

Bioclipse

Bioclipse

Bioclipse is a free and open source workbench for the life sciences.

Bioclipse is based on the Eclipse Rich Client Platform (RCP) which means that Bioclipse inherits a state-of-the-art plugin architecture, functionality, and visual interfaces from Eclipse, such as help system, software updates, preferences, cross-platform deployment etc.

Bioclipse provides advanced functionality in fields such as cheminformatics, bioinformatics, semantic web, spectrum analysis, drug discovery, safety assessment, and general chemistry education.

Bioclipse (2)

Bioclipse is developed as a collaboration between the Proteochemometric Group , Dept. of Pharmaceutical Biosciences, Uppsala University, Sweden, and the Cheminformatics and Metabolism Team at the European Bioinformatics Institute (EBI).

Bioclipse is released under Eclipse Public License (EPL) + exception, see the License Statement, putting no constraints on choice of backend and/or license for creating plugins for Bioclipse; it is totally open for both open source plugins as well as commercial.

Bioclipse (4)

Some examples of functionality is listed below:

Cheminformatics

- Cheminformatics in Bioclipse is mainly based on the Chemistry Development Kit (CDK), and contains a framework for managing and analyzing chemical compounds.
- Bioclipse supports editing in 2D, processing large collections of molecules in tables, calculation of various types of properties, and much more cheminformatics functionality.
- The Jmol application is integrated in Bioclipse as an editor, and provides advanced interactive 3D visualizations.

Bioclipse (5)

Illustration: Work with collections of molecular structures

Bioclipse

Bioclipse Navigator

ChEBI_complete.sdf

ola

ChEBI_complete.sdf

Sample Data

- 2D structures
- 3D Structures
- Javascripts
- PDB
- SDF

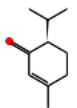
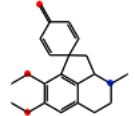
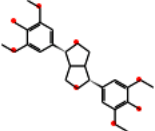
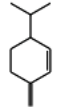
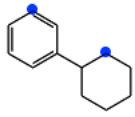
	2D-structure	IUPAC Names	SMILES	Last Modified
9		(6S)-3-methyl-6-(propan-2-yl)cyclo...	<chem>CC(C)[C@@H]1CCC(C)=CC...</chem>	25 Mar 2008
10		(8a'R)-5',6'-dimethoxy-1'-methyl-2',...	<chem>[H][C@]12CC3(C=CC(=O)C...</chem>	10 Mar 2008
11		(7alpha,7'alpha,8alpha,8'alpha)-3,3',...	<chem>[H][C@]12CO[C@H](c3cc(O...</chem>	06 May 2008
12		(4S)-p-mentha-1(7),2-diene(6S)-3-...	<chem>[H][C@]1(CCC(=C)C=C1)C(...</chem>	23 May 2008
13		3-[(2S)-piperidin-2-yl]pyridine	<chem>[H][C@]1(CCCC1)c1ccncc1</chem>	15 Jan 2008

Table | Single Molecule | Headers

Outline | Properties

Property | Value

General

- Has 2D Coords: yes
- Has 3D Coords: no
- Molecular Format: N/A
- Molecular Formula: C10H16H16
- Molecular Mass: 136.2344

Molecular Properties

- Beilstein Registry Number: 19027674229885
- CAS Registry Number: 498-15-7
- Charge: 0
- ChEBI ID: CHEBI:7
- ChEBI Name: (+)-car-3-ene
- Formulae: C10H16
- Gmelin Registry Number: 663435
- InChI: InChI=1/C10H2/c1-7-4-5
- InChI: InChI=1/C10H2/c1-7-4-5
- InChIKey: QTEBNBFBDXVOON-BDAKN
- InChIKey: QTEBNBFBDXVOON-BDAKN
- IUPAC Names: (1S,6R)-3,7,7-trimethylbicy
- KEGG COMPOUND Data: C11382
- Last Modified: 10 May 2006
- Mass: 136.23404
- PubChem Database Link: 11533292
- SMILES: CC1=CCC2C(C1)C2(C)(C)
- SMILES: CC1=CCC2C(C1)C2(C)(C)
- Synonyms: (+)-3-Carene(+)-Delta(3)-

Bioclipse (6)

Illustration: Editing of chemical structure

The screenshot displays the Bioclipse software interface. The central canvas shows the chemical structure of reserpine, a complex alkaloid with a pentacyclic core and a phosphate group. The interface is divided into several panels:

