PROTEIN MOTIF RETRIEVAL THROUGH SECONDARY STRUCTURE SPATIAL CO-OCCURRENCES
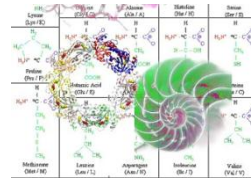
Virginio Cantoni, Department of Computer Engineering and Systems Science, University of Pavia, Pavia, Italy, virginio.cantoni@unipv.it

Alessio Ferone, Alfredo Petrosino, Department of Applied Science, University of Naples Parthenope, Naples, Italy, {alfredo.petrosino;alessio.ferone}@uniparthenope.it
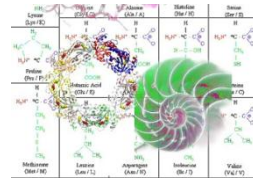
Protein Secondary Structure

- Structural building blocks
- Motifs, domains, fold…
  - common material
  - used by nature
  - to generate new sequences
- Many methods for
  - Identification
 - classification

Protein Secondary Structure
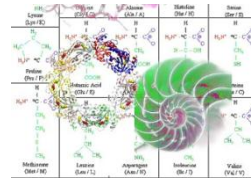
- ## A structural motif is
  - a three-dimensional structural element
  - which appears in a variety of molecules
  - usually consists of just a few elements
- ## Several motifs packed together to form
  - compact
  - local
  - semi-independent units
 - called domains

Protein Secondary Structure

- ## DSSP
  - Most used method for defining protein secondary structure

- ## Eight types of SS
  - Helix: 3_{10}-helix, α-helix and π-helix
  - Sheets or strands: extended strand (in parallel and/or anti-parallel b-sheet conformation)
  - Coils:hydrogen bonded turn, bend and unclassified residues
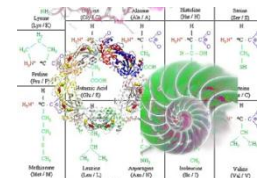
Generalized Hough Transform
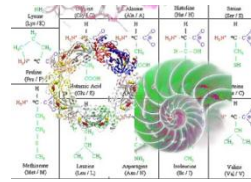
- Using the G-Hough for the comparison and the search for structural similarity between a given protein and the proteins of a data-base

- Search of a structural motif or a domain
  - detection and the statistical distribution of these components
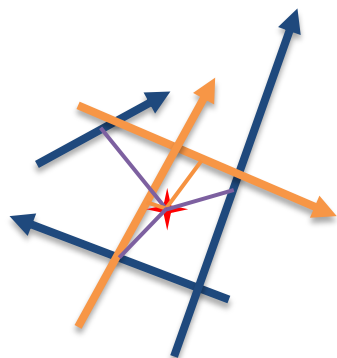
Generalized Hough Transform
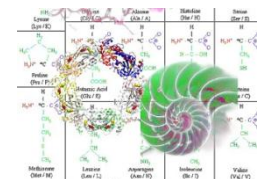
- Extract proteins similar to a given one
- Every element (e.g. a-helix or b-strand) is superposed through a rigid motion with each of the elements on the model.
- For each possible correspondence a vote is given to a particular candidate position of the model
- Every detail on the examined protein votes for a possible presence of the searched model
- Accumulation of all the contributions of all the secondary components of an unknown molecule
- If a particular attendance of the model obtains a sufficient number of contributions the similarity is detected
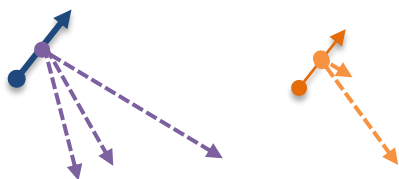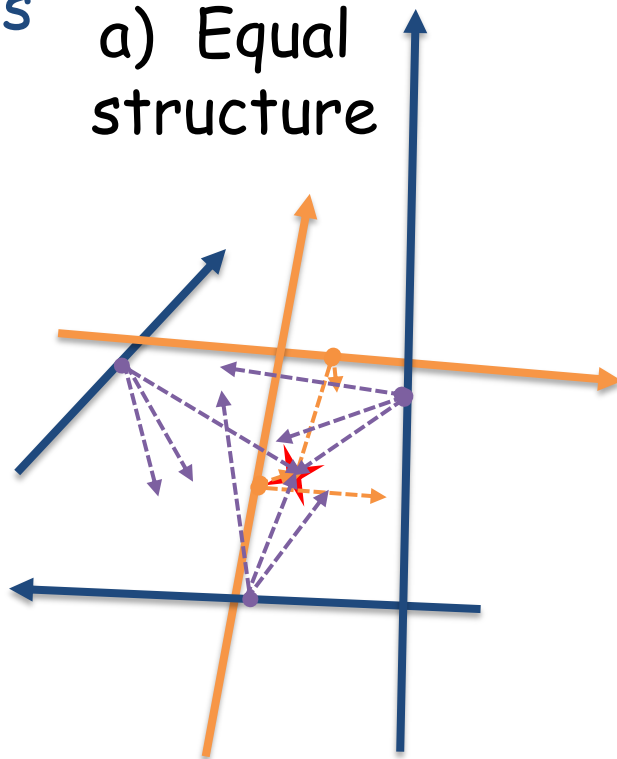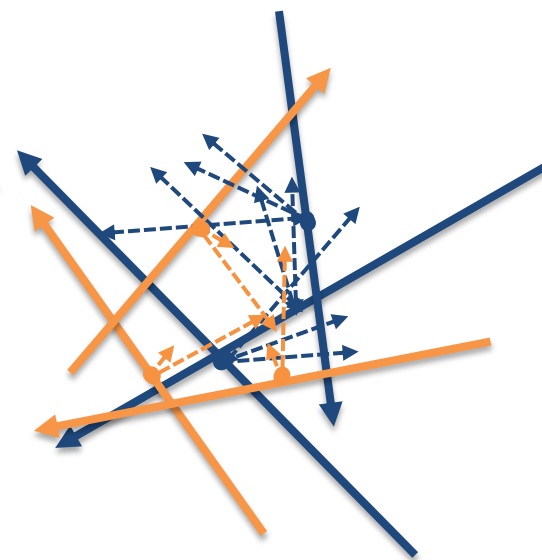
Generalized Hough Transform

