

Designing Data Repositories to Support Preservation and Publication for the Chemistry Community

Simon Coles
EPSRC National Crystallography Service
School of Chemistry
University of Southampton



Social Networks for Chemists



Google generation: new behaviour and approach

Sharing Rich Media



- Video + Paper = Pubcast

New Approaches to 'Sharing Experiments'



- Specialised domain-oriented innovations

Formation of Open Communities



- New approaches surfacing and growing FAST

Open Notebook Science

Useful Chemistry XML B subscribe with Bloglines UsefulChem molecules UsefulChem wiki

Google™ Custom Search Search

This is an open source science project in chemistry. Post specific problems in chemistry that you want to see solved. Post specific partial solutions to these problems. Or execute a suggested step. NOTE: ANYTHING POSTED HERE IS SUBJECT TO A SHARE-A-LIKE WITH ATTRIBUTION CREATIVE COMMONS LICENSE (see bottom of page)

UsefulChem Exp025 Protected page discussion history notify

Join this Space Recent Changes Manage Space Search

Navigation: All Experiments, Mailing List, Docking, Libraries, References, Experiment Format, Extra Credit, Paper01 Draft, Paper02 Draft, Isolated Compounds, Alicia's Masters Thesis, CombiUgi Project, To Do List

Objective
To convert [adrenaline](#) to [DOPAL](#) by acid catalysis

Procedure
A solution of [adrenaline](#) (1.0g 5.5mmoles) in 85% phosphoric acid was heated (116-118°C; 1h) in a heating mantle then removed from heat and allowed to cool. The solution was stirred for 1h, then saturated with NaCl. It was taken up in ethyl ether, and dried over anhydrous MgSO₄. The ether was removed to obtain DOPAL (80 mg 0.53mmol, 9.5% yield)

Characterization:
1. [TLC of 25A](#) in 3:1 MeCl₂/MeOH and in 6:1 MeCl₂/MeOH, and [stained with CAM](#)
2. [HNMR of 25A](#) in acetone-d₆, and the expansions of the prominent peak regions ([one](#) Varian inova). The integration is good enough to not require further purification.

Discussion
This is the first time that DOPAL was obtained pure (by NMR integration) immediately after the reaction under nitrogen and a careful temperature control. It is likely that the main impurity in previous attempts ([Exp016](#), [Exp023](#)) is the carboxylic acid ([Exp016HNMR](#), [Exp023HNMR](#)). The atmosphere should have prevented.

Conclusion
DOPAL can be obtained pure in 9.5% yield by heating adrenaline in 85% phosphoric acid followed by hydrolysis and extraction into ether.

Other Open Science Notebooks
Jean-Claude Bradley maintains a list of Open Science Notebooks. If you decide to make your lab notebook open, send him an email to be put on the list so others can find your work more easily.

Help me improve my notebook
I'm also trying to come up with some Open Notebook Science requirements and suggestions to help move towards an open notebook that is intelligible to people besides myself. I'd be interested for your comments/suggestions on what else should be included.

How I construct my notebook
A few people have written asking how I make the document itself. The document is made in latex (specifically pdflatex on a mac). I've posted a simple example electronic lab notebook in latex with the commands and functions that I frequently use. [sample.tex](#) is the main file; see this file to determine if you need to install any additional packages. [Introduction to Latex](#).

Open Science Isn't
Finally, I just want to...

RRRESEARCH
THINKING ABOUT OUR RESEARCH INTO THE MECHANISM, FUNCTION AND EVOLUTION OF DNA UPTAKE BY HAEMOPHILUS INFLUENZAE AND OTHER BACTERIA.

SATURDAY, JANUARY 19, 2008

New microarray data
The post-docs have finished the first-pass analysis of how *E. coli* gene expression is affected by both the *E. coli* Sxy and the *H. influenzae* Sxy proteins. I suppose I shouldn't be surprised that it's more complicated than I had hoped. For example, unlike the situation in *H. influenzae*, in *E. coli* there are also groups of genes whose expression goes down when Sxy is present.

One complication is that these cells are probably seriously OVER-producing Sxy. Unlike *H. influenzae*, where we've only done arrays of cells expressing a single-copy sxy gene under its natural promoter, these *E. coli* studies used a sxy gene on a high-copy plasmid and under a highly inducible promoter. We know that prolonged expression of Sxy from this plasmid produces large quantities of denatured Sxy (in inclusion bodies) and we don't know the extent to which even the 30-minute expression used for the array studies might create a situation unlike that of natural sxy expression.

POSTED BY ROSIE REDFIELD AT 6:25 PM

ABOUT ME
ROSIE REDFIELD CANADA
I run a microbiology research lab in the Life Sciences Centre at the University of British Columbia.
VIEW MY COMPLETE PROFILE

PREVIOUS POSTS
Thermodynamics of home heating
Gene transfer agent
Sorry, wrong link
Sorry for the paucity of posts
Genespring progress and problems
Data on *E. coli* protein