- Bioclipse Navigator:** Located on the left, it shows a tree view of the project files. Under "Sample Data", there are 2D structures (0037.cml, 0037.mdl, ATP.mol, polycarpol.mol, reserpine.mol, thiamin.mol) and 3D Structures (2-methylpropanal, pentan-1-ol.mol, pentanal.cml, propan-2-ol.cml, tetracosane.cml). There are also folders for Javascripts, PDB, and SDF (Fragments2.sdf).
- Properties:** Located at the bottom left, it displays the properties of the selected molecule (reserpine.mol). The table below shows the details:
- Outline:** Located on the right, it lists the atoms and bonds in the structure. The atoms are listed as Oxygen (O): O.sp3, Carbon (C): C.sp3, and Phosphorus (P): P.ate. The bonds are listed as N-C (single), C-N (single, aromatic), and C-C (single, aromatic).

Property	Value
Has 2D Coords	yes
Has 3D Coords	no
InChI	InChI=1/C10H19NSO13P3/c11-8-5-9(13-2-12-8)15(3-14-5)10-7(17)6(16)4
InChIKey	VIYUSENGHRDRMH-UHFFFAOYAU
Molecular Format	MDL Molfile (2D)
Molecular Formula	C10H19NSO13P3
Molecular Mass	510.2055
SMILES	O=P(O)(O)OP(=O)(O)(O)OP(=O)(O)OCC3OC(N2CNC1C(N)CNC12)C(O)C3(O)

Bioclipse (7)

Illustration: Compound visualized in 3D with Jmol

The screenshot displays the Bioclipse application window. The central 3D viewer shows a ball-and-stick model of a diazepam molecule, with a semi-transparent orange surface representing its molecular surface. The molecule features a benzodiazepine core with a diazepam substituent. The interface includes several panels:

- Bioclipse Navigator** (left): A tree view showing the chemical structure repository hierarchy. The path `BOdata/Blue Obelisk Chemical Structure Repository/drugs/diazepam.mol` is selected.
- Outline** (top right): A list of atoms in the molecule, including C1, C2, C3, C4, C5, C6, C7, N8, C9, C10, N11, C12, C13, C14, O15, C16, and C17.
- Properties** (bottom right): A table showing the properties of the selected molecule.
- JavaScript Console** (bottom): A text area showing the execution of JavaScript commands for molecule creation and mass calculation.

Properties Table:

Property	Value
derived	false
editable	true
last modified	February 13, 200
linked	false
location	/Users/ola/Libra
name	diazepam.mol
path	/BOdata/Blue Obe
size	2,419 bytes

JavaScript Console Output:

```
org.mozilla.javascript.Undefined@dee89b  
fromCml fromSMILES fromString  
  
> mol=cdk.fromSMILES("C1CCCCC1CCOCCO")  
CDKMolecule:C10H20O2  
  
> cdk.calculateMass(mol)  
152.10616882301883
```

Bioclipse (8)

Pharmacology and Drug Discovery

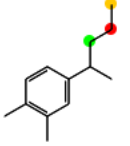
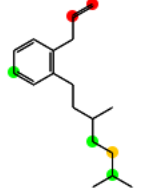
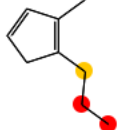
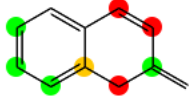
- Bioclipse is equipped with many features that simplify pharmacological research and drug discovery. Quantitative Structure-Activity Relationships (QSAR) is a methodology to relates the responses between several chemical structures and a target, for example to measure the toxicity of drugs in the human body, described by mathematical descriptors and modelled using statistical methods.
- QSAR models can for example predict if a compound is toxic, and scientists can get decision support for changing the chemical structure before even testing it in the wet lab.
- Bioclipse has many features to work with QSAR and similar fields. A QSAR Project is available with a Builder on a QSAR project file, completely describing the project model. A editor is available, with tabs to select chemical structures, choose mathematical descriptions (descriptors), and all other metadata for performing the analysis.

Bioclipse (9)

Illustration: Predicting site-of-metabolism on multiple molecules

The screenshot displays the Bioclipse software interface. On the left, the 'Bioclipse Navigator' shows a file tree with 'ola' as the root, containing various SDF and CML files. '20Mols2d.sdf' is highlighted. Below the navigator is the 'MetaPrint2D Report' section, which shows the results of the last MetaPrint2D run: Status OK, Database ALL, Operator DEFAULT, and Calculation time 49 ms.

