

Pipeline Pilot Web API

Exploitation via spotfire & excel

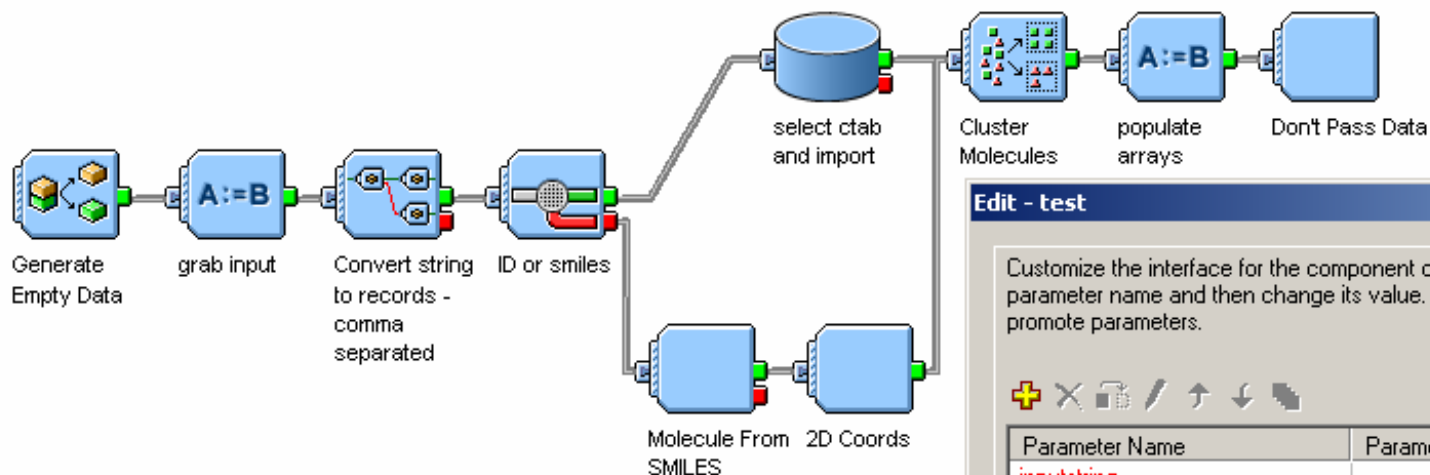
Outline – Spotfire and Excel PP Plugins

- **Background and motive**
 - Why target Spotfire and Excel
- **Examples**
 - **Spotfire**
 - Clustering
 - Fragmentation
 - Rgroup analysis
 - **Excel**
 - Enumeration
 - Property Calculation
- **Conclusions but Caveats**
 - Issues with software
 - Issues with strategy

Background

- **Spotfire & Excel – Popular for organising and visualising data**
- **Pipeline Pilot – used for rapid protocol development**
- **Global Web API license**
- **Increased Exploitation of Pipeline Pilot**
 - local use of the thick client is limited by license and expertise
 - Target applications already used by chemists and extend their functionality using PP

Simple Clustering Protocol



Edit - test

Customize the interface for the component or protocol. parameter name and then change its value. Use the promote parameters.



Parameter Name	Parameter Value
inputstring	
Tempfiles	
WebExports	Cluster_Array
RunToCompletion	False
AvgNumberPerCluster	
MaximumDistance	
NumberOfClusters	20
PredefinedSet	FCFP_4

- **Accepts ID or smiles string array from soap call and converts to data stream**
- **Clusters and returns cluster membership and centres as array to javascript callback function.**

Login Protocols Execution Results

Enter your username and click **Login** to login to Pipeline Pilot.

Login as:

Login



Login

Protocols

Execution

Results

Enter values for the parameters listed below, and click Run Protocol to begin its execution. Parameters displayed in red are required.

sdf to spotfire

A top-level component for building protocols.

Parameters:

sdffile




[Add a file](#)

