A novel computational approach for identifying protein relationships

Yael Mandel-Gutfreund

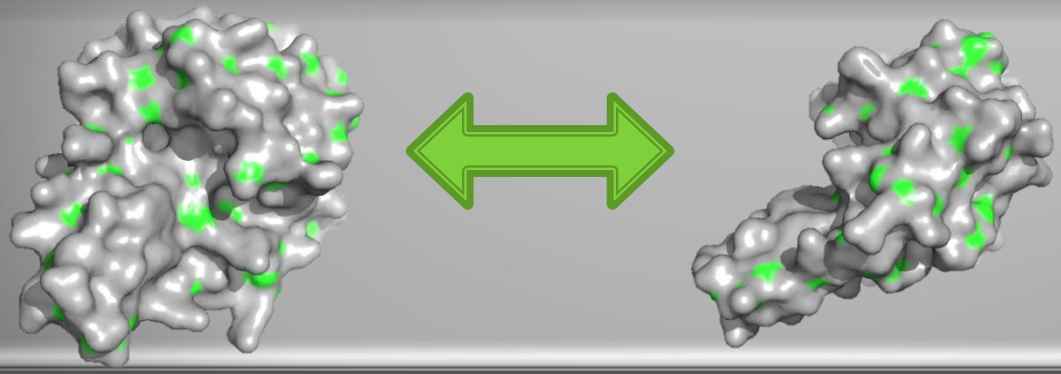
Faculty of Biology and Computer Science Department

Technion

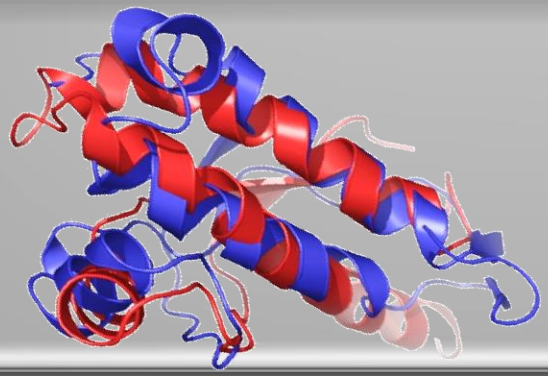
http://yaelab.technion.ac.il

Protein function is commonly inferred by homology

Surface



Structure

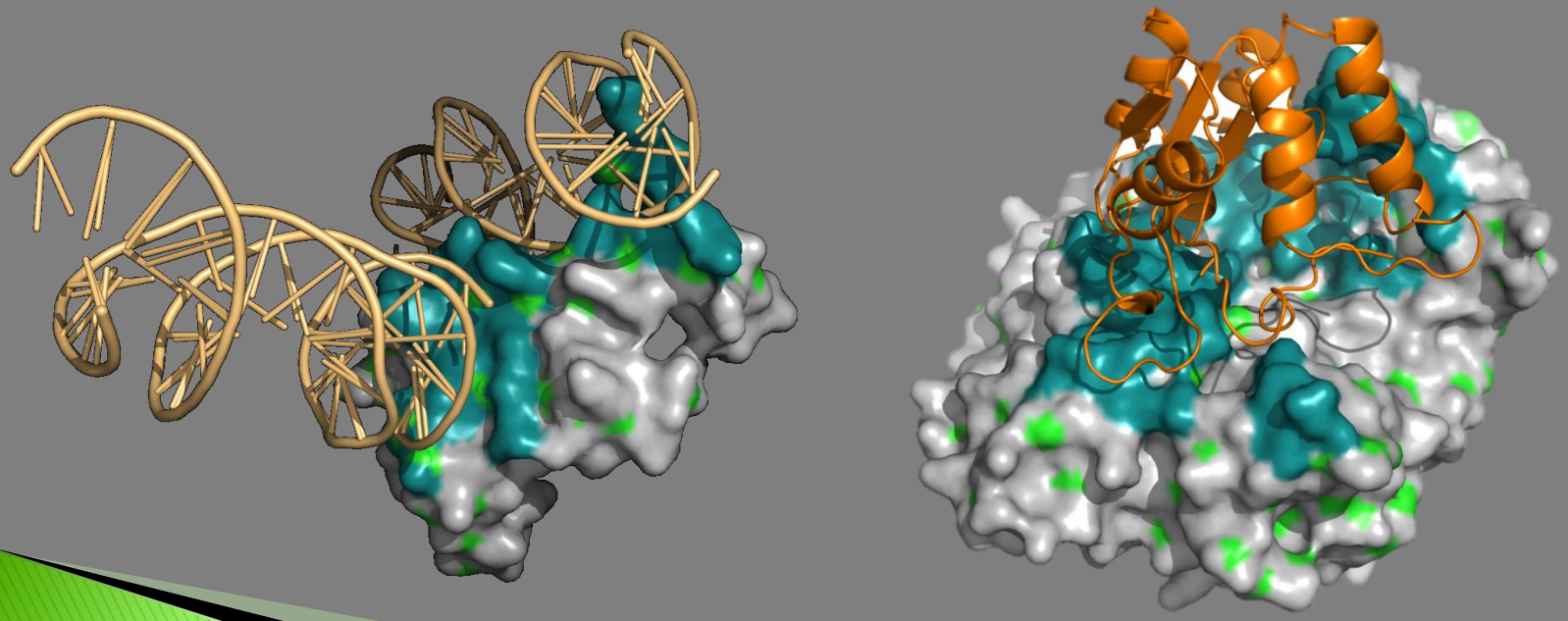


Sequence

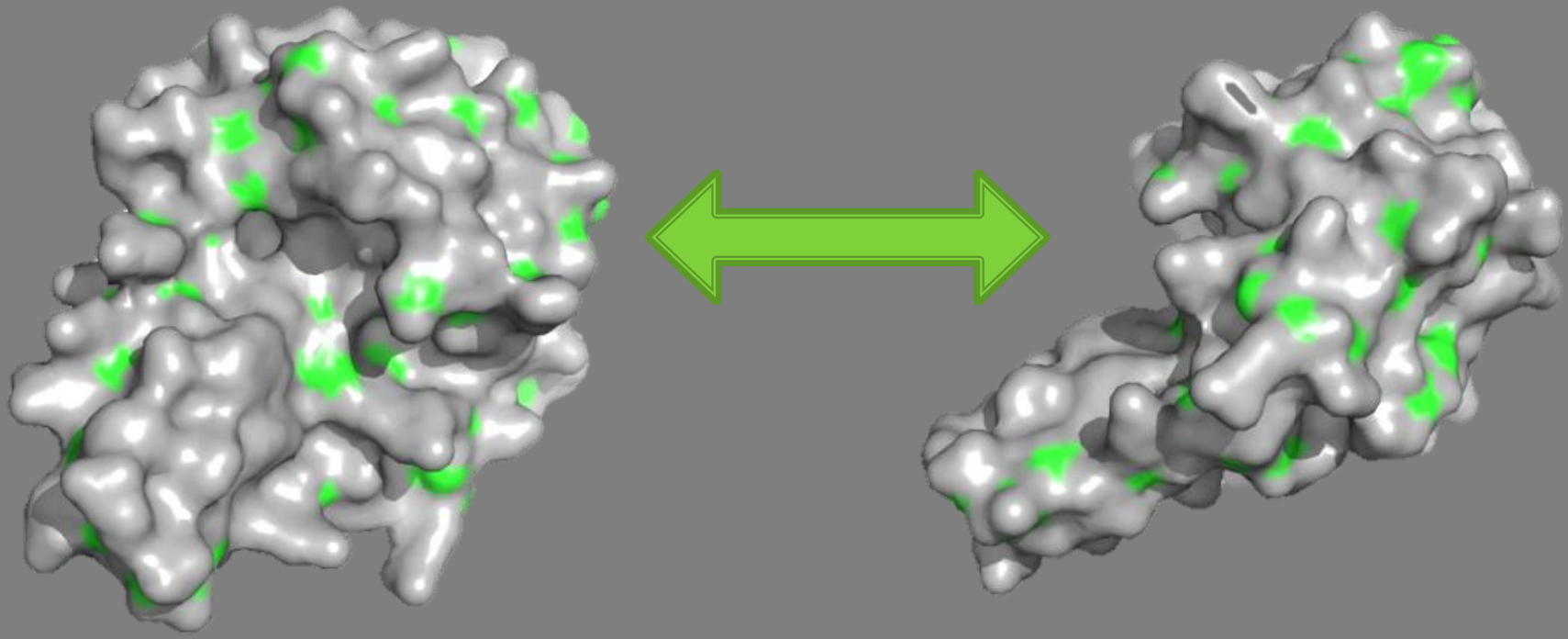
KKILKELEVVSPEEAAHQKAVVE

IKKELKNLSPEEASDQKAVVQKK

The protein surface can provide additional information regarding its interactions with other molecules



