



Helium: An EXCEL based User Tool for SAR analysis

September 15th, 2010

Today's presentation – Questions to be answered

- Why did GSK undertake development of Helium?
- How was the Helium idea conceived?
- How did GSK develop Helium to ensure user satisfaction and acceptance?
- What is the architecture of Helium?
- What functionality does Helium provide?
- What synergies does GSK see with combining Excel, JChem for Excel and Helium?
- What is Helium's current status and future plans?

Where we came from

IT Resources Supporting SAR in Discovery

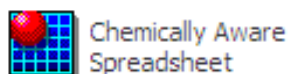
90's

00's

Today



Spotfire



Chemically Aware Spreadsheet




AutoStructure



accelrys



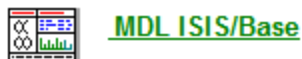
DISCOVERY
EXPLORER



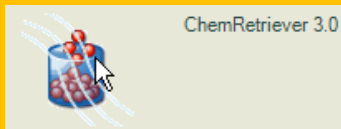
ISISNET



InforSense



MDL ISIS/Base



ChemRetriever 3.0



GSK CPDP
Compound Profiling Data Portal

Where we are headed

IT Resources Supporting SAR in Discovery

90's


00's

Today

Future?



 **Spotfire**


 Chemically Aware Spreadsheet

 **AutoStructure**


 **acclrys**


DISCOVERY

EXPLORER

 ISISNET

InforSense

 **MDL ISIS/Base**

 ChemRetriever 3.0

GSK CPDP
Compound Profiling Data Portal

 **Tibco Spotfire**

He

 **Instant JChem**

 **Excel 2007**

 **Jchem for Excel**

Simplification

The Birth of 'Helium'

- The “Swiss Army Knife” Concept
- One stop shopping for all of your SAR needs! (*and lower prices!*)
- Original tool was based on Spotfire
 - Spotfire had a Structure Viewer
 - Spotfire supported large data sets
 - At the time, cost was not a significant factor
- The results from the initial PoC were somewhat disappointing, because the average bench Researcher found Spotfire difficult to learn and use

The 'Cold Turkey'

- Two years ago, after significant consolidation in the Discovery IT Support area, the project was revisited using Agile Development Techniques as a Proof of Concept
 - Significant Improvement was made in usability
 - Still based on Spotfire container
 - Functionality and usability were received 'Very Positively'
 - Unsuccessfully tried to integrate a "forms" view into the PoC
 - Still received resistance from the average bench chemist due to Spotfire

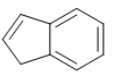
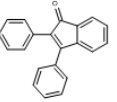
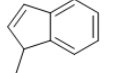
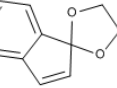
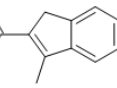
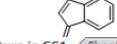
'Helium', the Sequel

- In 2009, GSK purchases the Chemaxon Suite of tools
 - JChem for Excel
 - Instant JChem
 - JChem cartridge
- And so was born “Helium in Excel”

Benefits

- JChem for Excel enhances a familiar and comfortable tool
- Instant JChem provides flexible forms tool
- JChem Cartridge underpins our chemistry web services

The screenshot shows the Microsoft Excel interface with the JChem ribbon active. The ribbon includes options for importing from databases or files, adding or showing structures, and converting between SMILES, shapes, and Marvin OLE shapes. A Helium Task Panel is open on the right, displaying a list of actions for the selected structure, such as 'Calculate Derived Properties', 'Calculate Simple Properties', 'Fragment R-groups', and 'Get Compound Number for Structure'. The main spreadsheet area shows a table with columns for External Id, Structure, and Database. The first row is highlighted, showing the External Id 'MFCD00003777' and the structure of indole.

External Id	Structure	Database
MFCD00003777		acd
MFCD00003784		acd
MFCD00039585		acd
MFCD00041457		acd
MFCD00044428		acd
MFCD00044428		

