

Protein Gaussian Image (PGI): a protein structural representation based on the spatial attitude of secondary structure

Virginio Cantoni, Department of Computer Engineering and Systems Science,
University of Pavia, Pavia, Italy, virginio.cantoni@unipv.it

Alessio Ferone, Department of Applied Science, University of Naples
Parthenope, Naples, Italy, alessio.ferone@uniparthenope.it

Romina Oliva, Department of Applied Science, University of Naples Parthenope,
Naples, Italy, romina.oliva@uniparthenope.it

Alfredo Petrosino, Department of Applied Science, University of Naples
Parthenope, Naples, Italy, alfredo.petrosino@uniparthenope.it

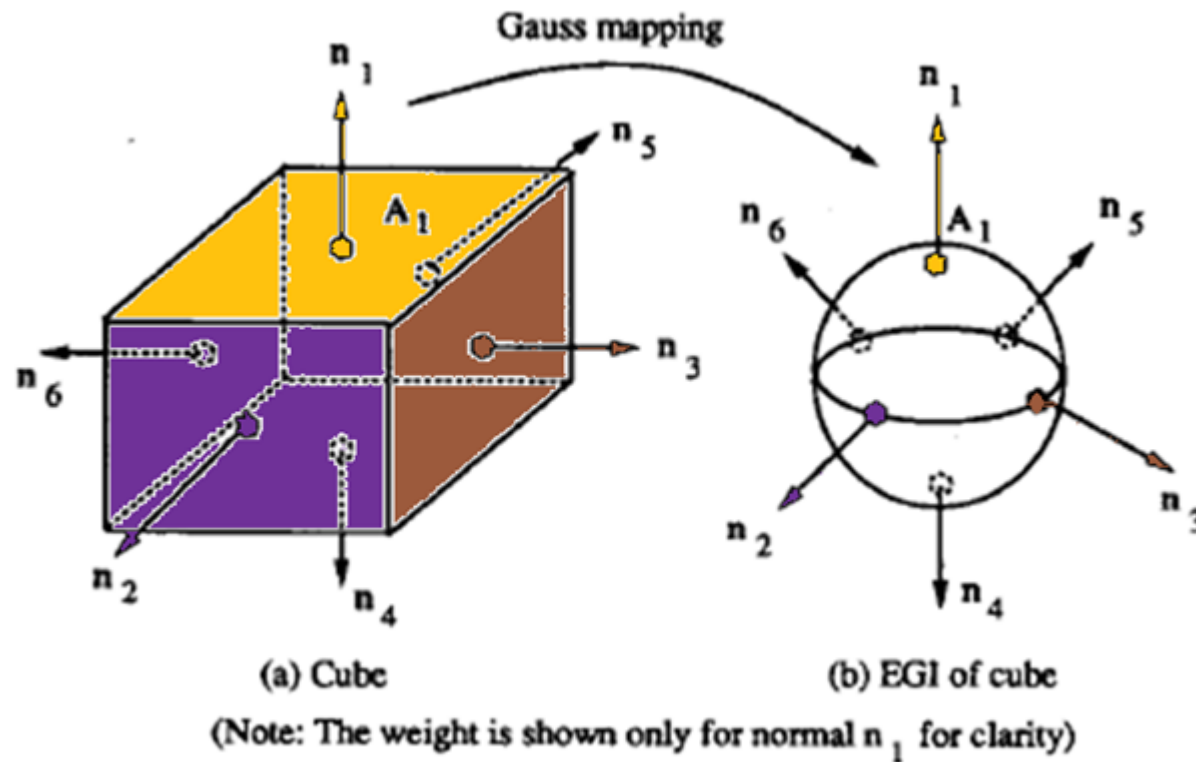
PGI

Protein Gaussian Image

- A new data structure for protein structural analysis
- Derived from two basic object representation approaches
- Developed by the pattern recognition community
- Extended Gaussian Image (EGI)
- Needle Maps (NMs)

Extended Gaussian Image

The EGI of a 3D object or shape is an orientation histogram that records the distribution of surface area with respect to surface orientation.