2 COMMENTS:
iayork said...
remember what you've said about sxy in the past. Can you make a null version of

- Immediate sharing of experimental information & data

New Information Exchange Environments



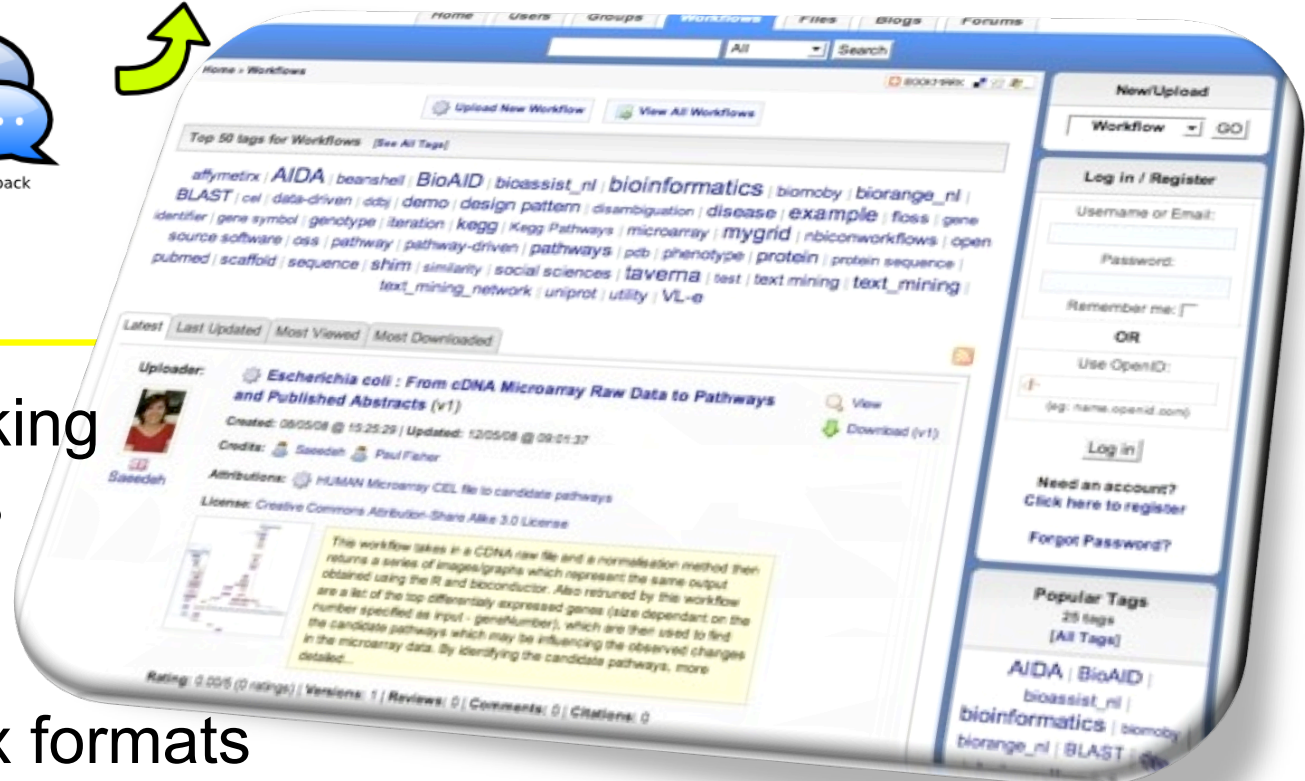
- Immersive alternative to conventional browsing & interaction

A Virtual Research Environment



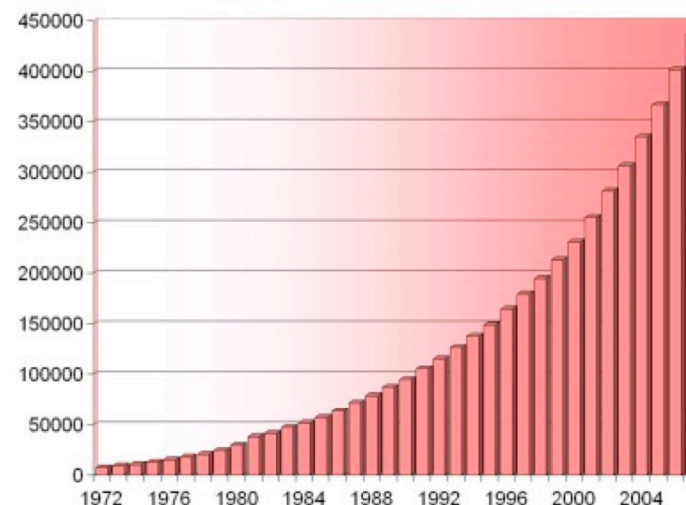
Adoption of “Web2.0” philosophy and methodology to promote exchange of research information:
The New eScience?!

- Community networking
- Targeted audiences
- Open or closed
- Discoverable
- Multimedia/complex formats



Data Deluge

- 40 years ago a PhD student would determine about 3 crystal structures during the course of their study – this can now be easily achieved in a day



- There are approx. 30 million known chemical compounds
- Approx. 2 million crystal structures have been determined
- There are less than 0.5 million published crystal structures residing in (licensed) curated databases
- There are just a few thousand 'open' crystal structures
- The primary cause of this is the current data publication process, which is tied to journal articles and peer review

Data Publishing & Open Access

- Short communications
- Electronic only
- Rapid publication
- Open access (01/2008)
- Highly cited
- Written in CIF
- Freely available tools
- **Still cannot keep up!**
- Journal 'article' format required
- Not all crystal structures are of primary importance to the underlying chemistry: by-products / unexpected results / tracking reactions



The Solution

Intellect & Interpretation
(Journal article, report, etc)

Underlying data
(Institutional data repository)

research papers

Acta Crystallographica Section B
Structural Science
ISSN 1502-3691

Serap Bedi,^a Simon J. Coles,^b David B. Davies,^a Michael B. Hurdhouse,^a Adem Kalc,^a Thomas A. Mayer,^b Robert A. Shaw,^a and Aydin Loh^a

Structural investigations of phosphorus–nitrogen compounds. 4. Steric and electronic effects in dibenzylamino derivatives of hexachlorocyclophosphazatriene and 4,4,6,6-tetrachloro-2,2-diphenylcyclotriphosphazatriene¹

