TECHNICAL NOTE

Integration of MoKa, VolSurf+, and MetaSite into Pipeline Pilot

Simon Cross, Gianluca Sforna, and Riccardo Vianello
Molecular Discovery Ltd, 215 Marsh Road, Pinner, Middlesex, London HA5 5NE, UK

Introduction

The GRID forcefield[1] was specifically designed to describe molecular interactions, and in particular with biomolecules such as drug targets. It is calibrated empirically and describes electrostatic, hydrophobic, and steric potentials around a target molecule through the interactions with over 60 chemical probes. Recently, the GRID Molecular Interaction Fields (MIFs) have been applied to generate a large number of molecular descriptors, modelling various physicochemical and ADME properties such as pKa, logP, logD, tautomerism, solubility, volume of distribution, Caco-2 permeation, blood-brain barrier permeation, metabolic stability, and site of metabolism prediction, in the software MoKa[2], VolSurf+[3], and MetaSite[4]. The close collaboration between Molecular Discovery and SciTegic® has led to the integration of these methods into Pipeline Pilot™, and this document aims to provide technical information about the integration and various examples are shown below to illustrate how to use the components.

General Integration Notes

Molecular Discovery's MoKa, VolSurf+, and MetaSite software is available for both the Microsoft® Windows® and Linux platforms, and a command-line version is available in addition to a self-contained graphical interface. The current implementation has been tested with MoKa 1.1, VolSurf+ 1.0, and MetaSite 3.0 versions. Where there is overlap with existing components (e.g. Pipeline Pilot native pK_a component), the Molecular Discovery implementation was designed to match the original input/output specification to facilitate any replacement in existing protocols.

Integration of MoKa into Pipeline Pilot

There are four components available for MoKa:

1. pKa
2. Ionize Molecule at pH
3. Enumerate Ionization States
4. Enumerate Tautomers

The first MoKa component directly replaces the SciTegic 'pKa' component and the example below illustrates piping structures from the SD Reader component into the MoKa pKa component before outputting all sites. The Atom Properties component then assigns the pKa values to the atom labels before the results are shown in a table viewer.

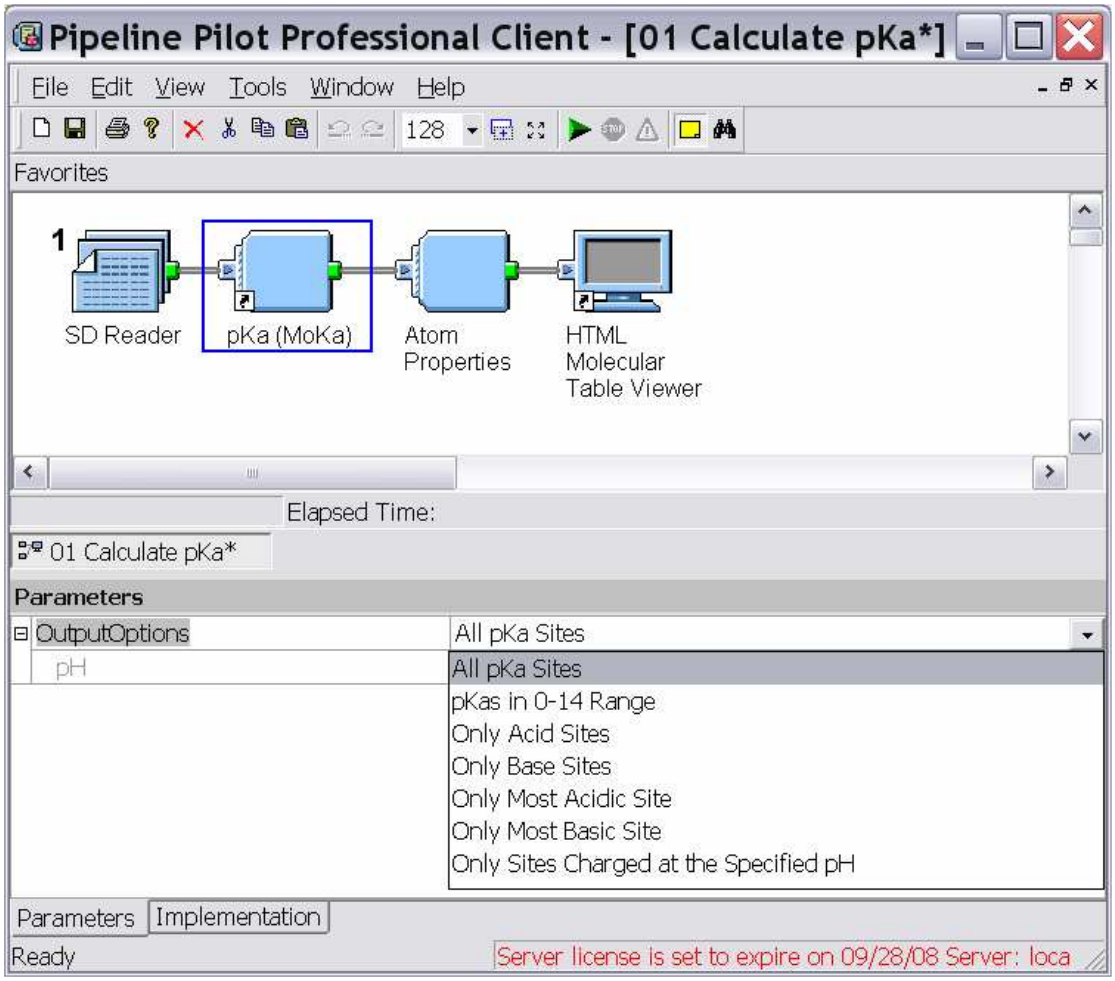


Figure 1. Pipeline illustrating how the MoKa pKa component can be used to generate an HTML table output with 2D structures and pKa values as atom labels.