A PatchBag approach for comparing protein surfaces

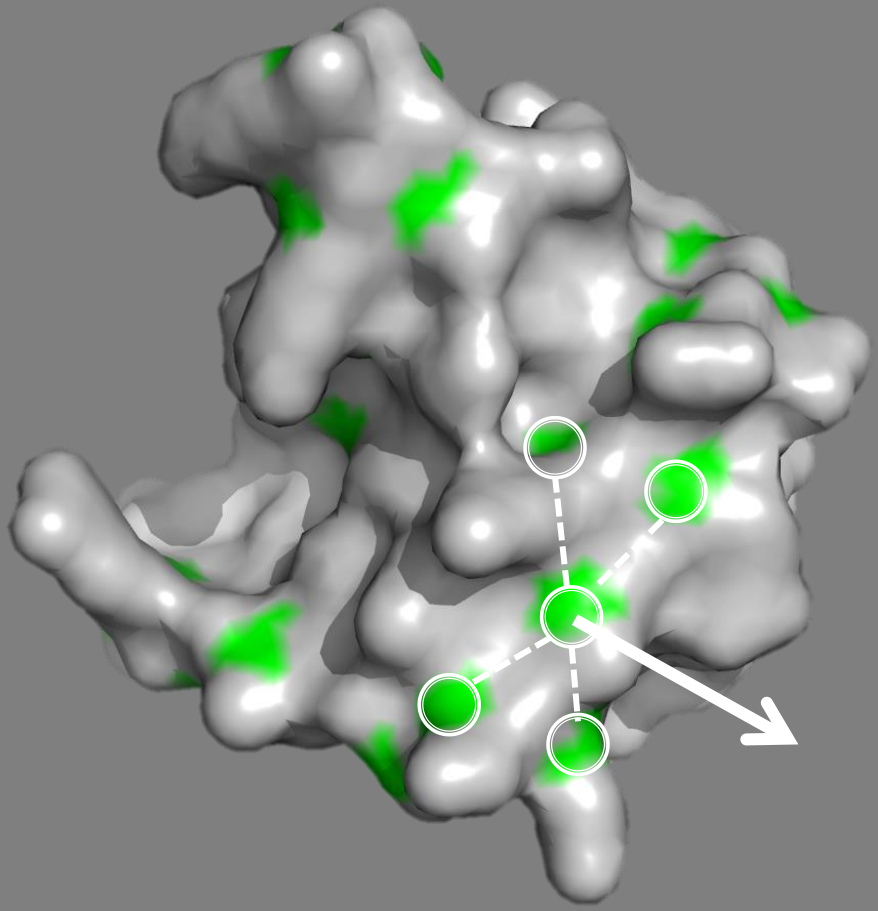


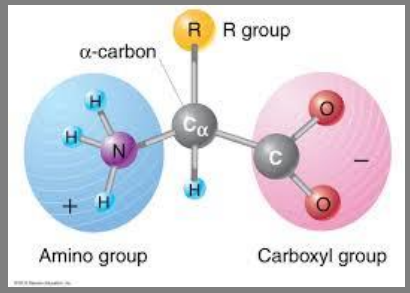
What is a PatchBag approach

PatchBag = Bag of Patches ~ Bag of Words

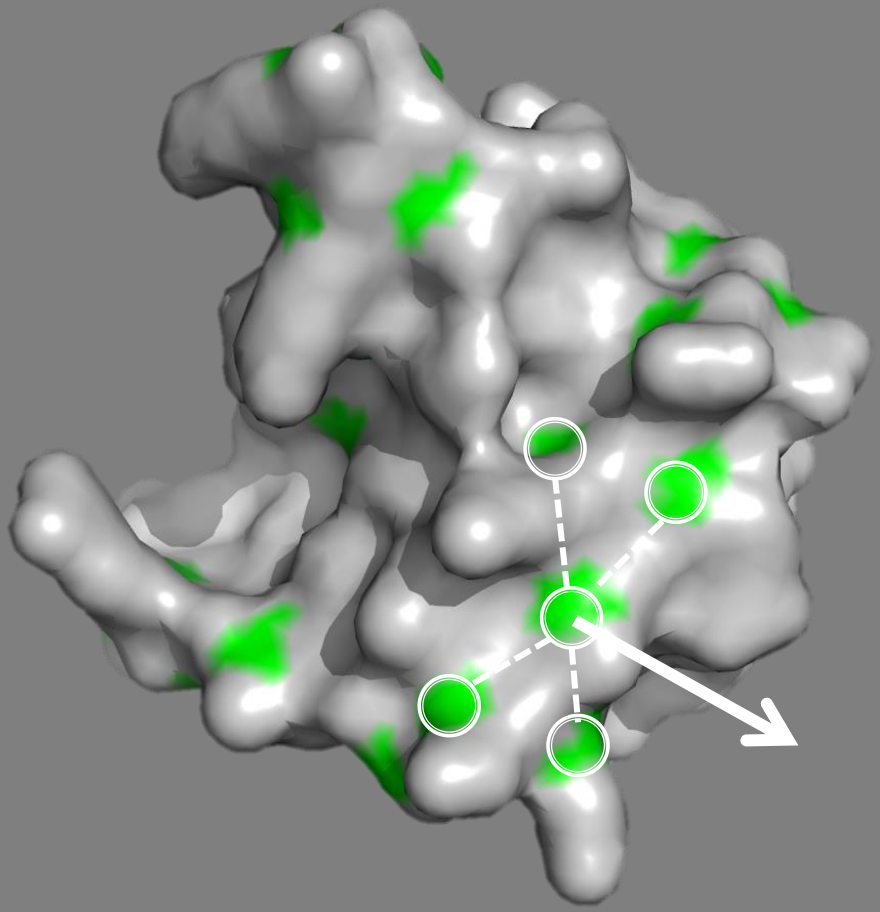
The **bag-of-words model** is a simplifying representation used in natural language processing and information retrieval (IR).
In this model, a text (such as a sentence or a document) is represented as the bag (multiset) of its words, disregarding grammar and even word order but keeping multiplicity.

How do we define words on protein surfaces ?

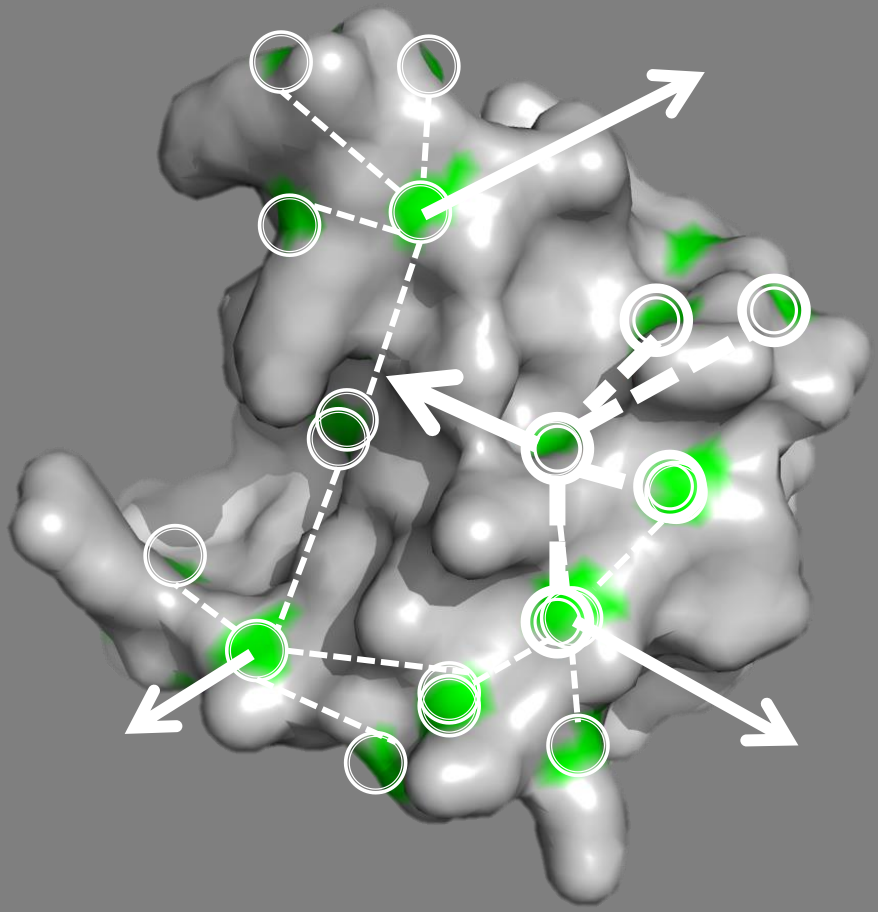




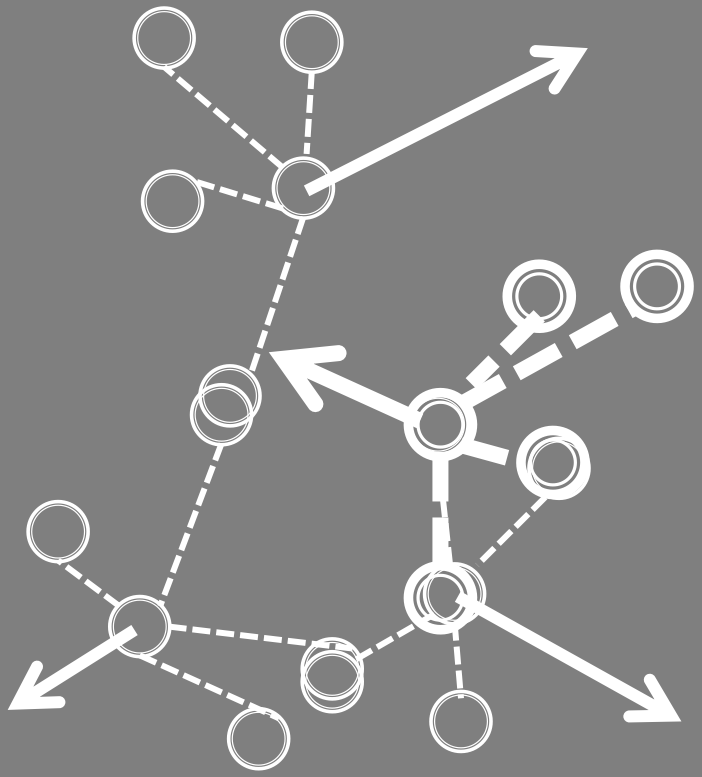
Surface Patch Definition



Surface Patch Definition

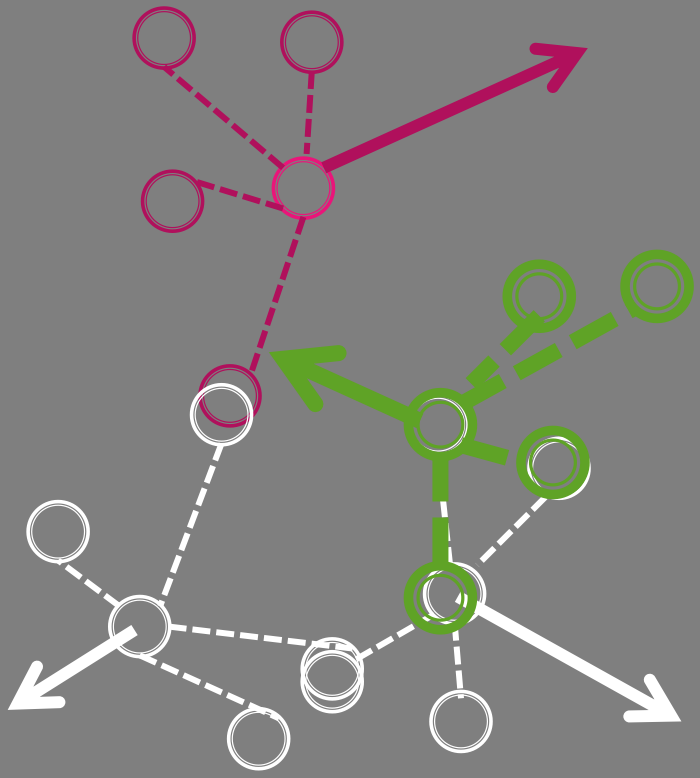


How can we compare Patches

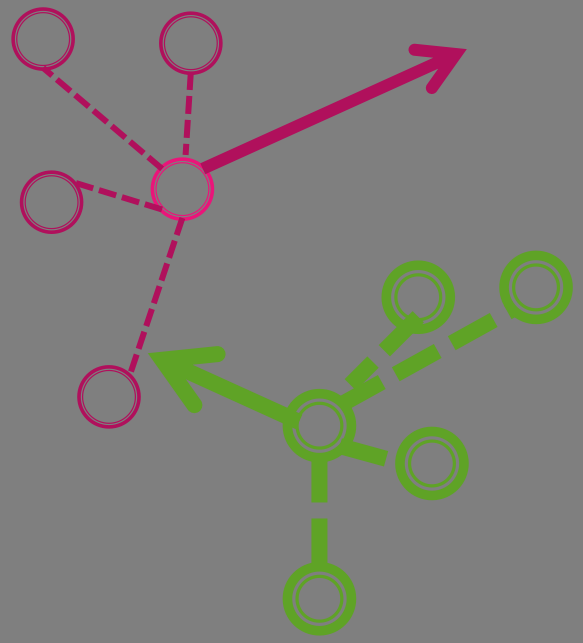


We need a patch dictionary (library) !

