A micro level find/ query tool for molecule viewers[1]

By V.Ganesh, Research student, University of Pune, INDIA

Introduction:

Molecule viewer is one of the tools that is frequently used by chemists, biologists and physicists alike. Though there are numerous molecule viewers available[2], none of them provide simple yet powerful find/ query tool to find microlevel information like bonding, distance etc. for a large molecule. Many of the modern molecule viewers do come with an inbuilt scripting interface, but to do simple things like searching and highlighting an atom requires you to write quite a few lines of script. Also this kind of information is extremely useful while developing new theoretical methods, analyzing results of calculations or experiments etc. In this article I describe simple but elegant and powerful find/ query tool built upon a simple one line interpreted language with immediate visual feedback.

The MolFind tool:

The tool is defined by a small language and closely coupled with a GUI for visual feedback (see following screen shot).
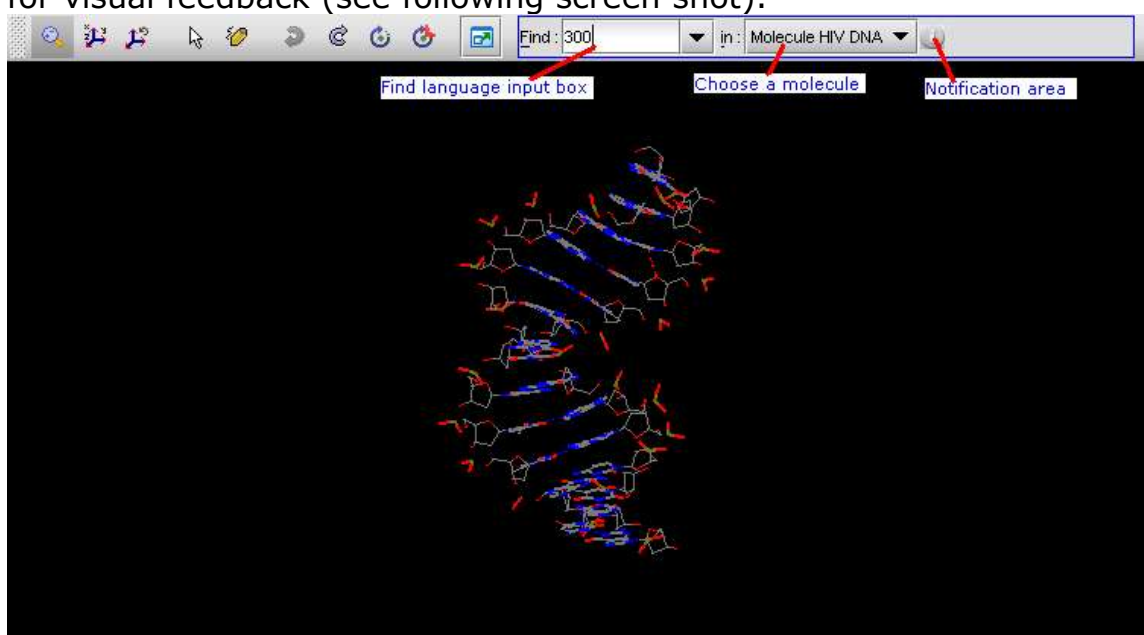
The tool GUI consists of three major parts:

1. The Find language input box, where the user enters short commands (described later) for querying micro level information about the selected molecule. This input box also keeps a history of all previous valid find commands entered by the user, so that they can be directly used at a later stage.
2. The tool can query multiple molecules; the molecule to be queried can be accessed via. the molecule chooser combo box.
3. A notification area for informing the user of any query results or possible errors in the find command that the user has issued.