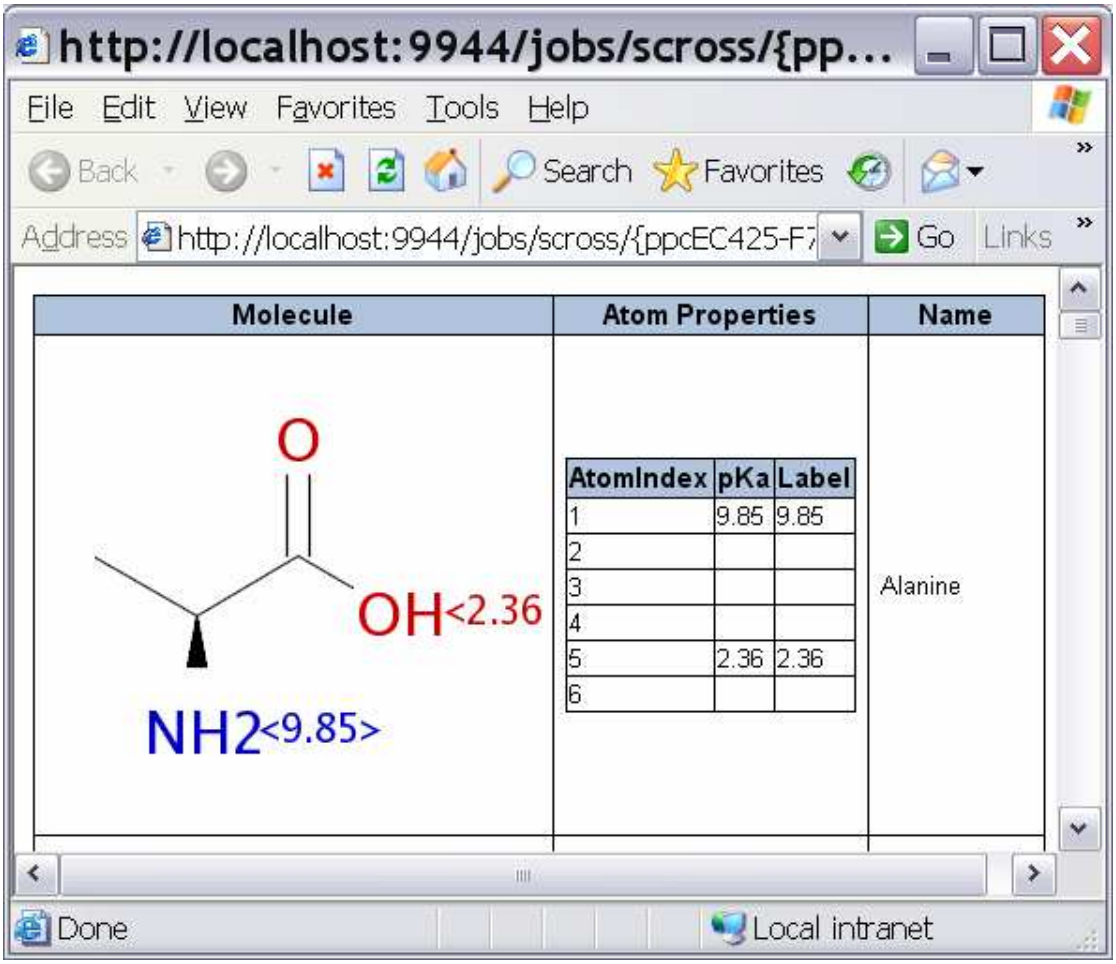


Figure 2. Example output generated by using the MoKa pKa pipeline shown in Figure 1.

The remaining MoKa components also directly replace the SciTegic equivalent 'Ionize Molecule at pH', 'Enumerate Ionization States', and 'Enumerate Tautomers' components and the example below illustrates how the 'Enumerate Tautomers (MoKa)' component can be used in a pipeline where structures are piped in through the SD Reader component, tautomers are enumerated, and then piped into an HTML results table.

