chem-bla-ics

OSRA: GPL-ed molecule drawing to SMILES convertor

Published July 20, 2007

Citation

V2lsbGlnaGFnZW4sIEUuICgyMDA3LCBKdWx5IDIwKS4gT1NSQTogR1BMLWVkIG1vbGVjdWxlIGRy
YXdpbmcgdG8gU01JTEVTIGNvbnZlcnRvci4gPGk+Q2hlbS1ibGEtaWNzPC9pPi4gaHR0cHM6Ly9k
b2kub3JnLzEwLjU5MzUwL2FyYzJqLTFoYTMy

Keywords

Cheminf, Openbabel

Abstract

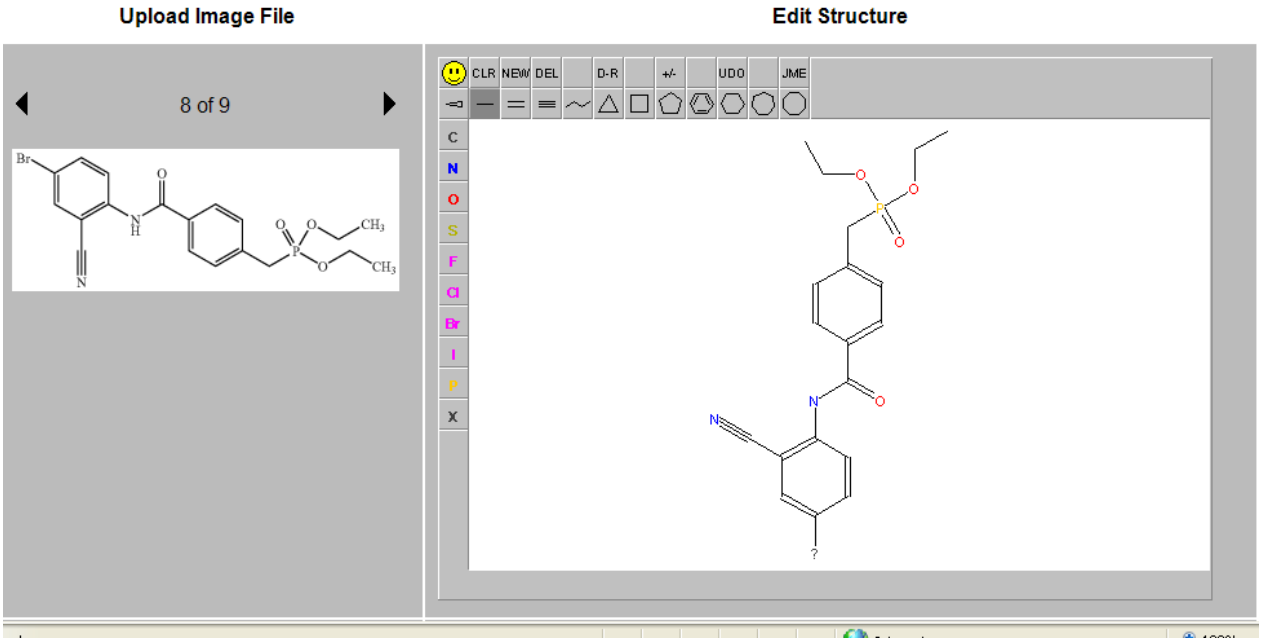
Igor wrote a message to the CCL mailing list about OSRA:

chem-bla-ics

Igor wrote a message to the CCL mailing list about OSRA:

> *We would like to announce a new addition to the set of chemoinformatics tools available from the Computer-Aided Drug Design Group at the NCI-Frederick. OSRA is a utility designed to convert graphical representations of chemical structures, such as they appear in journal articles, patent documents, textbooks, trade magazines etc., into SMILES.*
>
> *OSRA can read a document in any of the over 90 graphical formats parseable by ImageMagick (GIF, JPEG, PNG, TIFF, PDF, PS etc.) and generate the SMILES representation of the molecular structure images encountered within that document.*

The email does not give any information on the fail rate, but the demo they provide via the webinterface does show some minor glitches (the bromine is not recognized):



The source reuses OpenBabel and uses the GPL license. The value equal to that of text mining tools like OSCAR3 , and together they sounds like the Jordan and Pippen of mining chemical literature.