Helices and Strands

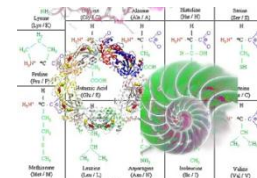a) Equal structure

b) Different structure

Query protein (scaled 0.5)

Mapping Rule

Votes Space

Votes Space

Generalized Hough Transform

- Instead of consider each SS isolated we can base our analysis on the co-occurrences of multiple SSs.
- Even with just two SSs the mapping rule is in general reduced to just one compatible location of the RP.
- Two SSs are characterized by a displacement defined by three parameters
  - axes angle θ
  - Midpoint Distance MD
  - Axes Distance AD
- Multiple location mappings are possible if there are couples having equal parameter terns
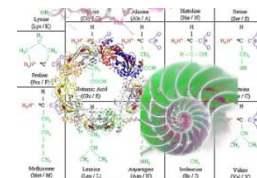
Generalized Hough Transform

a) helices

Reference Point

Axis angle θ

B

Midpoint distance

Axis distance

A
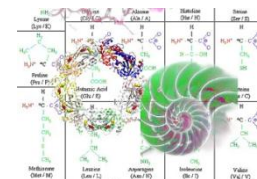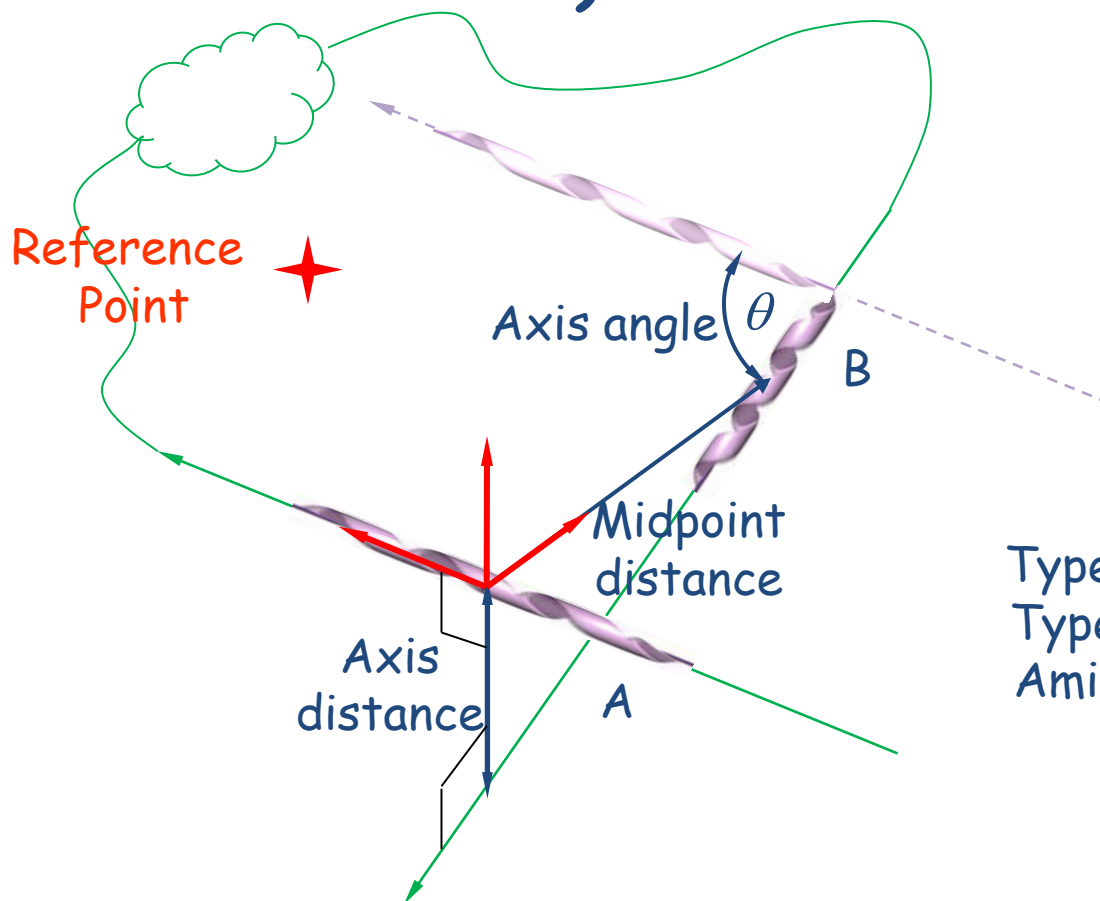
Type A: α-helix, l_1
Type B: π-helix, l_2
Amino acids: TD..

Generalized Hough Transform

- For each SS of the unknown molecule the neighborhood is investigated for co-occurrences

- The neighborhood is analyzed to discover if there are SSs compatible with the parameter terns of the Reference Table of the couples of SSs of the model