The main window displays a table with four rows, each representing a molecule. The columns are '2D-structure' and 'cdk:Title'. The molecules shown are adrenaline, alprenolol, clomethiazole, and coumarin. Each molecule's 2D structure is visualized in the '2D-structure' column, with atoms colored by their predicted site-of-metabolism: green for C, red for O, yellow for N, and blue for S.

	2D-structure	cdk:Title
1		adrenaline
2		alprenolol
3		clomethiazole
4		coumarin

At the bottom of the table, there are tabs for 'Table', 'Single Molecule', and 'Headers'. The status bar at the very bottom shows 'ola/20Mols2d.sdf'.

Bioclipse (10)

Bioinformatics

- Bioinformatics in Bioclipse concerns primarily the management and analysis of biological sequences (DNA, RNA, and protein).
- Bioinformatics in Bioclipse relies heavily on BioJava, which provides core bioinformatics functionality, and a graphical editor for sequence alignments.
- Various clients for Web services are also available to facilitate downloading of e.g. biological sequences and annotations, as well as for bioinformatics analysis.

Bioclipse (11)

Illustration: Sequence Editor displaying a wrapped alignment

The screenshot displays the Bioclipse Sequence Editor interface. The main window shows a wrapped alignment of three sequences: OPSD_FELCA, OPSD_SHEEP, and their Consensus. The alignment is displayed in a grid format with columns for each sequence and rows for the alignment segments. The sequences are color-coded by amino acid type. The Bioclipse Navigator on the left shows the project structure, including MyProject, MyQSARproject, and seqs. The JavaScript Console at the bottom shows a query for the OPSD_SHEEP protein.

Sequences | Source

Outline

Properties | Progress | JavaScript Console

```
> biows.queryUniProtKB("OPSD_SHEEP")
[Protein OPSD_SHEEP:
'mngtegnfyvpsnktgvrrspfeapqyilaepwqfsmloaymflilvgfpinftlyvtvghkkrlrtplnyilln
lavadlfmvfggftttlyslhgyfvfgptgcnlegffatlggeialwslvlaieryvvvckpmsnfrfgenhaimgv
aftwmalacaapplvgwsryipagmqscgalftlkpeinnesfviymfvhfsipliviffcygqlvftvkeaaaq
qesattkaekvtrmviimviaflcwlpyagvafyifthagsgdgpifmtipaffaksssvynpviymnkqfrn
cmlttlccgknpdgddeastvsktetsqvapa']
```

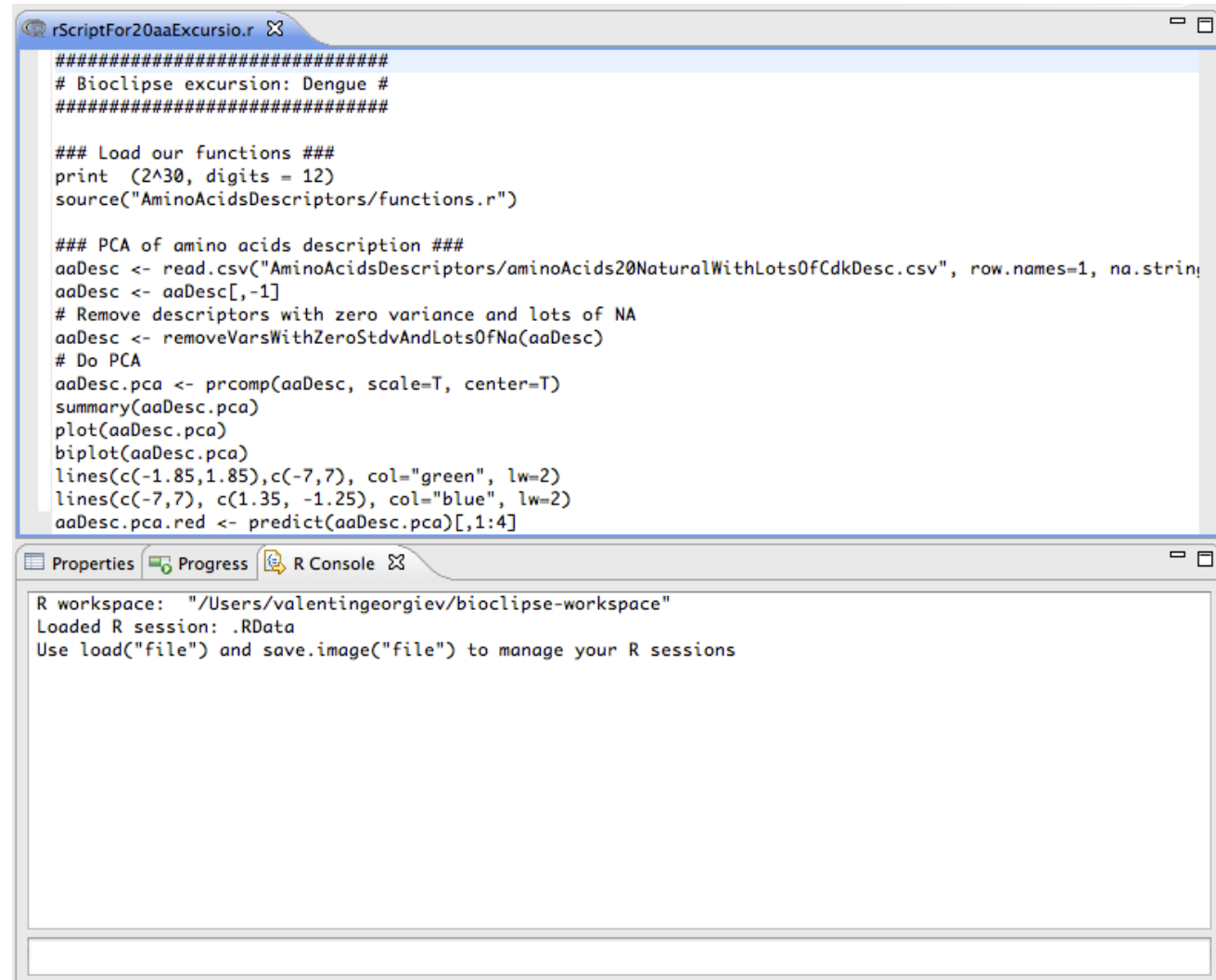
Bioclipse (12)

