

Evotec's Library Profiler tool Alistair Sedwell



Evotec Library Profiler

- Aim is to be able to select from a large virtual library either:
 - A combinatorial subset
 - Typically small focussed libraries
 - A non-combinatorial subset
 - Medicinal chemistry projects
- Desirable to allow access to all scientists
 - Creativity
 - Share ideas
 - Security aspect
- Interactive
- Subsets need to satisfy multiple criteria

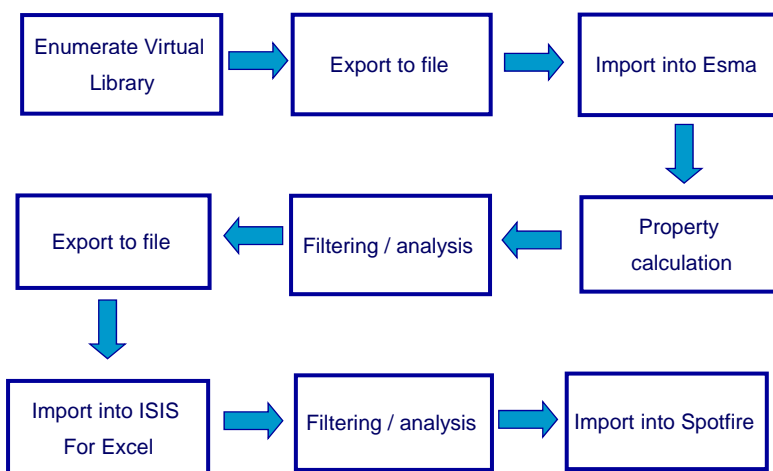
Application development at Evotec

- Team consists of:
 - 4 Developers
 - 1 DBA
 - 4 Chemical Data Scientists

- Technology preferences:
 - Java
 - Oracle
 - ChemAxon

- 1 FTE on this project for 6 months

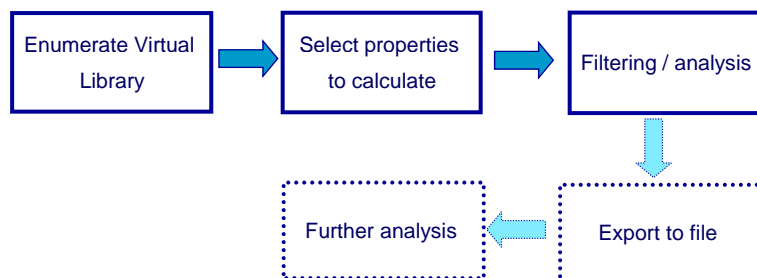
Workflow



Limitations

- Size of data sets
 - Current spreadsheet limited to 5-10k rows
 - Desirable to handle virtual libraries >100k
- File management
 - Systems generate many files in different formats
 - Duplicate data
- Limited licenses to existing tools
 - ISIS For Excel
 - Spotfire

Workflow using the Library Profiler



Library Selection

Library Profiler - Library Opener

Library Profiler Libraries \ ENUM Libraries

Route 1 \ Route 2 \

2272 products for this route.

Amine/Acid Condensation

[*(Amine):1][H][*];(Acid):2[*];\$(OH),\$(OC(=O)[C,C];99)>[*:1][*:2][*:99][H]

Open... Cancel

Property Selection

- Property selection
 - Drug likeness properties
 - QSAR models
- Selection for both product and reagent
- Select notification options

Library Import Wizard

Select which properties you would like to calculate for molecules being imported:

Property	Step 1 (Reagent 1) Structure	Step 2 (Reagent 2) Structure	Product Structure
HERG 2	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
HSA	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Hydrogen bond acceptors	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Hydrogen bond donors	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Lipinski Violation Count	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Lipinski Violations	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
MDCx	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Molecular weight	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Most Acidic pKa	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Most Basic pKa	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Non-H-Atom Count	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Number of Rotatable Bonds	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
pKa	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Protonated Atoms	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Solubility	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
TPSA	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

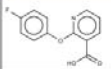
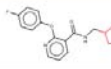
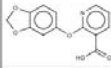
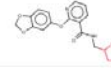
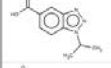
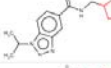
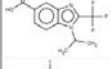
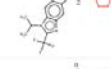
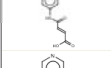

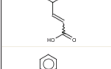
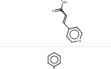
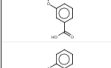
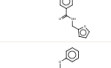