- for each co-occurrence a contribution is given for the possible existence of searched motif in the compatible location(s).
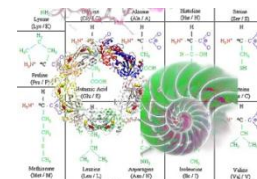
Generalized Hough Transform
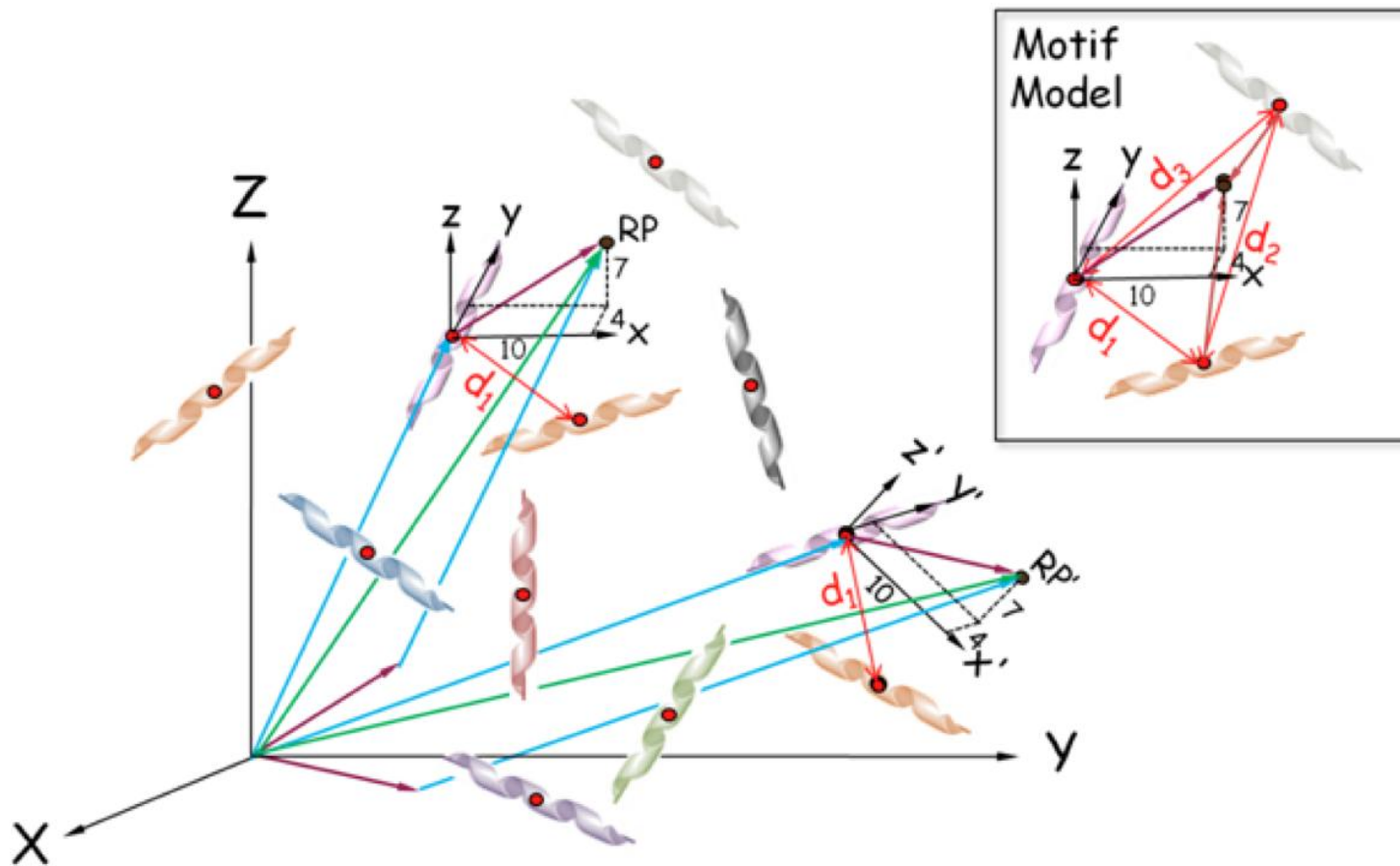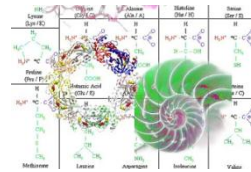
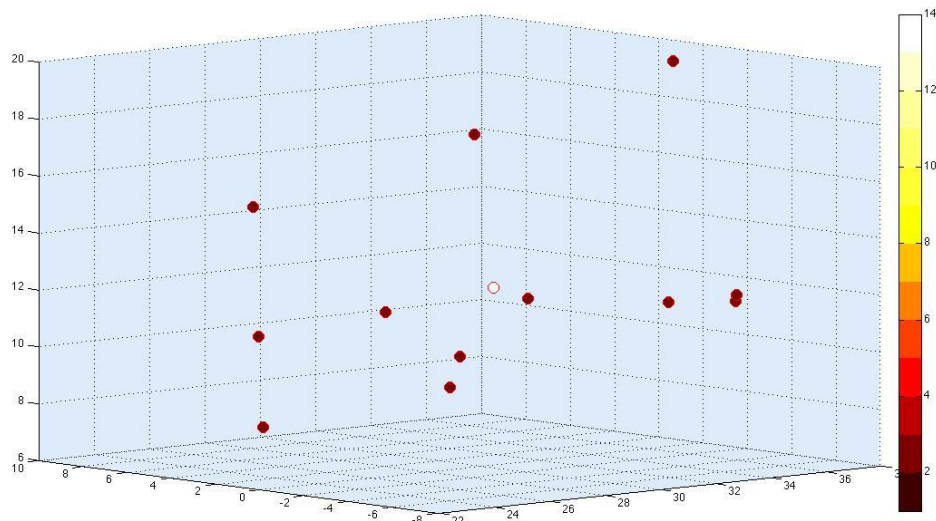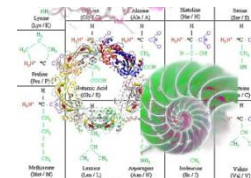