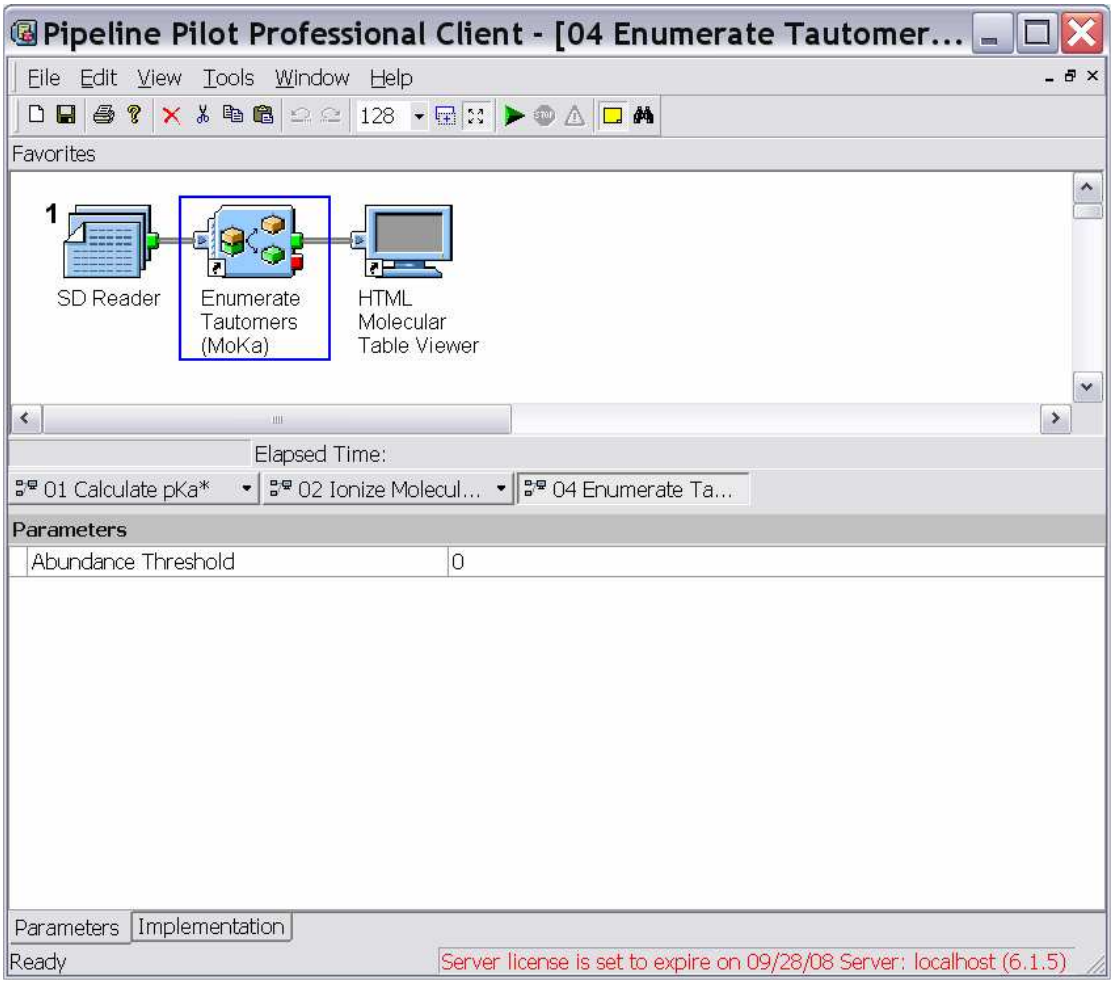


Figure 3. Example illustrating how structures can be piped from the SD Reader into the 'Enumerate Tautomers (MoKa)' component and then into an HTML table view. The abundance threshold percentage can be altered to filter out minor tautomers.