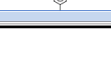
Apply defaults Clear

Cancel < Back Next > Import

Evotec Library Profiler

File Edit Workspace Windows Help

1011 rows for library Hamburg Amides (10042)

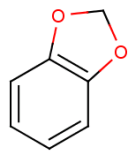
1 (Reagent 2) Structure	dLogP for Step 1 (Reagent 2)	Hydrogen bond acceptors for Step 1	TPSA for Step 1 (Reagent 2)	Molecular weight for Step 1 (Reagent 2)	Hydrogen bond donors for Step 1 (Reagent 2)	Product	dLogP for Product	BBB for Product	Hydrogen bond acceptors for Product	HEA for Product	Molecular weight for Product	TPSA for Product	AVERAGE Energy (kcal/mol) / Product
	2.61	4	59.42	233.19527	1		2.92	-0.3120915	5	0.27394637	316.32097		60.45
	2.032	6	77.00	259.2143	1		2.342	-1.200031	7	-0.06351126	342.346		70.91
	1.92	4	68.01	205.21236	1		1.66	-0.7910311	5	0.11273974	208.34506		69.04
	3.4	3	95.12	272.22324	1		3.237	0.10494234	4	0.3267661	395.35492		56.15
	1.386	4	66.4	209.17386	2		1.635	-0.7354481	5	0.10348304	292.30557		67.43
	0.742	3	50.19	149.14671	1		1.722	-0.42718893	3	0.02178751	244.31326		41.99
	3.983	3	46.53	214.2167	1		4.573	-0.15814606	3	0.6138523	309.38324		38.33
	3.983	3	46.53	214.2167	1		4.573	-0.17367998	3	0.614996	309.38324		38.33

Library Profiler Query Builder

Create a filter to define which rows should be removed from the current view of the Library:

Step 1 (Reagent 1) Structure \ Step 1 (Reagent 2) Structure \ Product

Product



Structure Search Type:
Substructure

Structure Operator:
equals

Similarity Threshold:
0.9

Edit Clear

Product Properties

Route Id equals

Product LNB equals

Calculable Properties:

AVERAGE Energy (kcal/mol) for Product equals Calculated

BBB for Product greater than 1 Calculated

BBB Distribution for Product equals Requires calculation

Bioavailability Score for Product equals Requires calculation

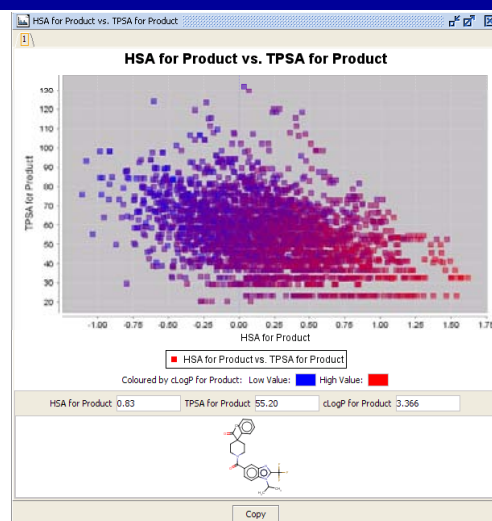
Caco-2 for Product equals Requires calculation

View as text... Load Query... Save Query... Clear OK Cancel

Filter History

Filter Step:	User:	Date:	Rows Removed:	View:	Query:	Delete:
1	Alistair Sedwell	22-Apr-2008 11:30:25	251			
2	Alistair Sedwell	22-Apr-2008 11:30:57	361			
3	Alistair Sedwell	22-Apr-2008 11:31:31	837			
4	Alistair Sedwell	22-Apr-2008 11:32:32	1044			