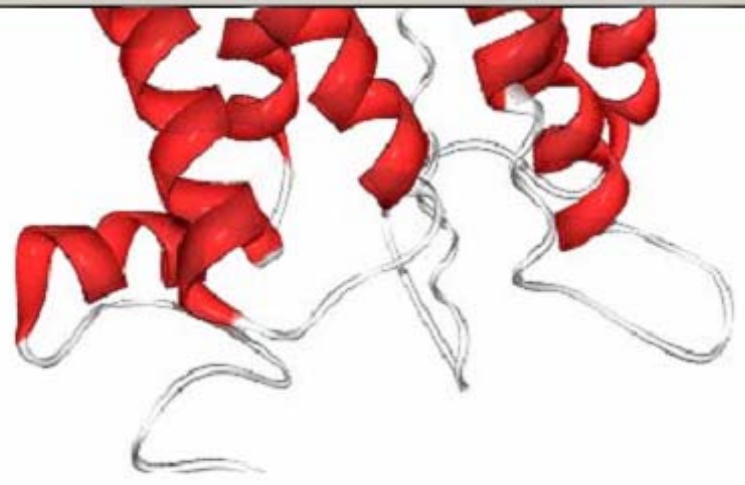
Maximum

calculate smiles

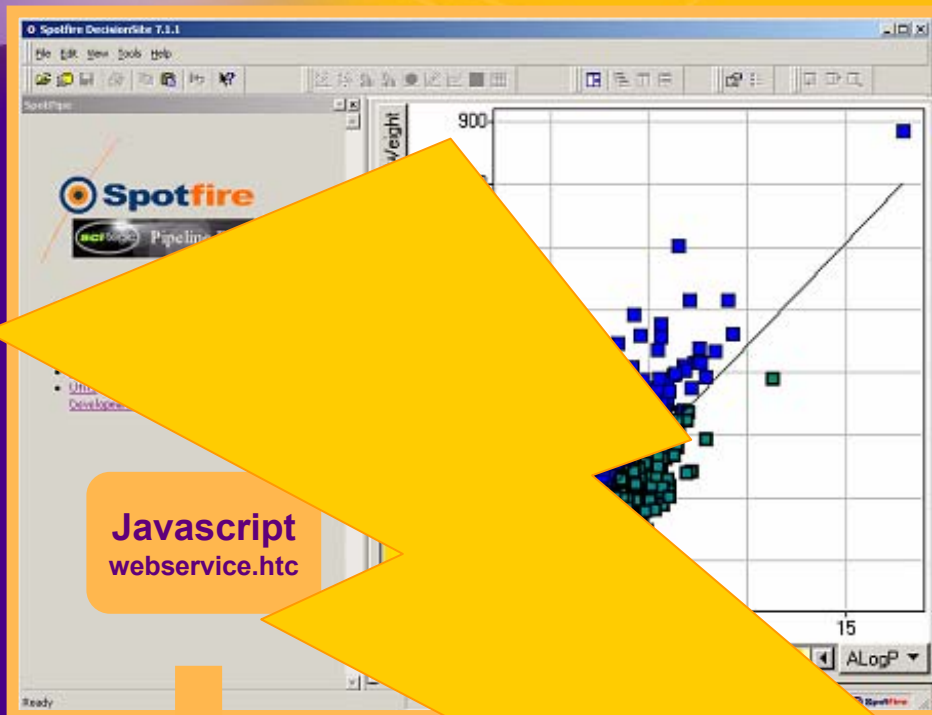
Run protocol

Running and completed protocols are listed below. Click on the name of any completed protocol to view its results and output files. You may click the Stop icon to halt execution of a running protocol or the Trash icon to discard the results of a completed protocol. Results are stored here only temporarily! - you must save your results files elsewhere to preserve them.

Protocol	Start Time	Status	Elapsed Time	Actions
sdf to spotfire	12:18:21 PM	Finished	0:09	
sdf to spotfire	12:18:43 PM	Finished	0:02	
sdf to spotfire	12:19:01 PM	Running	0:02	



Overview



```
winterm
#!/usr/local/bin/perl -w
require SDF_toolkit::MDL_sdf ;

open(OLDOUT,">&STDOUT");
open(OLDIN,"<&STDIN");
close(STDIN);
open(STDIN,"$ARGV[0]");
close(STDOUT);
open(STDOUT,">$ARGV[1]");

while(1) {

    $sdf_file = MDL_sdf_non_parsed_molecule->readFromInput();

    defined $sdf_file || last;

    $sdf_id = $sdf_file->data_for_field_name(EXTREG) ;
    $sdf_id = s/"A(\d+)/$1/;
    $sdf_id = pack("H*", "$sdf_id");
    $prop{EXTREG} = "$sdf_id";
    $propref = \%prop;
    $sdf_file->addReplaceProperties($propref);
    $sdf_file->write();
    next;
}
close(STDIN);
close(STDOUT);
open(STDOUT,">&OLDOUT");
open(STDIN,"<&OLDIN");

#!/usr/local/bin/perl
```

Pipeline Pilot

File Edit View Tool

Explorer

- Gareth
- Nick
- SciTegic Examples
- Spotpipe Addins
 - spotpipe_cluster
 - spotpipe_frags
 - spotpipe_getalogd
 - spotpipe_getconversion

SOAP

A screenshot of the Pipeline Pilot Explorer window showing a file tree structure. A yellow arrow points from the Spotfire window to this Explorer window. A pink SOAP icon is positioned near the arrow.

• **Get Parent ID** ⓘ• **Calculate Smiles** ⓘ• **Cluster Molecules** ⓘ• Cluster Cluster Name AvgNumberPerCluster MaximumDistance 20 NumberOfClustersNum_H_Acceptors PredefinedSet

Num_RotatableBond

ECFP_2

ECFP_4

ECFP_6

MDLPublicKeys

Num_Atoms

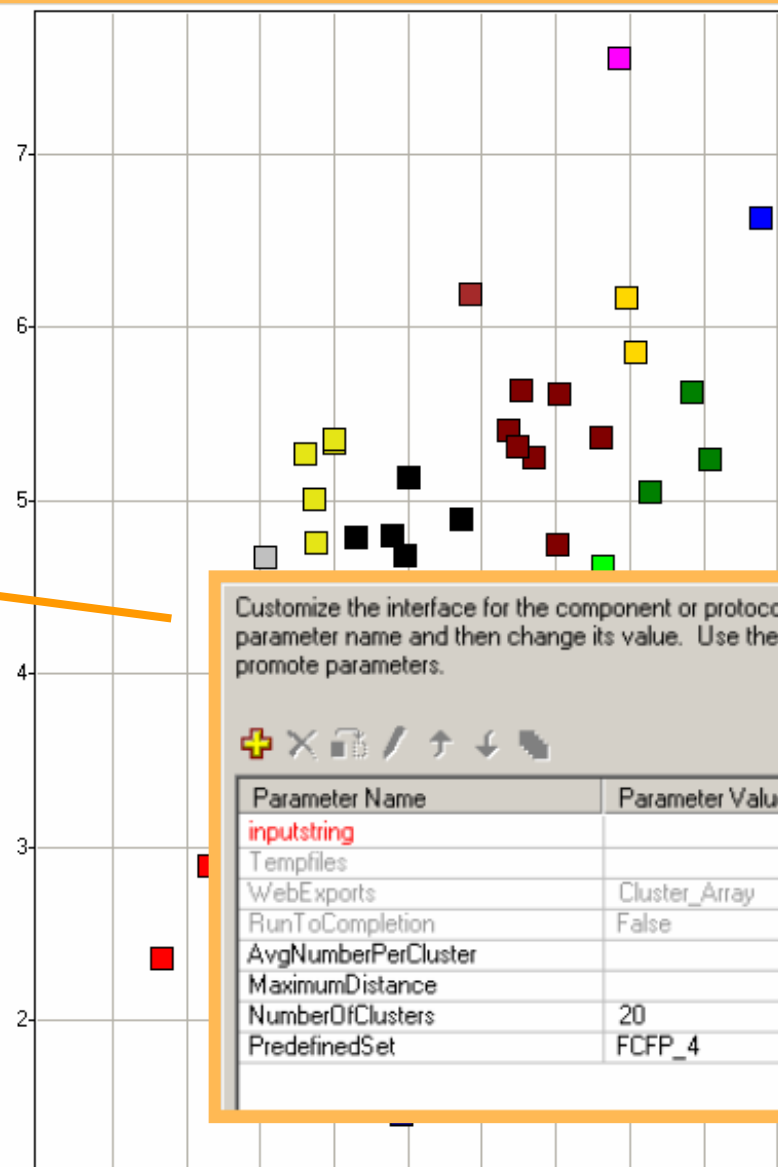
Num_Rings

Num_AromaticRings

Num_Fragments

• **Generate Maximal Common Substructure** ⓘ RequiredProportion• **Substructure Search** ⓘ• Subs_name Col name• **Get Structures** ⓘ• View Structures ⓘ