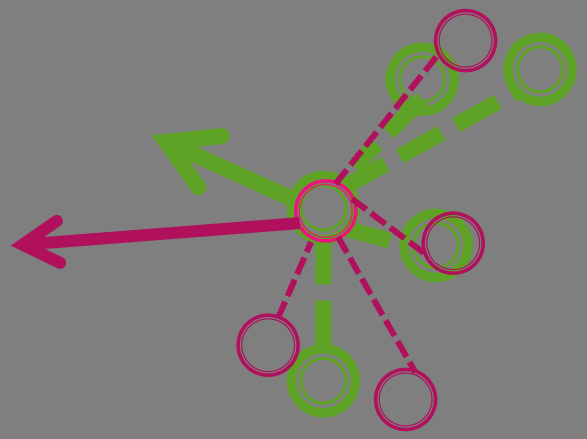
Similarity of 2 Surface Patches



Similarity of 2 Surface Patches

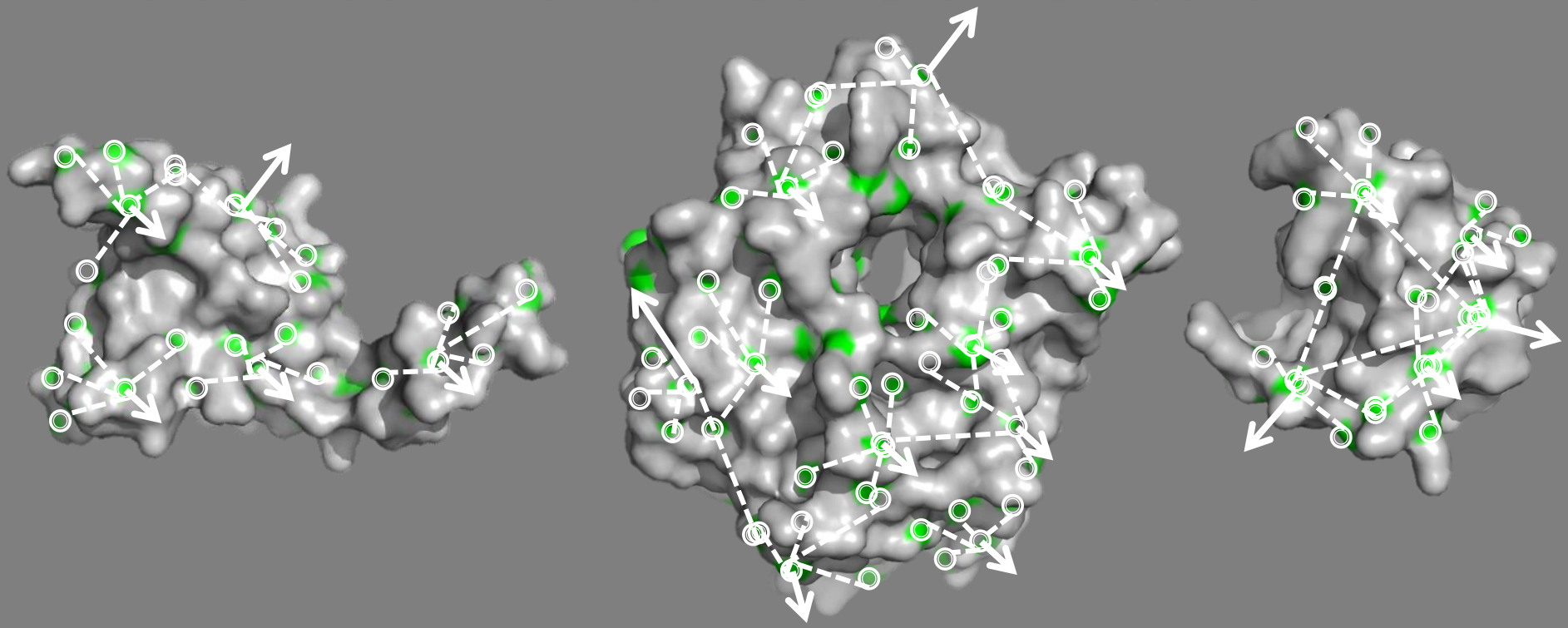


Similarity of 2 Surface Patches

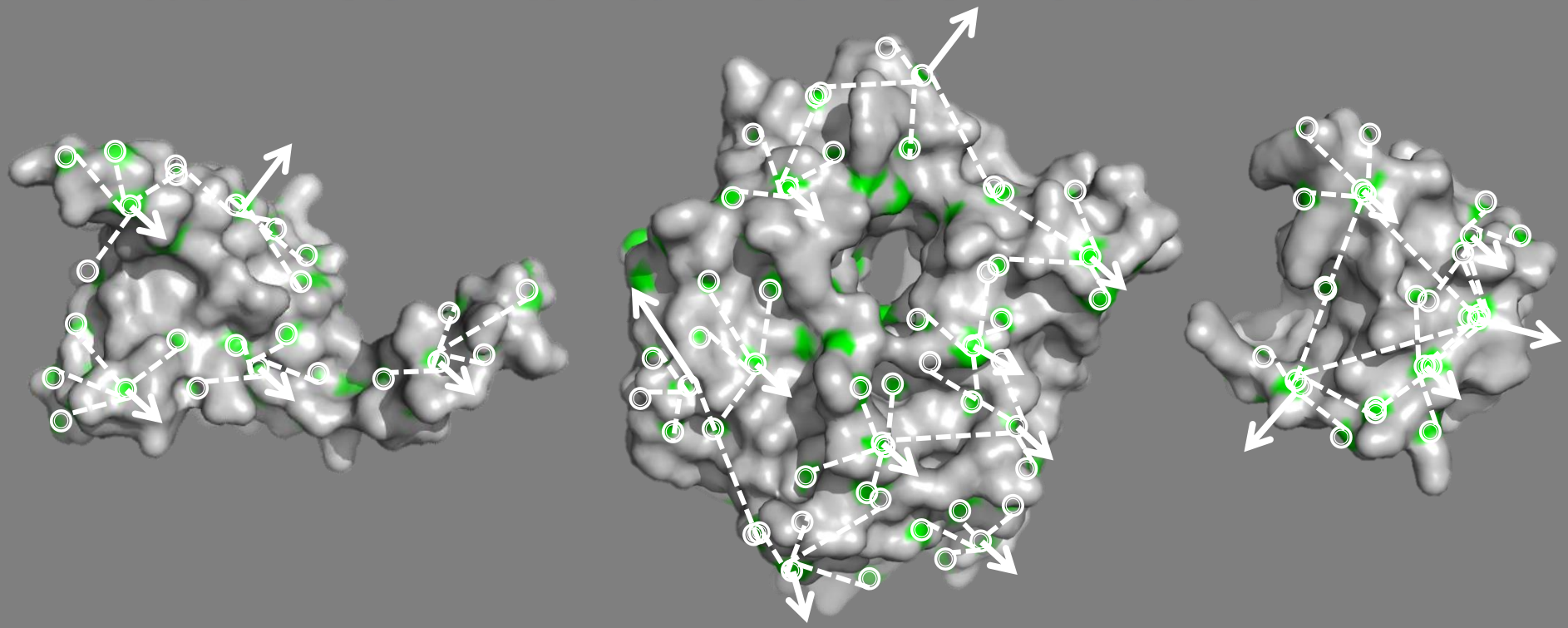


RMSD on the best alignment,
such that normal vectors “agree”

