Bioclipse 2:

towards integrated biocheminformatics


Ola Spjuth* and Jonathan Alvarsson

Department of Pharmaceutical Biosciences, Uppsala University, Uppsala, Sweden

* Corresponding author
email: ola.spjuth@farmbio.uu.se

Introduction and history

Bioclipse [1] is a free and open source workbench for the life sciences with advanced functionality in bioinformatics and chemioinformatics. It allows users to work with resources and entities in the life sciences, such as chemical structures, sequences, spectra, and alignments. Bioclipse 2, which was released in July 2009, constitutes a complete rewrite of the Bioclipse version published on EMBnet.news in 2007 [2] and provides more features and new graphical components that simplifies integrated life science research and development. The Bioclipse project has as of July 2009 accumulated over 28.000 downloads since its original release in 2007, and also been awarded 3 international prizes for its innovative architecture and intuitive interface.

Architecture

Bioclipse is built on Eclipse (http://www.eclipse.org), which is an open source framework that evolved from being an Integrated Development Environment (IDE) into a universal platform for constructing software applications. This provides Bioclipse with advanced plugin architecture, where all functionality is contributed via plugins. Bioclipse defines common interfaces for biological and chemical entities, such as IMolecule for chemical structures, and ISequence for biological sequences. Other plugins can operate on these entities without being aware of each other's existence, for example a tool that visualizes sequences graphically.
In Bioclipse, all functional source code contributed by plugins is collected in Bioclipse Managers; e.g. BioJava [3] contributes functionality via a

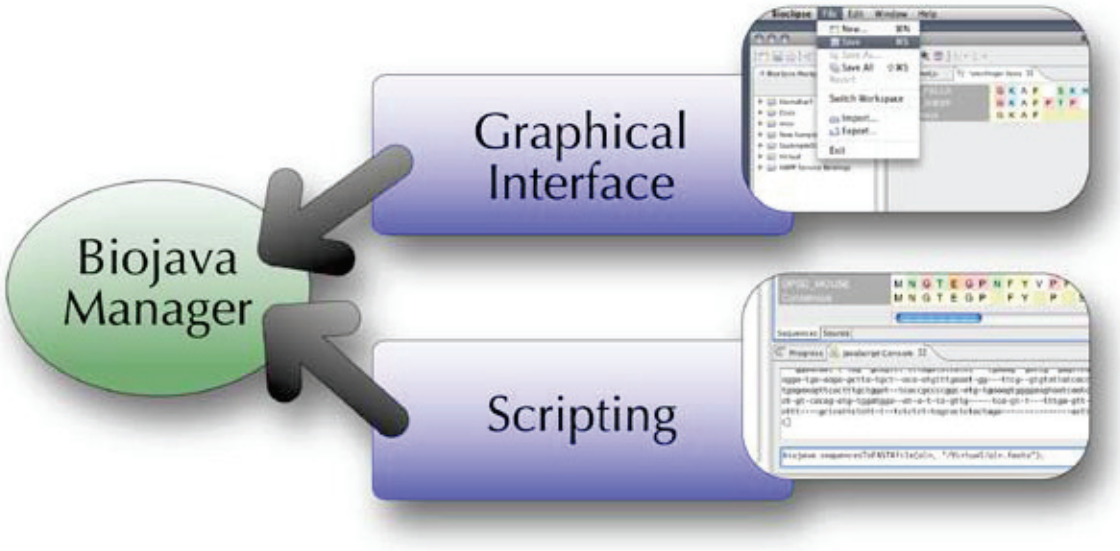


Figure 1. Overview of the Bioclipse architecture describing the use of Managers to collect functional code. The same manager is reachable both from the graphical interface and the JavaScript console, making all functionality available from the GUI and the Bioclipse Scripting Language.

```
seq1 = biojava.DNAfromPlainSequence("CTCGCTTAGAGATA", "myseq");

rna1  = biojava.DNAtoRNA(     seq1, "myRNA"     );
prot1 = biojava.DNAtoProtein( seq1, "myProtein" );

// "myseqs/sf.fasta" is a file with 2 proteins
seqs2 = biojava.proteinsFromFile("myseqs/sf.fasta");
```

Figure 2. Examples on creating and reading sequences in Bioclipse Scripting Language.