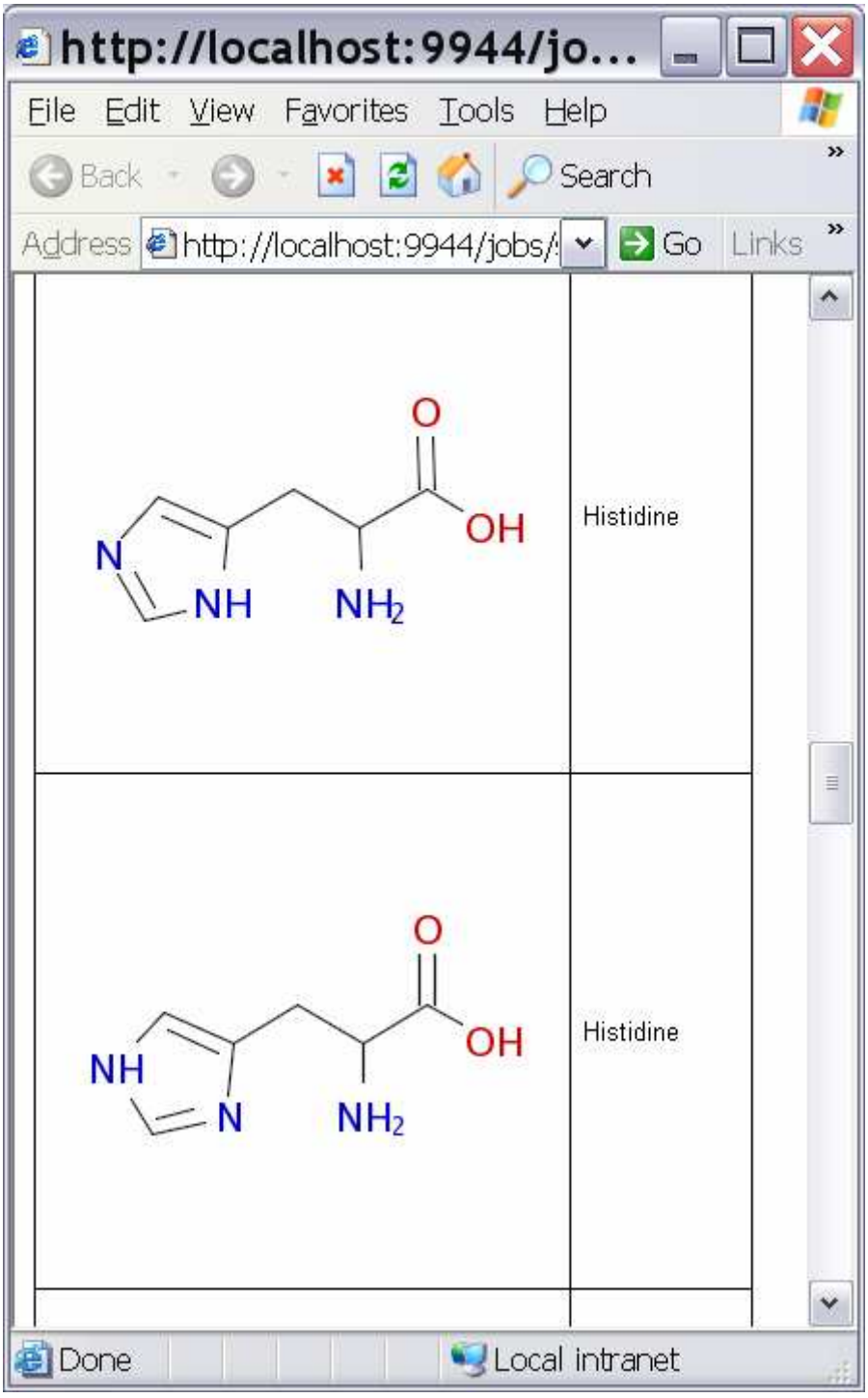


Figure 4. HTML output table produced from the pipeline shown in Figure 3, focusing on the Histidine tautomers. Note that changing the abundance threshold to 1 (%) filters out the minor tautomer shown in the top of the figure.

Integration of MetaSite into Pipeline Pilot

There are two components available for MetaSite:

1. MetaSite SoM Prediction
2. Enumerate Metabolites

The first component uses the MetaSite method to predict the 'site of metabolism' probability for each atom in each structure, according to a range of Cytochrome P450 enzymes. The specific Cytochrome can be changed under the CYP Models property, and options include: CYP1A1, CYP1A2, CYP2B6, CYP2C19, CYP2C9, CYP2D6, CYP2E1, CYP3A4, CYP3A5, LIVER, SKIN, BRAIN. Other options include changing the protonation policy on the molecule and the conformational sampling that is applied before the method is applied.