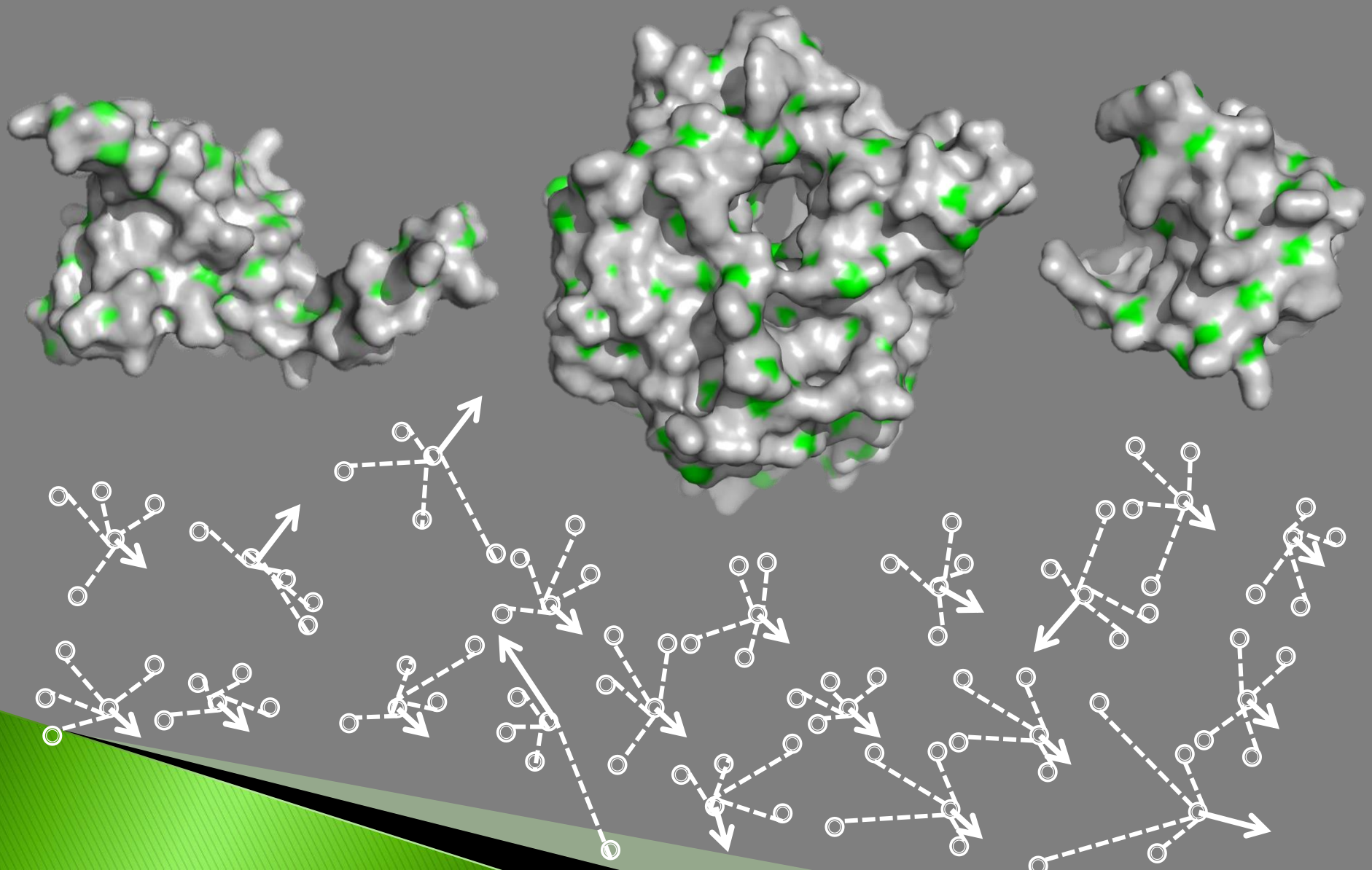
Patches Libraries Generation



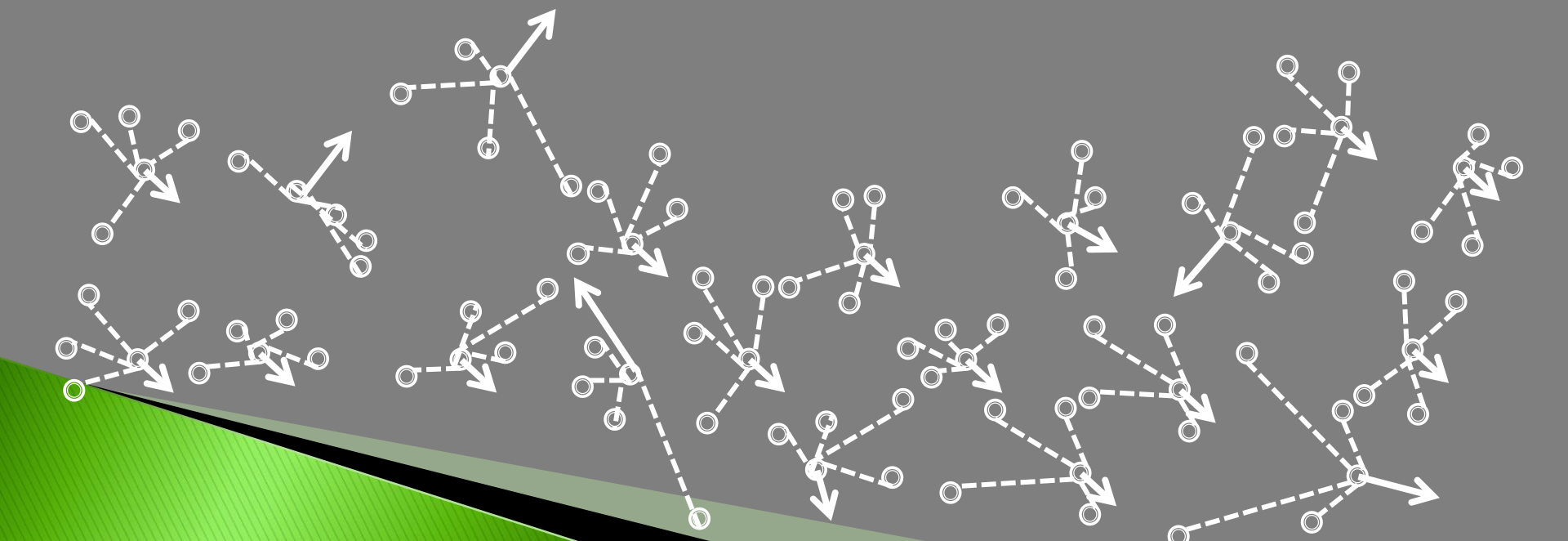
Patches Libraries Generation



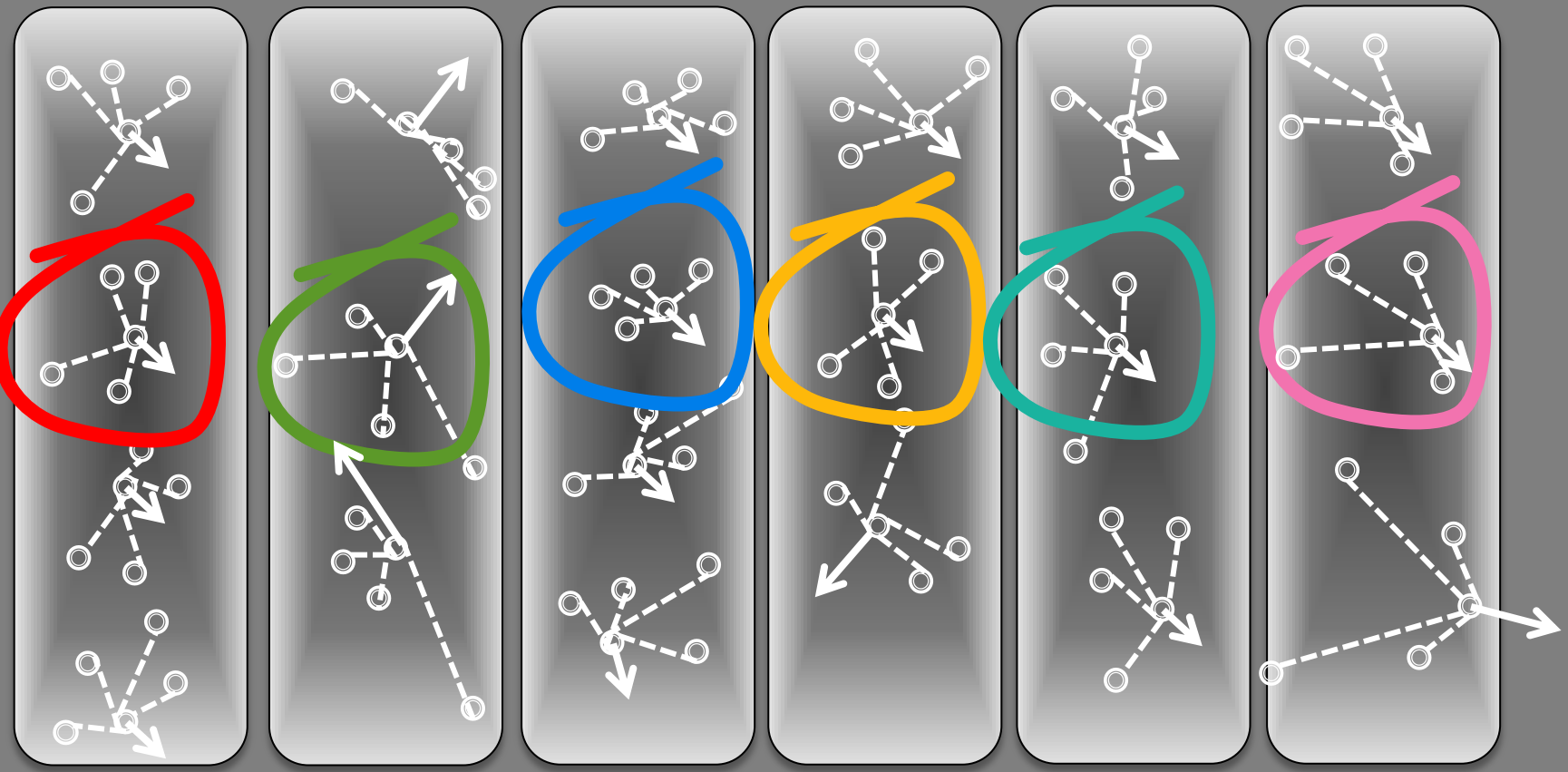
Patches Libraries Generation

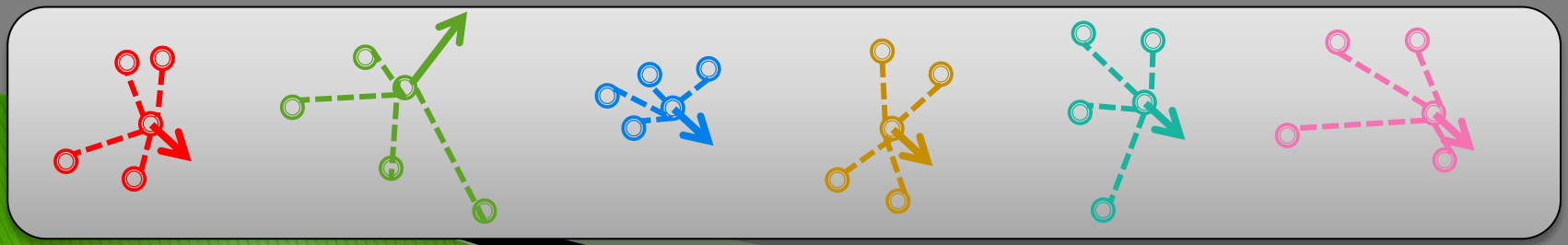


Patches Libraries Generation

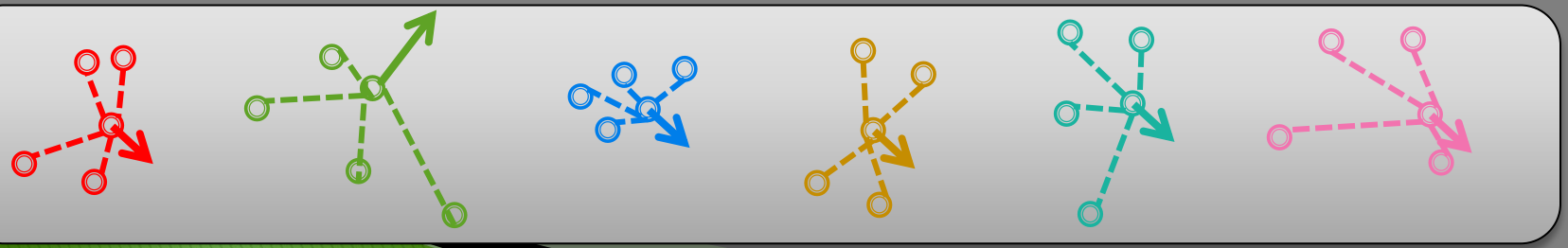