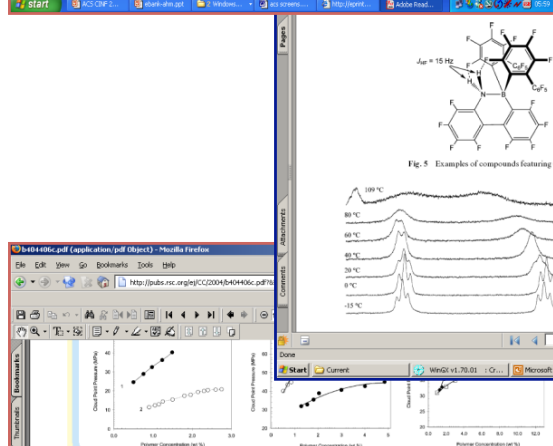
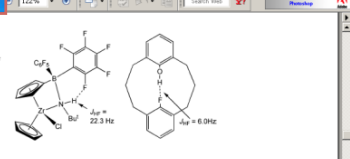
A systematic study is presented on the products of aminolysis of $N_2P_2Cl_4$ (1) and $N_2P_2P_2Cl_4$ (4) with dibenzylamine. Two series of mono- and disubstituted derivatives of compounds (1) and (4), namely $N_2P_2Cl_2[N(CH_2Ph)_2]$ (2) and $N_2P_2Cl_2[N(CH_2Ph)_2]$ (3) and $N_2P_2P_2Cl_2[N(CH_2Ph)_2]$ (5) and $N_2P_2P_2Cl_2[N(CH_2Ph)_2]$ (6) where (2), (3), (5) and (6) are new structures, are investigated in order to determine whether steric or electronic effects prevail in the formation of dibenzylamino-substituted cyclophosphazenes. The influence of an electron-releasing group (*i.e.* phenyl) on the stereochemistry and degree of substitution of the product is analysed by comparison of the above two series. The difference in unsymmetrically substituted acyclic P–N bond lengths, Δ , is used as a measure of the degree of the electronic contribution, in combination with basicity constants, to quantify the degree of the electron-releasing capacity of the *R* group. In order to compare general versus non-general substitution, a difunctional secondary amine was used to form the compound $N_2P_2Cl_2Me(CH_2CH_2NHMe)$ (7) (a re-investiga-

Received 13 December 2001
Accepted 27 February 2002

^aPart 3: Alkhalaf et al. (1998).

^bDepartment of Chemistry, Gates Institute of Technology, Gates, Turkey, ^cDepartment of Chemistry, University of Southampton, Highfield, Southampton SO17 1BJ, England, and ^dSchool of Biological and Chemical Sciences, Birkbeck College University of London, Gordon House, 29 Gordon Square, London WC1H 0PP, England

Correspondence e-mail: a.l.oh@birkbeck.ac.uk



particular, we noted that very small changes to structure led to dramatic and unpredictable changes in phase behavior, a frustrating situation. For example, it has been known for almost a decade¹ that poly(methyl acrylate (PMA) and poly(vinyl acetate (PVA) exhibit miscibility pressures in CO_2 that differ by hundreds of bar – this result would not be predicted by any group contribution thermodynamic model currently in use without purely empirical adjustments. Indeed, the PMA/PVA effect is preserved even when the material that is for all intents and purposes, insoluble. Again, traditional thermodynamic models provide no guidance here.

We believe that at least part of the answer to these puzzles lies in CO_2 's ability to act as both Lewis acid and Lewis base, coupled with subtle effects of neighboring substituents on the acidity of certain protons. For example, Wadden and colleagues^{2,3} in an analysis of interactions between acetate groups and CO_2 found that the acidity of the methyl acetate

$[P(O)Cl_2N_2][N(CN)BA]^+ [A]^-$ ($A^+ = C_2F_5$ (4a) or $C_2F_5C_2F_5$ (5a)). Although dyanamide has low lone pairs and is therefore theoretically capable of binding to three or possibly four borane molecules, only the diborates were obtained, even in the presence of excess $BA(N_2)$ (Scheme 5). Colourless crystals of the sodium salts 4a and 5a were grown from diethyl ether but proved unsuitable for X-ray diffraction studies. Further reaction of the sodium salts with $PhCl_3$ in dichloromethane afforded the triyl derivatives, $[P(O)Cl_2N_2CNBA(C_6H_5)]^+ [A']^-$ ($A' = C_2F_5$ (4b) and $C_2F_5C_2F_5$ (5b)). The structure of 4b is shown in Fig. 7. Long term storage of a solution of 5b eventually yielded orange crystals suitable for X-ray analysis which, however, once again turned out to be $[P(O)Cl_2N_2CNBA(C_6H_5)]^+ [A']^-$ (Fig. 7). The crystal structure of 4b (Fig. 7) confirms that the BCl_3F_3



