gsk do more feel better live longer

I think I've had this feeling of déjà-vu before!

Making the most of a structural id database

ACD Labs, Cambridge Symposium

27th September 2016

Steve Thomas

Challenges

How do I handle knowledge at the rate it is being generated today?

How do I best share my knowledge with my colleagues and gain access to theirs?

Outline

- Challenges in my research area
  - Data jigsaw
- Why we cannot rely on memory
  - Staff turnover
- Building a database
  - Building a data cube!
  - Metadata
- Embed and Grow
- Uptake and review
  - AAR
- Benefits
  - Accuracy and time
- Spread the word

David Jack Centre for R&D, Ware

do more
feel better
live longer

1898 - Allen and Hanburys, formerly Plough Court Pharmacy (est. 1715) built on the site of the old mill at Ware, Hertfordshire UK

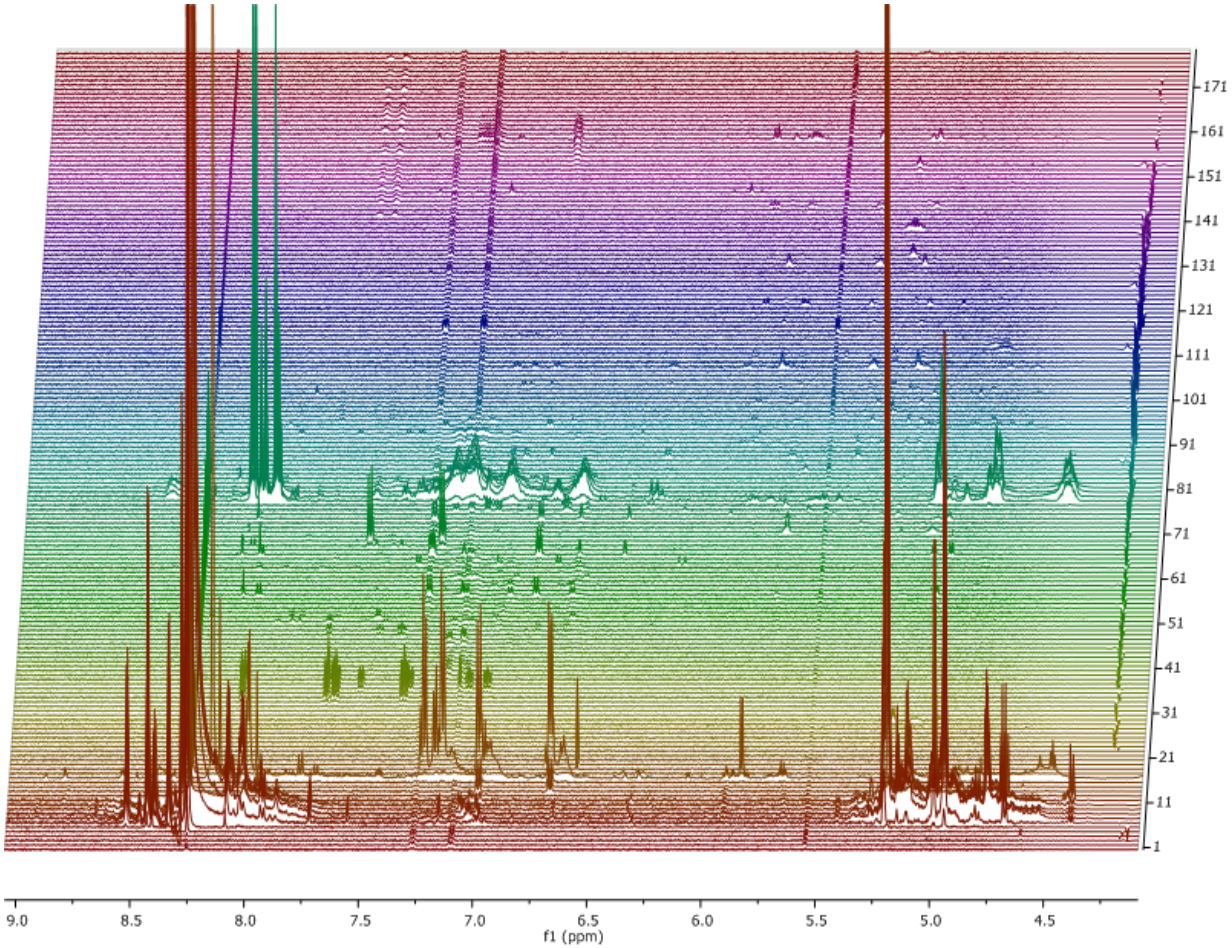
R&D → Platform Technology Science → Product Development and Supply → Global Spectroscopy

Metabolite ID

Blood

Urine

Bile



We are generating data faster than ever, how can we hope to efficiently use legacy knowledge?

Safety implications

The implications of passing a unsafe drug into the clinic are severe.

Getting it wrong is not an option.

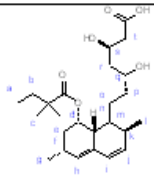
What's on our side?

- Top end equipment
- Experience
- Historical data
- Ability to smell a structural alert from 20 yards!

Historical data storage

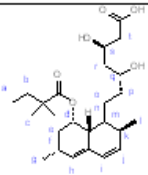
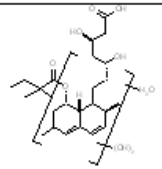
Study Number: Investigative Name: Steve Thomas Date:
Signature:

All data files have been reviewed and the data relevant to the report have been printed.

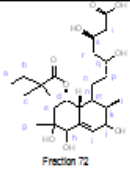
Peak ID	RT (min)	Proposed Structure	$[M+H]^+$ (m/z)	Diagnostic Fragment Ions (m/z)[1]	^{1}H NMR Chemical Shifts (ppm)[2]
Simvastatin Acid (SVA)	8.2		435 Pos ion 459 (Na^+ adduct)	 319: (F1) 301: 319 – H_2O (F2) 215: (F3) 159: (F4) 115: (F5) 85: (F6) 59: (F7)	a- 0.89 b- 1.61 c- 1.16 d- 5.38 e- 2.02 f- 2.48 g- 1.12 h- 5.52 i- 6.01 j- 5.83 k- 2.44 l- 0.94 m- 1.75 n- 2.35 o- 1.43 p- 1.59, 1.27 q- 3.73 r- 1.63 s- 4.13 t- 2.41, 2.31

Agreed by: Date:

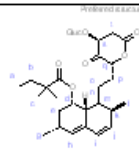
Study Number: Investigative

Peak ID	RT (min)	Proposed Structure	$[M+H]^+$ (m/z)	Diagnostic Fragment Ions (m/z)[1]	^{1}H NMR Chemical Shifts (ppm)[2]
P Simvastatin Acid	8.3	 Fraction 146	435 (P) $C_{25}H_{38}O_5$ Theory: 435.2747 Found: 435.2749 Diff: 0.5 ppm.	319: F1 215: F3 159: F4 115: F5 85: F6 59: F7	a- 0.89 b- Obscured c- 1.16 d- 5.38 e- 2.02 f- 2.45 g- 1.12 h- 5.54 i- 6.01 j- 5.84 k- Obscured l- 0.94 m- Obscured n- Obscured o- Obscured p- Obscured q- 3.73 r- Obscured s- 4.12 t- Obscured
M6 Hydration, dioxidation	3.7		485 (P+32+18) $C_{25}H_{41}O_8$ Theory: 485.2751 Found: 485.2770 Diff: 3.9 ppm.	369: F1+ 3[O] + 2[H] 247: F3 + 2[O] 115: F5 85: F6	NA