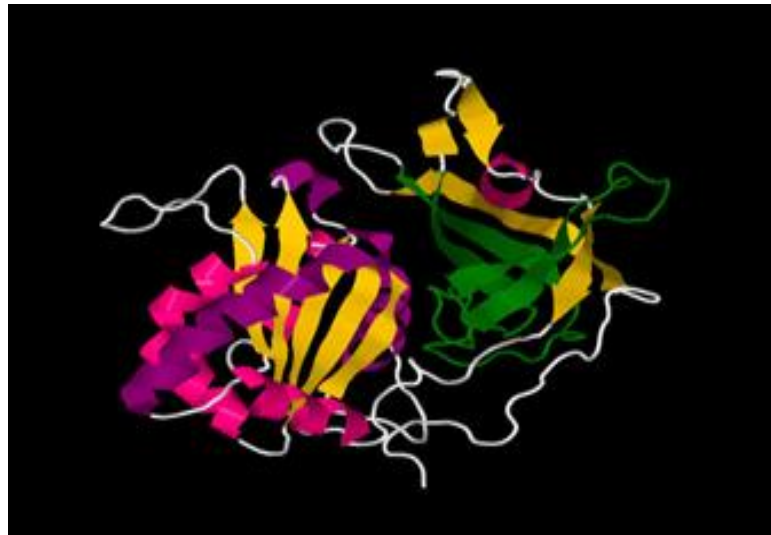# The Greek Key Motif
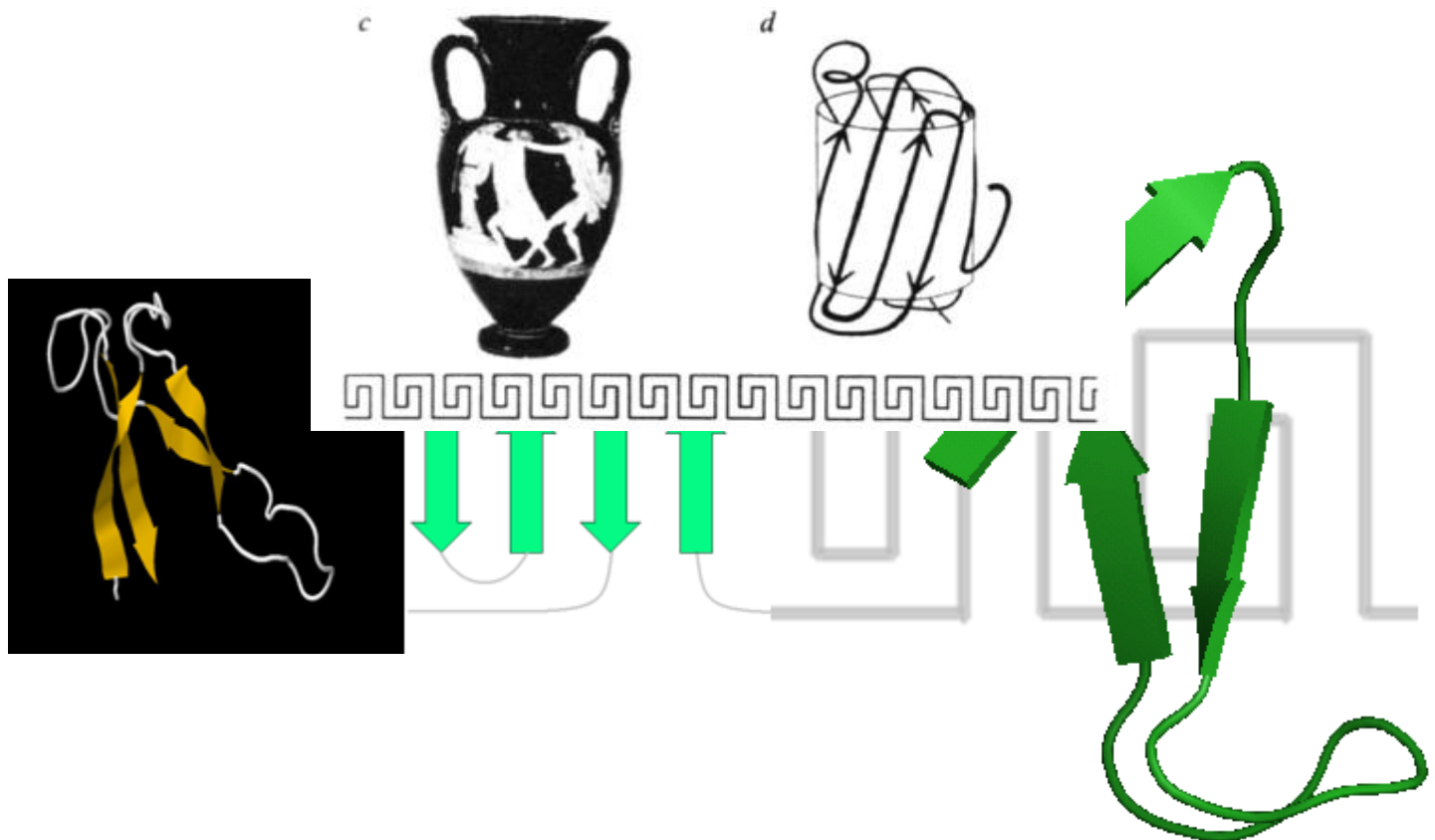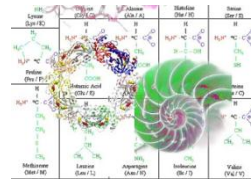
The Greek Key Motif

