WIP readme

Explanation of macromolecular structure file handling

CTS can load and parse .pdb and .cif formats as initial macromolecular inputs. Typically, each file will contain a single model, see advanced usage for how CTS can make use of multi-model inputs. After loading a file and parsing its data, CTS will attempt to save a .mat file with the same base name as the input file. This .mat file can be loaded as an input in place of the original input and loads much faster as it is reduced to only minimal necessary information.

Classification ID

Each model is assigned a class ID that determines how it is stored. During model generation, parallel to the model array itself each class has its array that stores the location of all of its members present in the whole model. Class IDs are extracted from the filename. Filenames are split into segments by . and \_\_ (double underscore), and for single-model files the first segment is used as the class ID. Otherwise, if there are 2 more segments than models the leading segments will become the class ID for the corresponding model. Anything else will use the first filename segment for all model IDs, appended by a number.

Examples:

actin.pdb -> class ID is actin if a single model, or actin\_1 actin\_2 … actin\_n for multiple models.

actin\_\_cofilactin.bundle.mmcif -> the first model is ID actin, the second cofilactin.

\*matlab quirk: due to internal handling, class IDs need to start with a standard letter. If they do not, fix\_ will be prepended to them to try to avoid malfunctions.

Multimodel file handling

CTS can generate a handful of organizations of macromolecules, dictated by the last part of the filename before the extension. The modeling process iteratively selects a files’ group for placement, and it can attempt to do so in a few ways. By default, it randomly selects one model from the set present from the original file and attempts placement. This default is the ‘single’ or ‘group’ method. Other methods require the method to be indicated in the filename, they are ‘bundle’ ‘cluster’ ‘complex’ and ‘assembly’.

Cluster attempts to place more random members of the file nearby

By default, the modeling process will randomly select a file’s group of particles to be placed next