Bioclipse-R Itegration

- The statistical computing language R is now integrated in Bioclipse.
- R is a free and open source software that runs on Windows, Mac OS, and a wide variety of UNIX platforms and similar systems (including FreeBSD and GNU/Linux).
- The Bioclipse-R feature provides a graphical R editor and a interactive R console for easy running of R scripts, snippets and commands, and the plotting capabilities of R.

Bioclipse (13)

Illustration: Integration of the R into Bioclipse workbench



```
rScriptFor20aaExcursio.r
#####
# Bioclipse excursion: Dengue #
#####

### Load our functions ###
print (2^30, digits = 12)
source("AminoAcidsDescriptors/functions.r")

### PCA of amino acids description ###
aaDesc <- read.csv("AminoAcidsDescriptors/aminoAcids20NaturalWithLotsOfCdkDesc.csv", row.names=1, na.string="")
aaDesc <- aaDesc[,-1]
# Remove descriptors with zero variance and lots of NA
aaDesc <- removeVarsWithZeroStdvAndLotsOfNa(aaDesc)
# Do PCA
aaDesc.pca <- prcomp(aaDesc, scale=T, center=T)
summary(aaDesc.pca)
plot(aaDesc.pca)
biplot(aaDesc.pca)
lines(c(-1.85,1.85),c(-7,7), col="green", lw=2)
lines(c(-7,7), c(1.35, -1.25), col="blue", lw=2)
aaDesc.pca.red <- predict(aaDesc.pca)[,1:4]
```

Properties Progress R Console

R workspace: "/Users/valentingeorgiev/bioclipse-workspace"
Loaded R session: .RData
Use load("file") and save.image("file") to manage your R sessions