Patches Libraries Generation

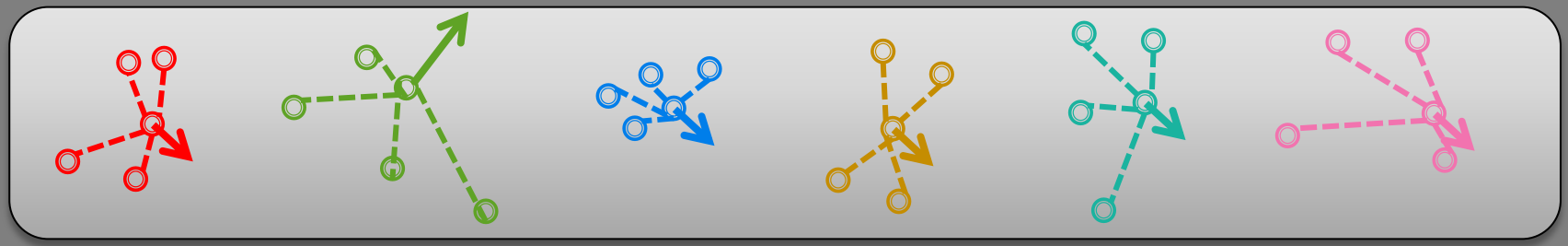


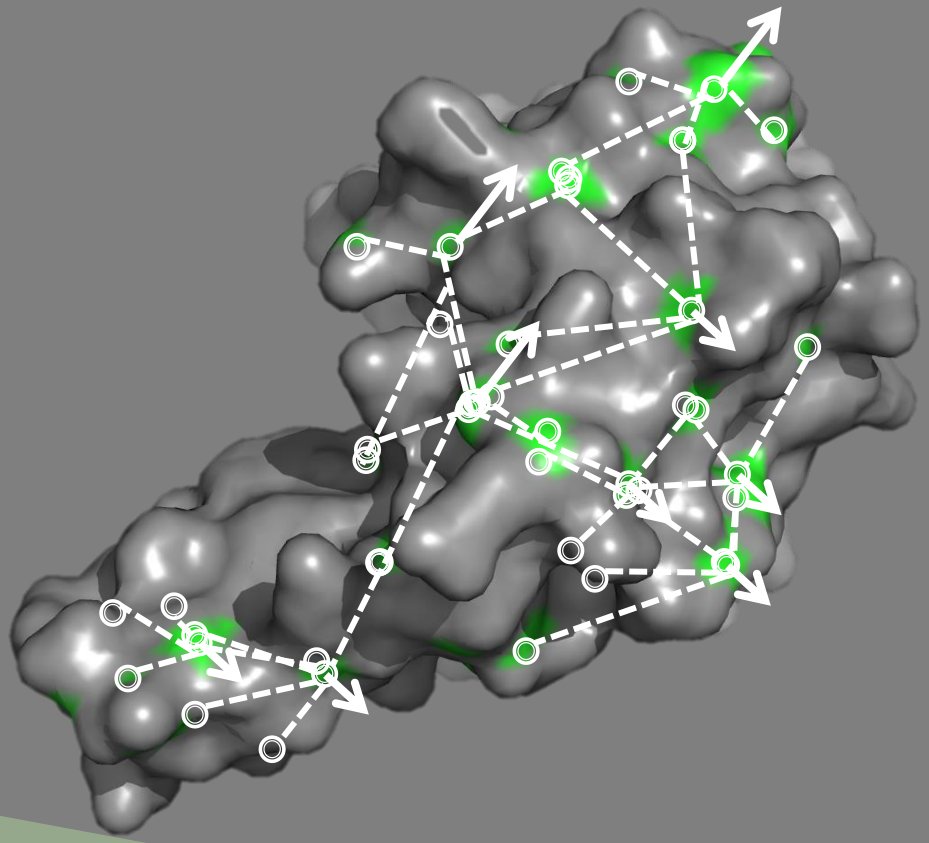


Patches Libraries Generation

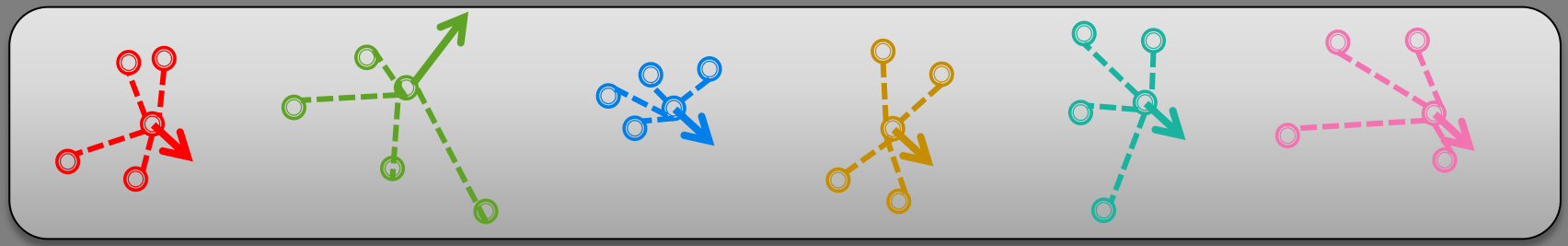


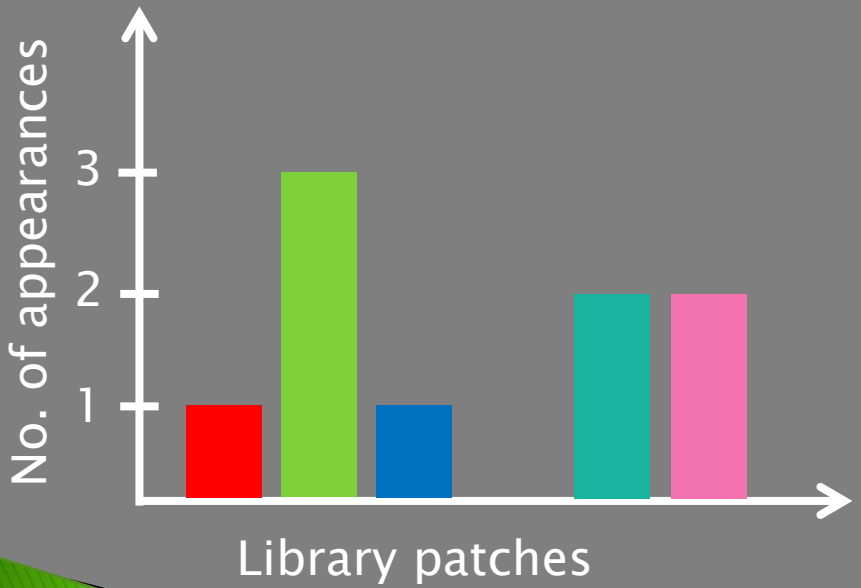
PatchBag

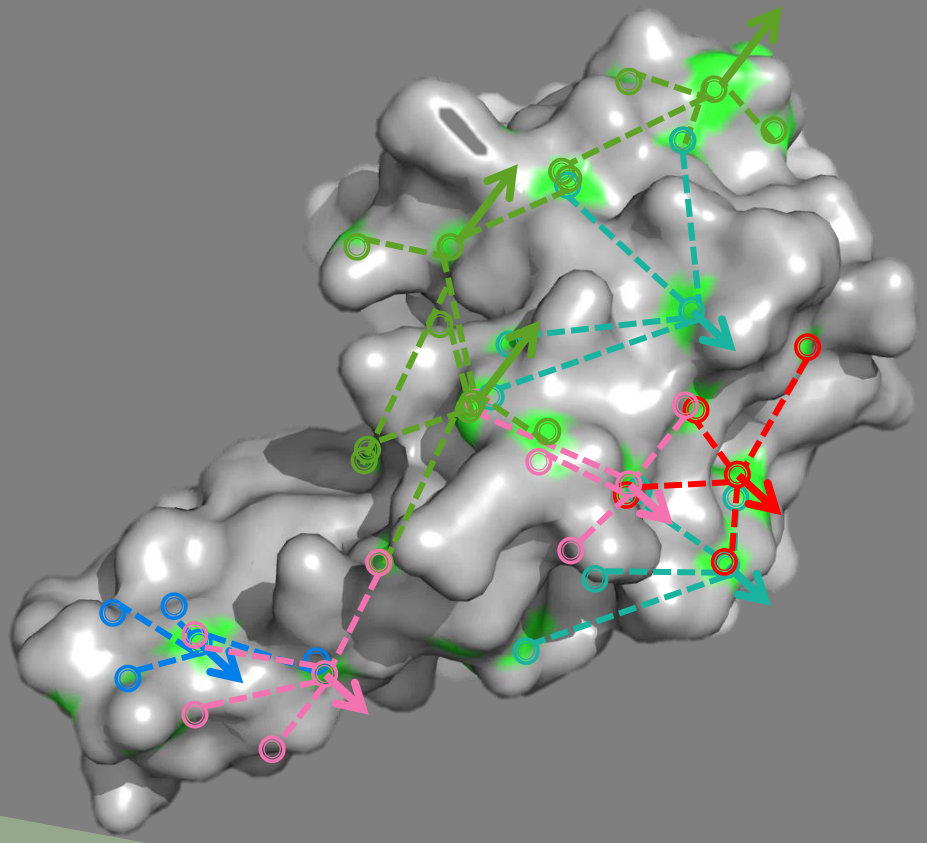




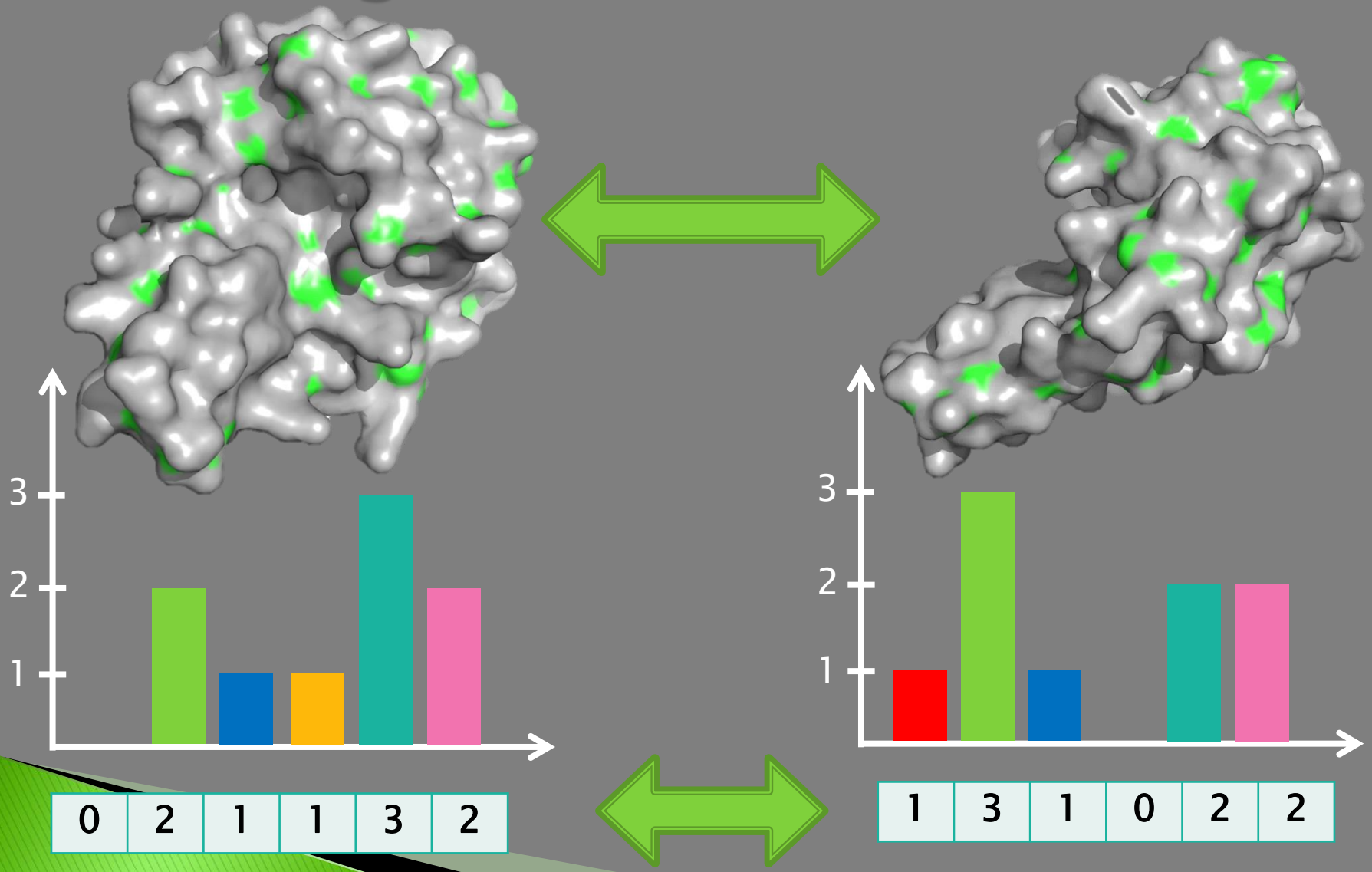
PatchBag



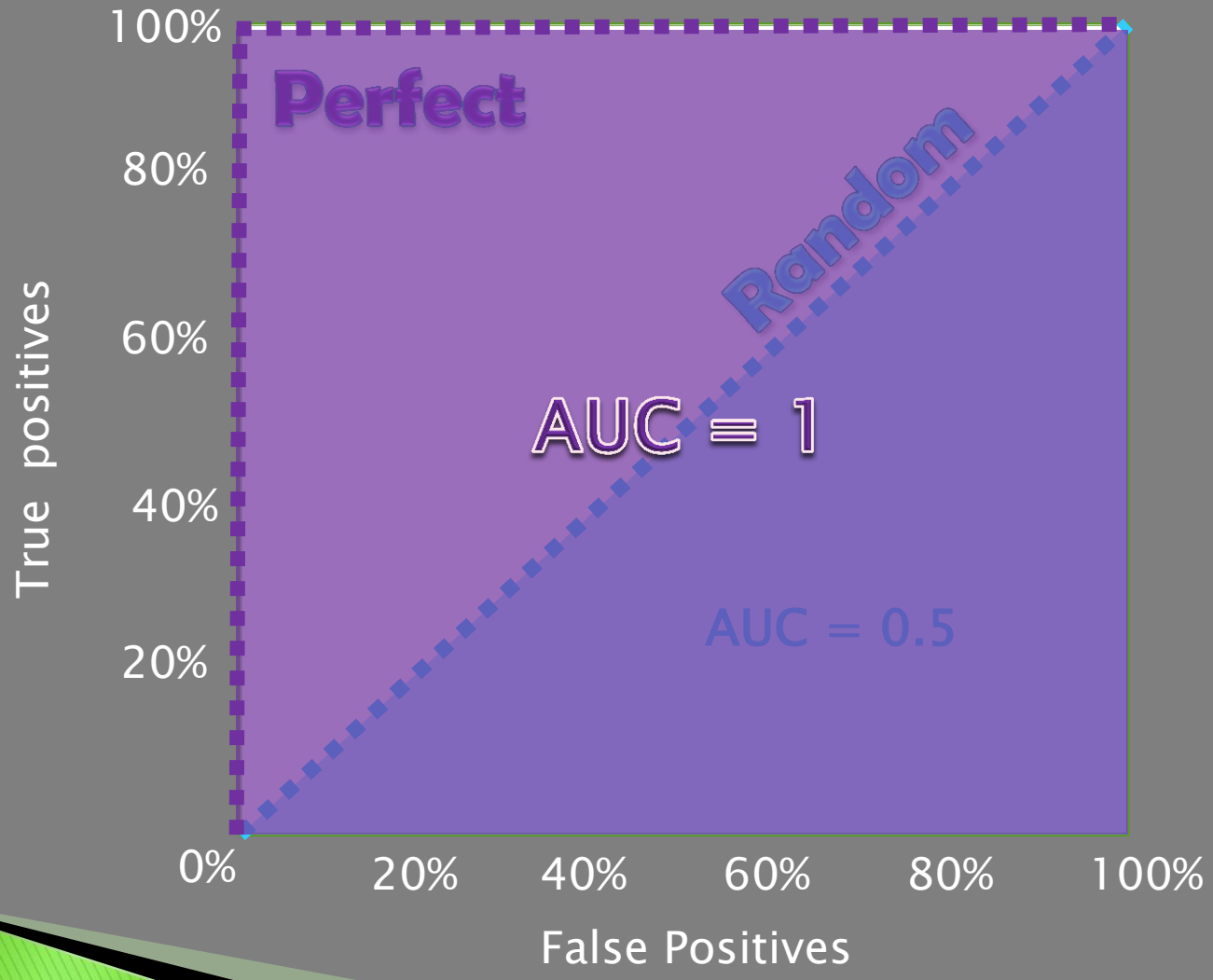