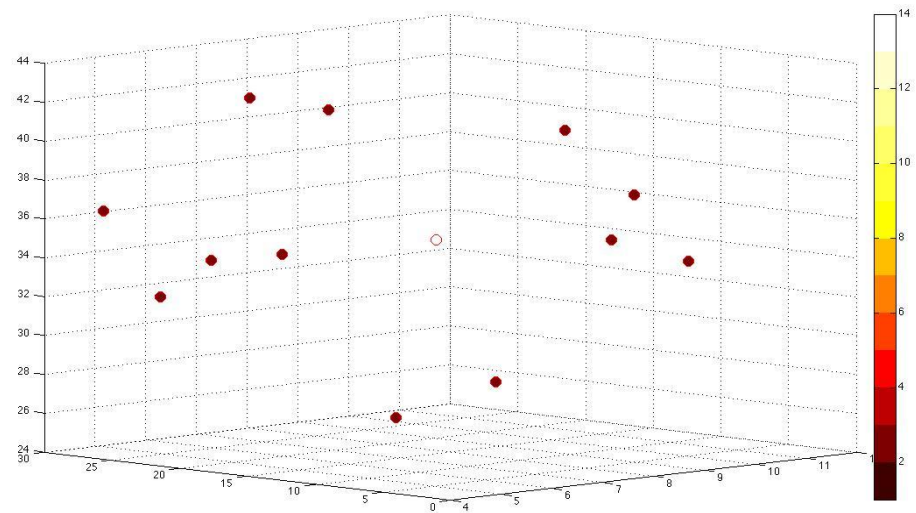Richardson, 1977) has compared Greek key motifs to the Greek keys found on a black Greek vase
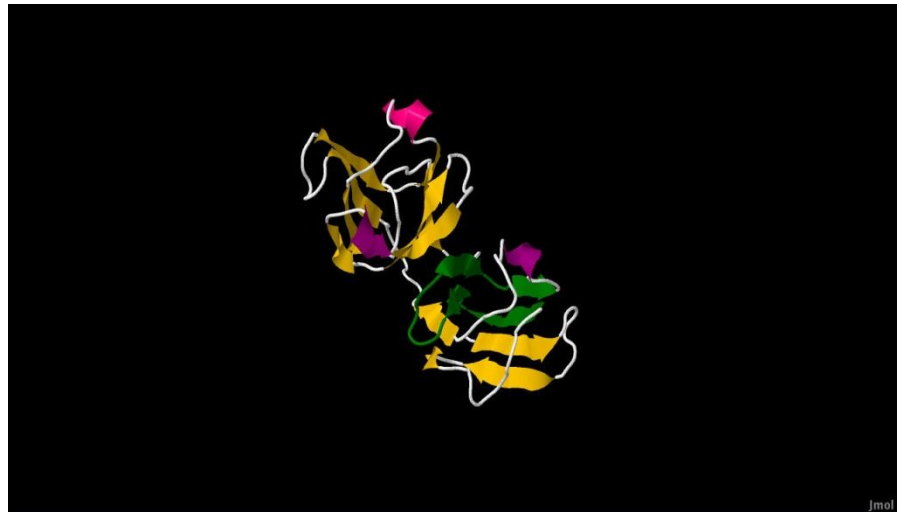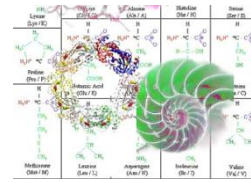
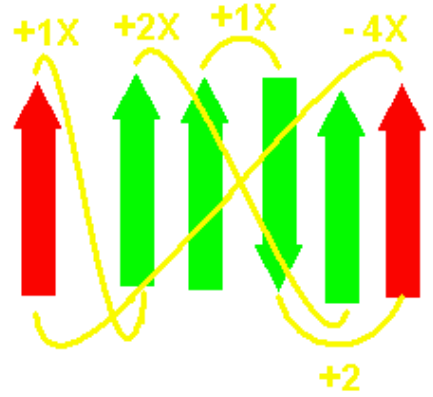The picture is generated by PyMOL on PDB file 4GCR for γ crystallin with residues 34-62 displayed and everything else masked
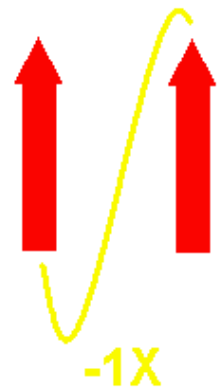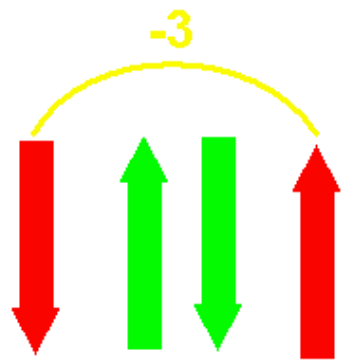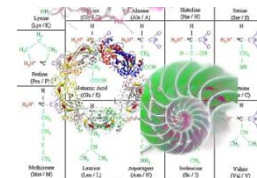
1FNB