ALogP_1063111427541

 4 5 6 7

Cluster_1063111478884

1

Cluster_centre_1063111478884

0

Cluster_1063111884507

1

Cluster_centre_1063111884507

0

Cluster_1063111891585

1

Cluster_centre_1063111891585

0

Details-on-Demand

Properties & Parameters

SpotPipe_dev

Query Devices
1000

Subs Rgroup Prop Frags Utils Reactions

- **Calculate Properties** (Subs Co...
Molecular_Formula
Molecular_PolarSurfaceArea
Molecular_SurfaceArea
Molecular_Volume
Prefix
 Calculate AlogP ⓘ
 View Structures ⓘ

Get
Ext
for
spo

Column for structure
(Parent ID,smiles or chime)
Substance ID

AR-C160671...
AR-C160674...

remove query

Edit - test

Customize the interface for the component or protocol. For existing parameters, select the parameter name and then change its value. Use the toolbar buttons to add, delete, and promote parameters.

+ × ↺ ↻ ↕ ↕ ↻

Parameter Name	Parameter Value
csFld_molstructure	
Tempfiles	
WebExports	Molecular_FormulaPPArrayString.Molecular_Pola...
RunToCompletion	False
inputstring	
Properties	<input type="checkbox"/> Molecular_Formula <input type="checkbox"/> Molecular_PolarSurfaceArea <input type="checkbox"/> Molecular_SurfaceArea <input type="checkbox"/> Molecular_Volume <input checked="" type="checkbox"/> Molecular_Weight <input type="checkbox"/> N_Count <input type="checkbox"/> O_Count <input type="checkbox"/> NPlusO_Count <input type="checkbox"/> Num_AromaticBonds <input type="checkbox"/> Num_AromaticRings <input type="checkbox"/> Num_Atoms <input type="checkbox"/> Num_Hydrogens <input type="checkbox"/> Num_Bonds <input type="checkbox"/> Num_Chains <input type="checkbox"/> Num_H_Acceptors
structuresource	

Help Text Interface Promote

OK Cancel Help

Export to Sdfile, Sybyl & Corina

ubs Rgroup Prop Frags Utils Reactions Home

- **Rename Column** ⓘ
- **Delete Column** ⓘ

Column to rename

Substance ID
TP 1249 Solid ~
TP 1249 Solid pIC50
TP 1291 Solid ~

New Name

ic50

- **Export Sdfile & Sybyl** (3D) ⓘ View Structures ⓘ

NC1CCCCC1

Column for structure
(Parent ID, smiles or chime)

Substance ID

Column for compound id

Substance ID

PipelinePlotPro

Substance ID	TP 1249 Solid ~	TP 1249 Solid pIC50	TP 1291 Solid ~
1	48	78	42
2	45	65	38
3	42	82	35
4	40	88	32
5	38	92	30
6	35	95	28
7	32	98	25
8	30	100	22
9	28	102	20
10	25	105	18
11	22	108	15
12	20	110	12
13	18	112	10
14	15	115	8
15	12	118	5
16	10	120	3
17	8	122	2
18	5	125	1
19	3	128	0
20	2	130	0

Fragmentation

Subs Rgroup Prop Frags Utils Reactions Home

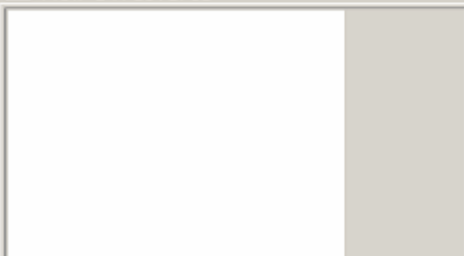
- Fragments (Alpha) ⓘ
- Get Most Frequent Frags ⓘ

Activity
ic50

Alpha

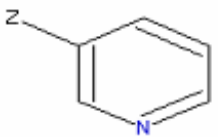
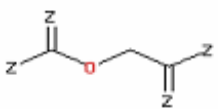
15 Max Frag Size 30 No. frags