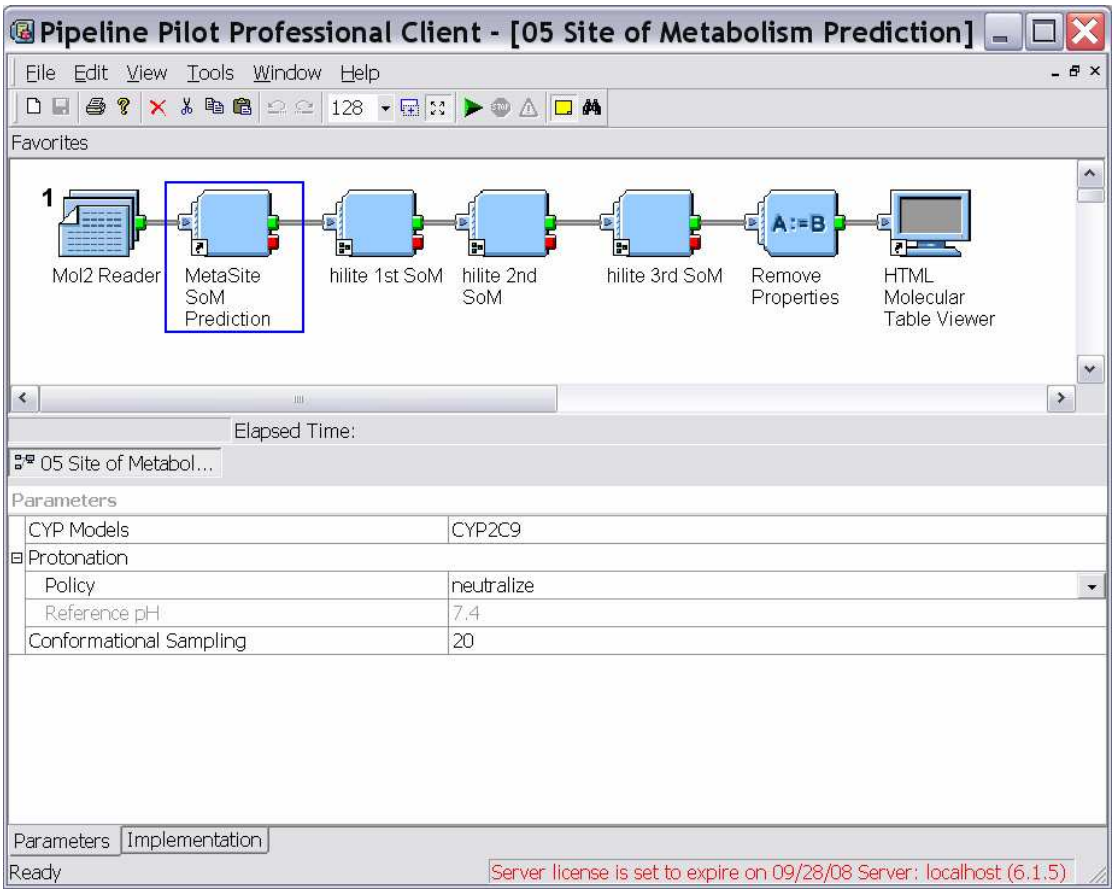
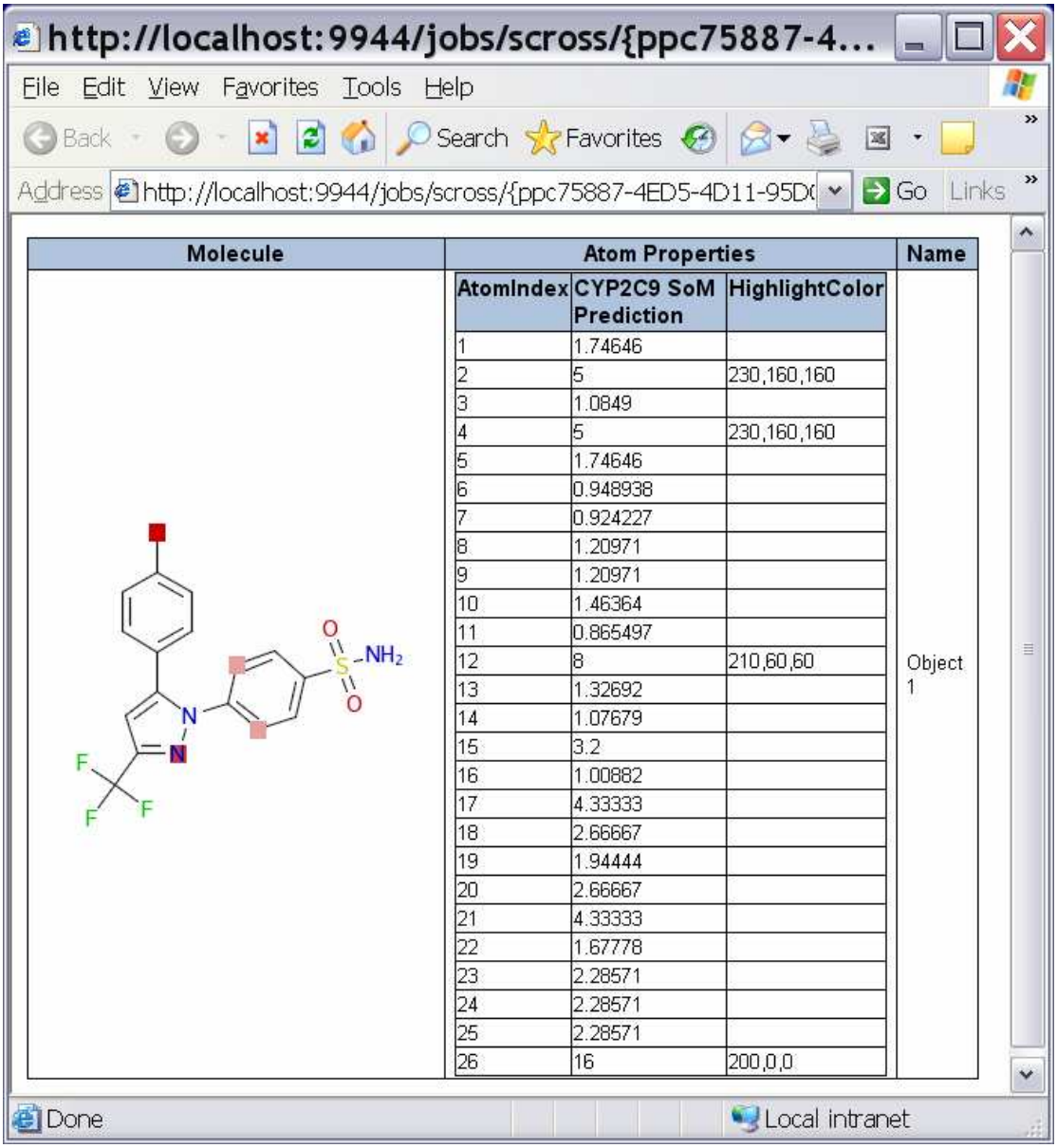


Figure 5. Example illustrating how the 'MetaSite SoM Prediction' component can be used in a pipeline. In this example, once the method is applied, some custom components are used to highlight the top three most likely atoms and the results are then displayed in an HTML molecular table.



Molecule	Atom Properties		Name
	AtomIndex	CYP2C9 SoM Prediction	HighlightColor	
	1	1.74646		
	2	5	230,160,160	
	3	1.0849		
	4	5	230,160,160	
	5	1.74646		
	6	0.948938		
	7	0.924227		
	8	1.20971		
	9	1.20971		
	10	1.46364		
	11	0.865497		
	12	8	210,60,60	Object 1
	13	1.32692		
	14	1.07679		
	15	3.2		
	16	1.00882		
	17	4.33333		
	18	2.66667		
	19	1.94444		
	20	2.66667		
	21	4.33333		
	22	1.67778		
	23	2.28571		
	24	2.28571		
	25	2.28571		
	26	16	200,0,0	

Figure 6. HTML molecular table produced by the pipeline in Figure 5. Celecoxib has been run through the pipeline prediction using the CYP2C9 model, and the results show that the most likely site of metabolism is the benzylic methyl atom 26, coloured red.