Charting – Scatter plot



Charting - Histogram

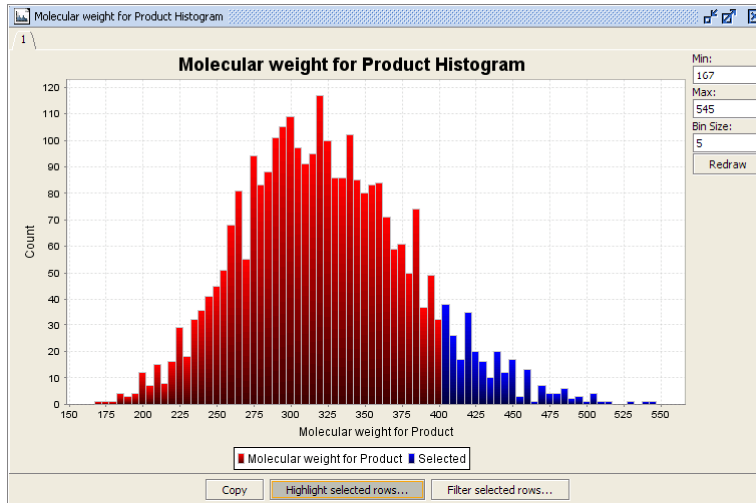
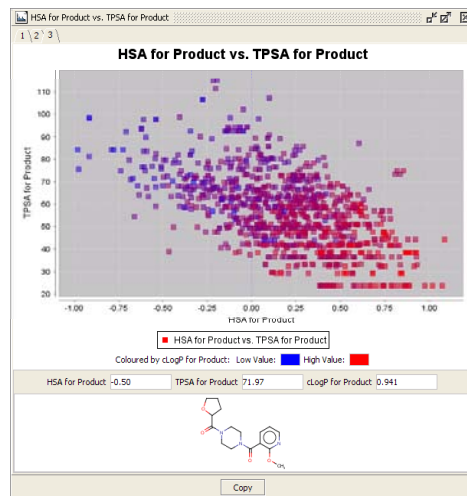
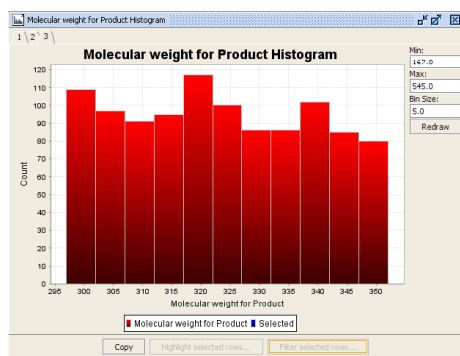