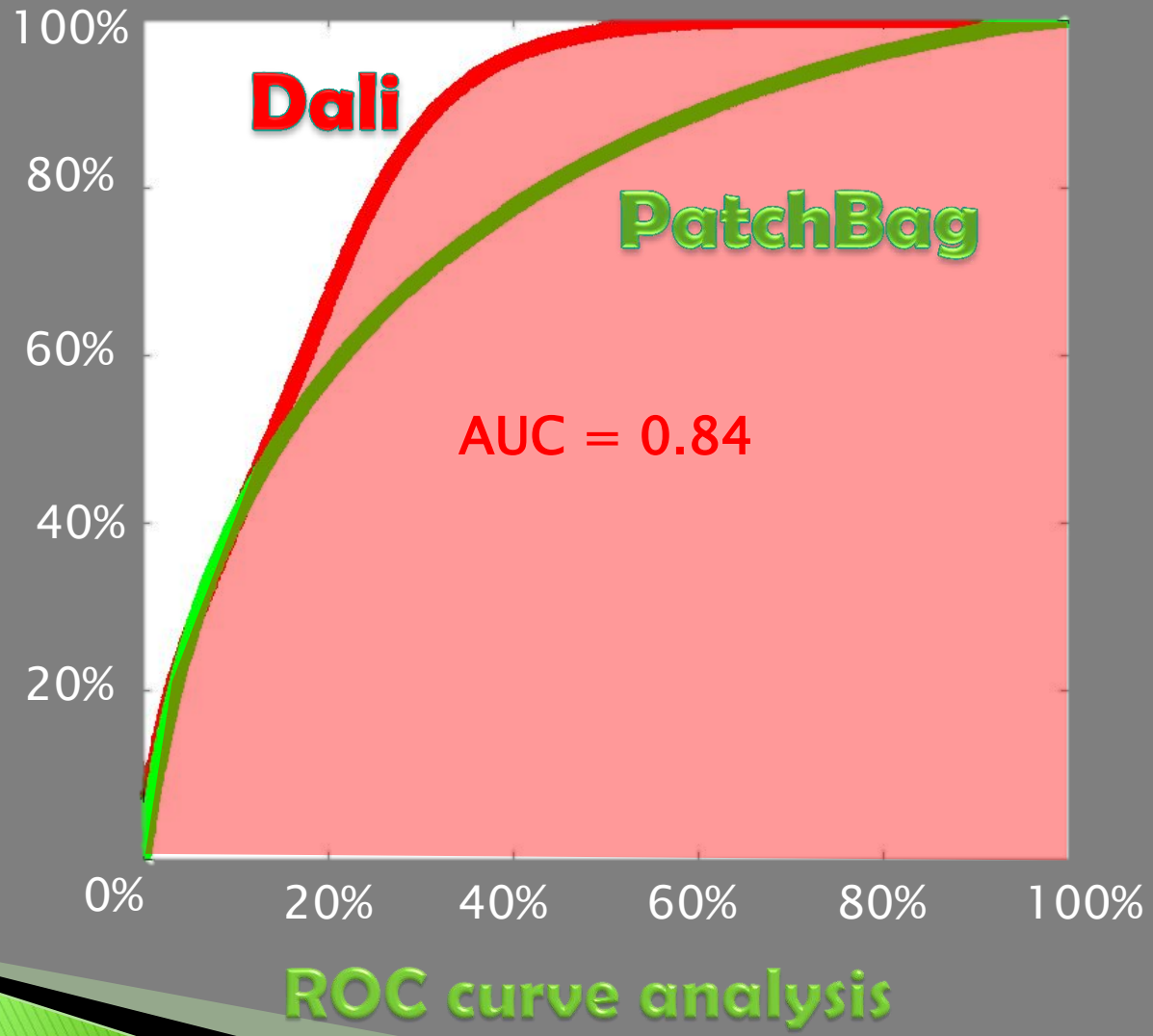
PatchBag



PatchBag detects structural similarities

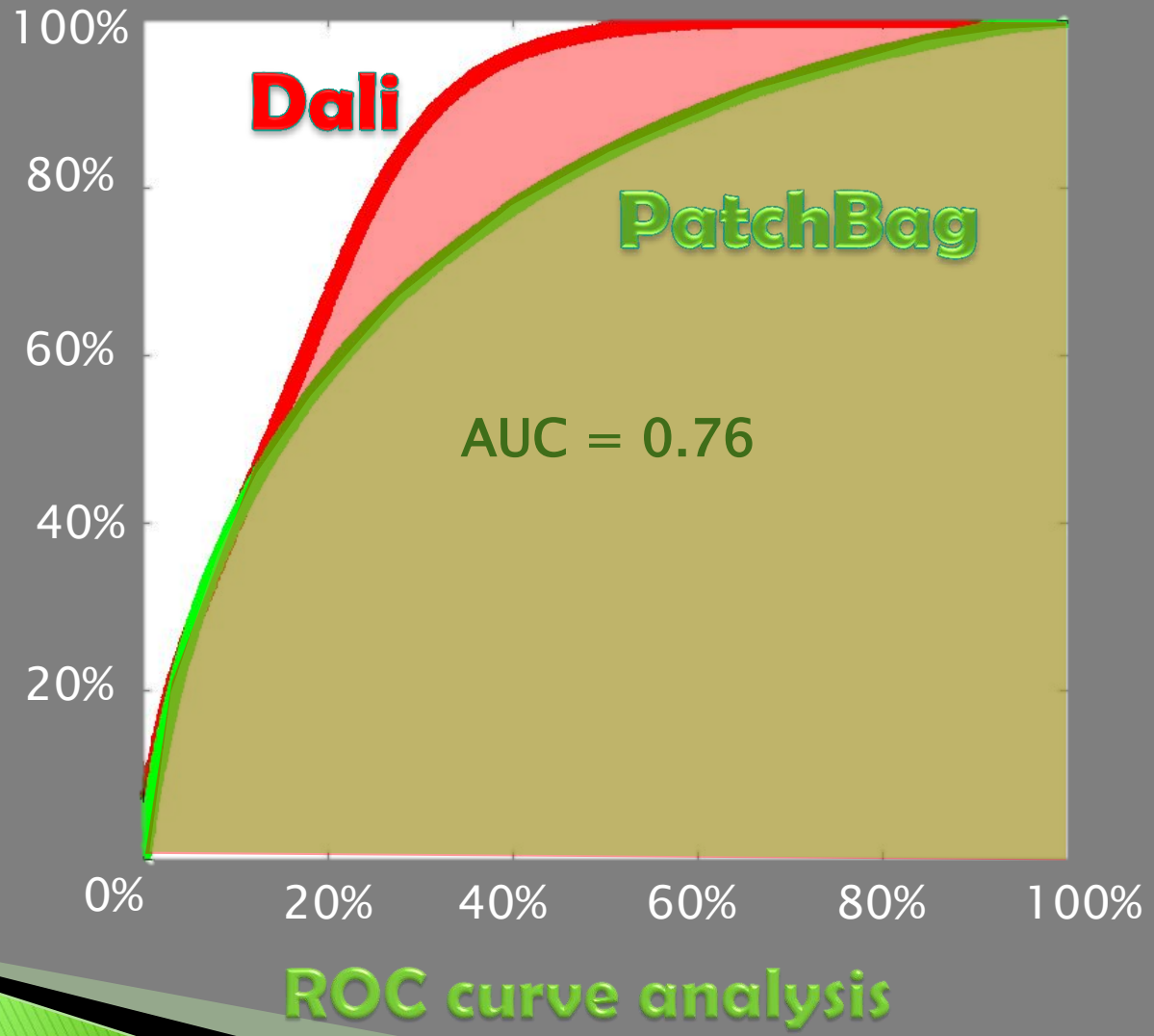


PatchBag detects structural similarities

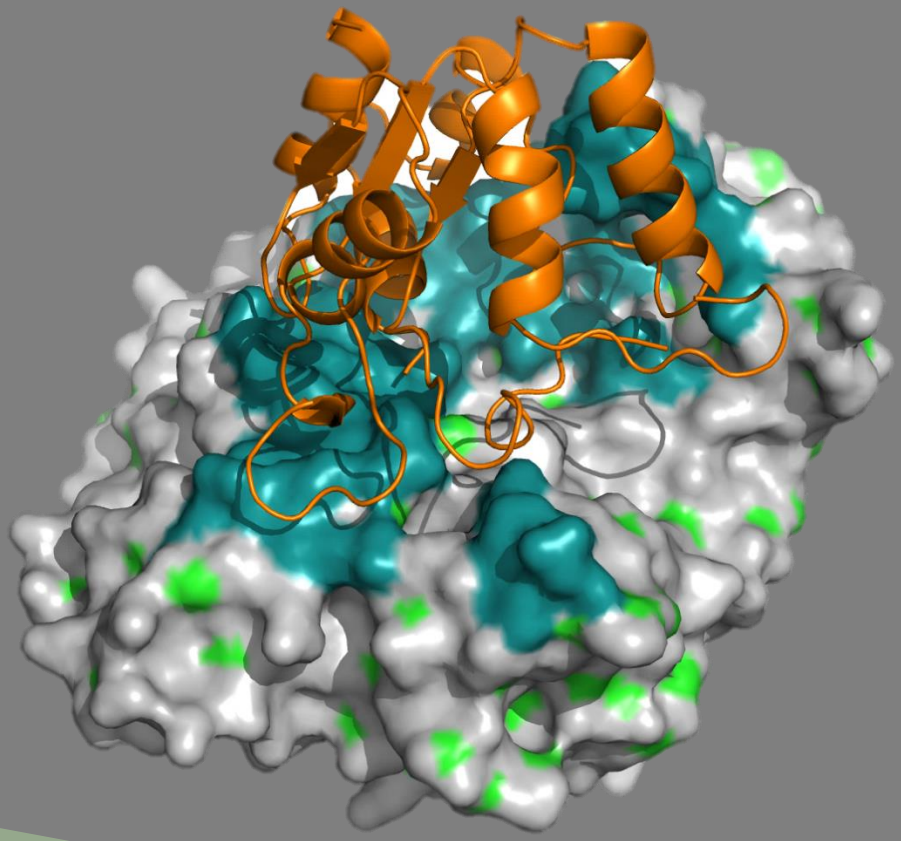


ROC curve analysis

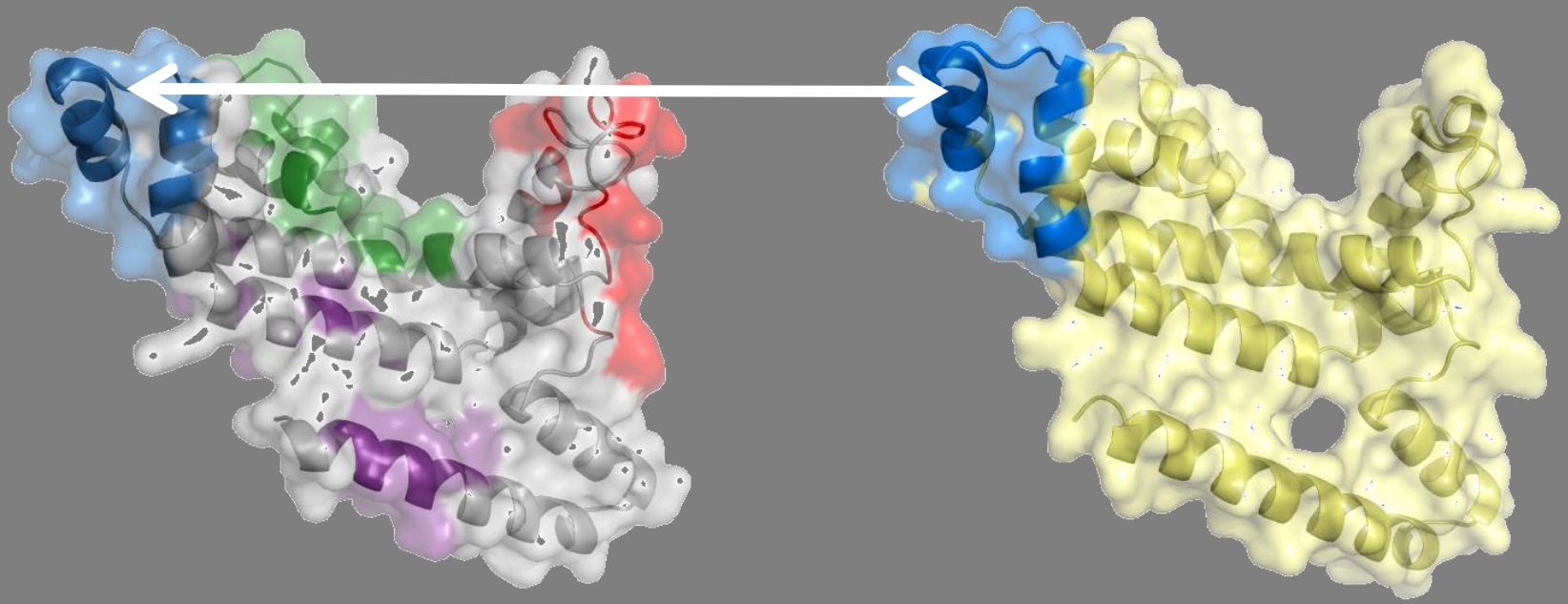
PatchBag detects structural similarities