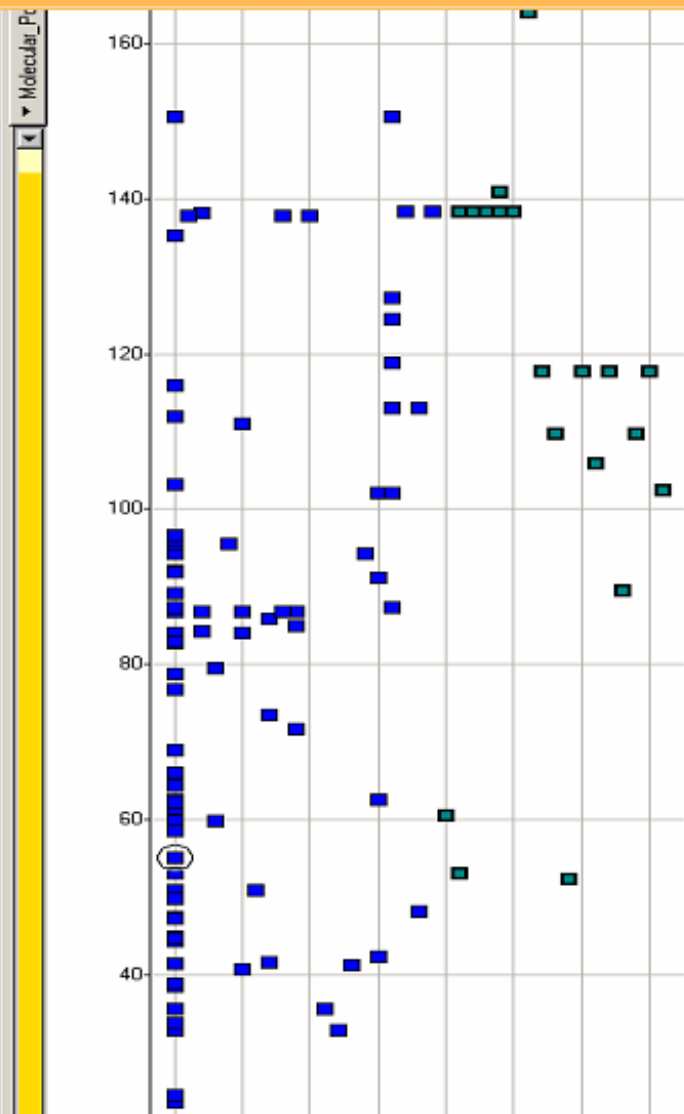
- View Structures ⓘ



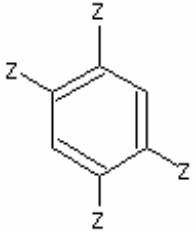
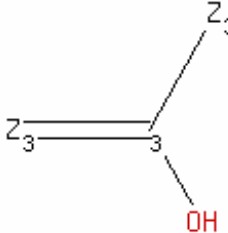
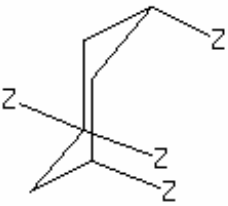
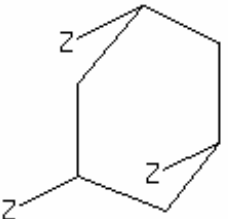
Column for structure
(Parent ID, smiles or chime)

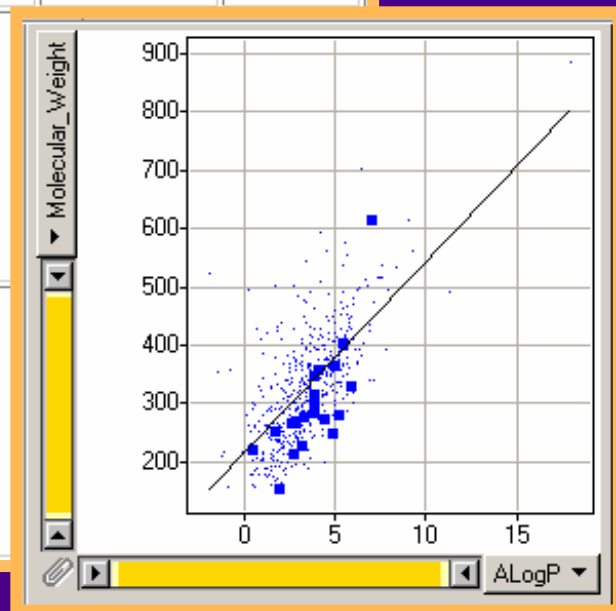
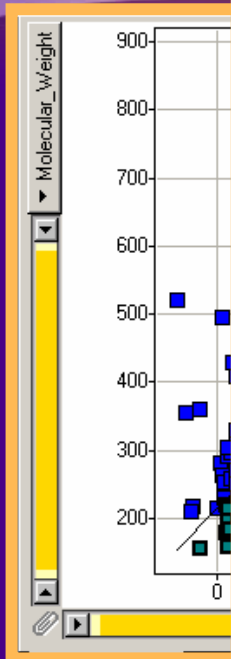
Substance ID

Molecule	Mean_Effect	Frequency
	1.99	2
	1.08	2



Fragmentation 2

Molecule	ratio	alloccurrence	markoccurrence	Index	No_Marked	No_Total
	0.16	6	6	1	262	483
	0.16	23	19	2	262	483
	0.15	5	5	3		
	0.14	4	4	4		