BioJavaManager. The Manager objects are built with the help of Spring (http://www.springsource.org) and published into the scripting environment. Hence, the same objects that are called from the GUI are also reachable from scripts (see Figure 1), which is named Bioclipse Scripting Language (BSL). The reference BSL is based on JavaScript, and users can invoke all functionality in Bioclipse by typing commands in the JavaScript Console. A JavaScript Editor is also included, which allows for scripting entire analyses. It is already an appreciated feature to use the graphical editors of Bioclipse together with the scripting language to solve biological problems.

Bioinformatics

The core framework for Bioinformatics in Bioclipse 2 is primarily based on Biojava [3], which is available from the BioJavaManager. Figure 2 shows some examples on how to create and read sequences on the JavaScript Console.

Bioclipse also has bioinformatics plugins that take advantage of remote functionality, such as Web services. Examples include WSDbfetch for retrieving data from public repositories, and Kalign for sequence alignments [4].

Bioclipse 2 also features new GUI components, such as a new Sequence Editor (see Figure 4) which allows for editing and visualization of sequences, including DNA, RNA, protein sequences, as well as pairwise and multiple alignments.

Cheminformatics

The core framework for Cheminformatics in Bioclipse 2 is primarily based on The Chemistry Development Kit (CDK) [5], which is available via the CDKManager. Figure 5a shows an example of how CDK can be used to create molecules via the JavaScript Console, and open them in the chemical editor JChemPaint. Bioclipse also includes features to query public repositories for chemical substances, for example PubChem. Figure 5b shows a script for querying PubChem for substances annotated with H1N1, downloading them, and visualizing them in the MoleculesTable. As in bioinformatics, the querying functionality is also available from the GUI using a wizard (see Figure 3c).

Bioclipse contains many graphical editors to empower scientists in cheminformatics. One example is the interactive 3D visualization tool Jmol (see Figure 6).

Conclusions

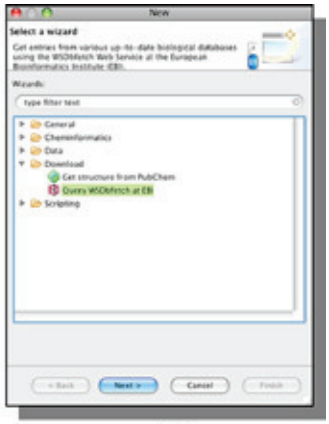
The Bioclipse project aims at providing a workbench with the commonly needed features in chem- and bioinformatics, and also to enable scientific research and development spanning multiple fields. There are ongoing projects to further develop the platform and existing features, but also many new initiatives that widen

```
biows.queryEMBL("X56734")
biows.queryRefseq("NM_000410")
biows.queryUniProtKB("INSR_HUMAN")
```

(a)

```
seqs = biows.queryEMBL("X56734,X56735");
aln = kalignws.alignDNA(seqs);
biojava.sequencesToFASTAfile(aln, "save here");
```

(b)



(c)

Figure 3. a) Three different commands to query public repositories for sequences. b) Script to query EMBL for two DNA sequences, align them using Kalign, and write the alignment to a FASTA file. c) The first page of a graphical wizard for executing the same queries as in a).

Figure 4. Screenshot from Bioclipse showing the Sequence Editor open with a file containing multiple sequences. The outline (middle frame) shows an overview of the sequences and allows for simple navigation. The JavaScript console (bottom frame) enables scripting of Bioclipse.

the scope of Bioclipse into more fields. The list includes toxicity assessment, site-of-metabolism predictions, integrated local and networked databases, and QSAR analysis. Social features such as integration with MyExperiment [6] are already available, as are features for working with semantic technologies like RDF/OWL. The Bioclipse Wiki (http://www.wiki.bioclipse.net) and the Bioclipse Blog (http://bioclipse.blogspot.com/) holds the most recent information regarding the Bioclipse development.