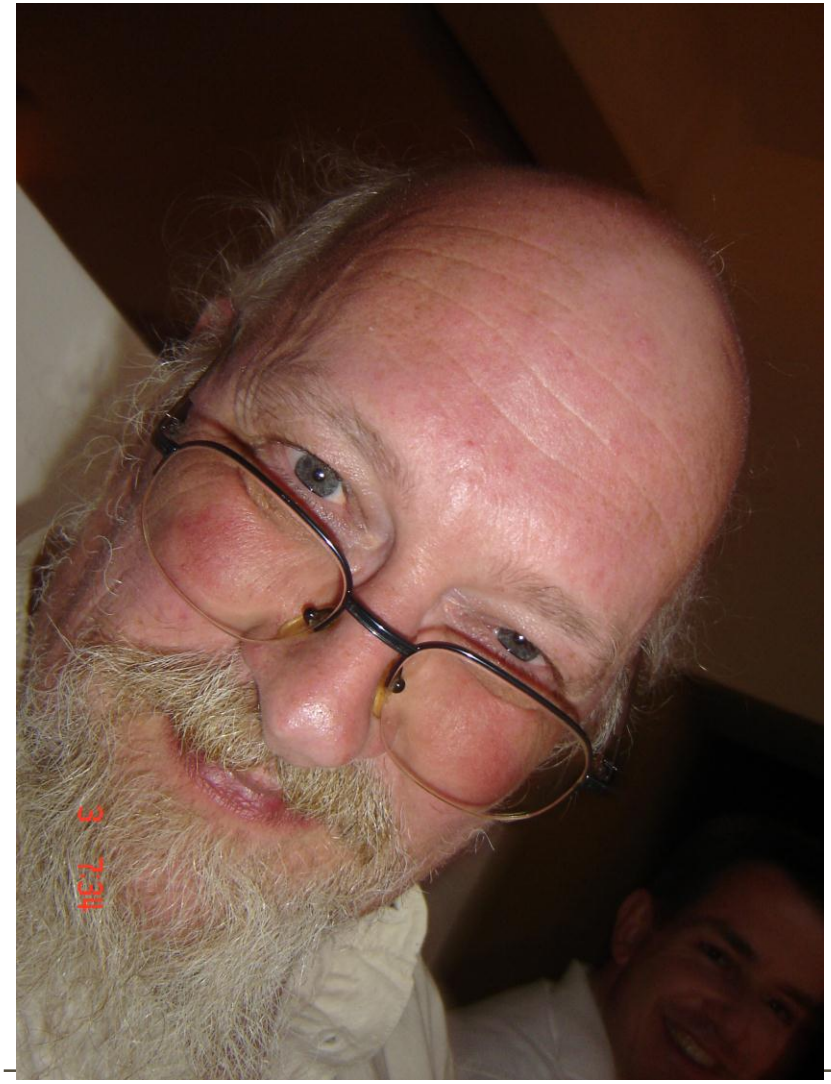
Study Number: Investigative

Peak ID	RT (min)	Proposed Structure	$[M+H]^+$ (m/z)	Diagnostic Fragment Ions (m/z)[1]	^{1}H NMR Chemical Shifts (ppm)[2]
M8 Hydration, dioxidation	3.9	 Fraction 72	485 (P+32+18) $C_{25}H_{41}O_8$ Theory: 485.2751 Found: 485.2770 Diff: 3.9 ppm.	369: F1+ 3[O] + 2[H] 351: m/z 369 – H_2O 333: m/z 369 – $2H_2O$  273: (F8) 247: F3 + 2[O] 229: m/z 247 – H_2O 115: F5 85: F6 59: F7	a- 0.78 b- 1.51 c- 1.09 d- 5.20 e- Obscured f- Obscured g- 1.16 h- 3.81 i- 5.83 j- 3.83 k- Obscured l- 0.78 m- Obscured n- Obscured o- Obscured p- Obscured q- 3.65 r- 1.52 s- 4.02 t- Obscured. 2.30 Spectrum in AQN/D2O

Study Number: Investigative

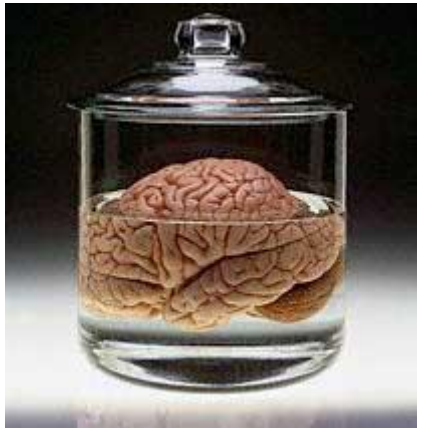
Peak ID	RT (min)	Proposed Structure	$[M+H]^+$ (m/z)	Diagnostic Fragment Ions (m/z)[1]	^{1}H NMR Chemical Shifts (ppm)[2]
M19 Lactonisation, oxidation, dehydrogenation, glucuronidation.	5.5	 Fraction 97	607 (P-18+14+176) $C_{31}H_{43}O_{12}$ Theory: 607.2755 Found: 607.2764 Diff: 1.5 ppm.	431: aglycone 319: F1 + [O] – 2[H] - [H_2O] 175, 113: glucuronic acid ions 115: F5 85: F6	a- 0.90 b- Obscured c- 1.18 d- 5.40 e- Obscured f- Obscured g- 1.15 h- 5.58 i- 6.05 j- 5.84 k- Obscured l- 0.95 m- Obscured n- Obscured o- Obscured p- Obscured q- Obscured r- Obscured s- Obscured t- Obscured Gluc anomeric- 5.02

A good memory



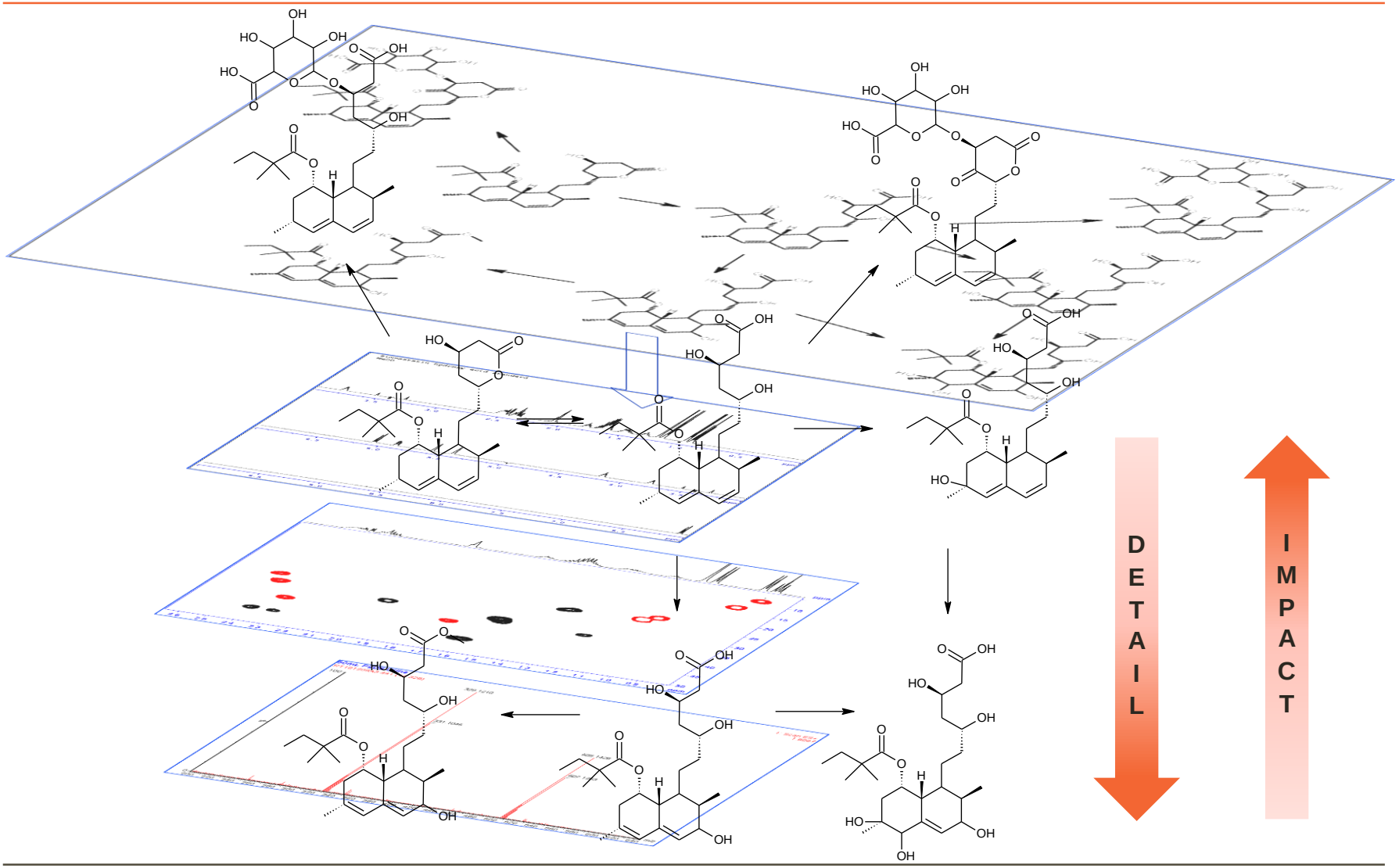
Grand wizard of metabolism – Frank

GSK employee:
Late Jurassic - Feb 2009

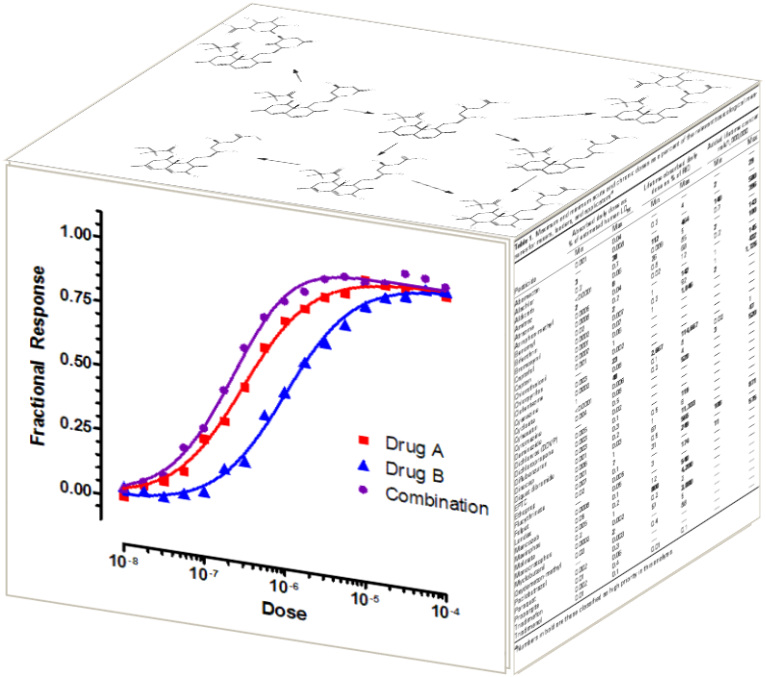


Contingency plan on Frank's retirement!

