It's a crime not to KNIME[1]

Open for Innovation KNIME saves time in drug discovery.

Bob Marmon and Inaki Morao

Evotec (UK) Ltd., 114 Innovation Drive, Milton Park, Abingdon, Oxfordshire OX14 4RZ, United Kingdom

Quantum chemistry in the 21st century

After decades of incremental development—accompanied by a healthy dose of scepticism from the experimental community—the ready availability of fast computers has ushered in a "golden age" of quantum chemistry.[2] Theoretical predictions about the structure and properties of molecules need to be more accurate and the traditional Molecular Mechanics (MM) approaches although fast can only achieve mediocre results that are no longer acceptable in comparison to quantum mechanics (QM) calculations.

NWChem is an Open Source High-Performance Computational Chemistry package that provides QM calculations that are scalable both in their ability to treat large problems efficiently, and in their use of parallel computing resources from high-performance supercomputers to conventional workstation clusters.[3]

NWChem Knime Node

In response to the demand for QM calculations in Computational Chemistry projects, we built a QM Knime node to access the NWChem package. The NWChem Knime node has proved to be extremely useful in expanding our offering of computational chemistry workflows (see below).

The node calls the NWChem package installed in our Linux farm over SSH using a command file constructed from options chosen by the user, runs the specified job(s) and brings back the whole NWChem output file. The node extracts typically required properties such us HOMO, LUMO, final energy and geometry, and appends these as new columns to the node output table. We have provided the user with a versatile configuration dialogue for the most common NWChem options such as charge, basis set, task and level of theory. The options can be overridden manually in the command panel or by an external variable which allows the user configure less standard QM calculations such as solvation models, NMR prediction or Fukui indexes.
The node is normally preceded by the OpenBabel node, in order to prepare the incoming molecule(s) in a NWChem input format.

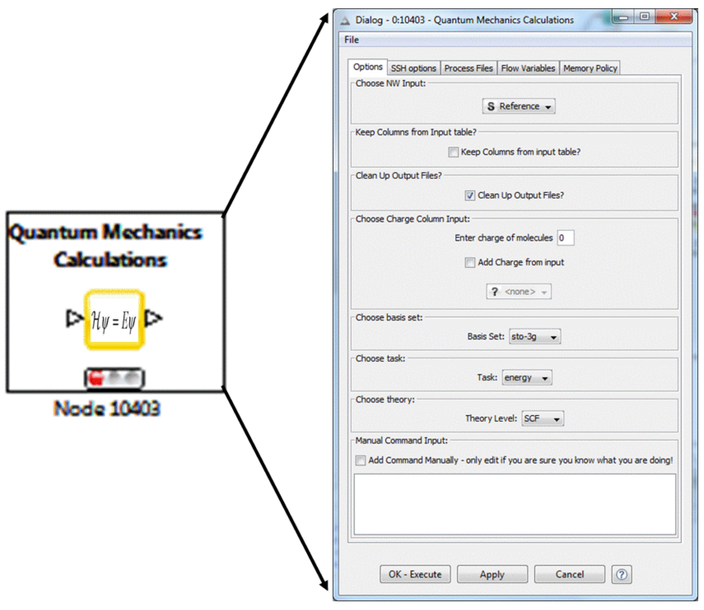
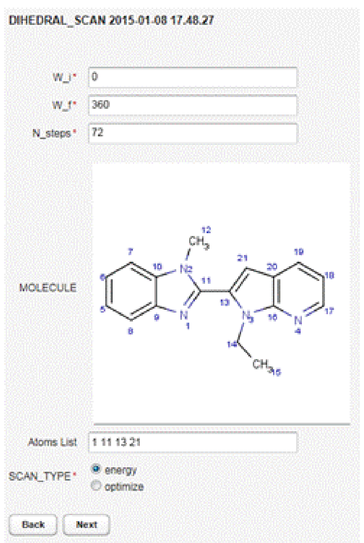


Figure 1: The QM Knime node based on NWChem software. The main dialogue panel is also displayed to show the options provided so the user can create the NWChem input without any knowledge of the NWChem documentation.

Two case studies using the NWChem node

In this section we will present two Knime workflows frequently used in drug discovery programmes where their key steps are QM calculations and are therefore perfect examples of the application and use of the NWChem node.

Computational chemistry workflows used within drug discovery programmes are accessed via our KNIME web portal.[1] User group permissions regulate access to confidential information and workflows. The workflow output is sent to the user in a report (generated using the BIRT reporting tool) via email.



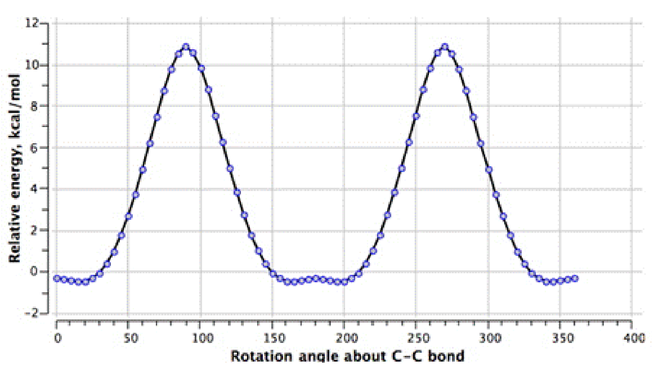


Figure 2: *(left)* Input form on the Knime web portal where all parameters needed to carry out the PES scan are specified and converted into a variable input for the NWChem node. *(right)* Output graph shows the energy penalty due to rotation about the specified bond.