License and Availability

Bioclipse 2 is released under the Eclipse Public License (EPL), a flexible open source license that

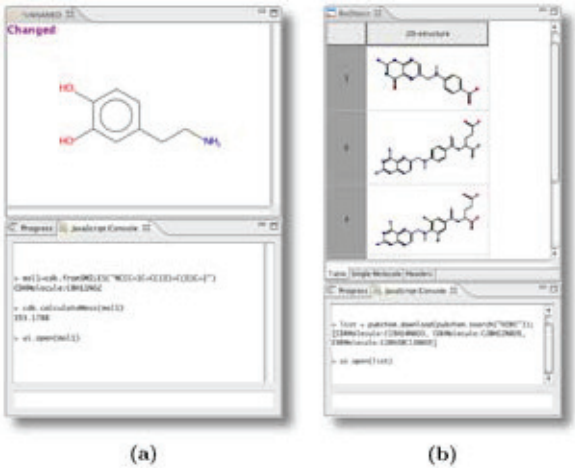


Figure 5. a) Screenshot from Bioclipse showing creation of the compound Dopamine from SMILES in the JavaScript Console, calculation of its mass, and opening of the molecule in JChemPaint. b) Screenshot showing a PubChem query with visualization of the resulting molecules in the MoleculesEditor.

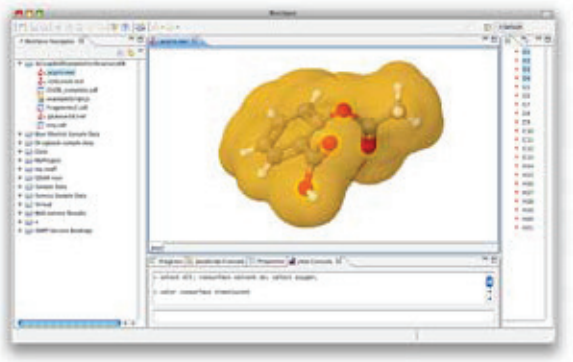


Figure 6. Bioclipse integrates advanced visualization components. Here the interactive 3D visualization tool Jmol is used for displaying an isosurface of aspirin. The Jmol Console (bottom) is used to enter Jmol commands to affect the visualization. Many of these commands are also available from the Jmol menu.

allows additional plugins to be of any license. Bioclipse 2 is implemented in Java and supported on all major platforms. Source code and binaries are freely available at http://www.bioclipse.net and development versions are available from http://pele.farmbio.uu.se/bioclipse-devel/.

References

[1] Spjuth O, Helmus T, Willighagen EL, Kuhn S, Eklund M, Wagener J, Murray-Rust P, Steinbeck C, Wikberg JES: Bioclipse: an open source workbench for chemo- and bioinformatics. BMC Bioinformatics 2007, 8:59.

[2] Spjuth O: Using Bioclipse to integrate bioinformatics functionality. EMBnet news 2007, 13 (1), 5-11]

[3] Holland RCG, Down TA, Pocock M, Prlic A, Huen D, James K, Foisy S, Drager A, Yates A, Heuer M, Schreiber MJ: BioJava: an open-source framework for bioinformatics. Bioinformatics 2008, 24(18):2096–2097.

[4] Labarga A, Valentin F, Anderson M, Lopez R: Web services at th European Bioinformatics Institute. Nucleic Acids Res 2007, 35(Web Server issue):W6–11.

[5] Steinbeck C, Hoppe C, Kuhn S, Floris M, Guha R, Willighagen EL: Recent developments of the Chemistry Development Kit (CDK) - an open-source Java library for chemo- and bioinformatics. Curr Pharm Des 2006, 12(17):2111–2120.

[6] De Roure, D., Goble, C. and Stevens, R. The Design and Realisation of the myExperiment Virtual Research Environment for Social Sharing of Workflows. Future Generation Computer Systems 2009, 25