Table 1
Experimental details

	(2)	(3)	(5)	(6)	(7)
Crystal data	$C_{12}H_{14}Cl_2N_2P_2$	$C_{12}H_{14}Cl_2N_2P_2$	$C_{12}H_{14}Cl_2N_2P_2$	$C_{12}H_{14}Cl_2N_2P_2$	$C_{12}H_{14}Cl_2N_2P_2$
Chemical formula	$C_{12}H_{14}Cl_2N_2P_2$	$C_{12}H_{14}Cl_2N_2P_2$	$C_{12}H_{14}Cl_2N_2P_2$	$C_{12}H_{14}Cl_2N_2P_2$	$C_{12}H_{14}Cl_2N_2P_2$
Chemical formula	$C_{12}H_{14}Cl_2N_2P_2$	$C_{12}H_{14}Cl_2N_2P_2$	$C_{12}H_{14}Cl_2N_2P_2$	$C_{12}H_{14}Cl_2N_2P_2$	$C_{12}H_{14}Cl_2N_2P_2$
Weight	350.47	350.47	350.47	350.47	350.47
Cell setting, space group	Monoclinic, $P2_1/n$	Orthorhombic, $Pbcm$	Monoclinic, $P2_1$	Monoclinic, $P2_1$	Orthorhombic, $P2_12_12_1$
a, b, c (Å)	11.225 (2), 8.187 (1), 18.294 (4)	11.077 (2), 8.938 (1), 10.203 (1)	11.077 (2), 8.938 (1), 10.203 (1)	11.077 (2), 8.938 (1), 10.203 (1)	11.077 (2), 8.938 (1), 10.203 (1)
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
V (Å ³)	2112.9 (7)	968.7 (3)	1370.5 (5)	1370.5 (5)	1470.5 (7)
Z	4	2	4	4	4
D_x (Mg m ⁻³)	1.888	1.405	1.408	1.408	1.405
Refinement type	Full data	Full data	Full data	Full data	Full data
R (intensity)	0.026	0.026	0.026	0.026	0.026
R (weight)	0.026	0.026	0.026	0.026	0.026
wR (intensity)	0.026	0.026	0.026	0.026	0.026
wR (weight)	0.026	0.026	0.026	0.026	0.026
Temperature (K)	120 (2)	120 (2)	120 (2)	120 (2)	120 (2)
Crystal form, colour	Block, colourless	Block, colourless	Block, colourless	Block, colourless	Block, colourless
Crystal size (mm)	0.35 × 0.25 × 0.15	0.42 × 0.15 × 0.15	0.42 × 0.15 × 0.15	0.42 × 0.15 × 0.15	0.30 × 0.15 × 0.15
Data collection	Nano Kappa CCD	Nano Kappa CCD	Nano Kappa CCD	Nano Kappa CCD	Nano Kappa CCD
Diffractometer	g and o scan	g and o scan	g and o scan	g and o scan	g and o scan
Data collection method	Multi-scan	Multi-scan	Multi-scan	Multi-scan	Multi-scan
Absorption correction	0.7066	0.7142	0.7066	0.7142	0.7066
T_{min}	0.8666	0.8666	0.8666	0.8666	0.8666
No. of measured, independent and observed reflections	11 436, 4796, 4054	22 246, 5347, 3033	13 855, 4508, 3152	42 554, 8051, 7100	7551, 3272, 3003
Observed reflections	11 436	22 246	13 855	42 554	7551
R (intensity)	0.026	0.026	0.026	0.026	0.026
R (weight)	0.026	0.026	0.026	0.026	0.026
Range of a, b, c (Å)	–14 → 14	–13 → 13	–13 → 13	–13 → 13	–14 → 14
Range of h, k, l (Å)	–8 → 8	–10 → 10	–10 → 10	–10 → 10	–10 → 10
Range of h, k, l (Å)	–10 → 10	–10 → 10	–10 → 10	–10 → 10	–10 → 10

Preparation of $C_2F_5N_2P_2Cl_2$ (1)
Phosphorus trichloride (12.5 g, 55 mmol) was ground to a powder using a pestle and mortar in a dry box. PBD (0.47 g, 5.5 mmol) and 50 mL diethyl ether were then added, and the mixture was heated to reflux for 12 h. The solvent was removed *in vacuo* to leave an off-white foam which was washed with warm hexane (50 mL) to give $C_2F_5N_2P_2Cl_2$ (1) as a white powder (0.495 g, 4.05 mmol). This solid was stirred with triphenylmethylmethane (0.15 g, 0.405 mmol) in dichloromethane (15 mL) for 2 h. The solution was filtered to remove KCl , concentrated to 5 mL, and cooled to –20 °C to give an orange crystalline solid, yield 0.324 g (0.315 mmol, 60% with respect to KCl). IR (neat): 2109 (s), 1719 (s), 1619 (s), 1519 (s), 1419 (s), 1319 (s), 1219 (s), 1119 (s), 1019 (s), 919 (s), 819 (s), 719 (s), 619 (s), 519 (s), 419 (s), 319 (s), 219 (s), 119 (s), 9 (s). 1H NMR ($CDCl_3$, 20 °C): 3.00 (t, 2H, $J = 7.5$ Hz, CH_2), 2.00 (t, 2H, $J = 7.5$ Hz, CH_2), 1.00 (t, 3H, $J = 7.5$ Hz, CH_3). ^{13}C NMR ($CDCl_3$, 20 °C): 140.3 (s, CH_2), 130.0 (s, CH_2), 125.0 (s, CH_2), 120.0 (s, CH_2), 115.0 (s, CH_2), 110.0 (s, CH_2), 105.0 (s, CH_2), 100.0 (s, CH_2), 95.0 (s, CH_2), 90.0 (s, CH_2), 85.0 (s, CH_2), 80.0 (s, CH_2), 75.0 (s, CH_2), 70.0 (s, CH_2), 65.0 (s, CH_2), 60.0 (s, CH_2), 55.0 (s, CH_2), 50.0 (s, CH_2), 45.0 (s, CH_2), 40.0 (s, CH_2), 35.0 (s, CH_2), 30.0 (s, CH_2), 25.0 (s, CH_2), 20.0 (s, CH_2), 15.0 (s, CH_2), 10.0 (s, CH_2), 5.0 (s, CH_2), 0.0 (s, CH_2). ^{31}P NMR ($CDCl_3$, 20 °C): 140.3 (s, CH_2), 130.0 (s, CH_2), 125.0 (s, CH_2), 120.0 (s, CH_2), 115.0 (s, CH_2), 110.0 (s, CH_2), 105.0 (s, CH_2), 100.0 (s, CH_2), 95.0 (s, CH_2), 90.0 (s, CH_2), 85.0 (s, CH_2), 80.0 (s, CH_2), 75.0 (s, CH_2), 70.0 (s, CH_2), 65.0 (s, CH_2), 60.0 (s, CH_2), 55.0 (s, CH_2), 50.0 (s, CH_2), 45.0 (s, CH_2), 40.0 (s, CH_2), 35.0 (s, CH_2), 30.0 (s, CH_2), 25.0 (s, CH_2), 20.0 (s, CH_2), 15.0 (s, CH_2), 10.0 (s, CH_2), 5.0 (s, CH_2), 0.0 (s, CH_2). ^{19}F NMR ($CDCl_3$, 20 °C): 140.3 (s, CH_2), 130.0 (s, CH_2), 125.0 (s, CH_2), 120.0 (s, CH_2), 115.0 (s, CH_2), 110.0 (s, CH_2), 105.0 (s, CH_2), 100.0 (s, CH_2), 95.0 (s, CH_2), 90.0 (s, CH_2), 85.0 (s, CH_2), 80.0 (s, CH_2), 75.0 (s, CH_2), 70.0 (s, CH_2), 65.0 (s, CH_2), 60.0 (s, CH_2), 55.0 (s, CH_2), 50.0 (s, CH_2), 45.0 (s, CH_2), 40.0 (s, CH_2), 35.0 (s, CH_2), 30.0 (s, CH_2), 25.0 (s, CH_2), 20.0 (s, CH_2), 15.0 (s, CH_2), 10.0 (s, CH_2), 5.0 (s, CH_2), 0.0 (s, CH_2).