Drill down from a biotransformation map



Not a database, a data cube
Search from any point of interest



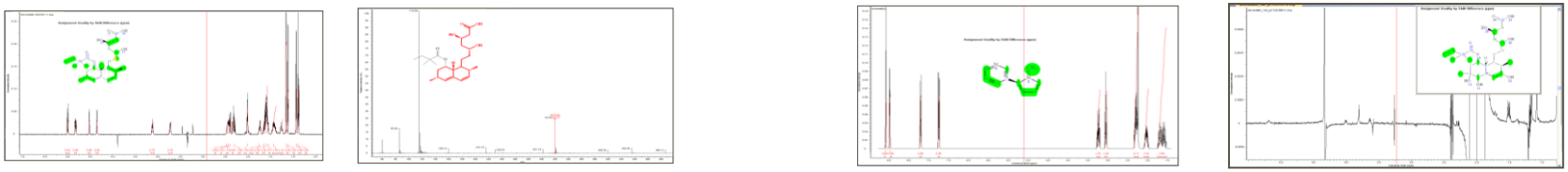
Variety of customers leads to a wide range of ideas as to which data is most important.

ACD/Chemfolder Enterprise

Initially designed as a tool for synthetic chemists:

– CFE allows users to database reactions
– Links to data supporting structure assignments
– Structures searchable
  –“Has this been made before?”
– Data searchable
  –“What have I made?”

A + B → C + D



To be of use in metabolism, the system must cope with 50+ products?
Your workflows?

Success criteria

1. How easy is it to get the data into the database?

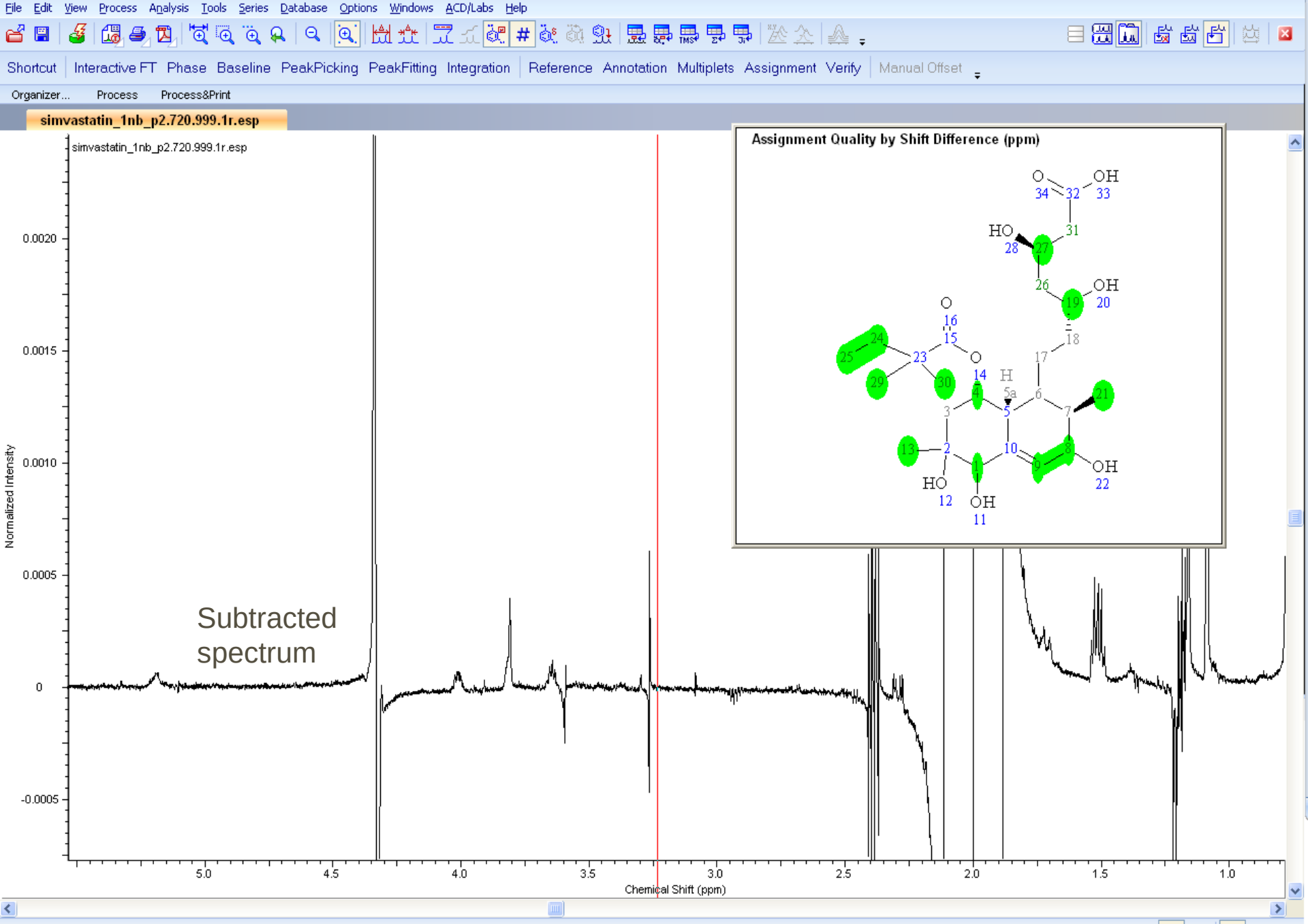
2. Can it help us move away from relying on the power of memory to solve tricky problems if seen before and gain access to other associated information?

3. Can we speed up the process of interpretation?

4. Can we have greater confidence in the elucidations and structural ID?

5. Can it help us avoid mistaken elucidations?

6. Will the proposed solution offer advantages in communication of data?


File Edit View Process Analysis Tools Series Database Options Windows ACD/Labs Help
Shortcut Interactive FT Phase Baseline PeakPicking PeakFitting Integration Reference Annotation Multiplets Assignment Verify Manual Offset
Organizer... Process Process&Print
simvastatin_1nb_p2.720.999.1r.esp
Assignment Quality by Shift Difference (ppm)
Subtracted spectrum
Normalized Intensity
Chemical Shift (ppm)
simvastatin_1nb_p2.720.999.1r.esp SPECTRUM 1H δ = 3.23 ppm
1-ChemSketch 2-Structure 3-Processor 4-Database 5-Search

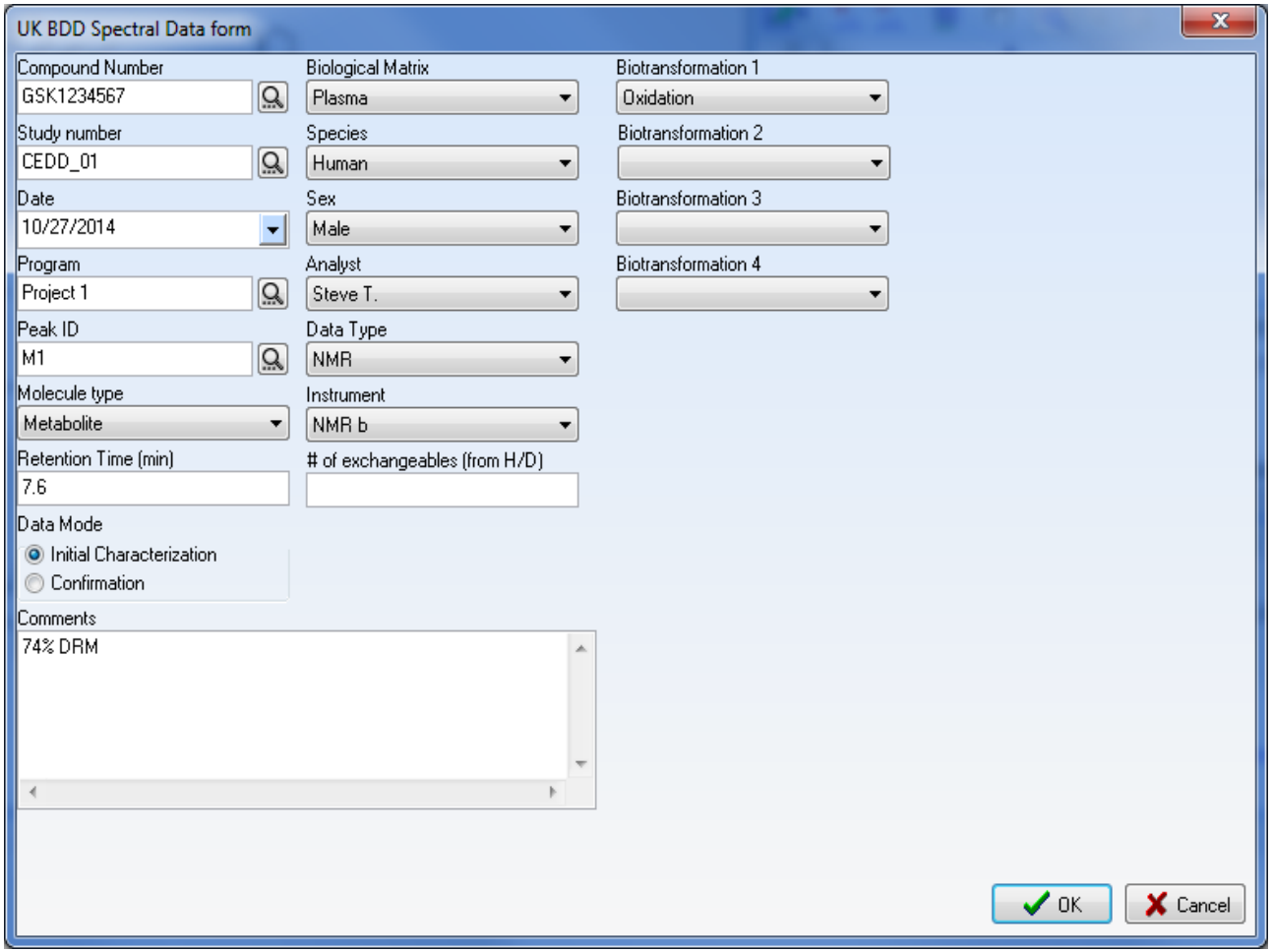
User data entry form

How to ensure your data is easily found.