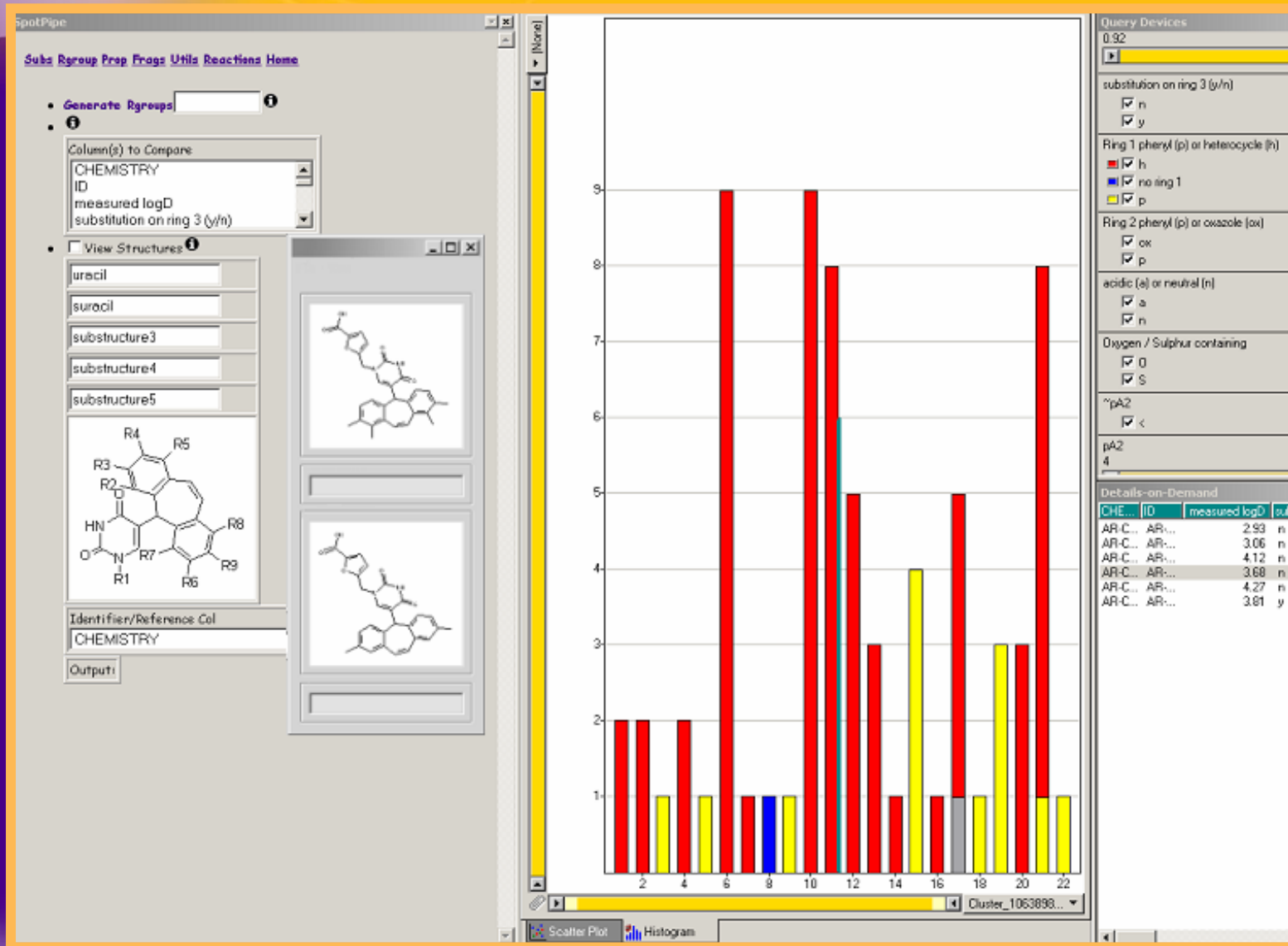


Reactions

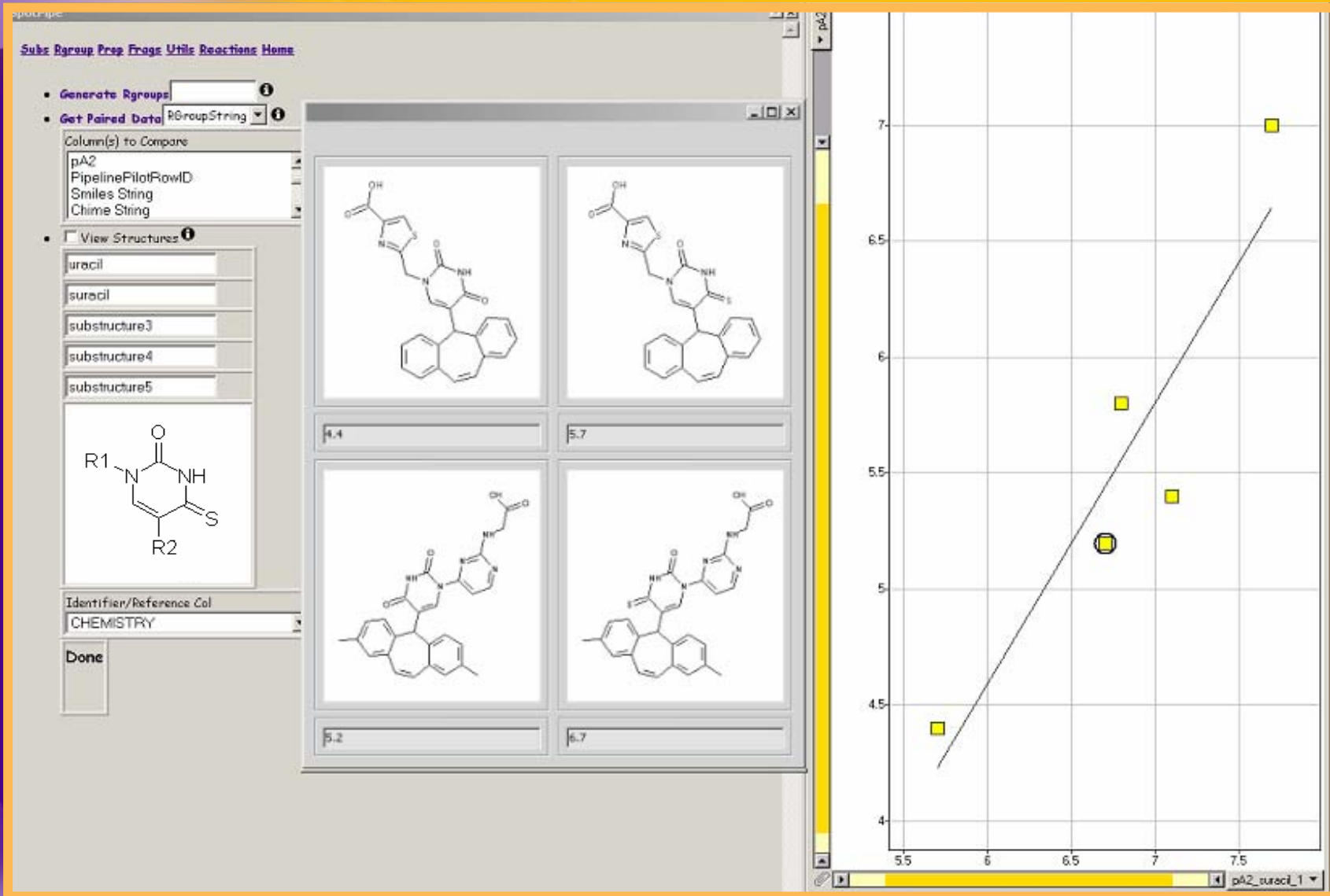
The screenshot shows the SpotPipe software interface. At the top is a menu bar with 'File', 'Edit', 'View', 'Visualization', 'Tools', 'Window', and 'Help'. Below the menu bar is a toolbar with various icons. The main window has a title bar 'SpotPipe' and a menu bar with 'Subs', 'Rgroup', 'Prop', 'Frgs', 'Utils', 'Reactions', and 'Home'. Under the 'Reactions' menu, there are two options: 'Generate Reaction' (with an information icon) and 'View Structures' (with an information icon and a checkbox). Below these options is a large white box containing a chemical reaction diagram. The diagram shows a carboxylic acid structure (a methyl group attached to a carbonyl group, which is also bonded to a hydroxyl group) reacting with an arrow to form an amide structure (a methyl group attached to a carbonyl group, which is also bonded to an amino group). Below the reaction diagram is a text input field with the label 'Column for structure (Parent ID, smiles or chime)'. The field contains the text 'ID' and has a dropdown arrow and an information icon to its right.