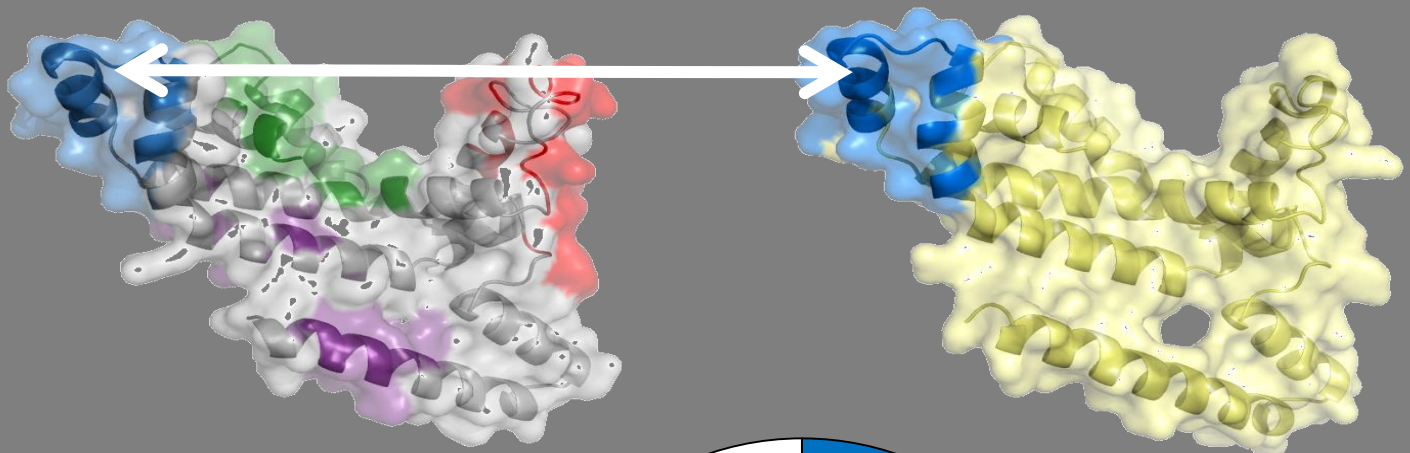
Predicting protein–protein interactions

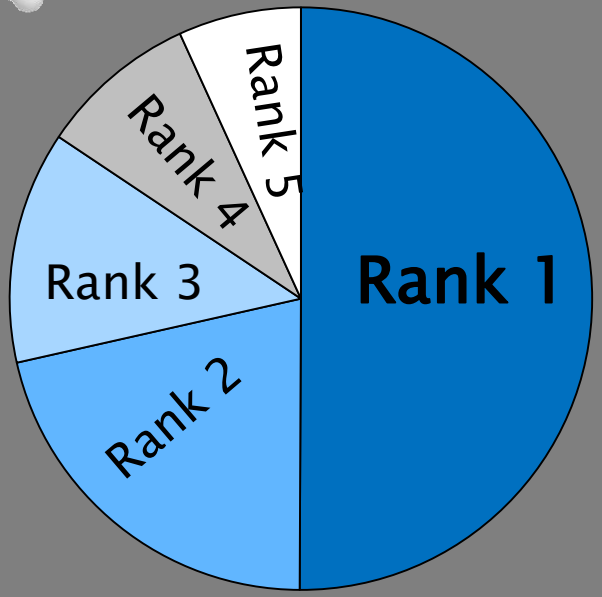


Can PatchBag be employed to detect similarities between sub-surfaces?



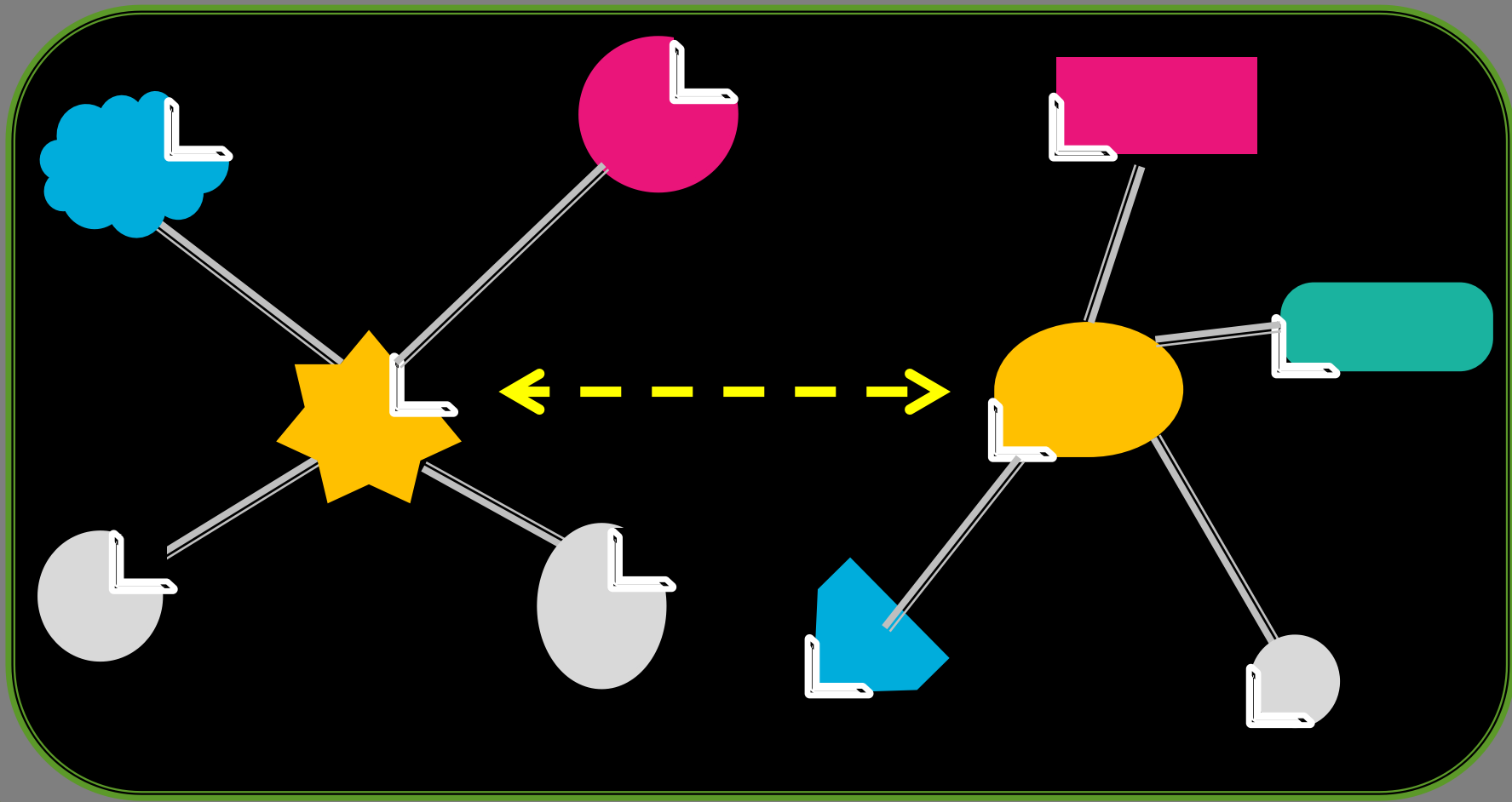
PatchBag can be employed to detect similarities between sub-surfaces





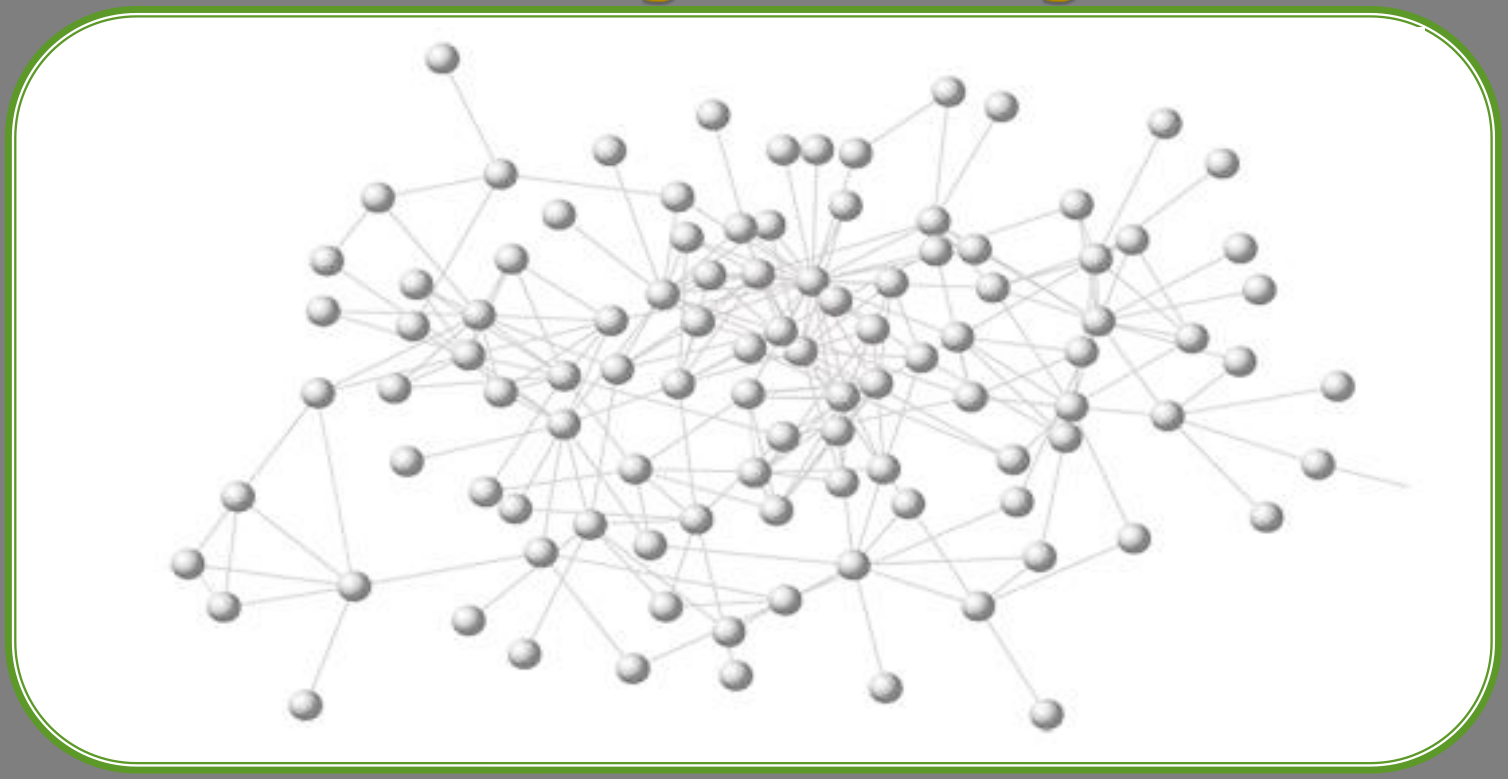
Predicting protein–protein interactions is not an easy task

Guilty by association



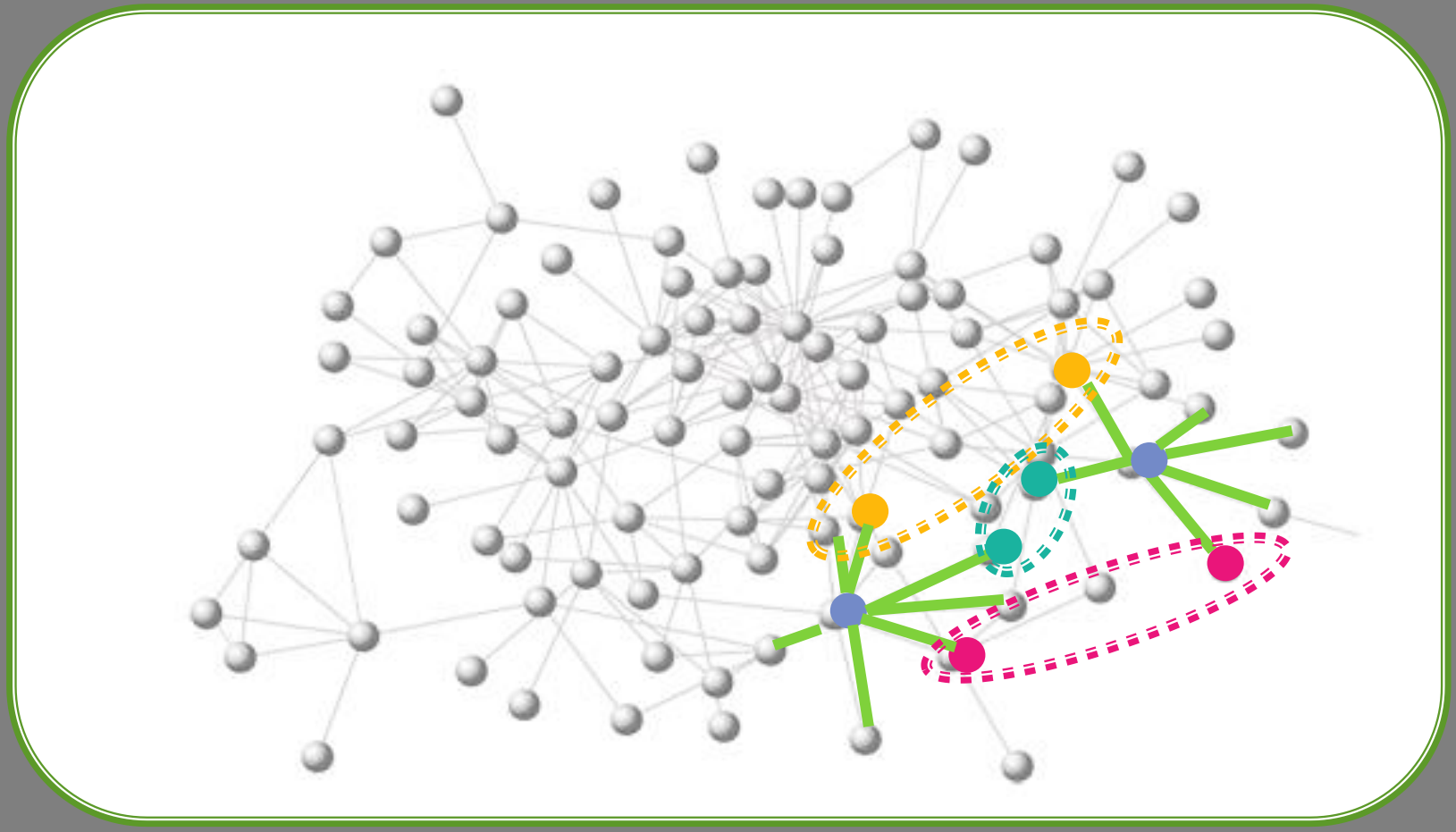
Step 1: Extract a database of known protein-protein interactions