4GCR

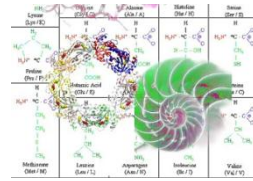# The Greek Key Motif

University of Waterloo

Shuo Xiang (Alex)
Dr. Ming Li

Preparatory Knowledge

Preparatory Knowledge

- Dr. Richardson's nomenclature of β-strand topologies may be summarized as:

- "+y": coil goes y β-strands to the right, starting β-strand and destination β-strand are anti-parallel to each other

- "-y": coil goes y β-strands to the left, starting β-strand and destination β-strand are anti-parallel to each other

- "+yX": coil goes y β-strands to the right, starting β-strand and destination β-strand are parallel to each other

- "-yX": coil goes y β-strands to the left, starting β-strand and destination β-strand are parallel to each other
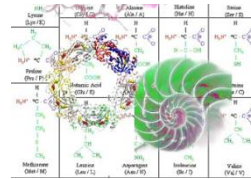
Formal Definition of Greek key
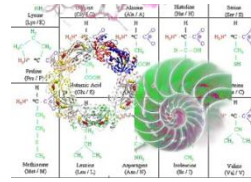
- With Dr. Richardson's nomenclature, Greek keys could now be formally defined as any set of 4 consecutive β-strands having the topology of "-3, +1, +1" or "+3,-1, -1" (Hutchinson and Thornton, 1993)
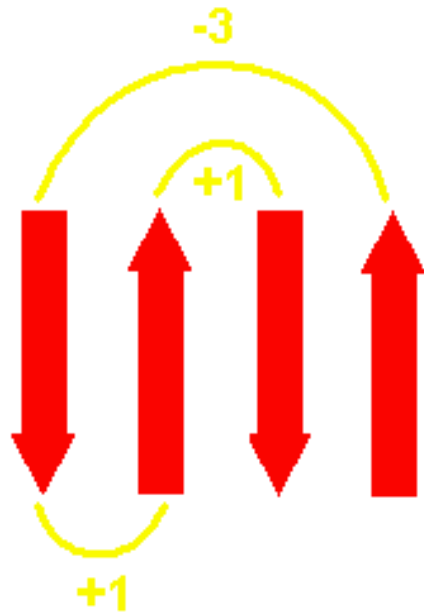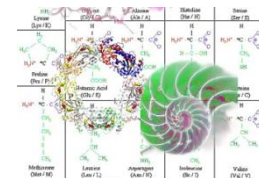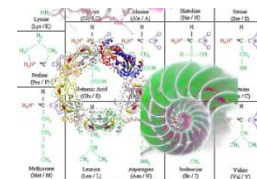
Classification of Greek key

- However, not all four β-strands of the Greek key falls within the same β-sheet.
- Hence there arises a need to classify Greek key structures according to their distribution of β-strands amongst β-sheet(s).
- Dr. Hutchinson and Dr. Thornton has given such a classification in (Hutchinson and Thornton, 1993)
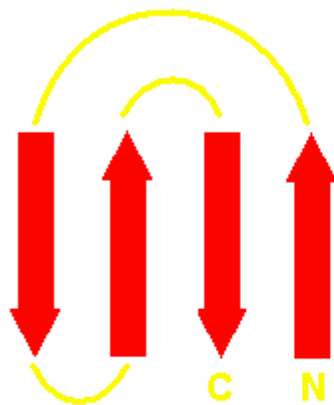
Classification of Greek key

- If all four β-strands of the Greek key lie in the same β-sheet, then it is called a (4,0) Greek key, meaning that there are four strands in one β-sheet and zero strands in the other β-sheet.