- **Generating homogeneous data set**
- **Ask what if questions**
- **Reagent selection**

R-group Analysis



Parallel SAR



Spotpipe

- **Spotfire is a commonly used and powerful visualisation environment**
- **PP Allows Rapid development & Delivery of analysis tools**

PPNavigator - Pipeline Pilot for Excel plug-in

- **Excel is where chemists keep their data**
- **SAR tables and R-group tables as used in physical organic chemistry – all excel based**
- **Further exploitation of PP**

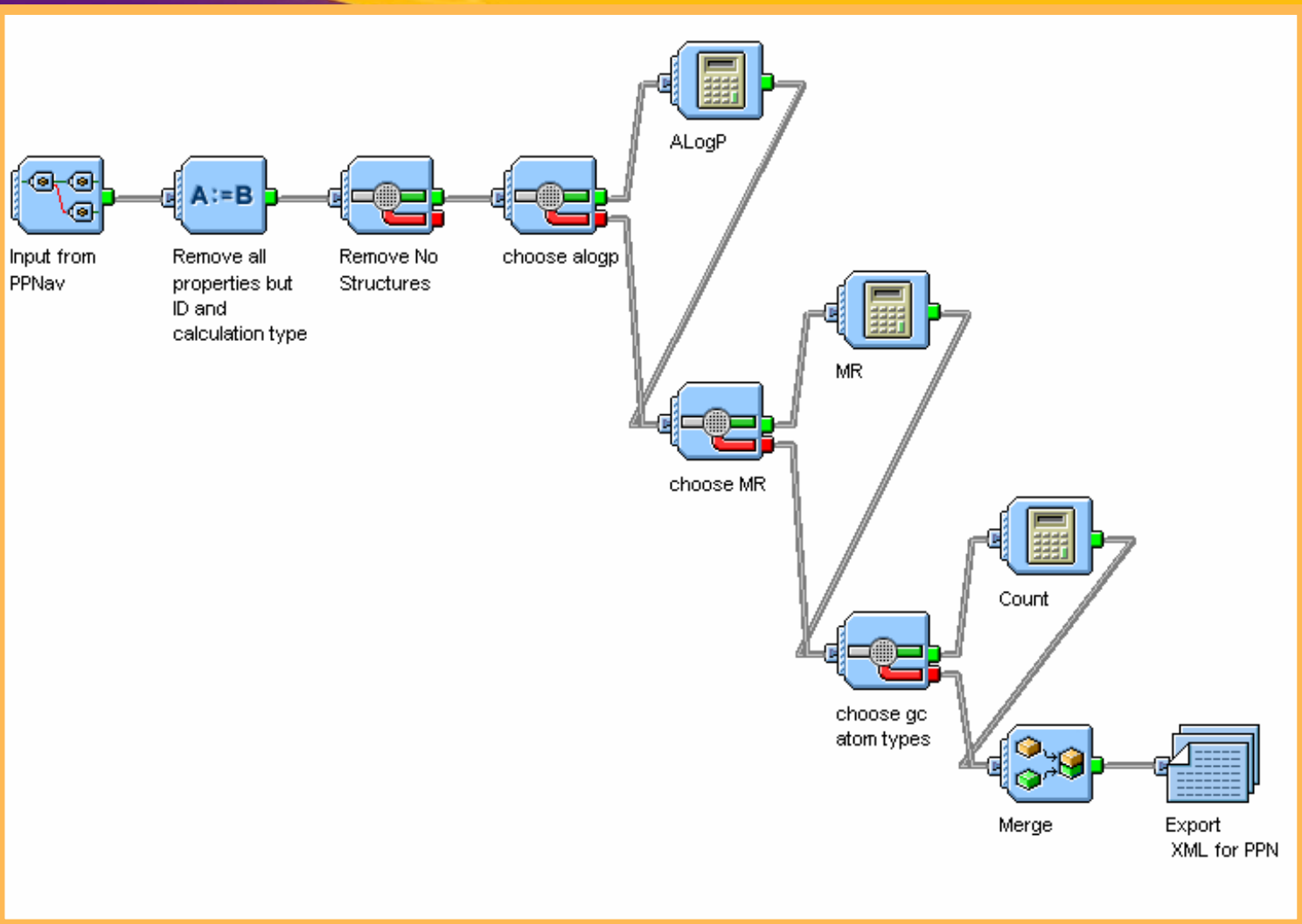
PP Navigator

The image shows a Microsoft Excel spreadsheet with the PP Navigator application window open. The Excel window title is "Microsoft Excel - Book1" and the menu bar includes File, Edit, View, Insert, Format, Tools, Data, Window, CobraSAR, Nicks Utilities, ChemInfo Utilities, Chemistry, and Help. The PP Navigator window has a title bar "PPNavigator ::" and a menu bar with "Server Log-In", "Browse Protocols", "Configure & Submit Jobs", "Job Status & Retrieval", and "Help". The main content of the PP Navigator window displays "Welcome To PipeLine Pilot Navigator" in blue italicized text, followed by "Version 1.3.00" and "Logged on to server ukcwdspipe01". There is a blue icon of a pipe with a compass needle inside. At the bottom of the window, there are buttons for "Server Log - In Status" and "User & Server Settings". A status bar at the very bottom shows "Connection GUID: {EF782AA6-BC21-40DD-8EFE-F09C9D9A42F0}" and "Connection to PP Server on ERROR established OK".

Browsing Protocols

The screenshot displays a software application window titled "Accord for Excel - Nick Presentation Data.xls". The interface includes a menu bar (File, Edit, View, Insert, Format, Tools, Data, Window) and a toolbar. A "ChemInfo Utilities" sidebar is visible on the right, listing various tools such as ChemInfo..., Accord Control, Excel Sheet Tools, Libraries Team Tools, SAR Tools, Import Structures, and ChemProp. The main workspace is a spreadsheet with columns A through N and rows 1 through 22. Row 1 contains the text "CHEMISTRY". Row 2 contains a chemical structure of a substituted benzene ring with a chlorine atom, a hydroxyl group, and a cyclopropane ring, labeled "Chemistry 1". Row 3 contains a similar chemical structure with a cyclopentane ring instead of cyclopropane, labeled "Chemistry 2". A "PP Navigator" window is overlaid on the spreadsheet, showing a tree view of protocols on a server. The tree view includes folders for Home, Go Up One Level, Clustering and Modelling, Database Searching, Enumerations, File Conversion, Filtering and Subsetting, Project Specific, Property Calculation, and Templates and Components. The "PP Navigator" window also displays a "Connection GUID" and a status message: "Connection to PP Server on Charnwood Development Pipe01 established OK". The Windows taskbar at the bottom shows the Start button, several open applications, and the system tray with the time 10:13.