The eCrystals Data Repository

- Quick & simple to deposit
- Software tools
- Laboratory archive
- Community involvement
- 'Embargo' facility
- Structured foundations
- Discoverable & harvestable



eCrystals
 Home | About | Browse by Year | Browse by People
 Login | Create Account

6,7,9,10,12,13,15,16-Octahydro-benzo-1,4,7,10,13-pentaoxacyclopentadecin

Sample Originator: Esther Rousay* and Jeremy G. Frey*
 Data Collection: Simon J. Coles*
 Structure Determination: Simon J. Coles* and Michael B. Hursthouse*
 University of Southampton*

C14H20O5
 InChI=1/C14H20O5/c1-2-4-14-13(3-1)18-11-9-16-7-5-15-6-8-17-10-12-19-14/h1-4H,5-12H2

Identification Number: 10.3737/ecrystals.chem.soton.ac.uk/145
 Controlled Keywords: crown ethers, crown

Date Created: 07 October 2004
 Deposited On: 21 Jan 2006 15:29
 Deposited By: Dr Simon J Coles

Available Files

Final Result

- [04sjc0831.cif](#)
- [04sjc0831.cml](#)
- [04sjc0831.fcf.txt](#)

Collection parameters

Chemical formula	C14 H20 O5
Crystallisation Solvent	
Crystal morphology	Plate
Crystal system	Orthorhombic
Space group symbol	Pbca
Cell length a	16.4963(18)
Cell length b	8.325(3)
Cell length c	20.061(6)
Cell angle alpha	90.00
Cell angle beta	90.00
Cell angle gamma	90.00
Data collection temperature	120(2)

Refinement results

Solution figure of merit	0.0409
R Factor (Obs)	0.0487
R Factor (All)	0.0977
Weighted R Factor (Obs)	0.1008
Weighted R Factor (All)	0.1192

Citation: Rousay, Esther and Frey, Jeremy G. and Coles, Simon J. and Hursthouse, Michael B. (2004) University of Southampton, Crystal Structure Report Archive.
 (doi:10.3737/ecrystals.chem.soton.ac.uk/145)

Files

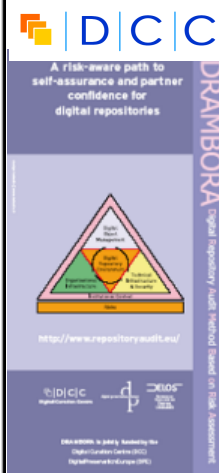
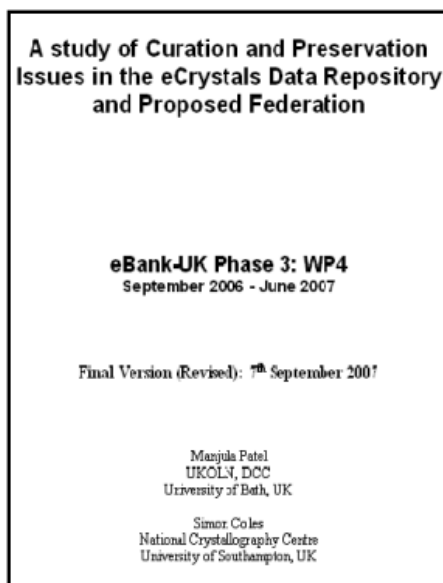
04sjc0831.res	6k
04sjc0831.xl.lst	34k
Solution	
04sjc0831.prp	6k
04sjc0831.xls.lst	39k
Processing	
04sjc0831.hkl	702k
04sjc0831.htm	10k
04sjc0831.0kl.jpg	57k
04sjc0831.h0l.jpg	85k
04sjc0831.hk0.jpg	88k
Data Collection	
04sjc0831.crystal.jpg	17k
Other Files	
04sjc0831.doc	78k
04sjc0831.ins	5k
04sjc0831.mol	3k
04sjc0831.p4p	1k
04sjc0831.pcf.txt	2k
04sjc0831.ellipsoid.gif	19k



A Thorough Approach to Dissemination

- Using simple Dublin Core protocol (OAI-PMH)
 - Crystal structure
 - Title (Systematic IUPAC Name)
 - Authors
 - Affiliation
 - Creation Date
- Additional **chemical** information through Qualified Dublin Core
 - Empirical formula
 - International Chemical Identifier (InChI)
 - Compound Class & Keywords
- Specifies which 'datasets' are present in an entry
- Application Profile <http://www.ukoln.ac.uk/projects/ebank-uk/schemas/>
- DOI links <http://dx.doi.org/10.1594/ecrystals.chem.soton.ac.uk/145>
- Rights & Citation <http://ecrystals.chem.soton.ac.uk/rights.html>

A Thorough Approach to Preservation



KEEPING RESEARCH DATA

SAFE

A COST MODEL AND GUIDANCE FOR UK UNIVERSITIES

Neil Beagrie, Julia Chruszcz, and Brian Lavoie

with case studies contributed by the Universities of Cambridge, Southampton, King's

College London, and the Archaeology Data Service University of York.