The find language is a simple one-line interpreted language which can be defined using the following regular expression:

$$<<atom\text{-}index><,>>?[<,>[keyword]]$$

Which is interpreted as one or more occurrences of valid atom indices in a molecule separated by a '**,**' followed by none or one valid *keyword*. If no keyword is specified, then the atom indices are visually highlighted, zoomed and translated for appropriate view (see the following screen shot).
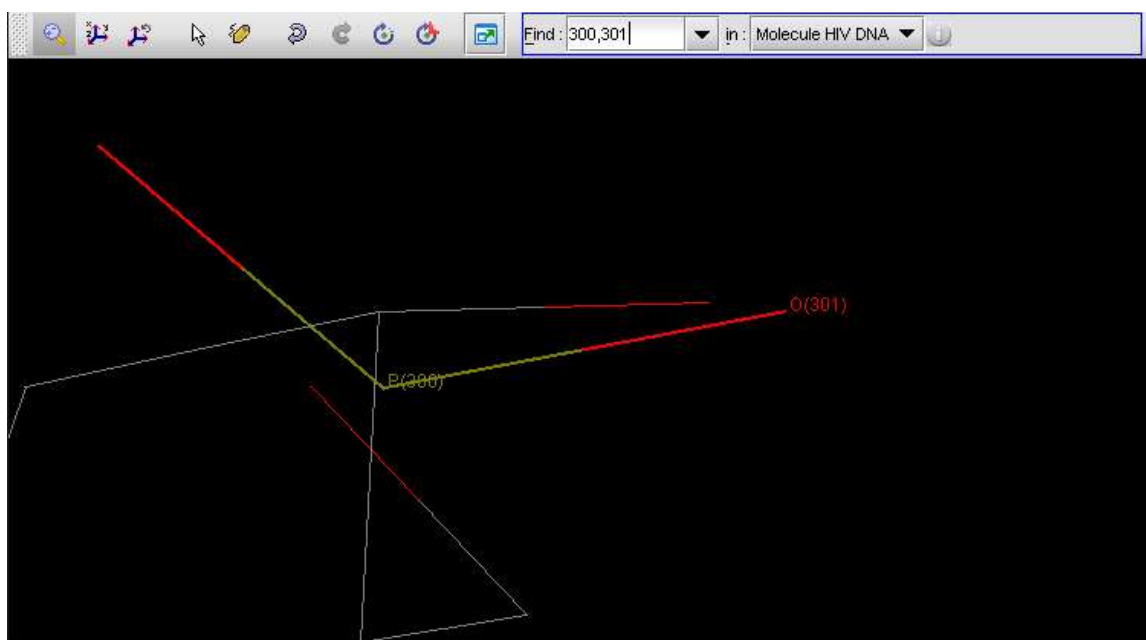
Note that in this specific case the valid atom indices for a molecule of N atoms is between 0 and N-1. Appropriate validations are performed by the find language interpreter to ensure the correctness of these indices.
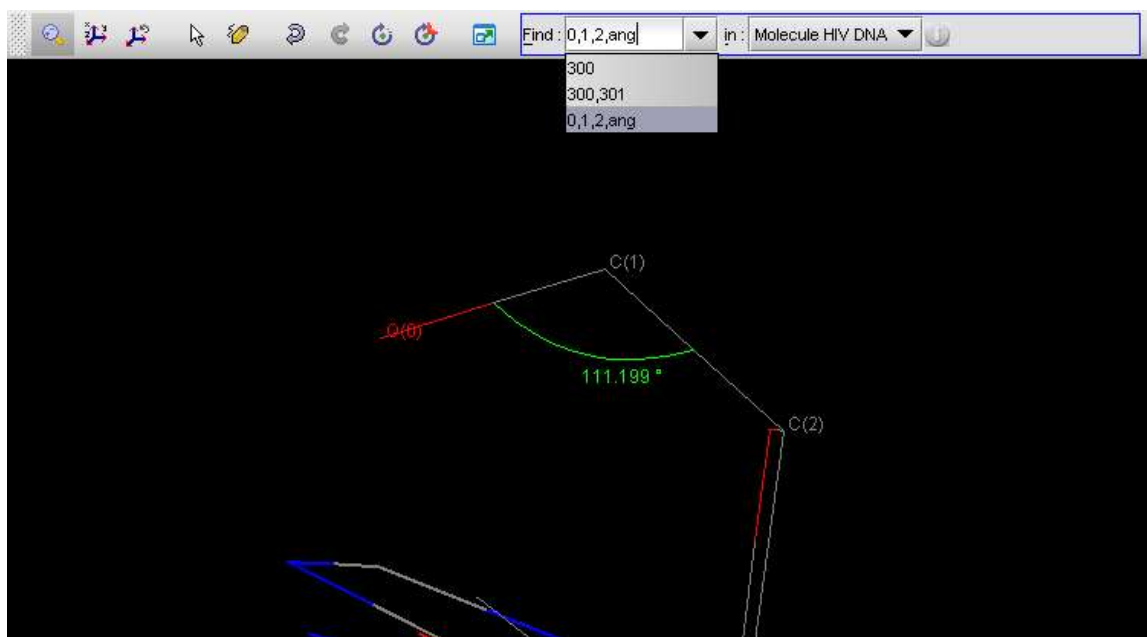
In a case where a *keyword* is specified, the atom indices are taken as argument to it. The *keywords* currently defined are as follows:

1. *<index1>*, **conn**: queries the connectivity of *index1*.
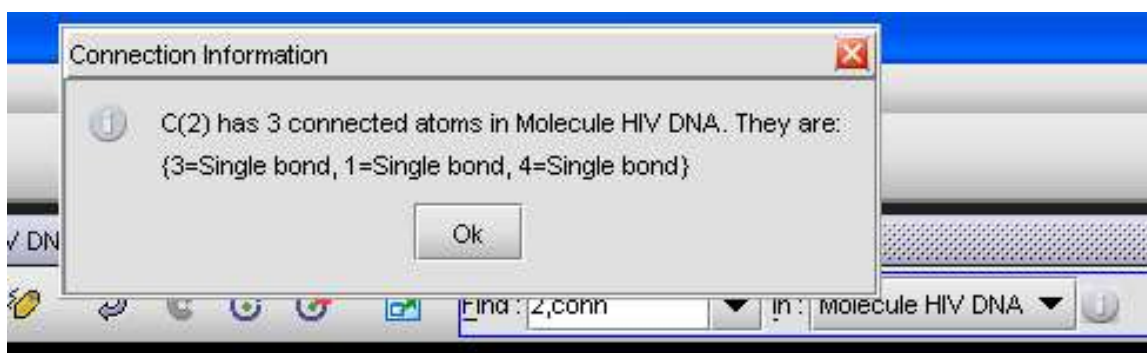
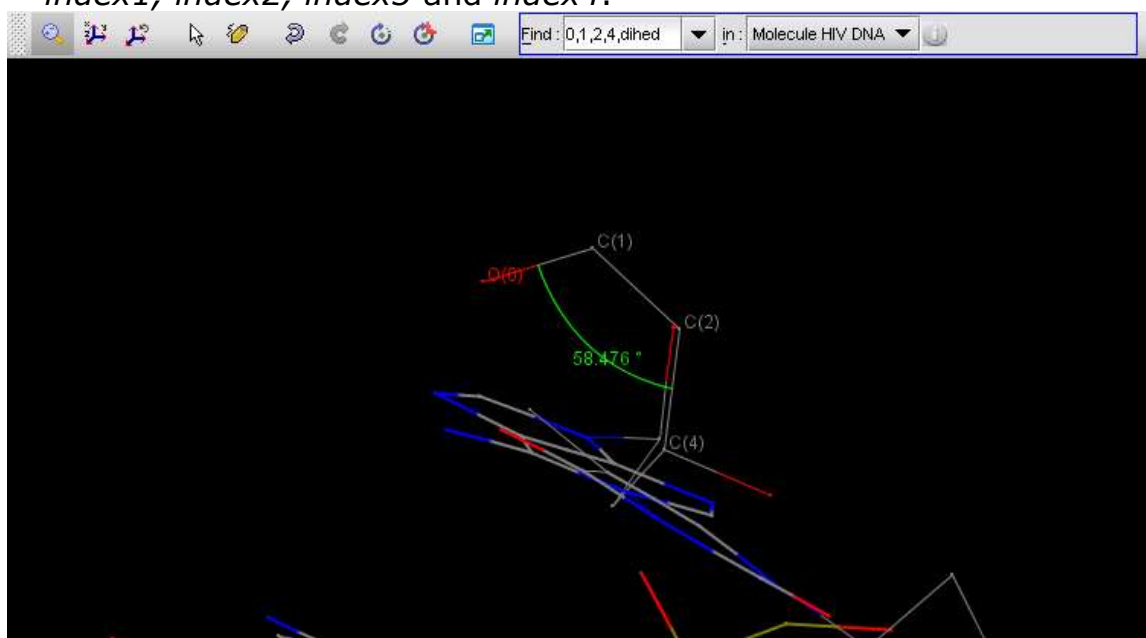2. *<index1, index2>*, **[dist/bond]**: queries the distance / bond between *index1* and *index2*.