- Note that β-strands of a Greek key can go into *at most* two β-sheets. More than two β-sheets would make it very hard to decide whether a Greek key exists instead of some other random β-structure.
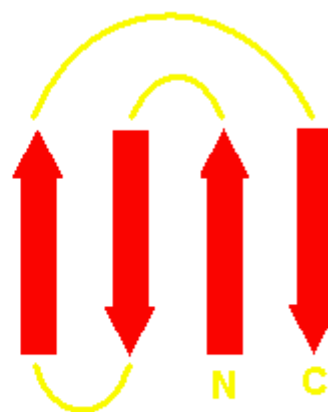
Classification of Greek key

- Furthermore, (4,0) Greek keys come in two flavours — an "N" version where the N-end of the Greek key is on the outside, and a "C" version where the C-end of the Greek key is on the outside. This is shown in the diagram below.
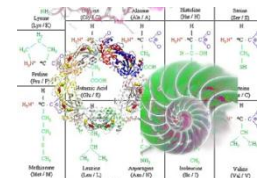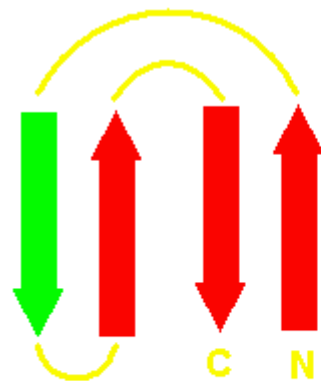

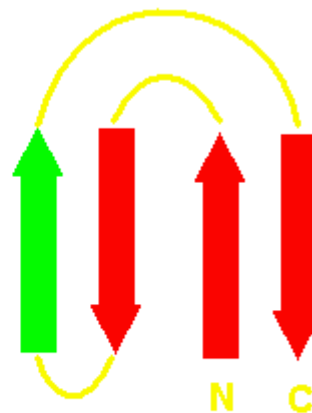
(4,0)N   (4,0)C

- Similarly, (Hutchinson and Thornton, 1993) classified the following as $(3,1)_N$ and $(3,1)_C$ Greek keys. Note that the green arrow represents β-strands from a different β-sheet.
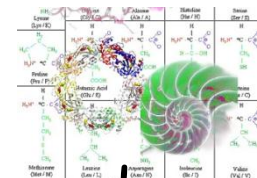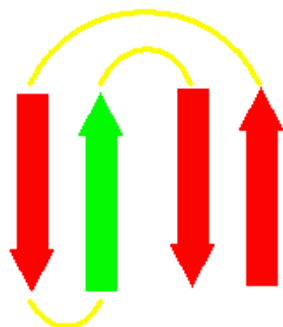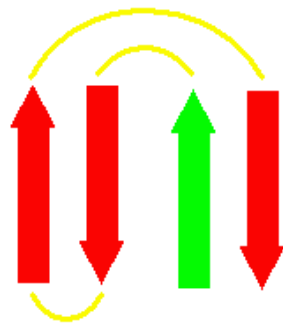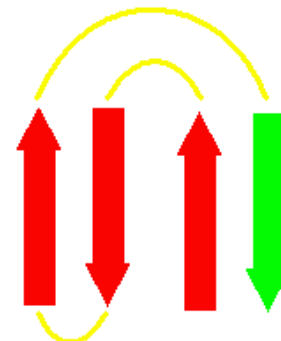

(3,1)N   (3,1)C

Classification of Greek key

- For this project the classification of (Hutchinson and Thornton, 1993) is extended to include the following additional combinations of four β-strands from two different β-sheets
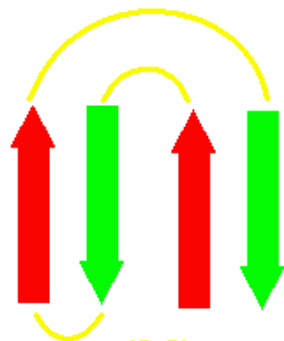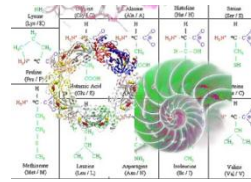
Conclusion

- G-Hough transform is suited for parallel implementation
- This method can only supply an approximate solution
- The results of this approach will identify a limited subset for a sub-sequent phase of refining
- Extended experimentation is now required to properly validate this new approach

General Hough transform approach to protein structure comparison

3C Vision

cues, contexts and channels
Elsevier (April 2011)
V. Cantoni, S. Levialdi, B. Zavidovique
Università di Pavia, Roma, Université de Paris XI