The second 'Enumerate Metabolites' component uses the MetaSite method to predict the site of metabolism probabilities for each atom, then the most likely sites are subjected to 22 reaction mechanisms to produce the enumerated metabolites. This component contains the additional parameter 'Limit' which filters the enumerated metabolites to output the top N most probable metabolites.

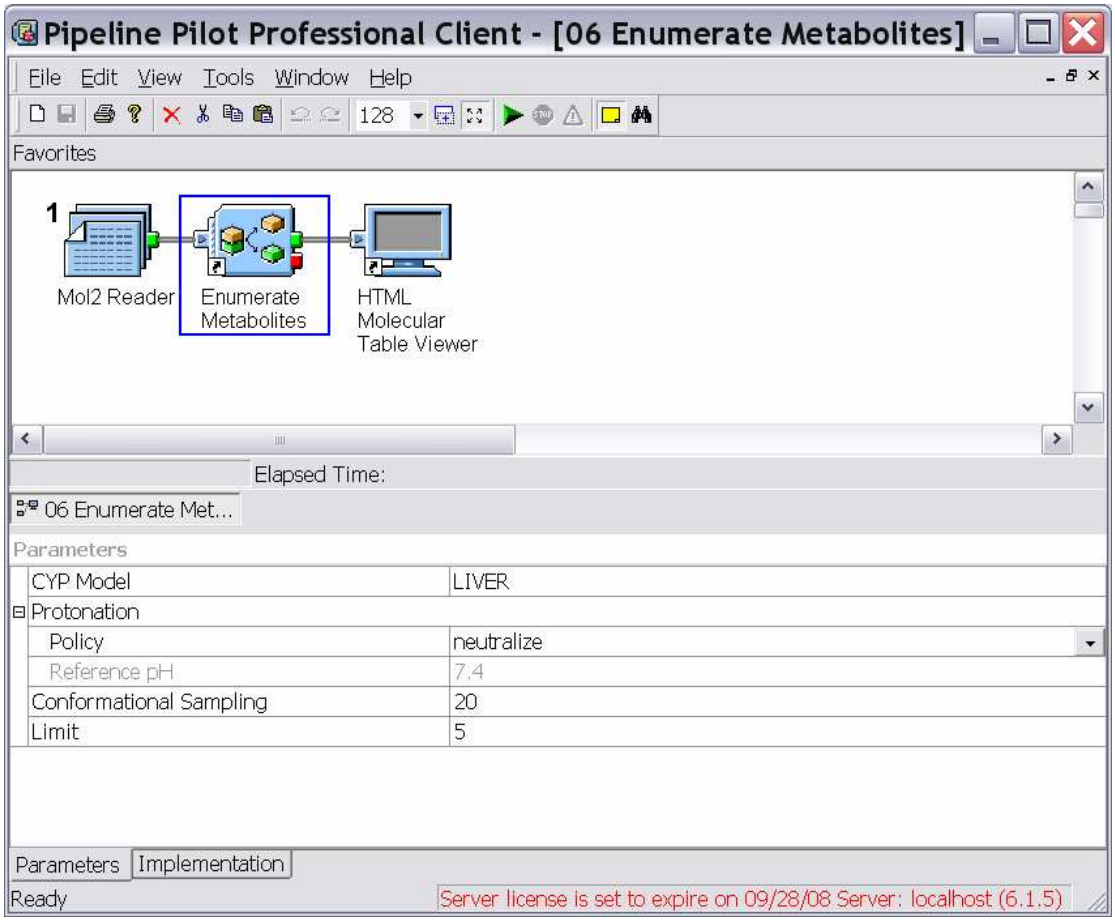
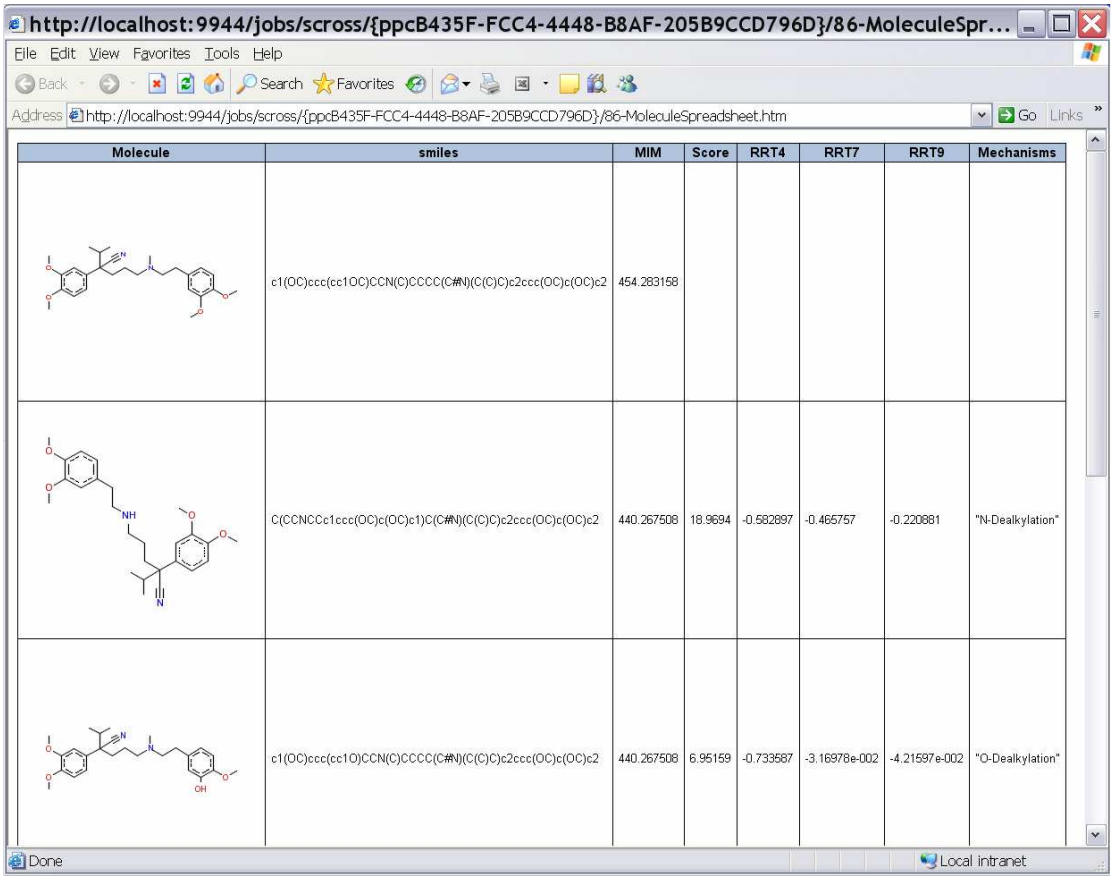