EGI

- Each surface patch is mapped to a point on the unit Gaussian sphere according to its surface normal
- A point on the Gaussian sphere is the total sum of area of all the surface patches that are of that surface normal
- EGI is in principle invariant to translation

Needle Map

- A needle map represents an object showing unit surface normals at points on the surface on a regular grid
- The needle-map is a valid representation for object recognition
 - midway between model (3-D) and appearance-based (2-D) recognition.
- Two values are sufficient to describe the direction of each normal
 - the third component may be determined from the other two
 - matching can be performed using 2-D vectors

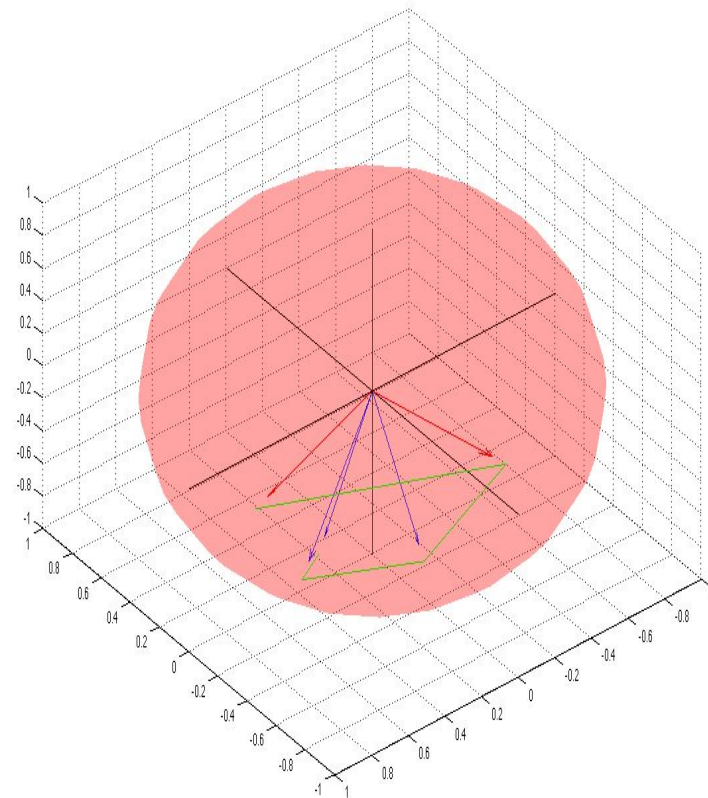
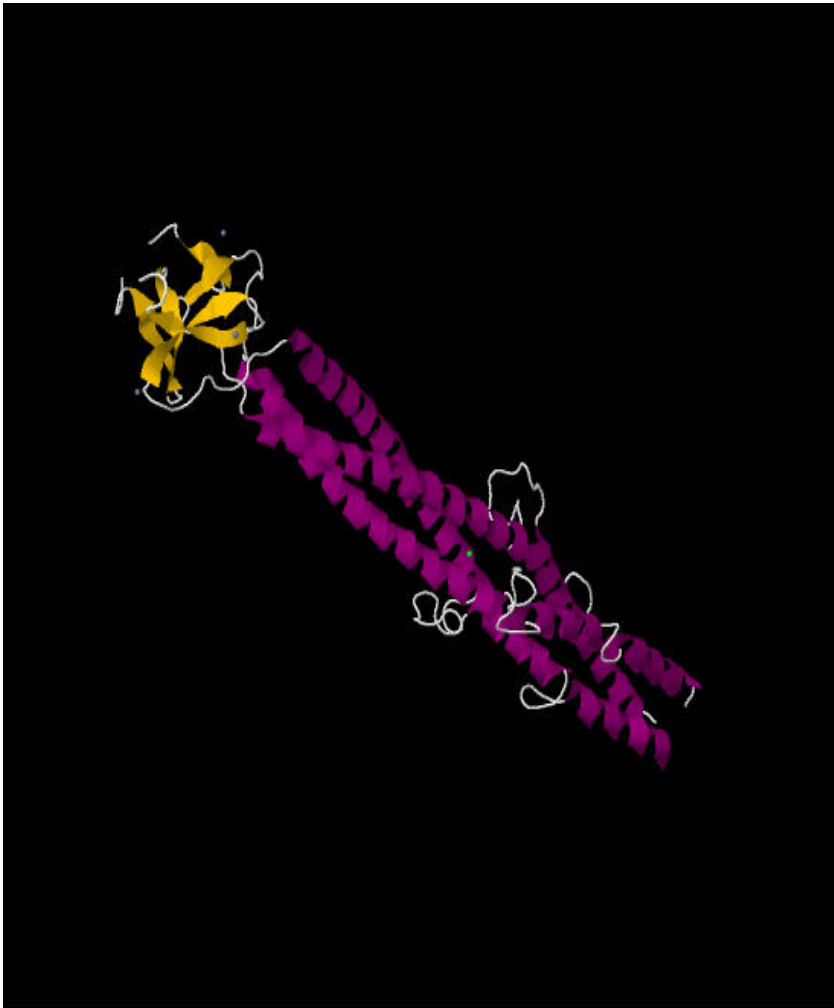
PGI

- PGI is a representation in the Gaussian image
- Each SS is mapped with a unit vector from the origin of the sphere having the orientation of the SS
- Each point of the sphere surface contains the data orientation
 - length
 - location of starting and ending residue
 - Etc...
- The chain sequence of SS is recorded as a list
 - mapped on the sphere surface