Chart based filtering



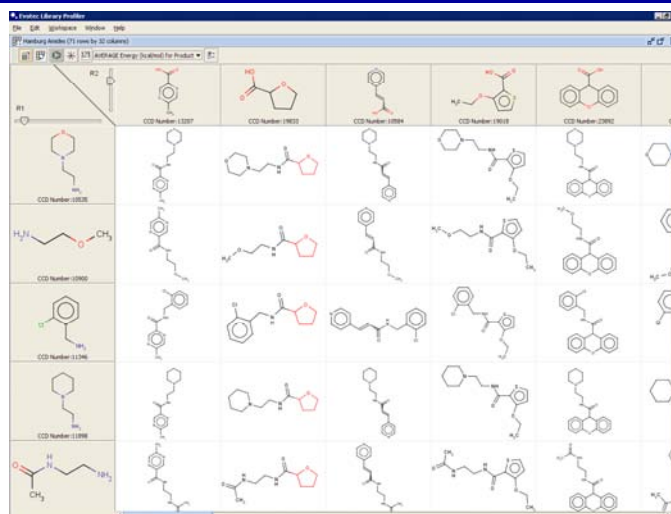
Charting

- Interactive integrated charting
 - Chart filtering
 - Chart versioning

- Filter history includes chart filters

- Row highlighting
 - Additive highlighting

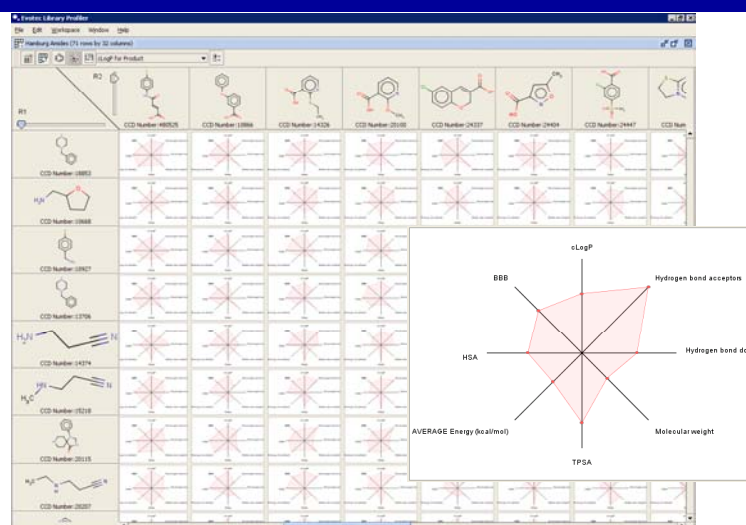
Pivot View - Structures



Pivot View - Properties

CCD Number	375	2,472	-887	-130	-874	722	1,446	785	286
CCD Number 14235	318	2,227	-891	-114	-1,766	876	1,388	-126	274
CCD Number 17626	2,041	3,663	1,113	1,832	196	1,891	2,432	1,885	1,459
CCD Number 18325	3,525	4,994	2,937	2,516	1,48	3,275	4,294	3,424	2,474
CCD Number 11345	902	2,818	284	-1,327	-883	822	1,827	396	286
CCD Number 11487	1,444	3,136	816	-476	-431	1,294	2,019	1,288	882
CCD Number 11898	1,184	2,766	-476	-895	-741	854	1,878	1,018	242
CCD Number 12319	1,255	2,751	387	206	-53	1,185	1,889	1,229	466
CCD Number 12866	1,754	4,523	797	1,874	-178	2,965	2,885	1,814	1,657
CCD Number 13953	-394	2,226	-1,001	-124	-2,676	768	1,058	318	-171
CCD Number 15669	3,289	3,417	2,262	3,129	1,277	4,12	4,41	3,269	2,446
CCD Number 14544	-394	2,226	-1,001	-124	-2,676	768	1,058	318	-171
CCD Number 19445	4,786	7,627	3,862	4,888	2,887	5,65	5,94	4,894	4,626
CCD Number 20444	2,458	3,874	1,449	812	2,368	2,262	2,482	1,629	1,629
CCD Number 18671	445	2,643	-215	1,537	1,206	1,596	305	822	
CCD Number 14235	1,315	3,733	-386	-53	1,185	1,889	1,229	466	
CCD Number 12788	1,57	3,48	1,44	-412	2,410	2,702	1,48	1,523	
CCD Number 13673	1,533	4,263	1,826	1,802	-189	2,794	3,084	2,042	1,886
CCD Number 13754	2,411	4,261	1,504	2,281	429	3,212	3,562	2,521	1,744
CCD Number 19445	887	3,224	384	876	-876	1,867	2,187	1,116	1,116

Pivot View – Spider plots



Other Features

- Library properties
 - Max, min, average, std dev of properties across library
- Workspace
 - Store all formatting
 - Private, Project, and Public workspace areas
- Export
 - Data sets
 - Reagent lists
- Preferences
- Tasks

Using ChemAxon tools

- High usage of Marvin View and Sketch
 - Easy to integrate
- JChem cartridge for filtering
 - Experience in using cartridge
- JChem tools for many of the property calculations
 - HBD, HBA, ROT, AMW, TPSA, Veber Bioavailability, BBB distribution, undesirable functional groups, Andrews AVERAGE energy, Bioavailability score, Ligand binding efficiency, PGP Substrate prediction, pKa, protonated atom count, non-H atom count

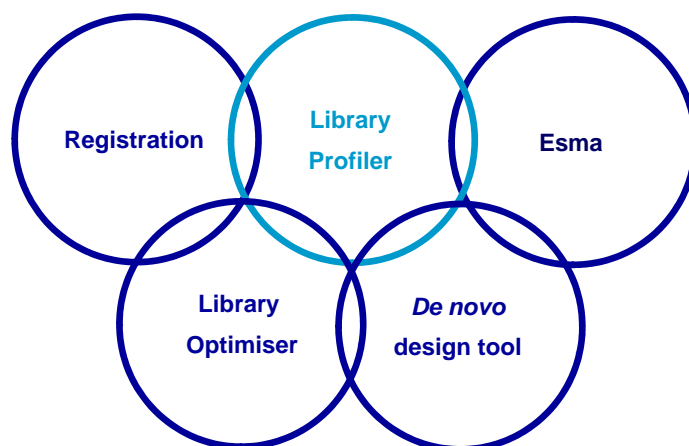
Future Features

- Clustering
 - Jarvis-Patrick
 - Scaffold

- Diversity selection
 - In-house diversity tools

- Further property calculation

Future – Integrated Drug Design Application



Thanks

- eScience Team
 - Phil Cochrane
 - Catherine Reisser
 - Paul Barrett
 - Rajeev Halder
 - Mike Potterton
 - Bob Marmon
 - Jane Couzens
 - Anne Pintat
- Mark Whittaker
- Mark Brewer
- Evotec chemists
- ChemAxon

Evotec's Library Profiler tool Alistair Sedwell

trust value
productivity innovation
commitment speed
synergistic partnerships
value commitment
innovation productivity
speed trust