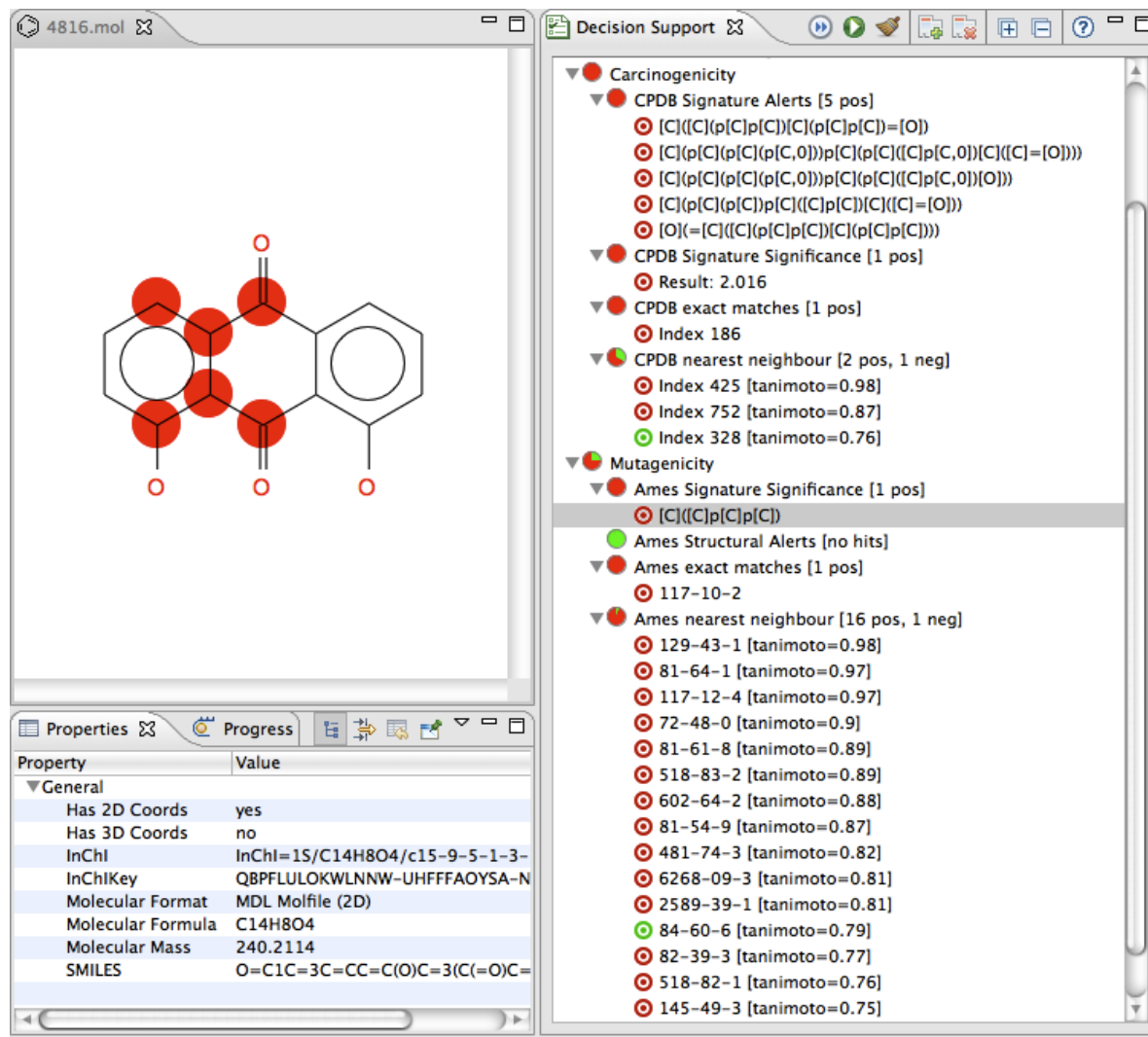
Bioclipse (14)

Decision support

- Bioclipse facilitates working with decision support systems, for example when predicting the susceptibility of a drug for certain patients. By sequencing the DNA of the patient (or e.g. a virus in the patient), it is possible to predict what drugs that would best attack the disease. An example of this is HIVPred, which is implemented as an XMPP cloud service and a plugin in Bioclipse to facilitate invocation.
- Bioclipse also has a feature for decision support in safety assessment with graphical views and editors for executing and integrating various computational models to predict the safety of chemical compounds.

Bioclipse (15)

Illustration: Chemical liability assessment in the Bioclipse workbench



Decision Support

- Carcinogenicity**
 - CPDB Signature Alerts [5 pos]
 - [C]([C](p[C]p[C])C)(p[C]p[C])=O
 - [C](p[C](p[C](p[C,0]))p[C](p[C]([C]p[C,0])[C]([C]=O)))
 - [C](p[C](p[C](p[C,0]))p[C](p[C]([C]p[C,0])[O]))
 - [C](p[C](p[C])p[C]([C]p[C])C)([C]=O)
 - [O]([C]([C]p[C]p[C])C)(p[C]p[C])
 - CPDB Signature Significance [1 pos]
 - Result: 2.016
 - CPDB exact matches [1 pos]
 - Index 186
 - CPDB nearest neighbour [2 pos, 1 neg]
 - Index 425 [tanimoto=0.98]
 - Index 752 [tanimoto=0.87]
 - Index 328 [tanimoto=0.76]
- Mutagenicity**
 - Ames Signature Significance [1 pos]
 - [C]([C]p[C]p[C])
 - Ames Structural Alerts [no hits]
 - Ames exact matches [1 pos]
 - 117-10-2
 - Ames nearest neighbour [16 pos, 1 neg]
 - 129-43-1 [tanimoto=0.98]
 - 81-64-1 [tanimoto=0.97]
 - 117-12-4 [tanimoto=0.97]
 - 72-48-0 [tanimoto=0.9]
 - 81-61-8 [tanimoto=0.89]
 - 518-83-2 [tanimoto=0.89]
 - 602-64-2 [tanimoto=0.88]
 - 81-54-9 [tanimoto=0.87]
 - 481-74-3 [tanimoto=0.82]
 - 6268-09-3 [tanimoto=0.81]
 - 2589-39-1 [tanimoto=0.81]
 - 84-60-6 [tanimoto=0.79]
 - 82-39-3 [tanimoto=0.77]
 - 518-82-1 [tanimoto=0.76]
 - 145-49-3 [tanimoto=0.75]