Final Report - April 2008

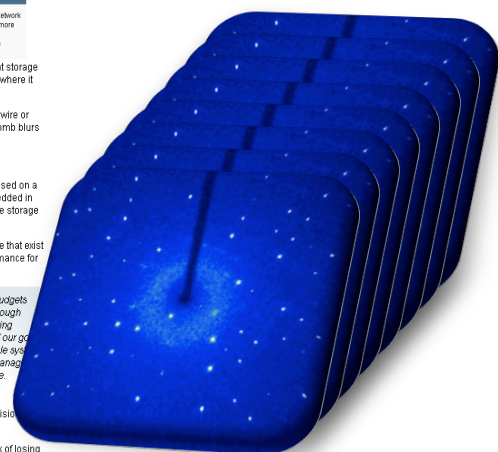
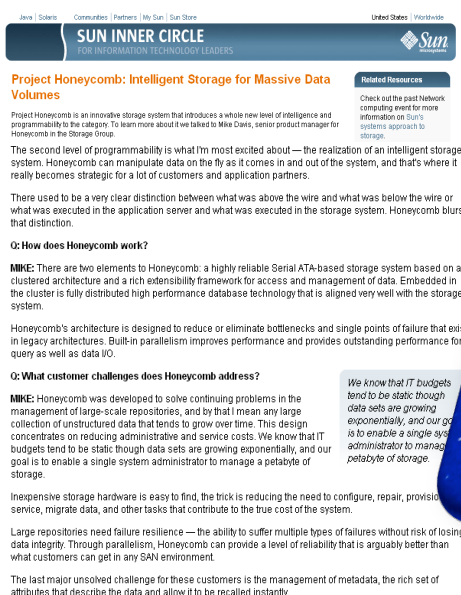
Prepared by:

Charles Beagrie Limited

www.beagrie.com

A study funded by

JISC



Scaling Up: A Community Solution

Interviews & analysis

Synthesis: IR Policy & Practice, Laboratory Practice & Workflows, Technical Interoperability & Standards, Metadata Schema & Application Profiles, Semantic Interoperability, Data Citation, Identifiers & Linking, Federation Architectures & Third Party Services, Rights & Licensing, Data Quality & Validation, Preservation, Curation & Sustainability

Recommendations, commentary

Matters Arising: Diverse lab practice, LIMS and proprietary formats, Data policy should reflect lab practice & institutional model, Data quality criteria/validation, “Prior publication” problem, We need scalable assignment of “terms” for data discovery, No discipline preservation model



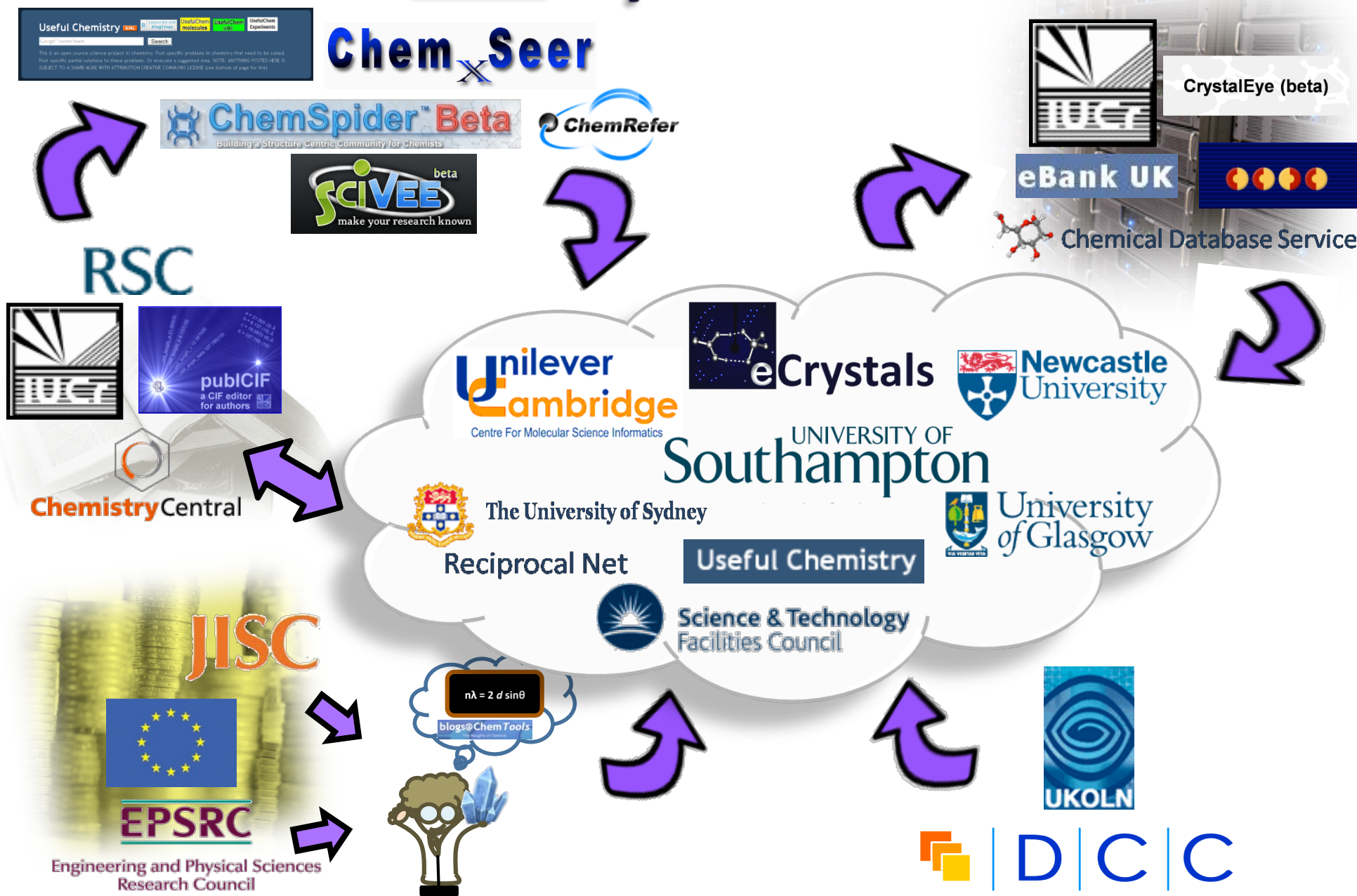
Scaling Up: Towards a Federation of Crystallography Data Repositories

Document details

Author:	Liz Lyon, Simon Coles, Monica Duke, Traugott Koch
Date:	12th May 2008
Version:	1.0 Final
Document Name:	ebank-phase3-report-final.doc
Notes:	

