

chem-bla-ics

CompLife2007, Utrecht/NL. Day 1 and 2

Egon Willighagen 

Published October 14, 2007

Citation

Willighagen, E. (2007, October 14). CompLife2007, Utrecht/NL. Day 1 and 2. *Chem-bla-ics*. <https://doi.org/10.59350/62aqh-f9y91>

Keywords

Knime, Chemspider

Copyright

Copyright © Egon Willighagen 2007. Distributed under the terms of the [Creative Commons Attribution 4.0 International License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

chem-bla-ics

CompLife 2007 was held 1.5 weeks ago in Utrecht, The Netherlands. The number of participants was much lower than last year in Cambridge. Ola and I gave a tutorial on Bioclipse, and Thorsten one on KNIME. Since a visit to Konstanz to meet the KNIME developers, I had not been able to develop a KNIME plugin, but this was a nice opportunity to finally do so. I managed to do so, and wrote up a plugin that takes InChIKeys and then goes of the ChemSpider to download MDL molfiles:

The screenshot shows the KNIME Eclipse SDK interface. The main workspace displays a workflow with two nodes: 'File Reader' (Node 2) and 'ChemSpider' (Node 1). The 'ChemSpider' node is selected, and its configuration dialog is open. The dialog shows the following details:

- ChemSpider**
- Insert long description here...
- Dialog Options**
- short name of first option (like in the dialog)
- description of first option
- short name of second option (like in the dialog)
- Ports**
- Data Input**
- 0 Description of first input port...
- Data Output**
- 0 Description of first output port...
- Views**
- name of first view Description of first view...

The 'Data Table' view shows the following data:

File	DataTableSpec	DataColumnProperties
1 Row, 2 Cols	S 1	S molFile
Row1	DEYFTQMOPDXOT-RERXVCS...	62853 -OEChem-02070708212D 84 86 0 0 0 0 0 0999 V2000 4

The console window at the bottom displays the following text:

```
*****  
*** Welcome to KNIME v1.2.1 - the Konstanz Information Miner ***  
*** Copyright, 2003-2007, Uni Konstanz and KNIME GmbH, Germany ***  
*****  
Log file is located at: /home/egonw/var/runtime-New_configuration/.metadata/knime/knime.
```

Why ChemSpider? Arbitrary. Done PubChem in the past already. Moreover, ChemSpider has the largest database of molecular structures and in that sense important to my research.

Why KNIME? Played with Taverna in the past, and expect to do much more work on Taverna in the coming year (see also [this](#) and [this](#)). Moreover, KNIME got a CDK plugin already, and the KNIME developers contributed valuable feedback to the CDK project in the last year. It was about time that I contributed something back, though the current functionality is quite limited. KNIME has a better architectural design than Taverna1, but will face tough competition with Taverna2, due next year.

The presentations

Heringa gave a presentation on network analysis, and discussed the scale-free network, hub nodes, etc, after which he gave an example on the 14-3-3 PPI family which both have promoting and inhibiting capabilities. Fraser presented work on improving microarray data analysis, by reducing non-random background noise. Schroeter presented the use of Gaussian process

chem-bla-ics

modeling in QSAR studies, which allows estimation of error bars (see DOI:[10.1002/cmdc.200700041](https://doi.org/10.1002/cmdc.200700041)). I did not feel the results were very convincing, though, but the method sounds interesting. Larhlimi presented research on network analysis of metabolic networks. His approach finds so-called minimal forward direction cuts, which identifies critical parts in the network if one is interested in repressing certain metabolic processes. Hofto presented some work on the use of DFT for proteins, and picked up that one has to do things critically to be able to reproduce binding affinities. Combinations of DFT or MM with QM are becoming popular to model binding sites. Van Lenthe presented such an approach of the second day of CompLife.

By far the most interesting talk at the conference, was the insightful presentation by [Paulien Hogeweg](#). She apparently coined the term *bioinformatics*. Anyway, she had a exciting presentation on feed-forward loops in relation to evolution, and showed correlation between jumps in FFL motifs with biodiversity. She also warned us for the Monster of Loch Ness syndrome, where computational models may indicate large underlying processes, which are not really existing. But that should be a problem that most of my readers should be aware of. She introduced evolutionary modeling, to put further restrictions on the models, to reduce the chance of finding monsters.

Hussong had an interesting presentation too, if one is interested in analysis of GC/MS or LC/MS data. He introduced a hard-modeling approach for proteomics data using wavelets technology. His angle on this was to use a wavelet that represents the isotopic pattern of a protein mass spectrum. Interestingly, the wavelet had negative intensities, something which one will never find in mass spectra. However, I seem to recall a mathematical restriction on wavelets that would forbid taking the squared version of the function. He indicated that the code is available via [OpenMS](#).

Jensen, finally, presented his work at the [UCC](#) on Markov models for protein folding, where he uses the *mean first passage time* as observable to analyze of processes in folding state space. This allows him to compare different modeling approaches and, for example, to predict how many time steps are needed to reach folding. Being able to measure characteristics of certain modeling methods, one is able to make a objective comparison. Something which allows [a fair competition](#).