Properties

Property	Value
General	
Has 2D Coords	yes
Has 3D Coords	no
InChI	InChI=1S/C14H8O4/c15-9-5-1-3-
InChIKey	QBPFLLULOKWLNW-UHFFFAOYSA-N
Molecular Format	MDL Molfile (2D)
Molecular Formula	C14H8O4
Molecular Mass	240.2114
SMILES	O=C1C=3C=CC=C(O)C=3(C(=O)C=

Bioclipse (16)

Illustration: Chemical liability assessment for multiple structures with results in spreadsheet

The screenshot displays the Bioclipse software interface. On the left is the 'Bioclipse Navigator' panel showing a file tree with folders like 'ames-test', 'balloontest', 'ChiralSignatures', 'Drugbank', 'ds-data', 'Gists', 'MyExperiment', 'myproject', 'olaQQQ', 'Sample Data', 'test', 'Virtual', 'wee', and 'XMPP Service Bindings'. The main window shows a spreadsheet titled 'part1.sdf' with columns for chemical structure, DrugBank ID, and various assay results. The spreadsheet contains 12 rows of data, each representing a different chemical structure and its corresponding assay results.

	2D-structure	DRUGBANK_ID	Ames Consensus	Ames exact mat...	Ames nearest n...	CPDB nearest n...	Ames Signature...	CPDB Signature...
5		DB00139	NEGATIVE [1 hits]	NEGATIVE [1 hits]	NEGATIVE [1 hits]	INCONCLUSIVE [...]	NEGATIVE [1 hits]	NEGATIVE [1 hits]
6		DB00140	POSITIVE [1 hits]	INCONCLUSIVE [...]	POSITIVE [1 hits]	INCONCLUSIVE [...]	POSITIVE [1 hits]	POSITIVE [1 hits]
7		DB00141	NEGATIVE [1 hits]	INCONCLUSIVE [...]	INCONCLUSIVE [...]	INCONCLUSIVE [...]	NEGATIVE [1 hits]	NEGATIVE [1 hits]
8		DB00142	NEGATIVE [1 hits]	NEGATIVE [1 hits]	NEGATIVE [1 hits]	NEGATIVE [1 hits]	NEGATIVE [1 hits]	NEGATIVE [1 hits]
9		DB00143	POSITIVE [1 hits]	POSITIVE [1 hits]	POSITIVE [1 hits]	INCONCLUSIVE [...]	POSITIVE [1 hits]	NEGATIVE [1 hits]
10		DB00144	NEGATIVE [1 hits]	INCONCLUSIVE [...]	INCONCLUSIVE [...]	INCONCLUSIVE [...]	NEGATIVE [1 hits]	NEGATIVE [1 hits]
11		DB00145	NEGATIVE [1 hits]	NEGATIVE [1 hits]	NEGATIVE [1 hits]	INCONCLUSIVE [...]	NEGATIVE [1 hits]	NEGATIVE [1 hits]
12		DB00146	NEGATIVE [1 hits]	INCONCLUSIVE [...]	NEGATIVE [1 hits]	NEGATIVE [1 hits]	NEGATIVE [1 hits]	NEGATIVE [1 hits]

At the bottom of the spreadsheet, there are tabs for 'Table', 'Single Molecule', and 'Headers'. The status bar at the bottom left indicates '0 items selected'.

Acknowledgments

Some material of the presentation is taken from
<http://research.microsoft.com/en-us/projects/chem4word/>

Some material of the presentation is taken from
<http://www.xml-cml.org/>

Some material of the presentation is taken from
<http://www.bioclipse.net/>