DSSP

- Most commonly used method to define secondary structure
- 8 types of secondary structures
 - helices
 - 3_{10} -helix
 - α -helix
 - π -helix
 - sheets or strands
 - extended strand (parallel and/or anti-parallel β -sheet)
 - Coils
 - Turn
 - bend
 - Unclassified amino acid

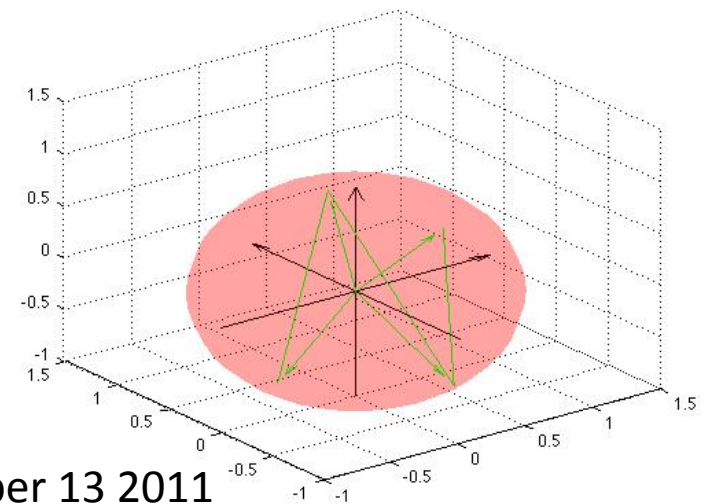
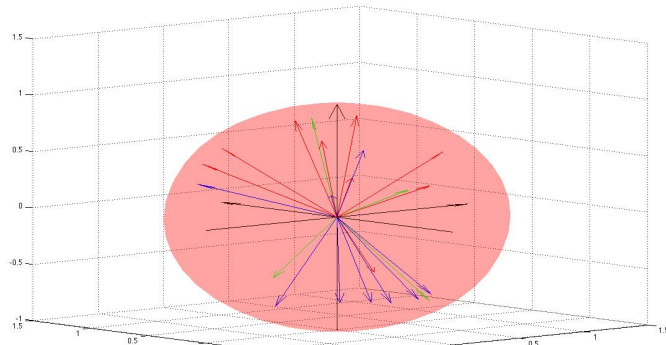
PGI (1aa01)



PGI

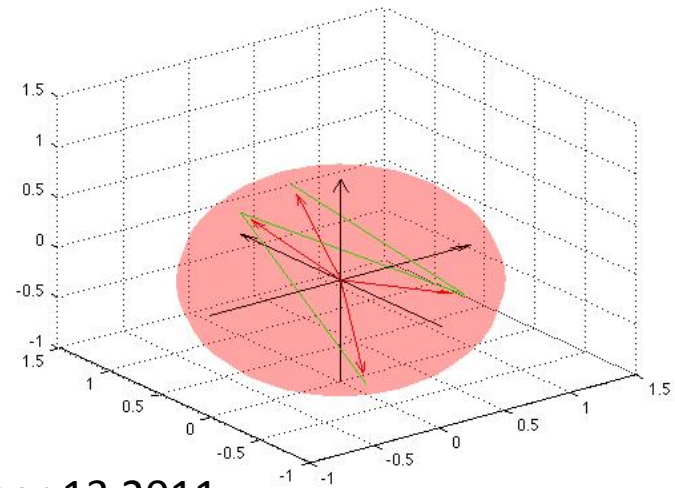
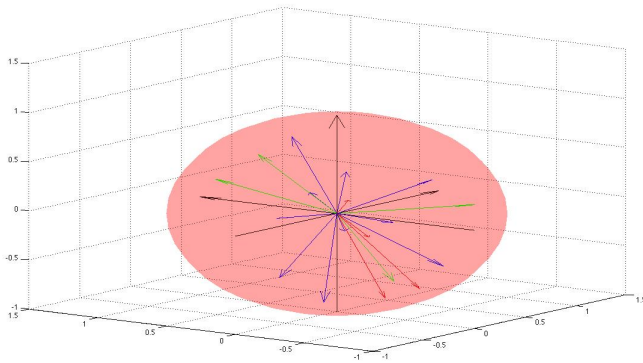
- This data structure is
 - complete
 - no information is lost for an analytic analysis
 - effective from the computational viewpoints
 - only two reference coordinates are needed
 - supports effectively the structural perception.

PGI (1FNB)



PR PS BB 2011 – September 13 2011

PGI (4GCR)



PR PS BB 2011 – September 13 2011

Conclusions

- A new data structure that supports both artificial and human analysis of protein structure.
- Analysis of the effectiveness of the representation for practical problems
- Structural block retrieval
 - Motifs
 - domains
 - units