Figure 7. Example showing how a structure is piped into the MetaSite 'Enumerate Metabolites' component and then the results are piped into an HTML molecular table viewer.



Molecule	smiles	MIM	Score	RRT4	RRT7	RRT9	Mechanisms
	c1(OC)ccc(cc1OC)CCN(C)CCCC(C#N)(C(C)C)c2ccc(OC)c(OC)c2	454.283158					
	C(CCNCCc1ccc(OC)c(OC)c1)C(C#N)(C(C)C)c2ccc(OC)c(OC)c2	440.267508	18.9694	-0.582897	-0.465757	-0.220881	"N-Dealkylation"
	c1(OC)ccc(cc1O)CCN(C)CCCC(C#N)(C(C)C)c2ccc(OC)c(OC)c2	440.267508	6.95159	-0.733587	-3.16978e-002	-4.21597e-002	"O-Dealkylation"

Figure 8. Results from the pipeline shown in Figure 7. The table shows the original compound, Verapamil, on the first row. The metabolites, enumerated using the LIVER consensus model, are listed on subsequent rows in order of likelihood, with norverapamil predicted as most probable. The other columns show the smiles, minimum isotopic mass, probability score, relative retention times at pH 4,7,9 (to aid with chromatographic analysis), and the reaction mechanism involved in producing the metabolite.

Integration of VolSurf+ into Pipeline Pilot

There is one component available for VolSurf+:

1. VolSurf+ Descriptors

The component uses VolSurf+ to calculate 128 descriptors that are designed for modelling ADME phenomena. Six of these descriptors are based on models of Protein Binding, Volume of Distribution, Caco2 cell permeability, Skin permeability, Blood-Brain Barrier permeation, and Metabolic Stability. The example below shows how these key models can be used to generate an ADME report from an SD file.

