Ligand GA Output Processing

Comment about directory processing: In the ligand dir, use of the Ligand\_GA\_Fitness\_Function… will result in a clearing of any directory or file with mol\* and a creation of molecule(index) directories, as many as are in the list sent to the fitness function. Then, in each of these molecule(index) directories there will be a stereoisomer sub-directory, as many as are there stereoisomers of the non-isomeric ally expressed molecule. The CCDC GOLD output of the docking of the individual stereoisomer is in this.

The parameters in the Ligand\_GA\_Config\_File have to be loaded for path declarations, general parameters. run(Ligand\_GA\_Config\_File) or copy/paste will do this.

From an output .mat file

1. Load the .mat file with ‘load \*\*\*.mat’

The population from each iteration is in a cell array GA\_Gen\_Save with ‘population’ and ‘fitness’. This set of molecules will have redundant molecules and molecules with different docking scores due to the indeterministic GA docking calculation. The best pose score is saved for each stereoisomer of the molecule and the variation is small but in some cases up to 10%.

1. Instead of loading the .mat file, the function Ligand\_GA\_Output\_Function can be used. Call it with the name of the file.mat file,

[sorted\_unique\_population,sorted\_unique\_fitness]=Ligand\_GA\_Output\_Function(“file.mat”)

This will give two cell arrays which are sorted according to -fitness and population, and all of which are unique non-isomeric molecules. Any different scores in the iteration set from a same non-isomeric molecule is kept at the maximum score in the different scores.

1. The stereoisomer fitnesses of a molecule chm=”” can be found from running one of these molecules through the fitness function,

Fitnesses=Ligand\_GA\_Fitness\_Function\_GOLD\_ADME(chm)

which will create a set of stereoisomer directories in the molecule1 subdirectory of ligand\_dir if one molecule is given.

If a one-column array is given instead of chm, then molecule1 to molecule$LENGTH$ subdirectories are created with all of their stereoisomer subdirectories. The function will return an array for Fitnesses, one for each molecule – and from the highest scoring from all of its stereoisomers. The calculations require the global ligand\_dir, GOLD path, and MGLTools path, and parameters. Loading of the Ligand\_GA\_Config file will set these for these directories/files. The fitness\_function will delete any molecule\* directories and create new ones for the individual molecules in the input list x<-chm.

1. In each stereoisomer(stereoisomer\_idx) subdirectory of the molecule(mol\_idx) directory, there is a file called “bestranking.lst” that will point to the m1\_?.mol2 file with the highest docking score – this is the highest scoring pose from the stereoisomer of the molecule(index). There is a mol2 file with the coordinates of this pose in “docked\_molecule … m1\_?.mol2”.

From an input list of non-isomeric molecules

1. Input a single one-column cell array list ‘x’ of molecules into the Ligand\_GA\_Fitness\_Function. The output fitnesses can be examined, non-stereoisomerically, and changed by whatever means.
2. The list can be added to with changes or deleted partially. The use of Ligand\_GA\_Output\_Function(x’) will sort. Multiple runs of this can find an average score for the best pose of the docked non-isomeric molecule then stereoisomer.
3. Iterate.