

Two new PyMOL tools: An ‘average3d’ Python module for computing & visualizing ensemble-averaged structures and ‘makeCGOplates’ for schematizing planar assemblies

Cameron Mura^{1*}

¹Department of Chemistry & Biochemistry; University of California, San Diego; La Jolla, CA 92093 USA;

*cmura@mccammon.ucsd.edu

August 2005

Implementation

Two recent utilities have been developed as Python modules for PyMOL.

One of these utilities, ‘makeCGOplates’, can be used to create and draw polygonal, plate-like representations using PyMOL’s *Compiled Graphics Object* (CGO) facility. This is particularly useful for highly schematic representations of polymers, fibrils, cyclic macromolecular assemblies, and other large (*e.g.*, cellular-scale) entities. Importantly, the tool automatically performs a geometric sort of the chains in the assembly (by walking through the adjacency matrix computed from the centers of geometry of the sub-entities), such that the path defined by the geometric centers corresponds to a convex polygon. Therefore, it is acceptable if the chain labels (PDB chainIDs/segIDs) are scrambled with respect to what should be the ‘correct’ cyclic order.

The other tool, termed ‘average3d’, can be used to process multiple 3D conformers of the same protein structure – *e.g.*, the bundles resulting from NMR structure determinations, structural snapshots along an MD trajectory, and so on. Given an ensemble of structures, one can use this tool to compute an average 3D structure and the following two types of RMSD measures:

1) A residue-specific RMSD measure with respect to the ‘average’ 3D structure ($\langle \mathcal{C} \rangle$) of the bundle of conformers (\mathcal{C}) at residue position i :

$$\text{RMSD}_{\langle \mathcal{C} \rangle}(i) = \sqrt{\frac{\sum_{c \in \mathcal{C}} \|\vec{r}(i) - \langle \vec{r}(i) \rangle\|^2}{|\mathcal{C}|}}$$

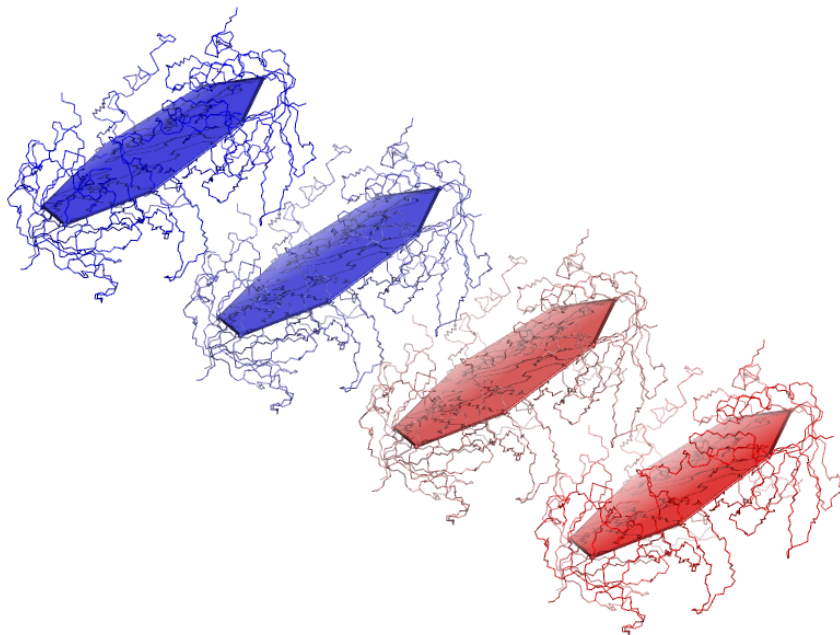
2) A residue-specific RMSD measure calculated as the average over all unique pairwise distances for matching residues i (note that this pairwise measure is systematically greater than the previous type of RMSD):

$$\text{RMSD}_{\langle \mathcal{C}_x, \mathcal{C}_y \rangle}^{(pairs)}(i) = \sqrt{\frac{\sum_{y>x}^{|\mathcal{C}|} \sum_{x=1}^{|\mathcal{C}|} \|\vec{r}_x(i) - \vec{r}_y(i)\|^2}{(|\mathcal{C}|)(|\mathcal{C}| - 1)/2}}$$

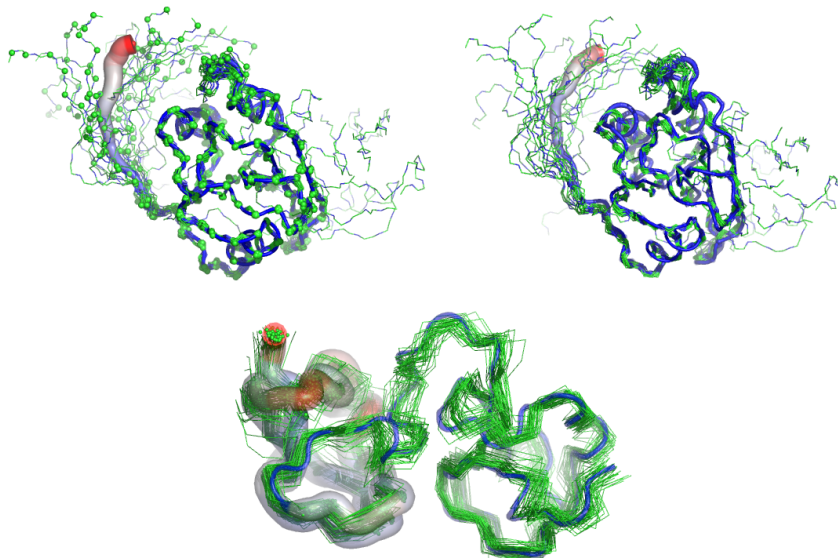
Usage

Both the CGO-related and structural averaging tools were implemented as Python modules which utilize PyMOL’s API, and are further described on my PyMOL page (<http://muralab.org/~cmura/PyMOL>). The following are two graphical illustrations of their usage.

Sample usage of ‘makeCGOplates.py’ (initially written for the SmAP fiber work):



Sample application of ‘average3d.py’:



Acknowledgements

This work was performed in the McCammon group at UCSD and was partly funded by a Sloan/DOE postdoctoral fellowship in computational molecular biology.