The Development Approach – Keys to Success

Internally

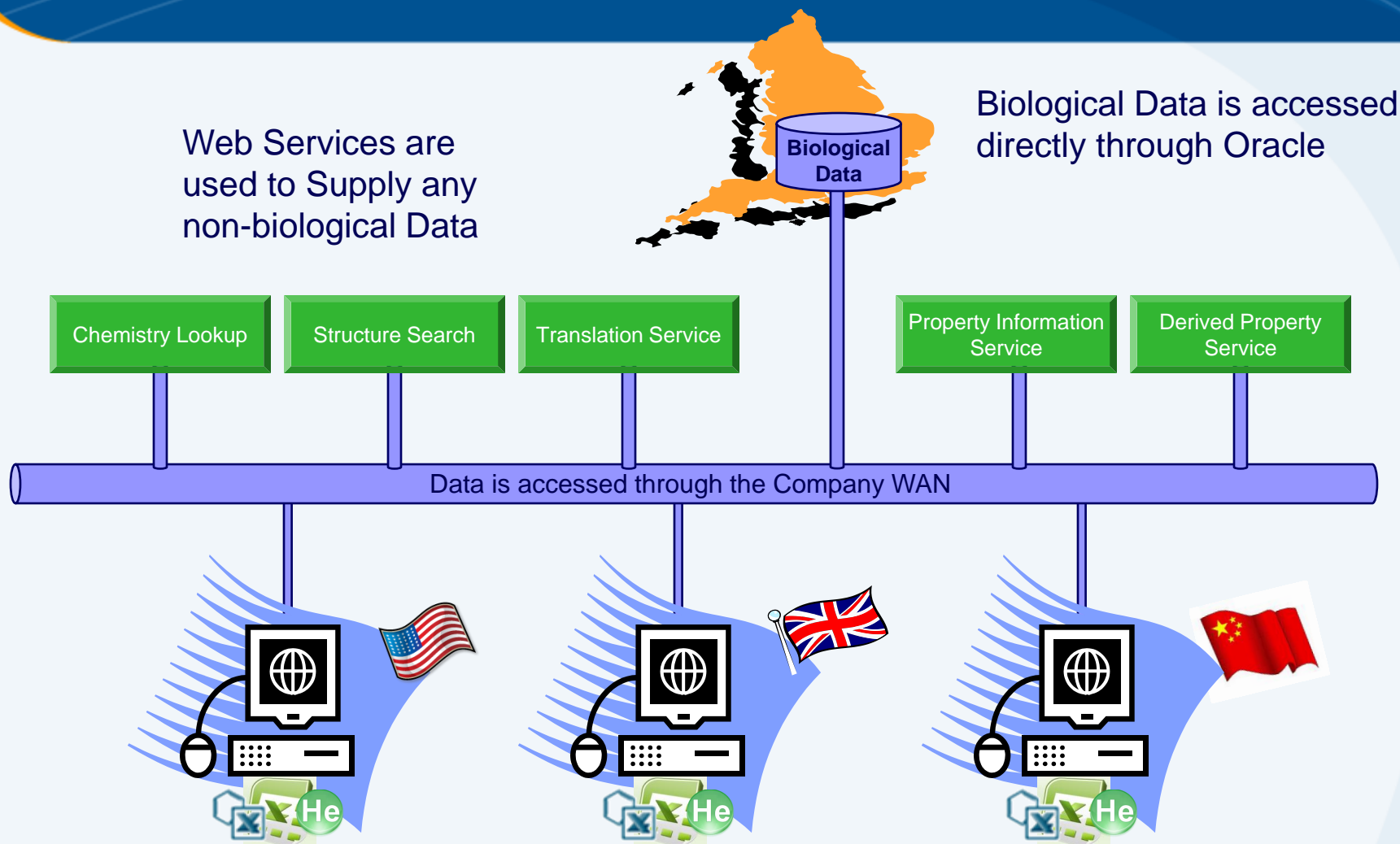
- User interaction and Ownership
- Identified Lead End User (Senior Researcher from Discovery)
- Interactive End User Group covering all disciplines and sites within R&D
- Agile development approach with regular deliverables
- Weekly End User Group meetings
- Extended End Users engaged at key points
- Accessed “Live” data, making the tool immediately useful
- ‘Viral’ release of the software to R&D

Externally

- Superb support and response from ChemAxon
- Collaborative relationship in solving problems
- Excellent communication between the two companies



Helium Architecture



- ❖ Clients can be in any R&D location, and can be from many different disciplines (Biology, Chemistry, Computational Chemistry, Compound Management)
- ❖ Dataset sizes can range from tens of compounds or structures to tens of thousands

Excel? How do we keep this manageable?

