**Ligand GA Penalty Constraints**

External Ligand GA functions can be used to constrain the evolution of the molecules. These are straightforward to implement in the fitness function. Two examples are shown here.

*Please do not waste my effort in showing you simple example constraints. There is tremendous structural information and chemical information that can be used and hydrogen bonding requirements are trivial in imposing in the molecule constructions. For lack of time I cannot show the very examples.*

The PLP\_fitness score is calculated for each molecule and stereoisomer in the GOLD fitness function, Ligand\_GA\_Fitness\_Function\_GOLD\_ADME. Figure 1 shows a molecule from a modified Aspirin run, tuned mutation probabilities [3,1,2,1,1,3,1,1,.02,.02,.02,.02] and atomic content CNOF. The docking score is GOLD PLP 74. However, it is unstable in water due to an enol ether group in the middle of it, which would be stable in an aromatic ring.

Figure 1



The presence of these sub-structure pieces in the molecules can be blocked by including penalty on the molecule.

In the fitness function include the line

PLP\_fitness = PLP\_fitness - NO\_ENOL\_ETHER(chm)

and the script NO\_ENOL\_ETHER.m,

function penalty = NO\_ENOL\_ETHER(chm)

penalty=0;

if isempty(strfind(chm,"C=CO")

penalty=20;

end

end

This will penalize the molecule and Ligand\_GA will eventually get rid of it in the population. Note that in this example the presence of an aromatic ring C1=CC=CC=C1 could also be checked and then not penalize the molecule, being stable. This simple function can generalize to any substructure type and can also be used not to penalize but enhance the molecular evolution towards a desired class of molecules.

Another example is in the use of a penalty of 3 member heterocyclic rings including aziridine, oxirane, and thiirane containing N,O, and S. Generally 3 member carbon rings are not favorable due to stability and reactivity, but are used in the design of drug-like molecules such as Simeprevir, which has 2. These rings can be penalized also by a short script to check if there are 3 atoms within the ring points, e.g. C1CC1, C1NC1, etc… The MoleculeStructure.m function generates gives topological bonding information, with atom types and bondings to exclude any ring types from entering into the Ligand GA evolution.

The number of constraints on molecules is limitless and a matter of short scripts. Furthermore, given computational resources structural similarity can also be included by scoring pattern matches with known lists of classes of molecules. The entire list of PubChem can be downloaded into a 2 column list text file with SMILES expression and compound ID (CID). After restricting the list to wanted classes of molecules, it can be grepped with structural similarity to a molecule in the Ligand GA population. Penalties or bonuses can then be added to the scoring in the fitness function, and in guiding the evolution.

Opioids are typically highly ringed and compact 3-dimensional molecules (not projectable onto a plane without intersecting lines), and density of atoms could be included. The crossover and mutation functions were originally tested in Ligand GA in finding high mass density molecules. The mass density fitness function is also included in the software, and can be used for molecules in general, not necessarily ones in involved in biochemistry.