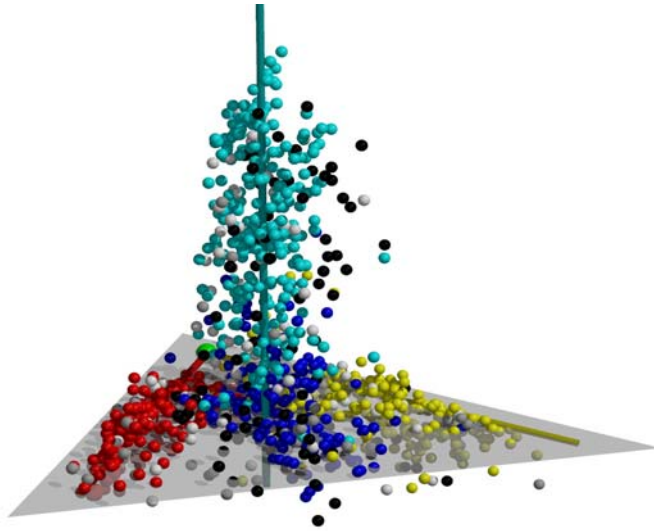


## Global Mapping of the Protein Structure Universe

**Contact Person:** Sung-Hou Kim; 510-486-4333; e-mail: shkim@lbl.gov

There are more than a trillion different proteins on earth, but they are made of combinatorial linkages and variations of much smaller number of protein structural motifs, the building blocks of protein structures. Each motif can be considered as the center of a protein structure “galaxy” consisting of many proteins of similar or related structures, and the protein structure universe can be represented in a “map” composed of a large number of such galaxies. We have mapped the protein structure universe represented by protein structural motifs of all the known protein structures experimentally determined. The map reveals four elongated clusters, that approximately correspond to four major protein fold classes. Some regions are more populated than others, and recent new protein structures are filling the unpopulated space in the map defined by the four regions. The map reveals features that may be interpretable in terms of evolution of protein structures, and provides a global view of the distribution of protein folds with respect to various molecular functions such as biochemical functions, metal binding, and others.

Funded by NIH (P50 GM62412) and NSF (DBI-0114707)



## References

- Hou, J., Sims, G., Zhang, C., Kim, S.-H. A global representation of the protein fold space. *Proc. of the Nat. Acad. Sciences*. **100**: 2386-2390. (2003).
- Choi, I.-G., Kwon, J., Kim, S.-H. Local feature frequency profile: A method to measure structural similarity in proteins. *Proc. of the Nat. Acad. Sciences*. **101**: (2004).