PP Navigator Protocol



Enumeration

Accord for Excel - Nick Presentation Data.xls

File Edit View Insert Format Tools Data Window ChemInfo Utilities Chemistry Help

A3 = "Chemistry 2"

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	CHEMISTRY													
2	Chemistry 1													
3	Chemistry 2													
4	Chemistry 3													
5	Chemistry 4													
6	Chemistry 5													
7	Chemistry 6													
8	Chemistry 7													
9	Chemistry 8													
10	Chemistry 9													
11	Chemistry 10													
12	Chemistry 11													
13	Chemistry 12													
14	Chemistry 13													
15	Chemistry 14													
16	Chemistry 15													
17	Chemistry 16													
18	Chemistry 17													
19	Chemistry 18													
20	Chemistry 19													
21	Chemistry 20													
22	Chemistry 21													
23	Chemistry 22													
24	Chemistry 23													

Chemical structures shown in the spreadsheet:

- Row 1: CHEMISTRY
- Row 2: Chemistry 1 (Structure: 4-chlorobenzamide derivative)
- Row 3: Chemistry 2 (Structure: 4-chlorobenzamide derivative)

PPNavigator ::

Server Log-In | Browse Protocols | Configure & Submit Jobs | Job Status & Retrieval | Help

Protocols on Server

- Protocols\Web Services\Excel\Enumerations
- Home
- Go Up One Level
- Development
- DEV - Generic Amide Coupling - ACCEPT
- Generic Amide Coupling - ACCEPT

AboutProtocol

Generic Amide Coupling - ACCEPT

Description:
Enumerate A Reductive Amination Reaction

Usage:
This reaction requires an SD file containing Acids and an SD file containing secondary amines. You can then choose whether to output the complete reaction or a subset, and whether to filter this via common criteria before it is saved.

Comments:
This reaction requires an SD file containing

Configure & Run Protocol

Connection GUID: {6316F68B-4DC5-4214-8A0C-E5623B90D73C}
Connection to PP Server on Charnwood Development Pipe01 established OK

Start | Aromatic Acids / Mixed Amines /

http://www.google.co.uk/ 10:14

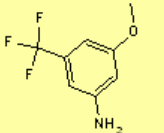
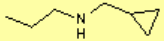
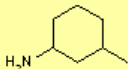
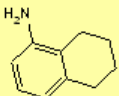
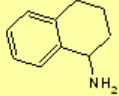
Exceed | Book5 | Rings3... | Inbox -... | Pipelin... | Micros... | http://... | ARC_P... | \ukcw... | Nick Pr... | Micros... | Google...

Output Options

Accord for Excel - Nick Presentation Data.xls

File Edit View Insert Format Tools Data Window ChemInfo Utilities Chemistry Help

A1 = CHEMISTRY

	A	B	C	D	E	F	G	H	I	J
1	CHEMISTRY	Molname								
2	 Chemistry 1	ID2_Cher								
3	 Chemistry 2	ID3_Cher								
4	 Chemistry 3	ID4_Cher								
5	 Chemistry 4	ID5_Chemistry_4								
6	 Chemistry 5	ID6_Chemistry_5								

PPNavigator ::

Server Log-In | Browse Protocols | Configure & Submit Jobs | Job Status & Retrieval | Help

Configure Protocol : Generic Amide Coupling - ACCEPT

Amines: \\ukcwsdpipe01\l_users\PPNavigator\wilden_g\8.sdf

SD file Containing Amines

Output_Subset: Diverse_Subset

Output All the molecules, or

Output_All

Diverse_Subset

Representative_Subset

OutputNumber

Number of Molecules In Subset

Filter: TRUE

Filter Using Lipinski Properties

Acceptors: 10

Number of H-Bond Acceptors

Run Protocol : Generic Amide Coupling - ACCEPT

Molecule ID not Unique : Creating Unique Molname Column

Generic Amide Coupling - ACCEPT has been submitted to the server with a job ID of {E1}

Ready

Start | Aromatic Acids | Mixed Amines

NUM

14:36

\\ukcwsdpipe01\l_users\

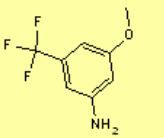
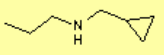
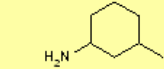
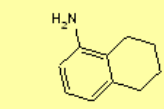
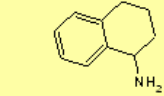
Inbox - Mic... | C:\ | Pipeline Pilot... | \\ukcwsdpipe... | Exceed | Accord for ... | Microsoft Vis... | Microsoft Po... | Digital Plus b...

Lipinski-like Filter Criteria

Accord for Excel - Nick Presentation Data.xls

File Edit View Insert Format Tools Data Window ChemInfo Utilities Chemistry Help

A1 = CHEMISTRY

	A	B	C	D	E	F	G	H	I	J
1	CHEMISTRY	Molname								
2	 Chemistry 1	ID2_Cher								
3	 Chemistry 2	ID3_Cher								
4	 Chemistry 3	ID4_Cher								
5	 Chemistry 4	ID5_Chemistry_4								
6	 Chemistry 5	ID6_Chemistry_5								

PPNavigator ::

Server Log-In | Browse Protocols | Configure & Submit Jobs | Job Status & Retrieval | Help

Configure Protocol : Generic Amide Coupling - ACCEPT

Filter TRUE

Filter Using Lipinski Properties

Acceptors 10

Number of H-Bond Acceptors

Donors 5

Number of H-Bond Donors

AlogPval 5

AlogP cut-off

MW 450

Molecular Weight

Run Protocol : Generic Amide Coupling - ACCEPT

Molecule ID not Unique : Creating Unique Molname Column

Molecule ID not Unique : Creating Unique Molname Column

Ready

Start | Aromatic Acids | Mixed Amines

\\ukwdspipe01\users\

14:20

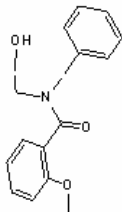
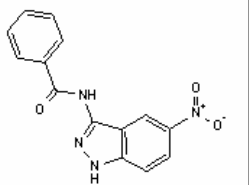
Inbox - Mi... | C:\ | Pipeline Pil... | \\ukwdsp... | Exceed | Accord fo... | Microsoft ... | Microsoft ...