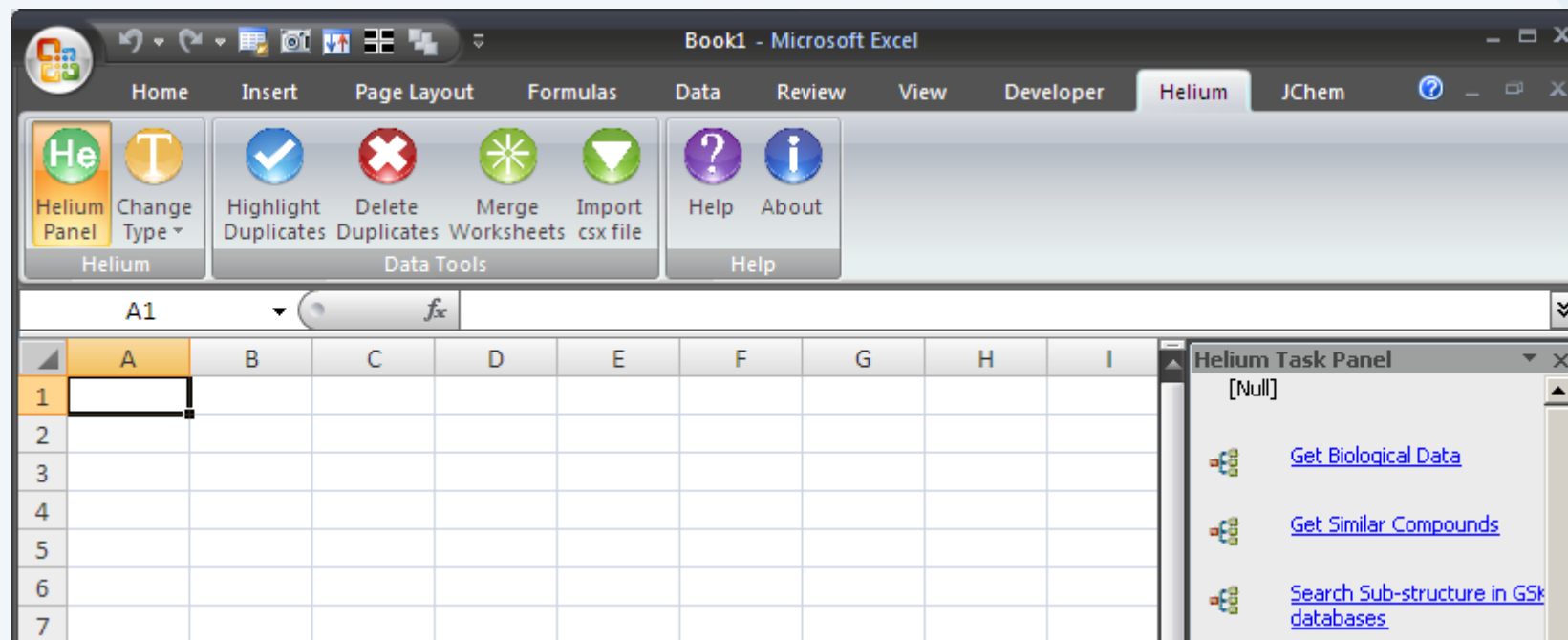
- An Excel Customization for 1000 Researchers?
 - Microsoft ClickOnce for managing the customization
 - Client code regularly checks for updates and installs
 - Insures that update installation is simple
 - Keeps installations of the software from “falling behind”
 - Collaboration and partnering with ChemAxon
- To date, the approach has been very successful

The components of Helium

❖ Microsoft Excel 2007

❖ JChem for Excel

❖ Helium



- General functionality for formatting, sorting, presentation and creating equations
- Chemical Structure presentation, file import and export, etc.
- Access to GSK web services, GSK biological data and consistency with legacy data

Helium User Interface

Book1 - Microsoft Excel

Home Insert Page Layout Formulas Data Review View Developer Helium JChem

Helium Panel Change Type Highlight Duplicates Delete Duplicates Merge Worksheets Import csx file Help About

Helium Ribbon – for non data specific tasks

JChem Ribbon to access JC4XL Functionality

Worksheet with datatyped Helium Data

External Id	Structure	Database
MFCD00003777	<chem>C1=CC=C2C=CC=CC12</chem>	acd
MFCD00003784	<chem>O=C1C=CC(=C1)C2=CC=CC=C2</chem>	acd
MFCD00039585	<chem>CC1=CC=C2C=CC=CC12</chem>	acd
MFCD00041457	<chem>C1=CC=C2C(=C1)OC2</chem>	acd