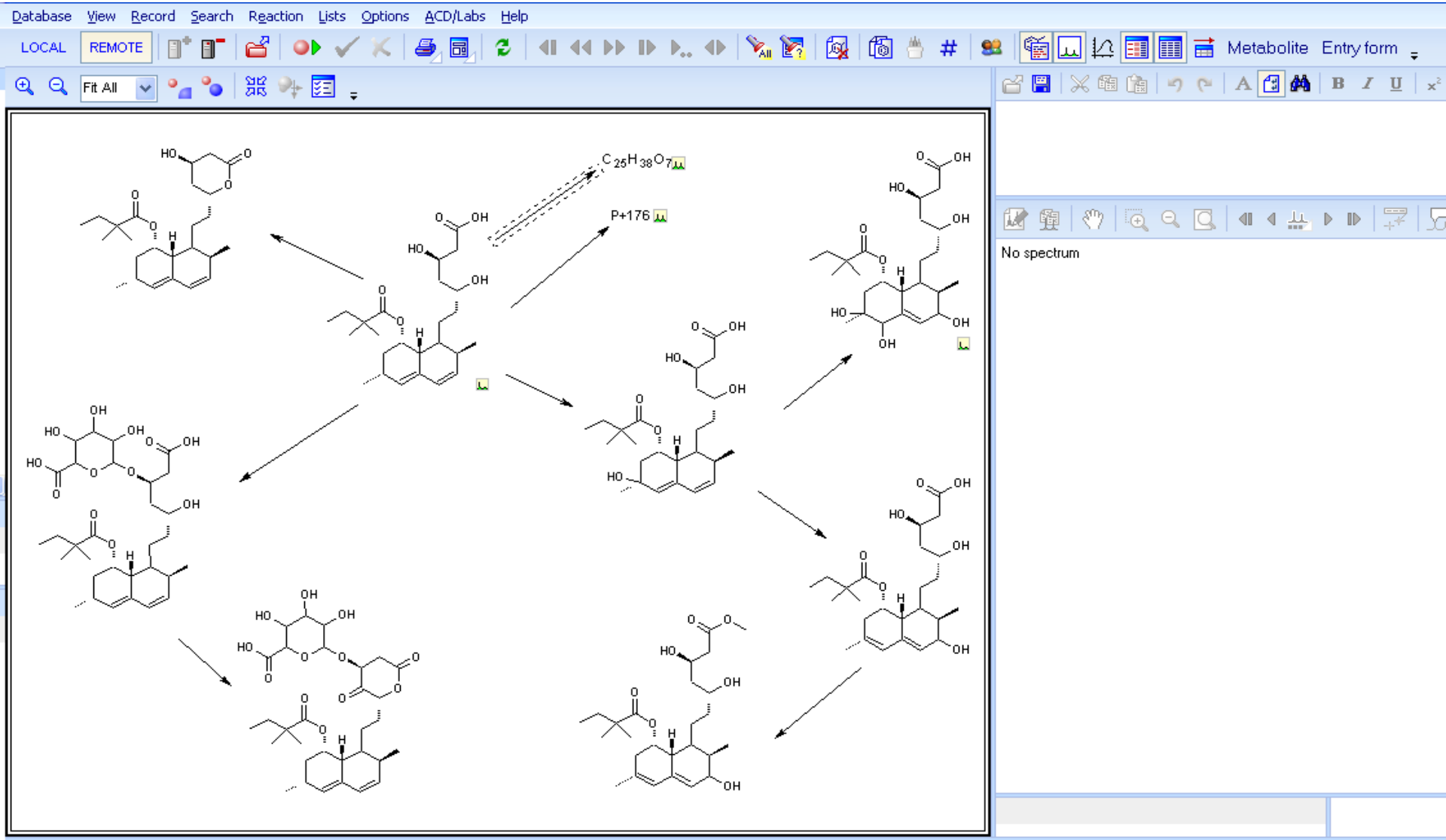
Auto triggered on update.

Fully customisable

- Auto fill e.g. Date
- Combiboxes
- Radio (choice)
- Mandatory entry
- Free text



Screen space and unknowns



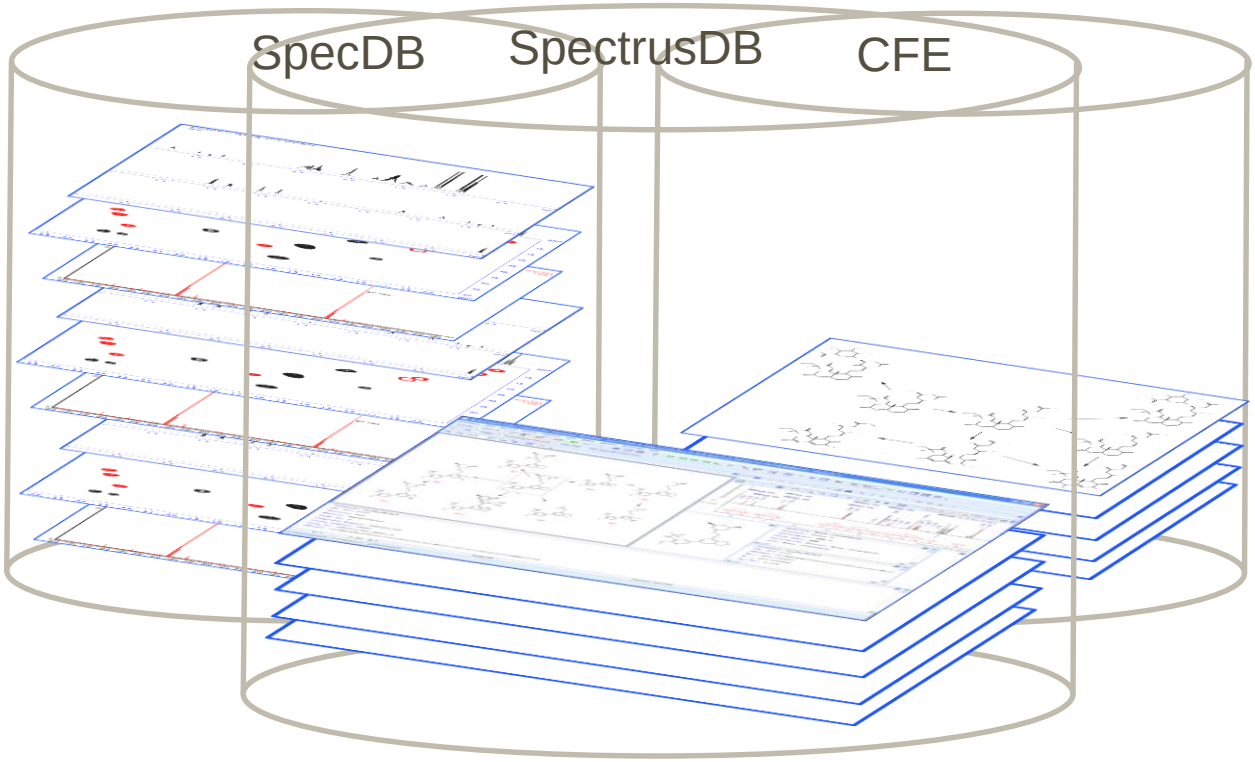
Reaction Mass Difference: 13.9793 Da ($-H_2$, +0)