Results

Accord for Excel - Book2

File Edit View Insert Format Tools Data Window ChemInfo Utilities Chemistry Help

A3 = "Chemistry 2"

	A	B	C	D	E	F	G	H	I	J	K
1	CHEMISTRY	Molname	LUMO_MOPAC	DIPOLE_MOPAC	HF_MOPAC	HOMO_MOPAC					
											
2	Chemistry 1	Chemistry-406	-0.24972	4.346	226.99583	-9.42141					
											
3	Chemistry 2	Chemistry-403	-1.19232								
4	Chemistry 3	Chemistry-175	-0.74026								
5	Chemistry 4	Chemistry-153	-0.83262								
6	Chemistry 5	Chemistry-403	-0.3752								
7	Chemistry 6	Chemistry-10	-0.35999								
8	Chemistry 7	Chemistry-458	-0.77682								
9	Chemistry 8	Chemistry-6	-0.47333								
10	Chemistry 9	Chemistry-118	-1.11403								
11	Chemistry 10	Chemistry-13	-0.31622								
12	Chemistry 11	Chemistry-455	-0.33525								
13	Chemistry 12	Chemistry-92	-0.93196								
14	Chemistry 13	Chemistry-459	-0.75147								

PPNavigator ::

Server Log-In | Browse Protocols | Configure & Submit Jobs | Job Status & Retrieval | Help

Server Job: Calculate Cerius2 Descriptors retrieved from Archive.
Submitted on 15:59:36 11/09/2003
5 Files Are Available To Retrieve

- 75A-MoleculeSpreadsheet.htm
- 75B.tmp
- Results.csv
- Results.ppn
- Results.scf

New Accord Book

Jobs Found On Server | Retrieve Job Results

Connection to PP Server on Charnwood Development Pipe01 established OK
Results imported to New Accord book

Start | \\ukcwsd... | Exceed | Microsoft ... | Pipeline Pil... | Slides For ... | Inbox - Mi... | Nick Prese... | Book2

NUM

08:52

Property Calculation

The screenshot shows an Excel spreadsheet titled "Accord for Excel - Nick Presentation Data.xls". The spreadsheet has columns A through K. The data in columns A and B is as follows:

Row	Column A	Column B
656	Chemistry 655	ID656_Chemistry_655
657	Chemistry 656	ID657_Chemistry_656
658	Chemistry 657	ID658_Chemistry_657
659	Chemistry 658	ID659_Che
660	Chemistry 659	ID660_Che
661	Chemistry 660	ID661_Che
662	Chemistry 661	ID662_Che
663	Chemistry 662	ID663_Che
664	Chemistry 663	ID664_Che
665	Chemistry 664	ID665_Che
666	Chemistry 665	ID666_Che
667	Chemistry 666	ID667_Che
668	Chemistry 667	ID668_Che
669	Chemistry 668	ID669_Che
670	Chemistry 669	ID670_Che
671	Chemistry 670	ID671_Che
672	Chemistry 671	ID672_Che
673	Chemistry 672	ID673_Che
674	Chemistry 673	ID674_Che
675	Chemistry 674	ID675_Che
676	Chemistry 675	ID676_Che
677	Chemistry 676	ID677_Che
678	Chemistry 677	ID678_Che
679	Chemistry 678	ID679_Chemistry_678
680	Chemistry 679	ID680_Chemistry_679
681	Chemistry 680	ID681_Chemistry_680
682	Chemistry 681	ID682_Chemistry_681
683	Chemistry 682	ID683_Chemistry_682
684	Chemistry 683	ID684_Chemistry_683
685	Chemistry 684	ID685_Chemistry_684
686	Chemistry 685	ID686_Chemistry_685
687	Chemistry 686	ID687_Chemistry_686
688	Chemistry 687	ID688_Chemistry_687

The PPNavigator dialog box is open, showing the "Configure Protocol : Calculate Cerius2 Descriptors" window. The "Descriptors" list includes:

- ISIS_Keys
- InfoContent
- HOMO
- LUMO
- LUMO_MOPAC
- DIPOLE_MOPAC
- HF_MOPAC
- HOMO_MOPAC
- Receptor_energies
- RadOfGyration
- Jurs_Descriptors
- Shadow Indices

The "Run Protocol : Calculate Cerius2 Descriptors" button is visible at the bottom of the dialog box. Below the button, the connection details are displayed:

Connection GUID: {70E56F8B-F4D8-4131-85E9-EE1D4143B465}
Connection to PP Server on Charnwood Development Pipe01 established OK

Spotpipe & PPNavigator

- **Familiar client framework**
- **Rapid development & deployment of stable & robust protocols**
- **Low maintenance overhead – updates become available as they appear on the server**

Issues

- **Web exports not dynamic – wsdl support**
 - Hard to code up flexible servers
- **Spotfire requires data types**
- **Scalability?**
 - Expense of multi-cpu license
 - Integration with queuing/load balancing systems
 - PVM and SOAP – stability/security?
- **Debugging – more layers means new debugging tools are needed**

In Conclusion

- Some components of value in their own right.
 - Alogp
 - Blisteringly fast and (reasonably) accurate
 - Enumerator (reaction based) – very fast and flexible
 - Bayesian models (can be educational!)
 - Fingerprints
- Pipelining and Web API infrastructure
 - Pipeline pilot provides an easy to use and stable framework to build around.
 - Could be done without PP.
 - Web API via Spotfire and excel provide a familiar environment for rapid deployment of advanced tools.

Acknowledgements

- **Phil Cochrane**
- **Mark Duffield**
- **Kevin Daniels**
- **Gareth Wilden**
- **Alan Tinker**