Helium Task Panel [Structure]

- Calculate Derived Properties
- Calculate Simple Properties
- Fragment R-groups
- Get Compound Number for Structure
- Get Similar Compounds
- Remove Isotopes
- Remove Salts / Solvates
- Remove Stereochemistry
- Search Sub-structure in GSK databases
- Standardise Valences
- Sub-structure Search within table

JChem Ribbon to access JC4XL Functionality

Datatype Sensitive Task Panel. This display appears when a "structure" data type is selected,

Worksheet with datatyped Helium Data

Helium Functionality

The screenshot shows the Microsoft Excel interface with the Helium Panel ribbon selected. The ribbon includes buttons for Helium Panel, Change Type, Highlight Duplicates, Delete Duplicates, Merge Worksheets, Import csx file, Help, and About. A dropdown menu is open for the 'Change Type' button, listing the following options: Compound Number, External Id, Null, Parent Compound Number, Prep Lnb Ref, Project Id, SMILES, Structure (which is selected and has a green checkmark), and User Id. The background shows a spreadsheet with a chemical structure in cell B2 and the text 'Database' in cell C2.

- Datotyping assigned by regular expressions
- Generic data types(e.g. Project Id) can be manually set

Helium Functionality List

Data Type	Task
Project ID	<i>Get Compound Numbers</i>
Project ID	<i>Validate Project ID</i>
Compound Number	<i>Get Biological Data</i>
Compound Number	<i>Get Critical Results</i>
Compound Number	<i>Get LNB Refs</i>
Compound Number	<i>Get Parent</i>
Compound Number	<i>Get SMILES</i>
Compound Number	<i>Get Structure</i>
Compound Number	<i>Get Structure for Parent Compound</i>
Compound Number	<i>Get Versions</i>
Compound Number	<i>Identify duplicate structures</i>
Compound Number	<i>Validate Compounds</i>
SMILES	<i>Calculate Derived Properties</i>
SMILES	<i>Calculate Simple Properties</i>
SMILES	<i>Canonicalize</i>
SMILES	<i>Get Exact Structures</i>
SMILES	<i>Get Similar Compounds</i>
SMILES	<i>Remove Isotopes</i>
SMILES	<i>Remove Salts / Solvates</i>
SMILES	<i>Remove Stereochemistry</i>
SMILES	<i>Search Sub-structure in GSK databases</i>
SMILES	<i>Standardize Valences</i>
SMILES	<i>Sub-structure search within table</i>
External ID	<i>Get SMILES</i>
External ID	<i>Get Structure</i>

Data Type	Task
<i>Lnb Ref</i>	<i>Get Compounds Numbers</i>
Lnb Ref	<i>Get Registration Information</i>
Lnb Ref	<i>Get SMILES</i>
Lnb Ref	<i>Get Structure</i>
Lnb Ref	<i>Get Versions</i>
Lnb Ref	<i>Validate LnbRefs</i>
User Id	<i>Get Registered Compounds LNB Refs</i>
Structure	<i>Calculate Derived Properties</i>
Structure	<i>Calculate Simple Properties</i>
Structure	<i>Get Exact Structures</i>
Structure	<i>Get Similar Compounds</i>
Structure	<i>Remove Isotopes</i>
Structure	<i>Remove Salts / Solvates</i>
Structure	<i>Remove Stereochemistry</i>
Structure	<i>Sub-structure Search in GSK databases</i>
Structure	<i>Standardize Valences</i>
Structure	<i>Sub-structure Search within table</i>
Parent Compound Number	<i>Get Parent</i>
Parent Compound Number	<i>Get SMILES</i>
Parent Compound Number	<i>Get Structures</i>
Parent Compound Number	<i>Get Versions</i>
Parent Compound Number	<i>Validate Compounds</i>
Null	<i>Get Biological Data</i>
Null	<i>Get Similar Compounds</i>
Null	<i>Search Sub-structure in GSK databases</i>

Helium Performance (worst case conditions)