Script Information: CFE record does not contain data

Enzyme: **CYP3A4**

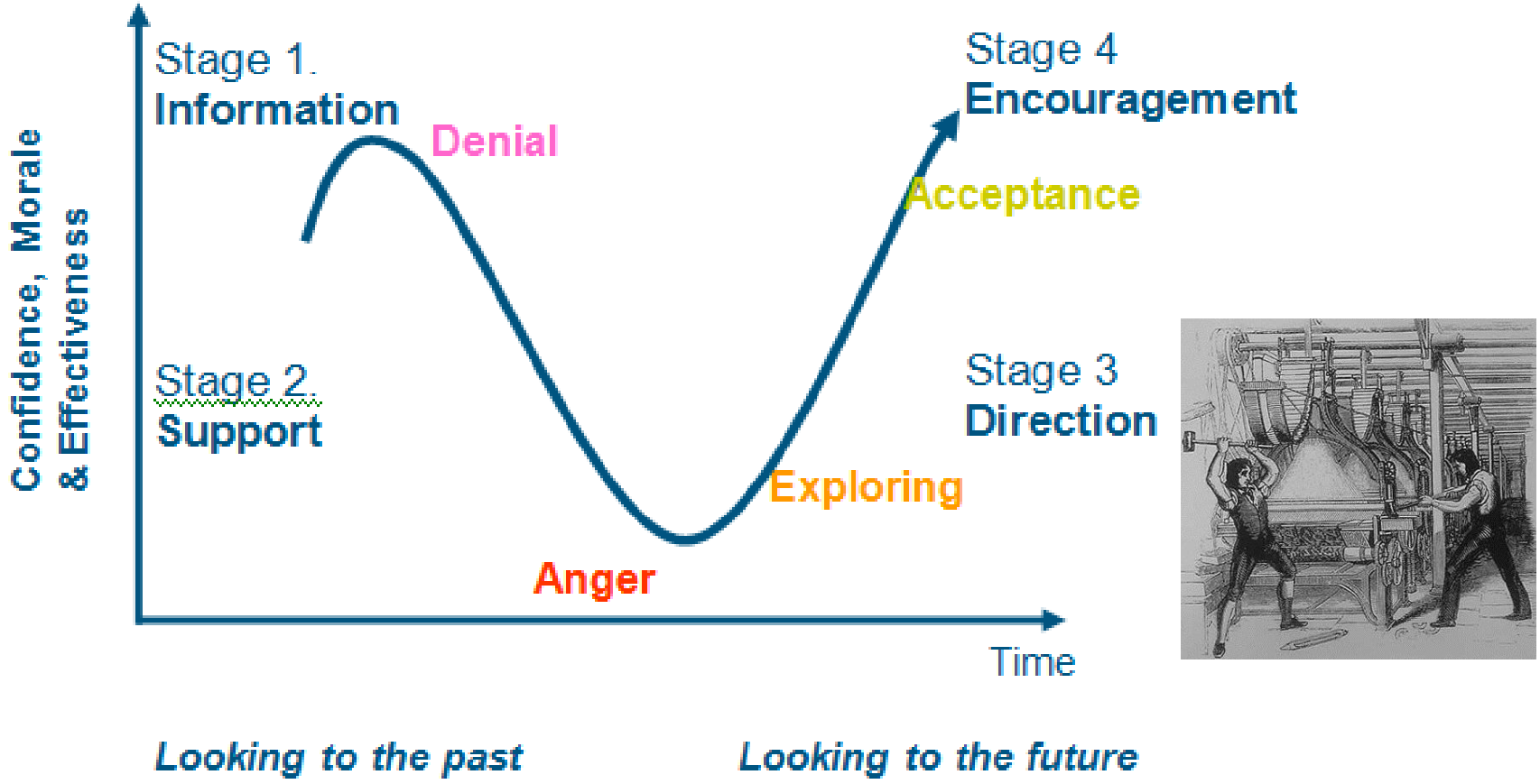
Category	Formula	FW
Scheme		
Reaction 1		

Merger of ACD/Chemfolder and ACD/SpecDB

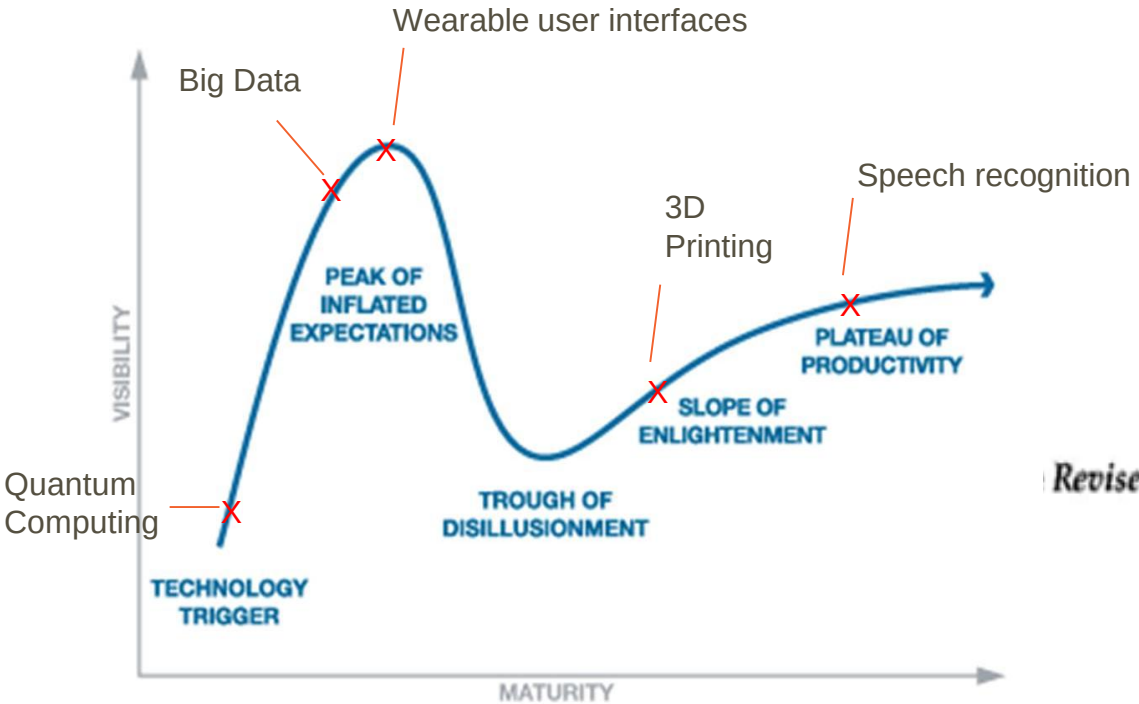


ACD/Spectrus Processor and ACD/Chemfolder Enterprise now together as a single Spectrus Database.

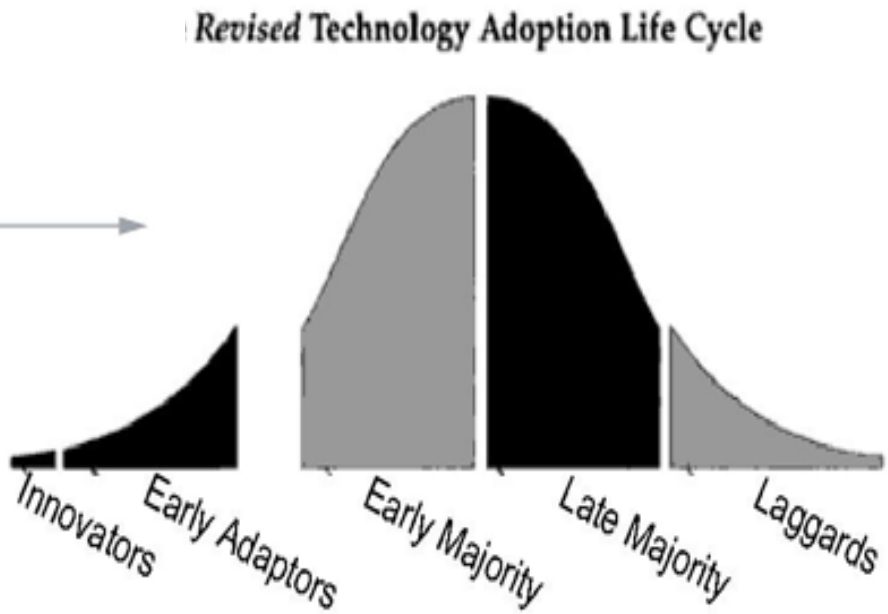
Change management curve



Momentum and speed bumps



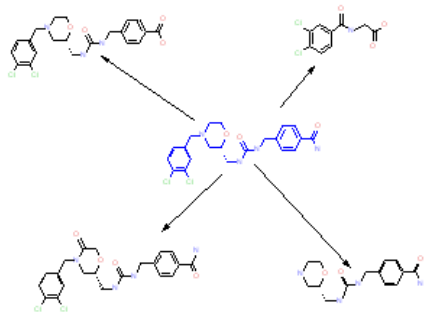
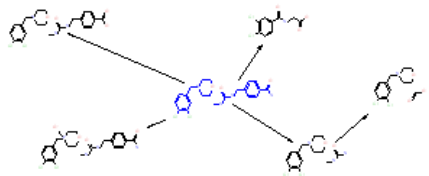
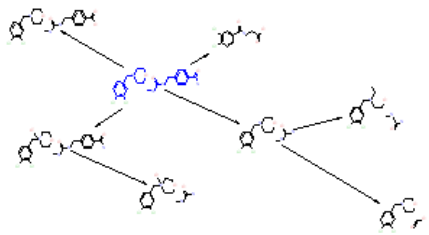
"The most frequent point of failure is the chasm between early adopters '***The visionaries***' and early majority, '***The pragmatists***'." John Trigg



Gartner "**Hype cycle**"