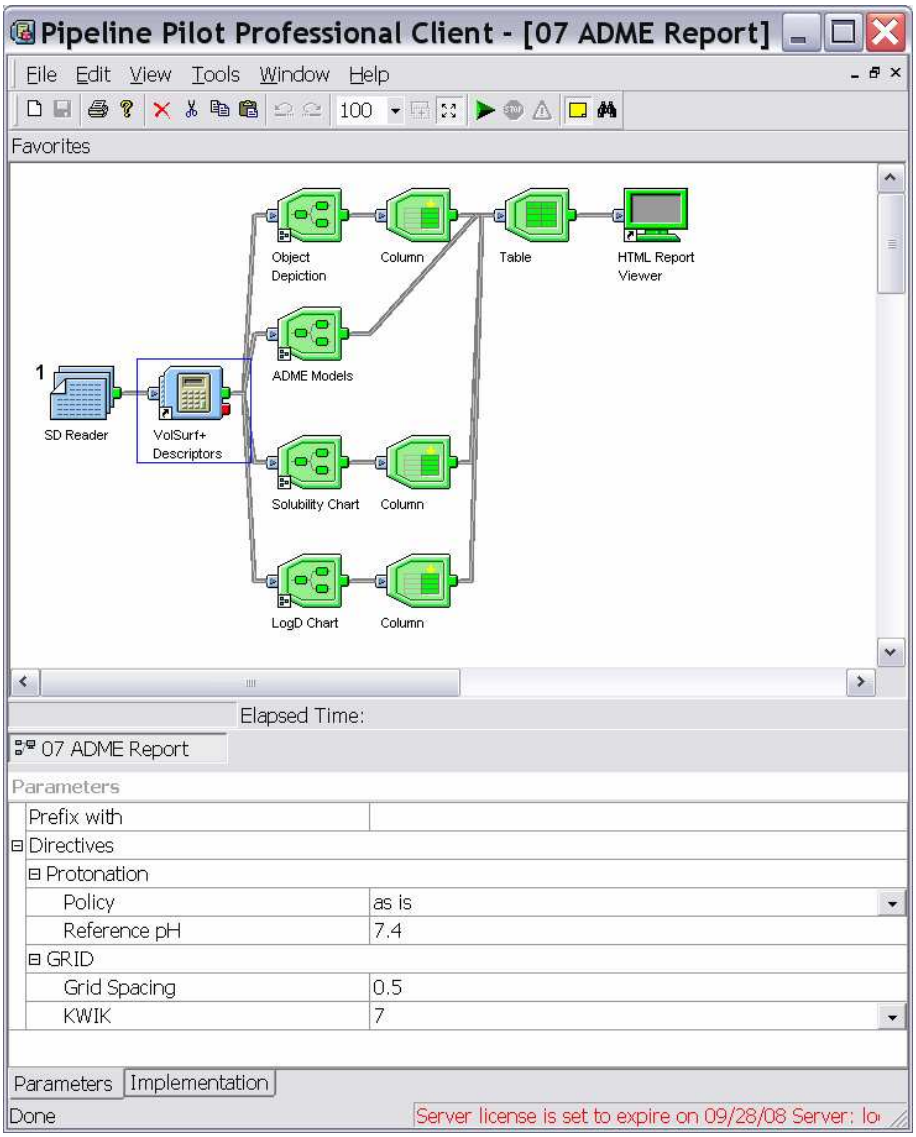


Figure 9. Example illustrating how structures from an SD file are piped into the VolSurf+ Descriptors component to create an HTML ADME report.

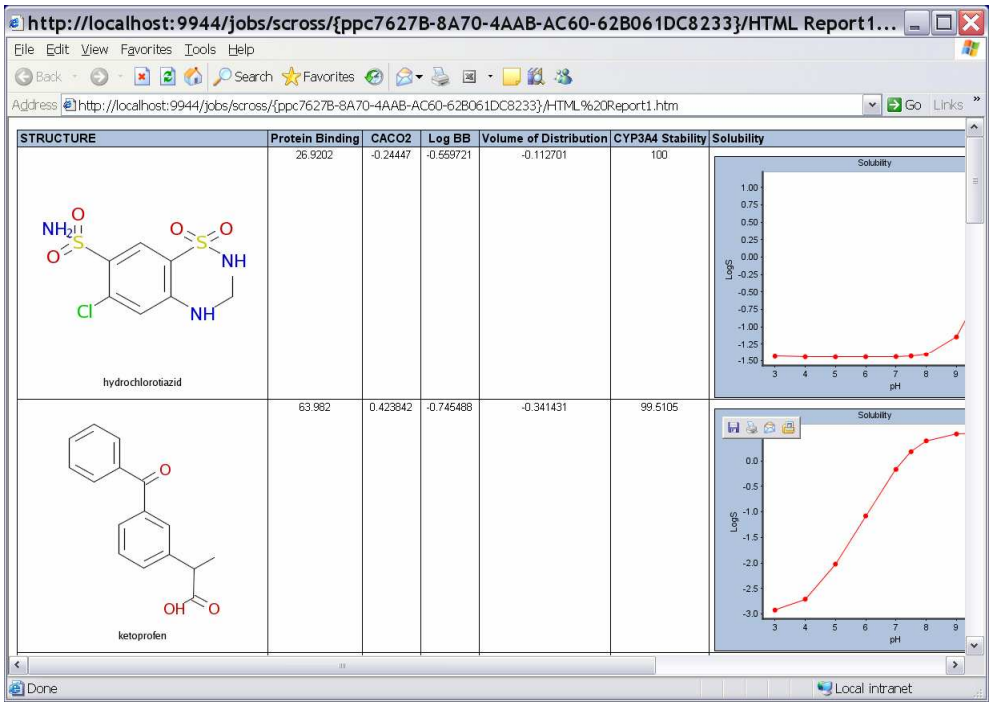


Figure 10. Results from the pipeline shown in Figure 9 illustrating an ADME report. In addition to the 2D structures, Protein Binding, Caco2 cell permeation, Log Blood-Brain Barrier permeation, Volume of Distribution, CYP3A4 Stability values are shown, followed by solubility and logD profiles by varying pH.

Summary

This technical note has described the integration of Molecular Discovery's scientific software, with a particular focus on physicochemical and ADME relevant prediction. The components described follow the same input/output rules, where possible, as those already present in Pipeline Pilot, enabling them to be used seamlessly with existing protocols. The methods are described in greater detail in the scientific references, and also our website www.moldiscovery.com. If you wish to enquire further about how you can access these methods, please contact Simon Cross by email at simon@moldiscovery.com.

1. Goodford, P.J. (1985) A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.*, 28, 849-857

2. Milletti, F. *et al.* (2007) New and Original pKa Prediction Method Using Grid Molecular Interaction Fields. J. Chem. Inf. Model., 47, 2172-2181

3. Cruciani G. *et al.* (2000) VolSurf: a new tool for the pharmacokinetic optimization of lead compounds. *Eur. J. Pharm. Sci.*, 11, S29–S39

4. Cruciani, G. *et al.* (2005) MetaSite: Understanding Metabolism in Human Cytochromes from the Perspective of the Chemist. J. Med. Chem. 2005, 48, 6970-6979