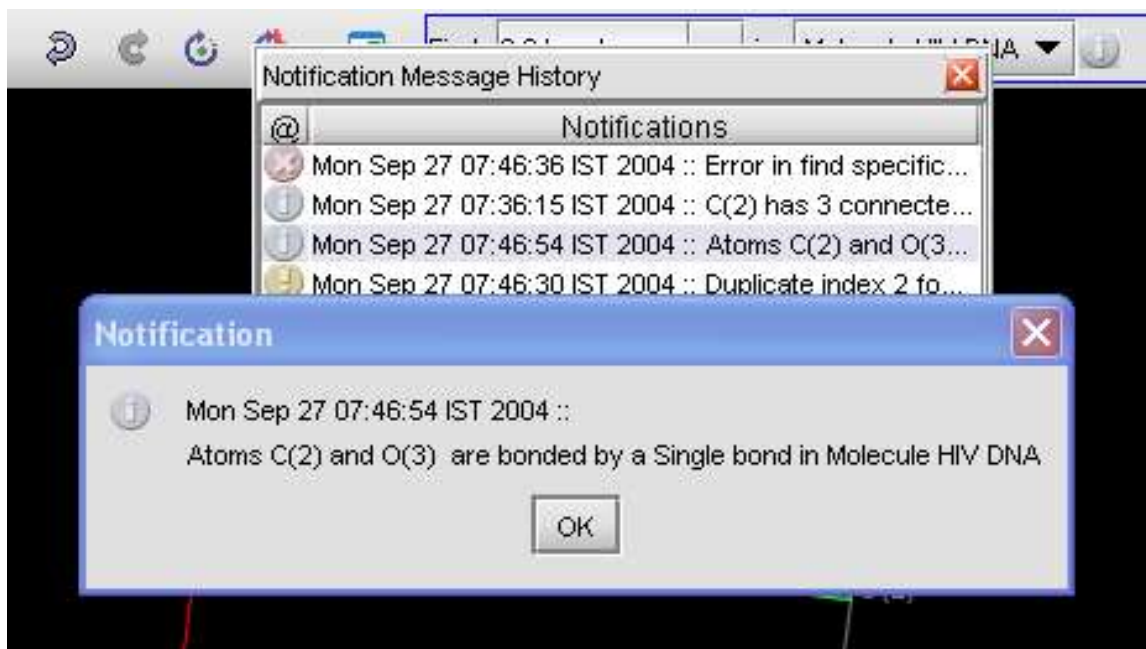3. *<index1, index2, index3>*, **ang**: queries the angle between *index1*, *index2* and *index3*.

4. *<index1, index2, index3, index4>*, **dihed**: queries dihedral between *index1, index2, index3* and *index4*.

The results of most of the user queries, warnings and errors form the find interpreter are logged into the notification area, which can be accessed by clicking on the notification icon on the toolbar (see the following screen shot).

Conclusions:

The tool described here has been implemented into MeTA Studio v2.0, under development by the author using the Java™ platform.

The amount of information that can accessed via the language and the tool described here is very limited but the idea can be easily extended to variety of other query options depending on the type of molecular data being viewed. For eg. if a iso-surface is being viewed this tool can provide query for possible critical points etc. Or in the case of a PDB, search for occurrences of peptides say alanine etc.

In essence, as more and more large molecules are handled by biologists and chemists the framework currently described here can be used as an elegant and simple way to interactively and visually query micro level molecular information in conjugation with other available tools like the tree representation of the molecular data etc. The current implementation only provides an "immediate" visual feedback and not a "real time" feedback, though this can be trivially done, the repainting may be expensive on lower end machines.