Expectations of novel technology over time

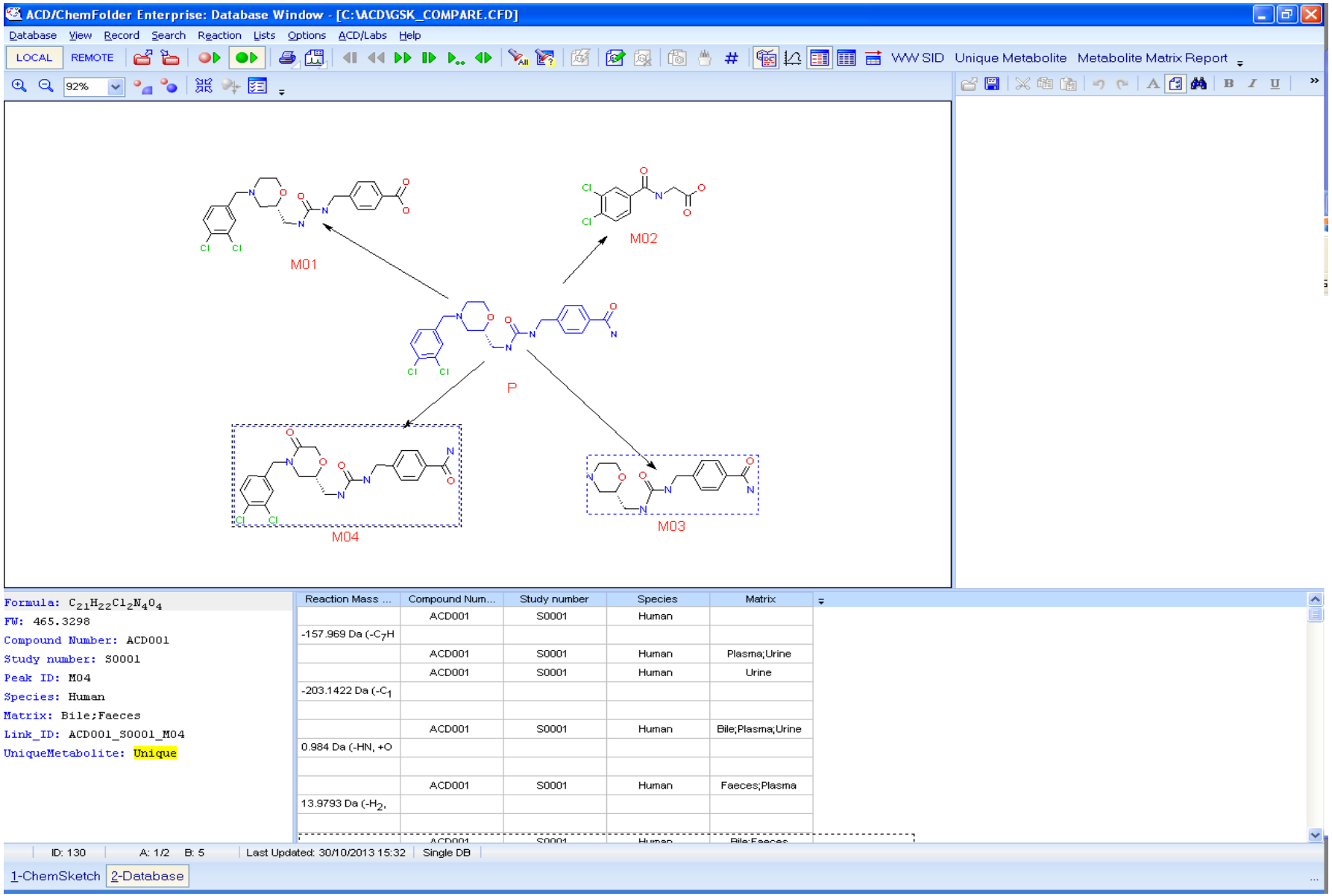
Scripting 1. - Map comparison

#ID	Structure	Species	Compound Number	Study number
130		Human	ACD001	S0001
146		Dog	ACD001	S0001
177		Rat	ACD001	S0001

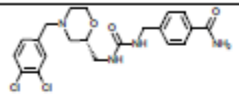
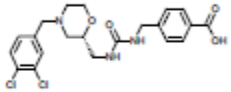
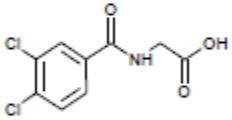
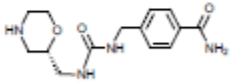
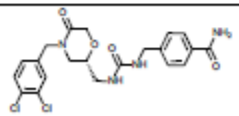
Safety cover?

Rat and dog data for illustrative purposes only

Scripting 1. - Map comparison

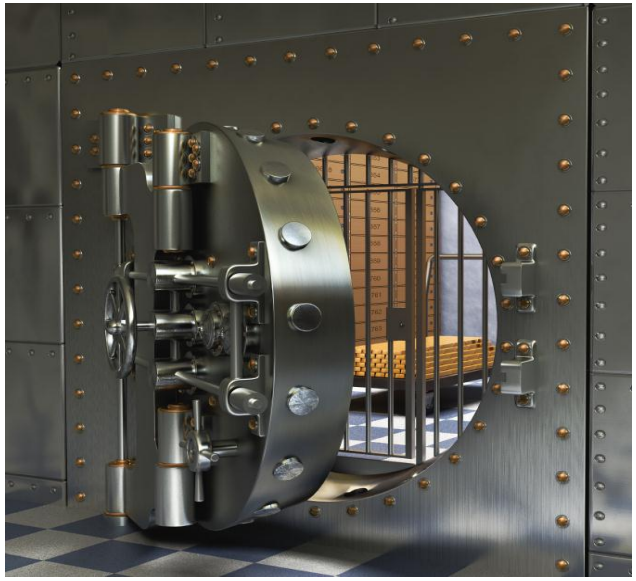


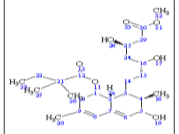
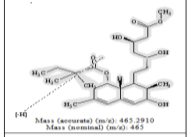
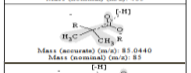
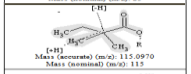
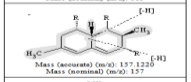
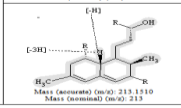
Scripting 2. - Matrix report

Peak ID	Link_ID	Structure	Bile	Faeces	Plasma	Urine
P	ACD001_S0001_P		ND	ND	X	X
M01	ACD001_S0001_M01		ND	X	X	ND
M02	ACD001_S0001_M02		X	ND	X	X
M03	ACD001_S0001_M03		ND	ND	ND	X
M04	ACD001_S0001_M04		X	X	ND	ND

Scripted table in word summarising where each metabolite was found

Outreach – disseminating the knowledge



Peak ID	RT (min)	Proposed Structure	Mass (m/z)	Diagnostic Fragment Ions (m/z)	1H NMR Chemical Shifts (ppm)
	0		M [H-] (accurate): 465.2910 M [H-] (nominal): 465	 Mass (accurate) (m/z): 465.2910 Mass (nominal) (m/z): 465  Mass (accurate) (m/z): 83.0440 Mass (nominal) (m/z): 83  Mass (accurate) (m/z): 115.0970 Mass (nominal) (m/z): 115  Mass (accurate) (m/z): 157.1220 Mass (nominal) (m/z): 157  Mass (accurate) (m/z): 213.1510 Mass (nominal) (m/z): 213	18 - 0.82 23 - 0.82 28 - 1.11 27 - 1.15 13 - 1.36 22 - 1.55 24 - 1.65 20 - 1.77 4 - 2.02 10eq - 2.26 29eq - 2.32 29ax - 2.38 10ax - 2.50 32 - 3.60 16 - 3.80 9 - 3.96 25 - 4.13 1 - 5.41 6 - 5.57 8 - 5.95

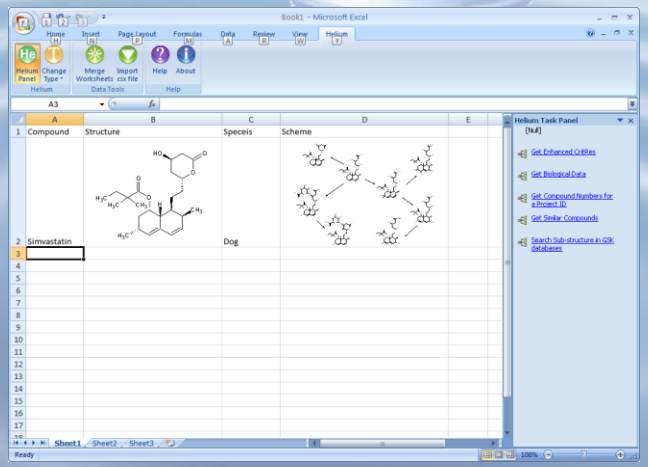
Avoid “Silo” mentality

- Users typically used to facile, free access will not want access restricted by logistics or cost.
- Without barriers our data will reach everyone it can help

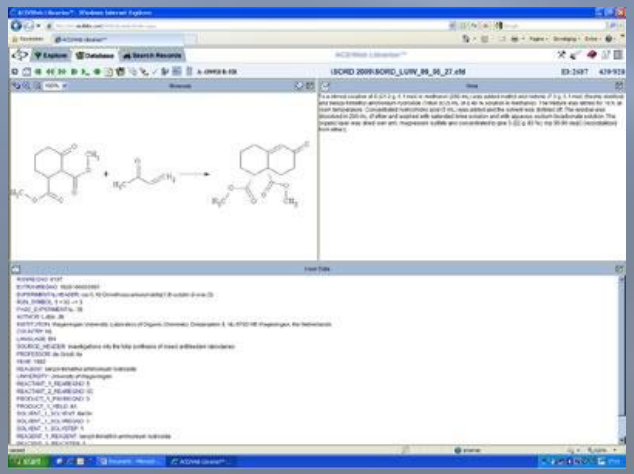
Dissemination strategies

Local super users

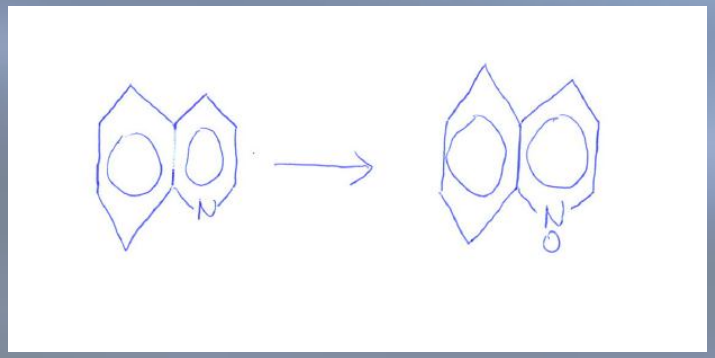
Helium



ACD/Web Librarian

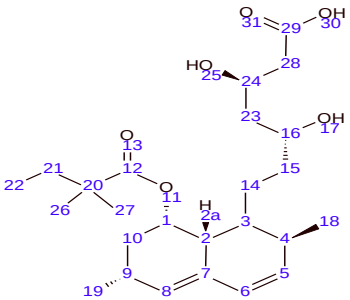


Requests!