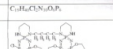
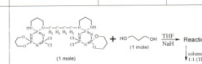
<http://wiki.ecrystals.chem.soton.ac.uk>

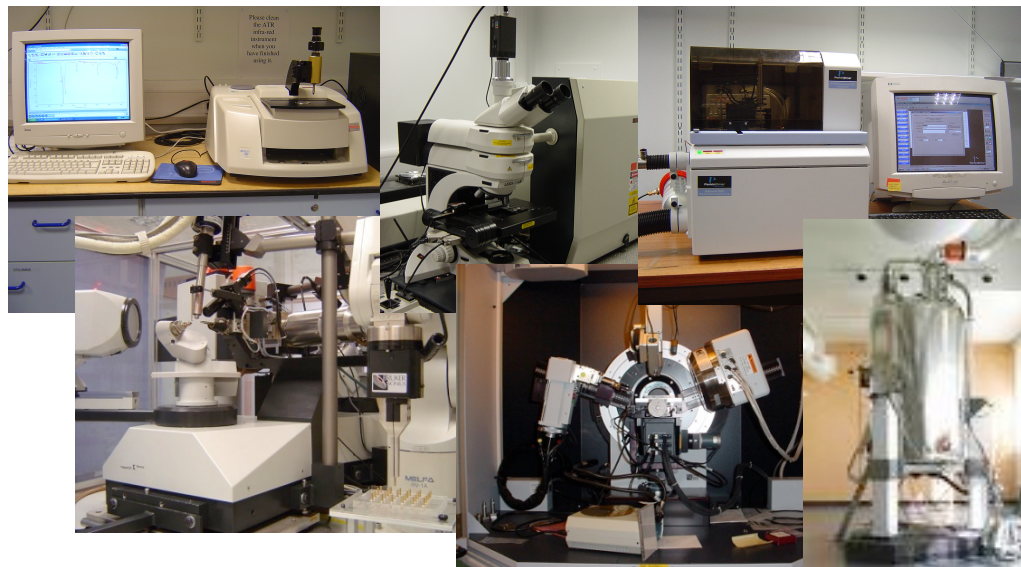
The eCrystals Federation



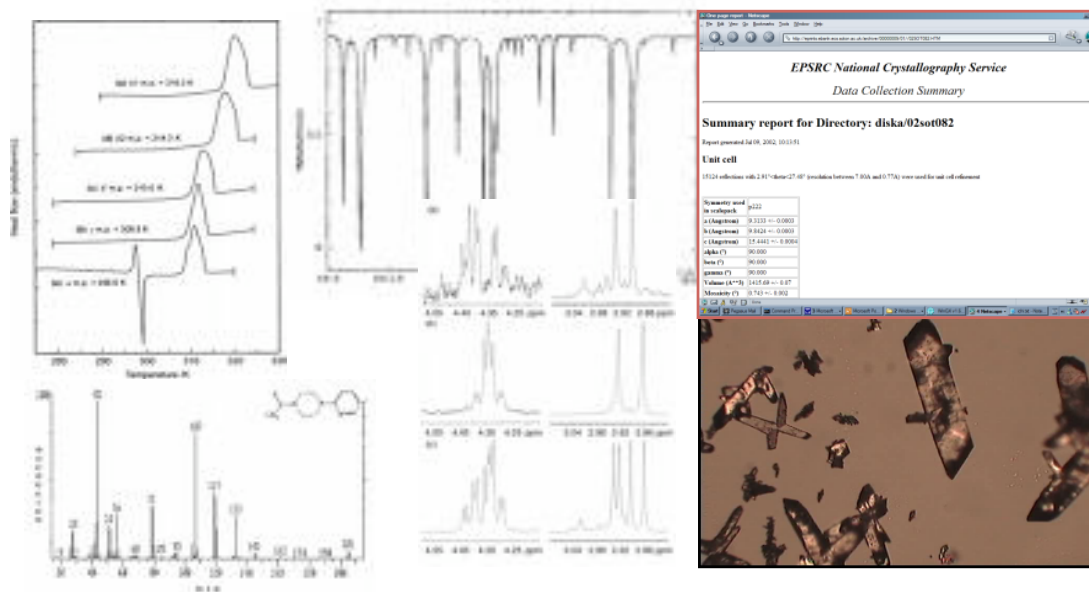
General Chemistry Issues: Data Generation

Characterisation

Compound Details		For X-ray								
Investigator:	Goual YENILMEZ CIP/CI									
Code of Compound:	303-489a									
Empirical Formula:	$C_{12}H_{16}ClN_2O_4$									
Proposed Structural Formula:										
Melting Point (°C):	303-489a (mp 132-133°C)									
Preparation (and including all solvents):	1,3-propanediol, tetrahydrofuran, NaH, dichloromethane, Hexane BLO									
General Properties (e.g., light, thermal stability, solubility, etc.):	This product is soluble in CH_2Cl_2 , $CHCl_3$									
Depository:	OCLC, Hexane BLO (3-1-12)									
Crystallisation Solvent:	Hexane									
Appearance of Compound:	Red, colorless									
Mass Calc. (g/mol):	Cal: 303.489-a (760)									
Elemental Analysis (Calcd):	<table><tr><td>C, %</td><td>H, %</td><td>N, %</td><td>O, %</td></tr><tr><td>76.1</td><td>7.8</td><td>7.8</td><td>76.3</td></tr></table>		C, %	H, %	N, %	O, %	76.1	7.8	7.8	76.3
C, %	H, %	N, %	O, %							
76.1	7.8	7.8	76.3							
Reaction Scheme:										
NMR Data:	<table><tr><td>1H</td><td>^{13}C</td><td>^{15}N</td></tr><tr><td></td><td></td><td></td></tr></table>		1H	^{13}C	^{15}N					
1H	^{13}C	^{15}N								



Synthesis



Shortfalls in Data Management

“Data from experiments conducted as recently as six months ago might be suddenly deemed important, but those researchers may never find those numbers – or if they did might not know what those numbers meant”