Activity	Time
Retrieve 33,885 Compound Numbers from 110 Project Ids	28"
Retrieve SMILES from 33,885 Compound Numbers	5' 12"
Retrieve Structures from 33,885 Compound Numbers	26' 28"
Canonicalize 33,885 Structures	8' 05"
Remove Salts and Solvates from 33,885 Structures	9' 35"
Remove Stereochemistry from 33,885 Structures	10' 05"
Get Lot Ids from 33,885 Compound Numbers	11' 40"
Validate 33,885 Compound Numbers	11' 38"
Check for Duplicate Structures for 33,885 Compound Numbers	9' 12"
Retrieve Parent Compound Number for 33,885 Compound Numbers	8' 30"
Retrieve MF, MW, BEW for 33,885 Compound Numbers	37' 47"

*This performance level represents a significant improvement over previous tools!
Typical SAR datasets are much smaller*

Helium, JChem and Excel – the Possibilities

- With the various tools available within Excel 2007, Helium and JChem for Excel, the researcher has an extensive, flexible tool for interrogating data:

The screenshot displays an Excel spreadsheet with columns A through M. Column A contains IDs (e.g., MFCD00039585), column B contains chemical structures, and columns C through M contain numerical data. The 'Conditional Formatting' task pane is open, showing options like 'Highlight Cells Rules', 'Top/Bottom Rules', 'Data Bars', and 'Color Scale'. The 'Color Scale' option is highlighted, and red arrows point from the text on the right to this menu item and the data cells.

	A	B	C	D	E	J	K	L	M				
4	MFCD00039585	<chem>C1=CC=C2C=CC=CC12</chem>	-6.8	0.003	1								
5	MFCD00041457	<chem>C1=CC=C2C(=O)OC12</chem>	-3.8	0.003	1								
6	MFCD00044428	<chem>CC(=O)C1=CC=C2C=CC=C12</chem>	0.35	0.591	1								
7	MFCD00046319	<chem>C1=CC=C2C=CC=CC12</chem>	-2.8	0.003	2	0.252	20	3.258	140	5	0	0	230.0
8	MFCD00051614	<chem>C1=CC=C2C(=O)C=C1C2=O</chem>	-1	0.18	2	0.992	19	2.373	212	0	1	1	211.3
9	MFCD00057817	<chem>C1=CC=C2C=CC=CC12</chem>	6.3	0.591	2	1.283	27	3.157	288	7	3	0	356.4
10	MFCD00060877	<chem>C1=CC=C2C=CC=CC12</chem>	-5.2	0.003	1	0.192	13	4.408	141	0	0	0	158.3
11	MFCD00061044	<chem>C1=CC=C2C=CC=CC12</chem>	-6	0.003	1	0.196	12	3.889	132	0	0	0	144.2

Excel conditional formatting options give great highlighting options for identifying patterns in your data.

Helium, JChem and Excel – the Possibilities

- With the various tools available within Excel 2007, Helium and JChem for Excel, the researcher has an extensive, flexible tool for interrogating data:

External Id	Structure	Core found	R1	R2	R3	G	H	I
MFC00072162		Yes	CN(C[*])CC[*]	[*]C1=CC=CC=	CN(C[*])CC[*]	30.8	0.919	2
MFC00083115		Yes	[*][H]	[*]CCC#N		-3.05	0.059	1
MFC00085585		Yes	[*]C1=CC=CC=	[*]C1=CC=CC=	[*]C1=CC=CC=	5.05	0.348	4

Easy column filtering

For Example after R-Group decomposition, you can convert structure to SMILES and then use the SMILES as filters in Excel to segregate your data

Helium, JChem and Excel - the Benefits

- Bringing new capabilities to be included in SAR
- Easily allows comparison of assay data vs. off target activity vs. liability
- A single, non-threatening interface to Discovery data
- Most researchers are familiar with Excel and it's functionality
- Excel files are universally portable and support data interchange
- One application to support and maintain vs. the current list of multiple tools
- We benefit from the advancement of Excel in the future (Excel 2010)

Current Deployment Status

- Helium for Excel, Version 1 is available currently with a reduced function set (~150 user installs) targeting 1000 users by end of year.
- Approaching final production release – expected at end of September/early October
- Legacy tools will be decommissioned approx 4 weeks after official release of the final version
- Currently working on integration of the Helium add-in with Spotfire. This integration is expected to be available in late 2010.

Future Plans

- Working with Chemaxon to improve copy / paste functionality of OLE objects (especially ChemDraw) to other Microsoft applications
- Looking at how Excel Training for Scientists can be implemented economically
- Implementing Instant JChem to replace ISIS H-Views for forms based data delivery
- Exploring the options of linking IJC Project data with JChem for Excel Database access

Thank you for your attention!

Questions?