Scripting 3. – Automated reporting (scripting)

Peak ID P
Retention Time 8.6 min
Mol Weight: 436.3



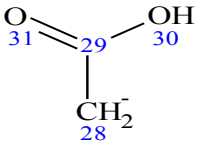
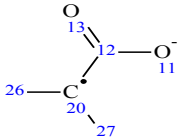
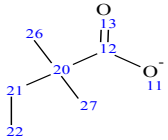
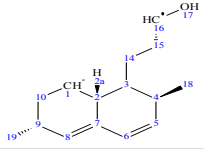
Formula	*m/z* Calc.	Difference (Da)	*m/z* Exp.
C25H39O6	435.2752	-0.1155	435.1597

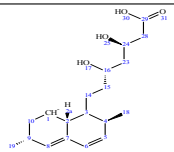
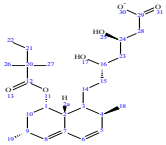
1H NMR:

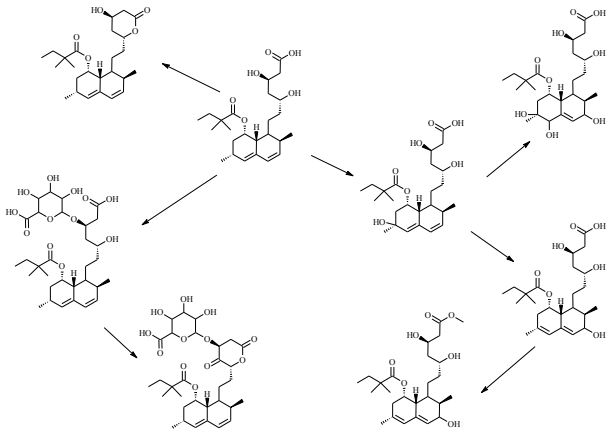
Atom	Exp. Shift (ppm)
18	0.94
19	1.12
15	1.16
27	1.16
26	1.16
15	1.27
10	1.43
23	1.43
23	1.57
21	1.61
14	1.61
21	1.66
3	1.74
10	2.02

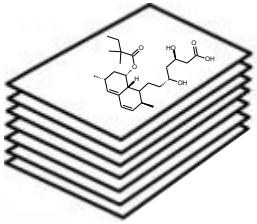
Atom	Exp. Shift (ppm)
4	2.31
9	2.35
2a	2.41
28	2.45
28	2.46
16	3.74
24	4.13
1	5.36
8	5.53
5	5.83
6	6.01

MS Fragmentation:

Structure	*m/z* Exp.
	59
	85
	115
	215

Structure	*m/z* Exp.
	319
	435





Comments:

Success criteria

1. **How easy is it to get the data into the database?**
   Variable e.g MassLynx ACD/Labs button, Xcaliber, Unify

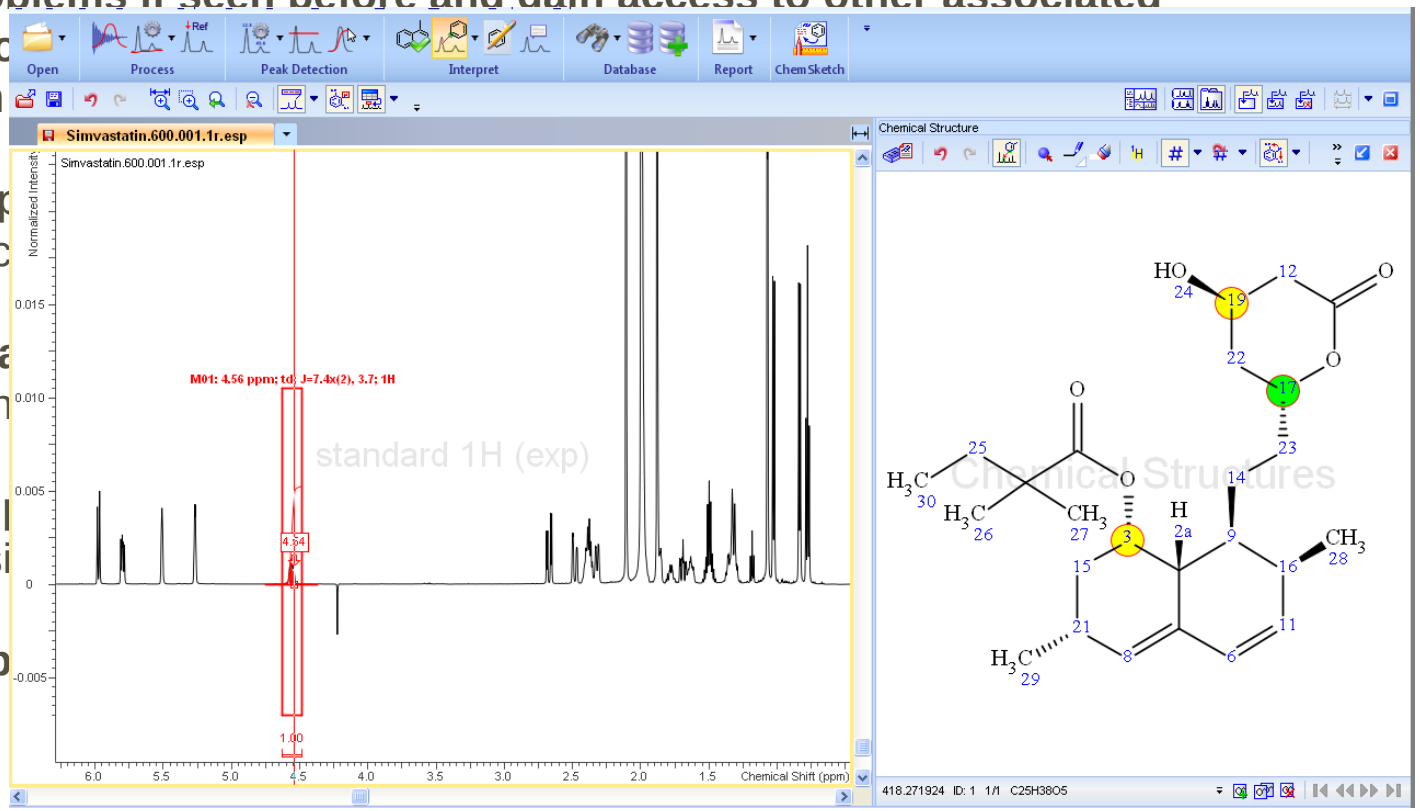
2. **Can it help us move away from relying on the power of memory to solve tricky problems if seen before and gain access to other associated informatic**
   Search on

3. **Can we sp**
   Ready acc

4. **Can we ha**
   Has anyon

5. **Can it hel**
   Smart assi

6. **Will the p**



Empowering the users – After action review

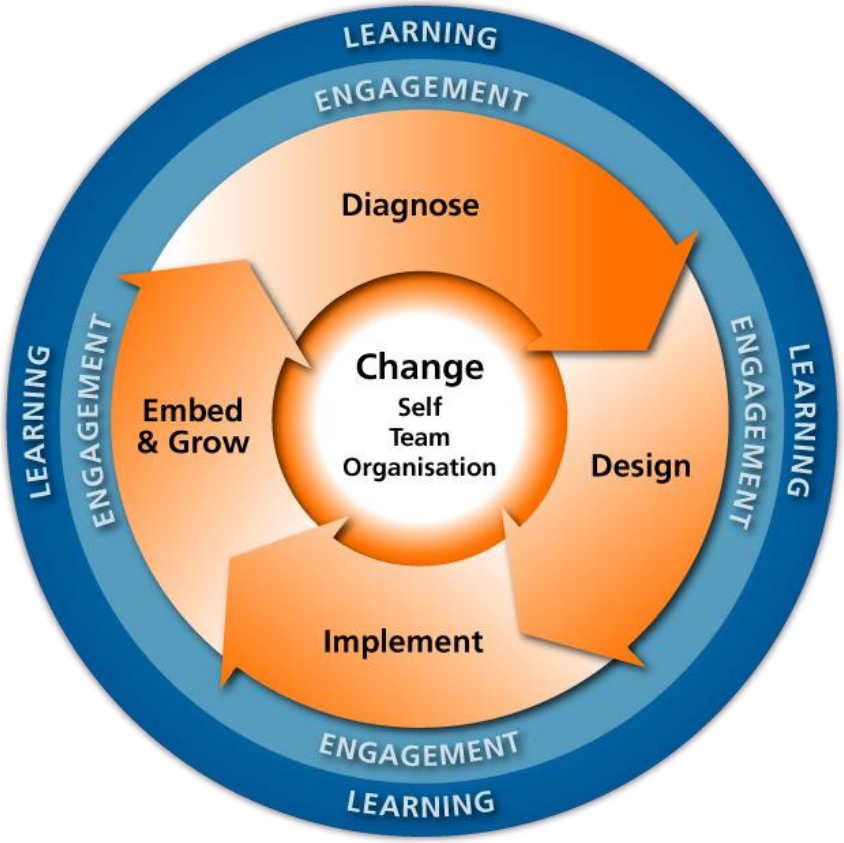
What went well?

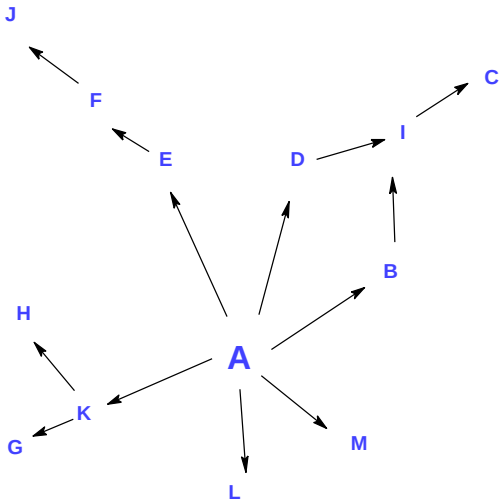
1. Communication between sites has been good on occasions
2. 1:1 training has been excellent
3. Regular up-dates
4. The database searchability is starting to show promise – some early adopters
5. Summer students – data entry (good use of resource)
6. Communicated well across boundaries and teams

What could be improved?