Step 2: Construct a network connecting proteins based on their surface similarity using PatchBag



- Every node in the network is a protein in the 3did database
- An edge in the network represents a significant PatchBag score

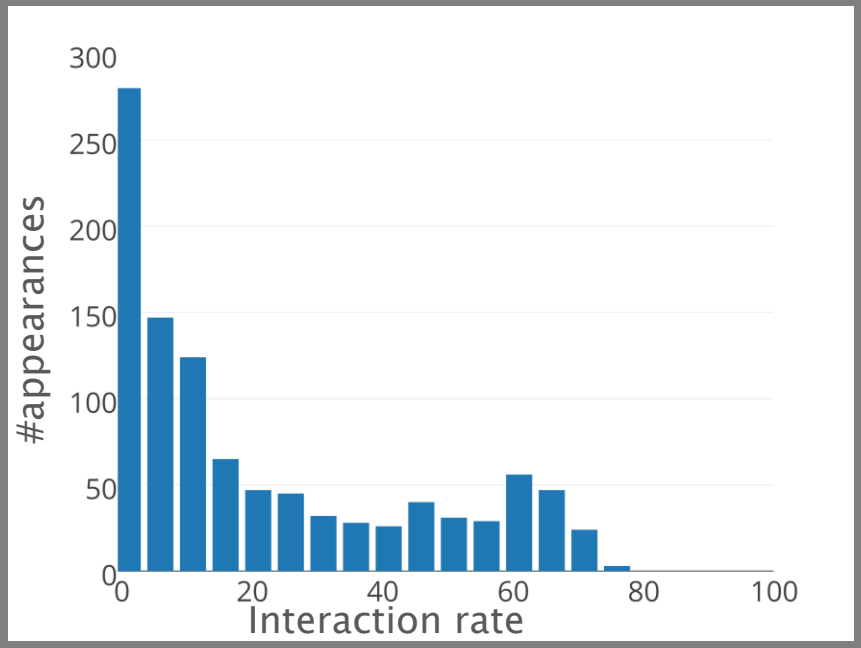
Step 3: Evaluate the percent of known protein pairs among highly connected sub networks



$$\text{Interaction rate} = \frac{\text{\# of known pairs}}{\text{Degree of the sub network}}$$

Interaction rate distribution

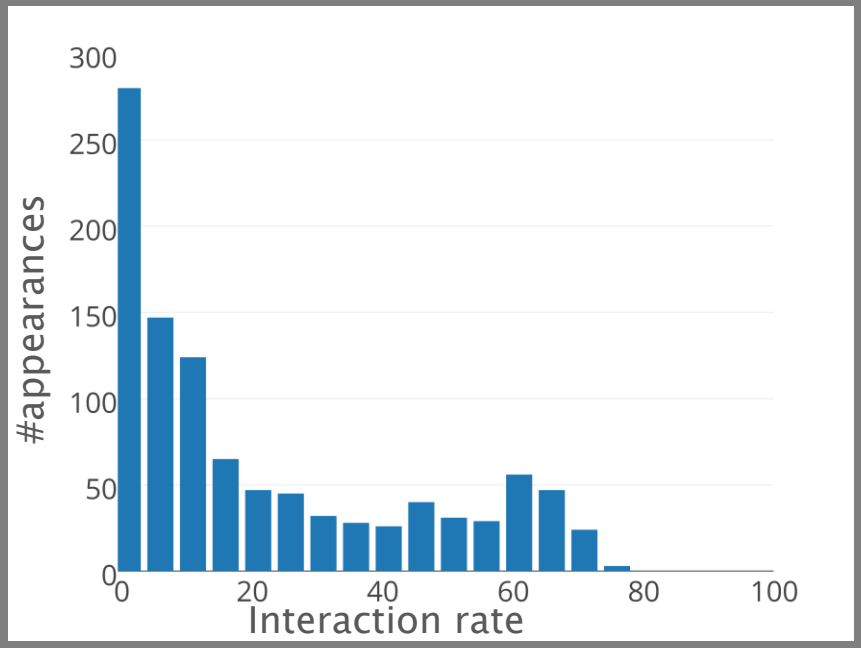
PatchBag on Interface Residues



Mean interaction rate: 35.7

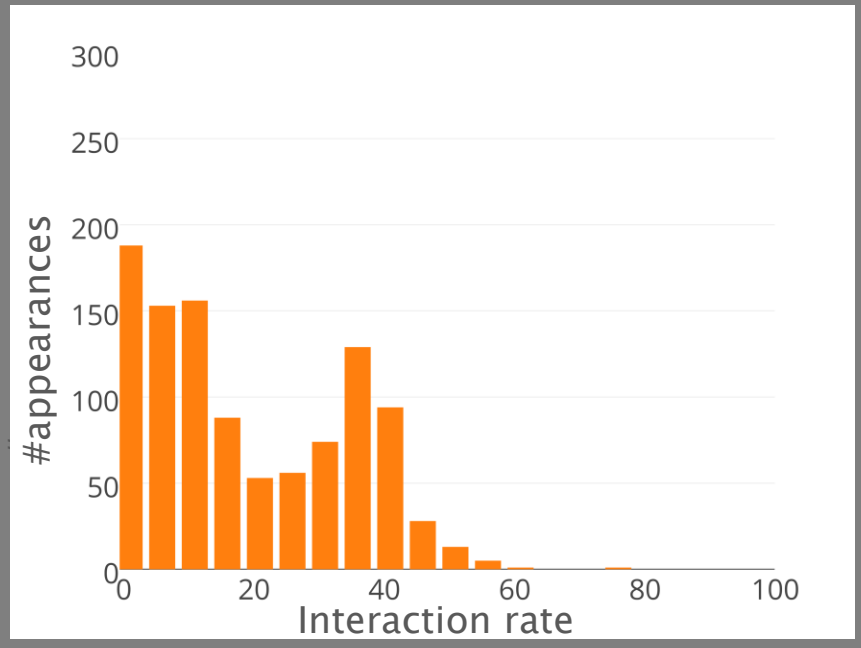
PatchBag network has significantly higher interaction rates relative to random network

PatchBag on Interface Residues



Mean interaction rate: 35.7

Random Network

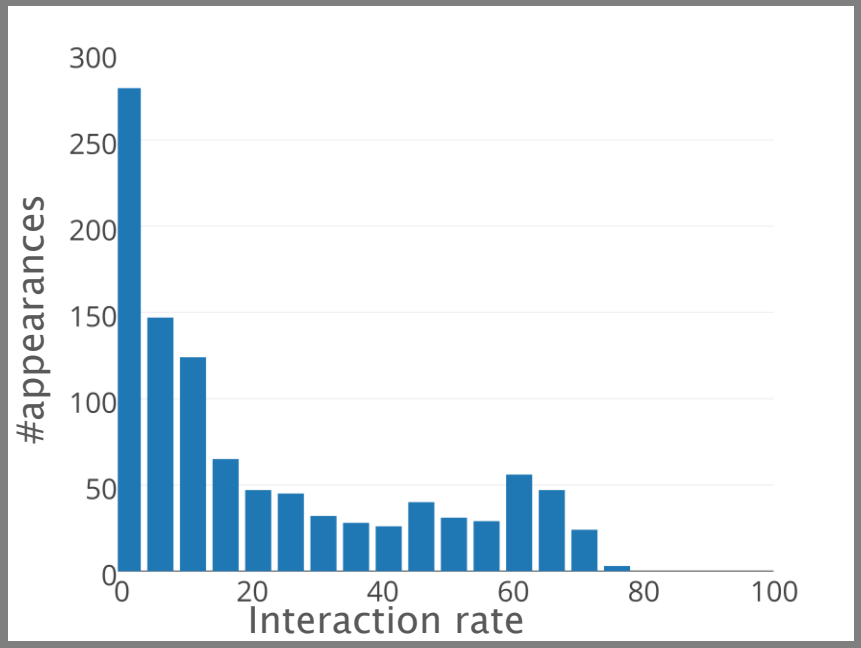


Mean interaction rate: 21.8

Mann-Whitney U test P-value: 2.52×10^{-37}

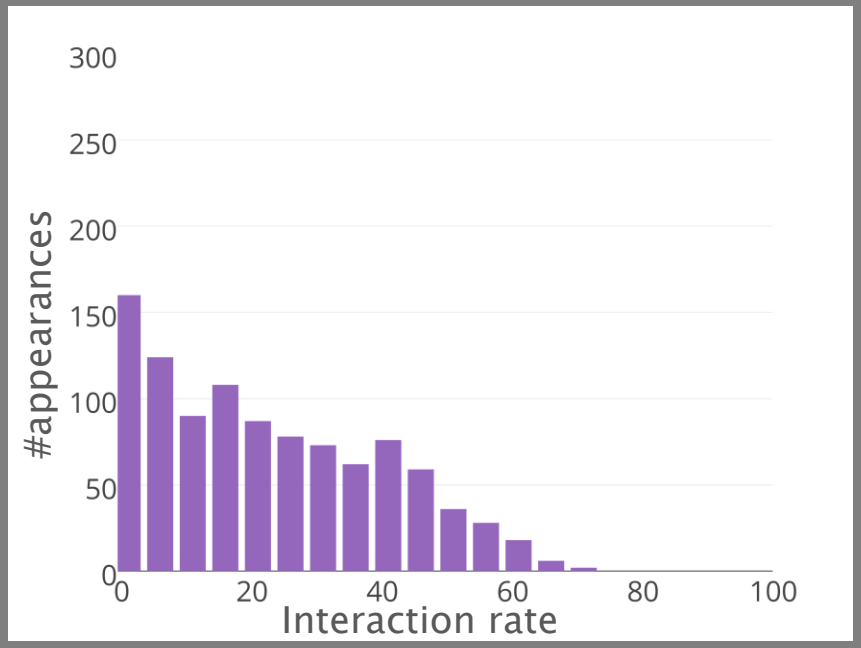
PatchBag network has significantly higher interaction rates relative to scattered network

PatchBag on Interface Residues



Mean interaction rate: 35.7

PatchBag on Scattered Surface Residues



Mean interaction rate: 26.8

Mann-Whitney U test P-value: 1.15×10^{-14}

Take home

- We introduce PatchBag, a fast and accurate method for identifying similar surfaces and interfaces on proteins
- PatchBag can be used for identifying novel molecular interacting surfaces
- PatchBag can be used to predict protein-protein interaction

Thanks to :

Inbal Budowski-Tal

Dr. Rachel Kolodny (Haifa University)

http://yaelab.technion.ac.il