1. Rigid and Relaxed Potential Energy Surface (PES) Scan

This Knime workflow computes the optimised *(relaxed)* or single point *(rigid)* energy calculations of the submitted molecule at different relative angles by rotating a specified single bond. This rotational bond is defined by the four atom numbers making up the dihedral angle associated to it (see Fig. 2).

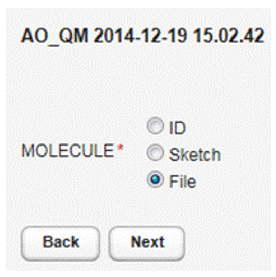
Other parameters such as initial and final values of the dihedral under study together with the number of steps between those values need to be provided by the user. Finally, the user must choose between the type of scan.

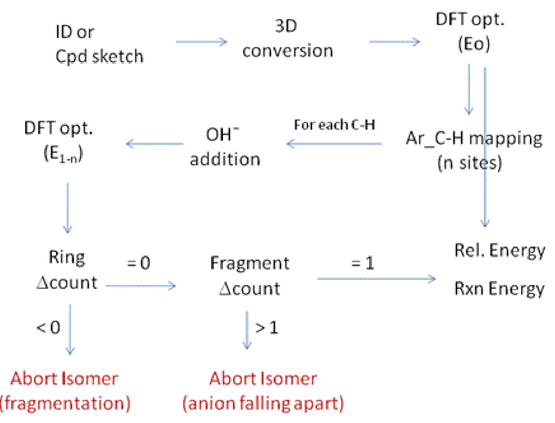
2. Aldehyde Oxidase (AO) Site of Metabolism

Aldehyde oxidase is a molybdenum hydroxylase that catalyses the oxidation of aldehydes and nitrogen containing heterocycles. This Knime workflow performs the Density Functional Theory (DFT) geometry optimization of all possible tetrahedral hydroxyl intermediates at aromatic C-H bonds of the submitted molecule to predict both the likelihood and the site of metabolism. Previous DFT calculations[5] have confirmed that the lowest energy tetrahedral intermediate from the initial substrate corresponds to the observed metabolite.

Structural checkers have been incorporated into the workflow to avoid artefacts that would lead to false conclusions. For instance, during the optimisation step the molecule can fragment after attempting hydroxyl addition or the oxidative addition may just not take place. Simply counting the number of aromatic rings and the number of fragments before and after the optimisation step allows us to remove these artefacts without further in depth analysis. The outcome from both checkers are also included in the report table (see Fig. 3).

In the Knime web portal dialogue the user can choose between three methods for introducing the molecule. The first and most popular option used by the chemists is the unique corporate ID which invokes a web service to carry out a federated search of our databases for the corresponding molecular structure. Chemical drawing on the fly or uploading the corresponding file are also valid alternatives





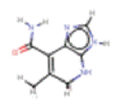
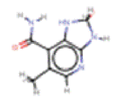
index	metabolite	E_diff	E_reaction	Ring_checker	OH_Addition_Checker
1		24.007	-66.251	ok	ok
2		0	-90.259	ok	ok

Figure 3: *(left)* Input form on the web portal where three options are given to the user to provide the molecule under investigation. *(right)* Schematic representation of the workflow behind the web portal. (middle bottom) Output table shows the potential metabolites, relative energy, reaction energy and additional checkers.

Summary

A new Knime node to carry out QM calculations (based on the NWChem software package) has been presented. KNIME provides a flexible and expandable platform to efficiently accommodate unusual types of tasks. The complexity of the presented cheminformatic workflows is alleviated by compartmentalising and further concatenation of the individual tasks. Behind the KNIME facade there are powerful computational packages that can run a significant amount of the computational chemistry tools needed in any drug discovery programme.[6]

References

1) Berthold, M., et al., *KNIME: The Konstanz Information Miner, Proceedings Studies in Classification, Data Analysis, and Knowledge Organization*, Freiburg, Germany, Springer-Verlag, 2007 (see also http://www.knime.org/)
2) *Pure Appl. Chem.,* 2000, 72, 8, 1405
3) *Comput. Phys. Commun.*, 2010, 181, 1477
4) *Appl. Magn. Reson.,* 2014, 45, 10, 981
5) *J. Med. Chem.*, 2007, 50, 19, 4642
6) *Mazanetz*, M.P., Marmon, R.J., Reisser, C.B., Morao, I. Drug discovery applications for KNIME: an open source data mining platform, *Curr. Top. Med. Chem.*, 2012, 2, 18, 1965-79

Acknowledgments

A special thank you goes to Research Informatics people at Evotec who have assisted with this development work.