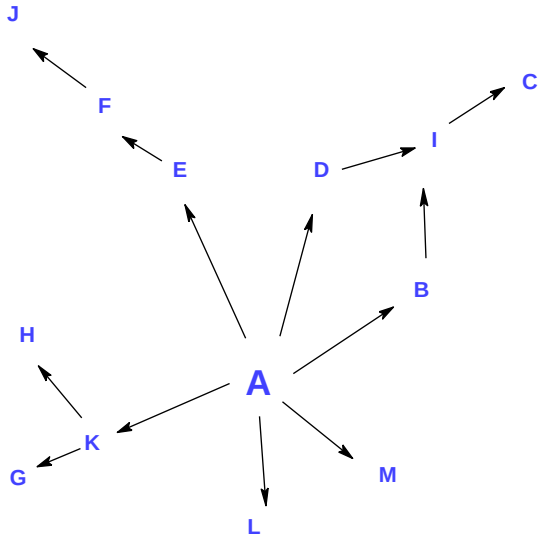
1. **Communication** between sites has been poor on occasions
2. **Training guide** is obsolete - needs up-dating (seen as vital)
3. "Our site didn't have a training guide, or at least weren't aware one existed "(**communication**)
4. Lack of t**raining** has hindered uptake
5. Clarity and visibility on business rules (**communication**)
6. Lack of energy at times – felt like lost momentum , because Spectrus upgrade was looming
7. Not easy to enter data - lack of **training** options?
8. Issues with processing software (automatic assignment) when looking at real metabolites
9. Same for MS (duplication)
10. Big issue with re-learning (as long time gaps between use) – comes back to provision of adequate **training** guide
11. Uncertainty about Spectrus – when, where and how (**communication**)
12. Better visibility to a project plan – when, where, how and who (**communication**)
13. **Training** set/tutorial should be available
14. Reporting and exporting issues – more energy in this space
15. Clarity around met id data reporting e.g. in report Appendix
16. Inconsistent message from ACD – different viewpoints from different ACD experts/sales folk?

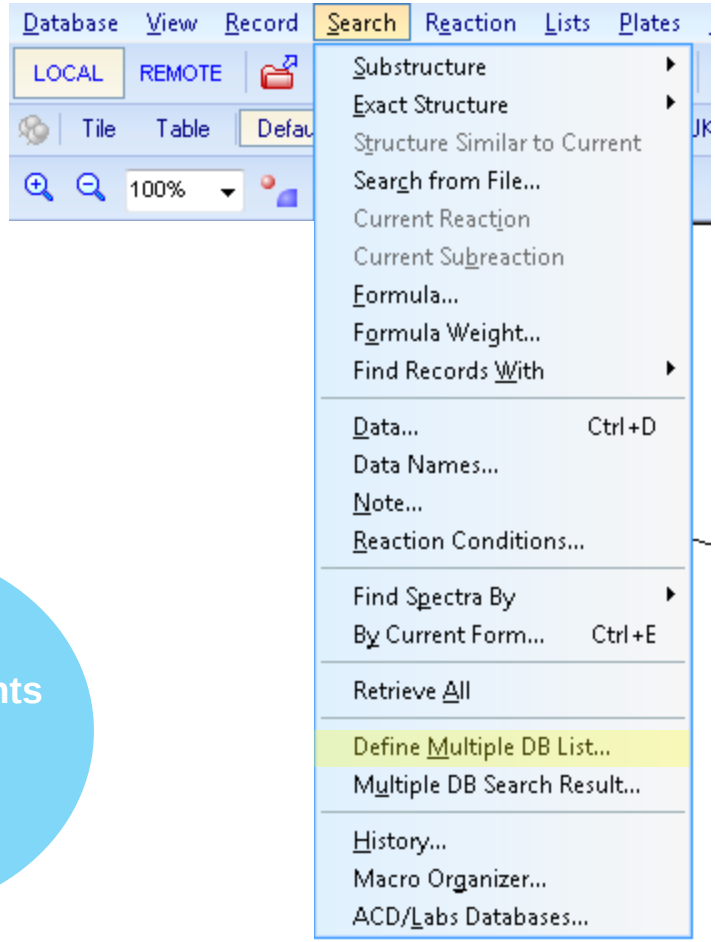
Spread the word
- Other groups facing similar challengess

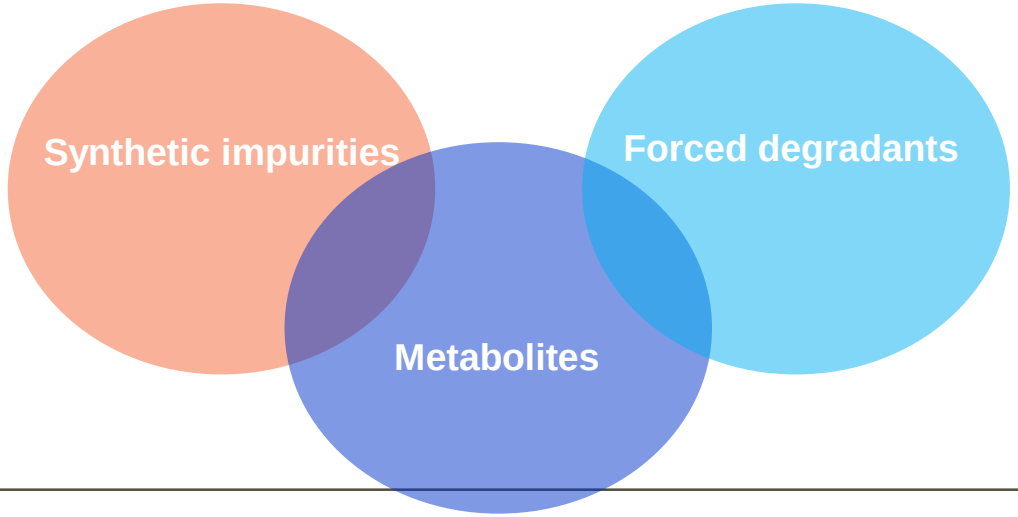




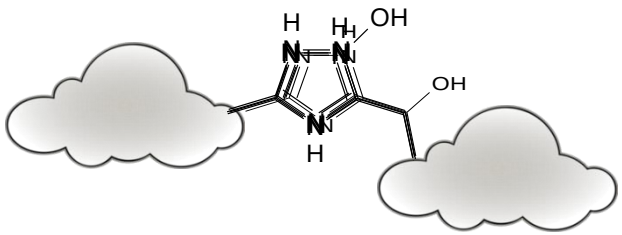
Working across boundaries

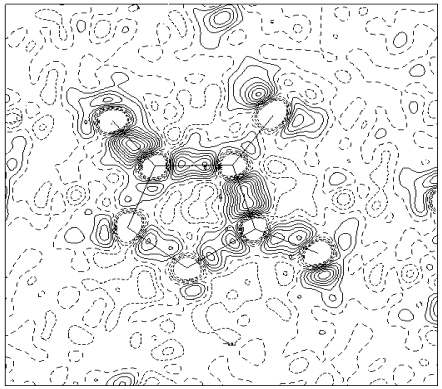


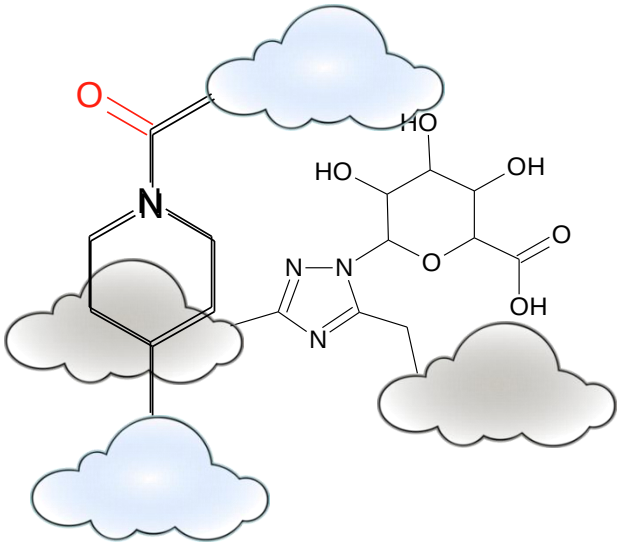


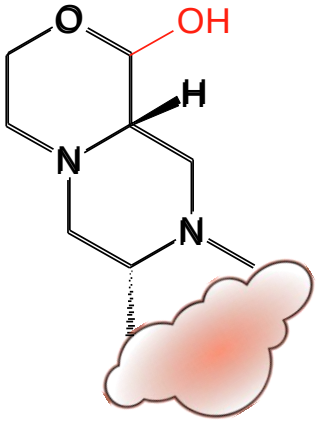


Synergy – Metabolism and Degradation









Summary

- Get everyone on-board. Community of empowered users.
  - Training, user guides
  - Hooks
- Valuable way to store, search and share sets of related molecules
  - Body (metabolites)
  - Environment (stability)
  - Chemist (reactions / impurities)
- Scripting capability
  - Automated reporting
- Further enhancements constantly delivered
  - ACD respond to new feature requests

Thank you