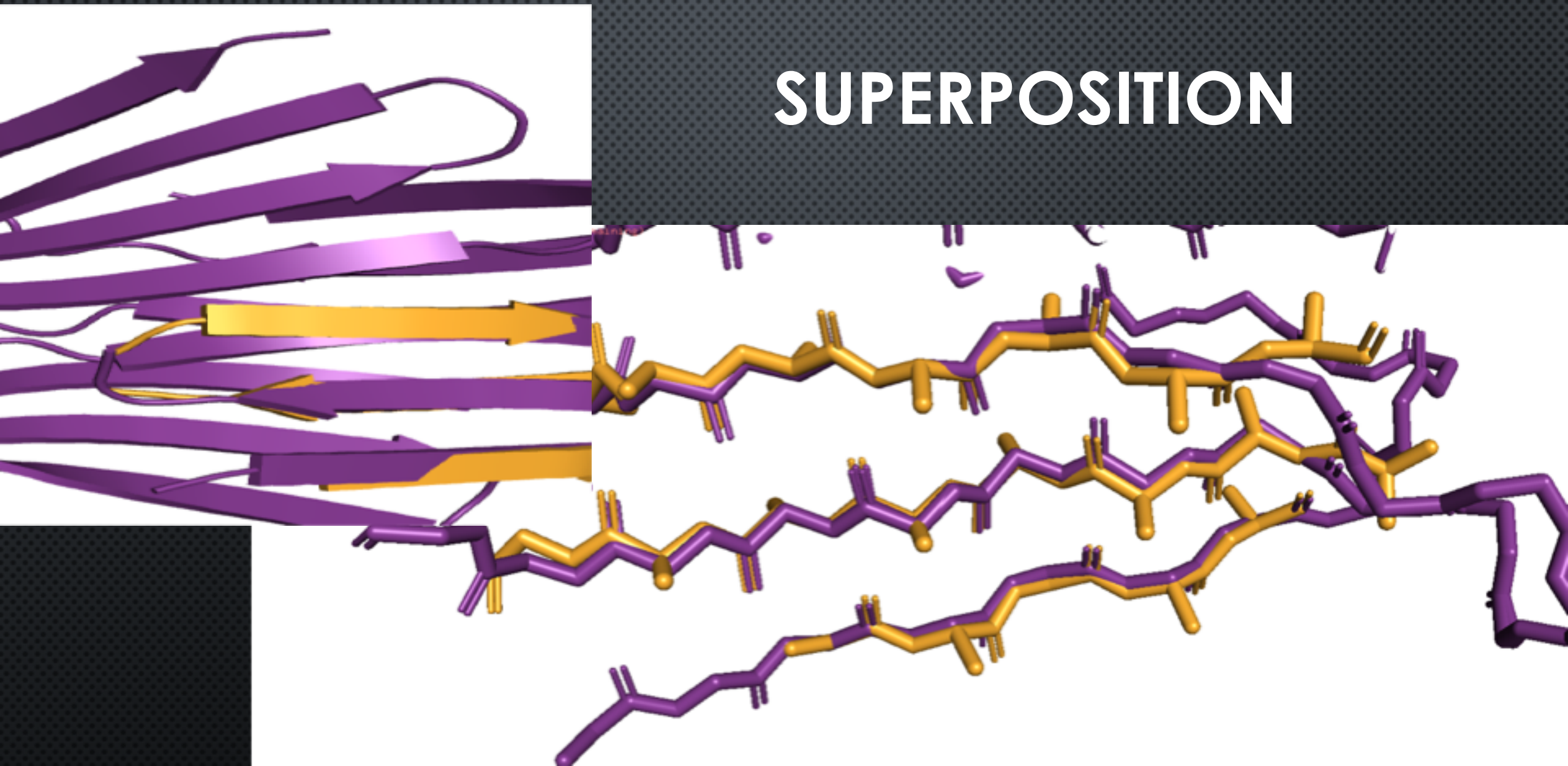
A hierarchical clustering allows to schematically describe relations among subfolds.



PDB ID: 1M99

The meta information retrieved can be used to establish a signature (fingerprint) for any structure that can be stored in databases.

SUPERPOSITION



Conclusions:

- ALEPH is a bioinformatics tool to generate models for MR optimizing user control.
- Annotation mode is meant to classify secondary and tertiary structure with different degrees of abstraction.
- Decomposition mode is useful to discover local groups of compact folds
- Library generation mode creates fragment based models for MR.
- Superposition mode is to identify fragment based models in the final target structure.

Pip, Python 3, PyQt5

Strictness

ARCIMBOLDO_SHREDDER spheres

ARCIMBOLDO_BORGES

Post mortem analysis

Randy Read's group,
Cambridge



AIRLIE MCCOY

COLLABORATORS
CONTRIBUTING
PROBLEM DATA AND
COMPUTING
RESOURCES



Isabel Usón's group,
Barcelona



George Sheldrick's group,
Göttingen