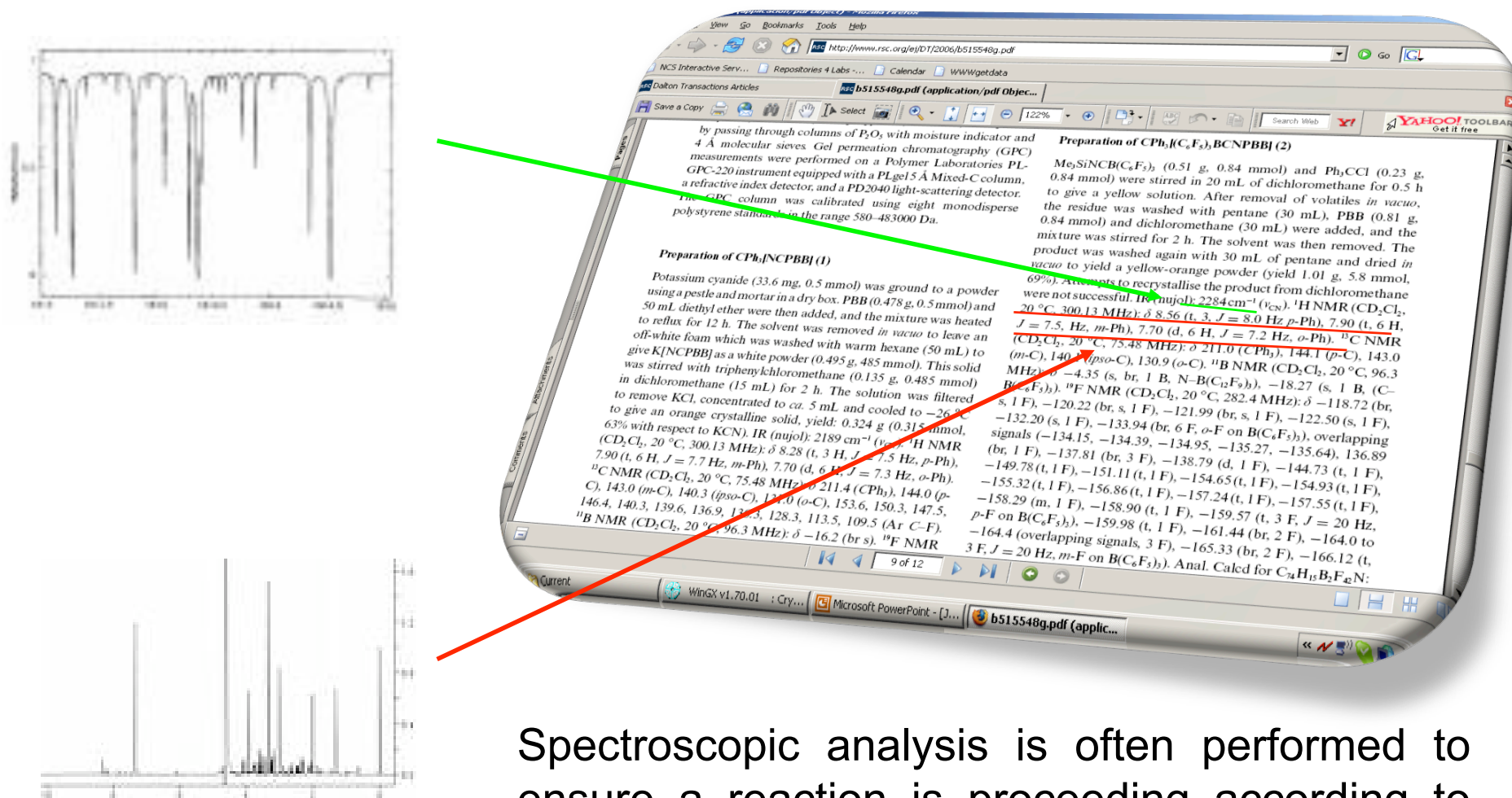
“Lost in some research assistant’s computer, the data are often irretrievable or an undecipherable string of digits”

“To vet experiments, correct errors, or find new breakthroughs, scientists desperately need better ways to store and retrieve research data”

“Data from Big Science is ... easier to handle, understand and archive. Small Science is horribly heterogeneous and far more vast. In time Small Science will generate 2-3 times more data than Big Science.”

‘Lost in a Sea of Science Data’ S.Carlson, The Chronicle of Higher Education (23/06/2006)

Analysis Data Publication & Information Loss



Spectroscopic analysis is often performed to ensure a reaction is proceeding according to plan – as a result <5% are published (via a process with heavy information loss)

the **S**mart **T**ea **P**roject



UNIVERSITY OF EAST LANCIA

COSHH ASSESSMENT FOR USE OF HAZARDOUS SUBSTANCES IN LABORATORIES

A written assessment of the risks must be prepared for all categories of work with hazardous chemical substances which are defined in the USA Code of Practice for Written Assessments under the COSHH Regulations.

1. **SCHOOL, School of Chemical Sciences.** **ASSESSMENT No.** Do not fill this box

2. **TITLE OF EXPERIMENT/PROCEDURE.** $2-(\text{H}_3\text{C})_2\text{CH}(\text{CH}_2)_2\text{CH}_2\text{CH}_2(\text{C}_6\text{H}_5)$

3. **NAME OF RESEARCH SUPERVISOR.** Dr. J. Z. JAMES

4. **NAME(S) OF ALL PARTICIPANTS (graduate students, technicians etc.).** P. J. JAMES

5. **LOCATION OF EXPERIMENTAL WORK.**

6. **HAZARDOUS SUBSTANCES TO BE USED.**

Name of substance	Principal hazard	Quantity	GHS #	Risk
1. $2-(\text{H}_3\text{C})_2\text{CH}(\text{CH}_2)_2\text{CH}_2\text{CH}_2(\text{C}_6\text{H}_5)$	Flammable	40 ml	H228	
2. H_2SO_4	Corrosive	100 ml	H314	
3. H_2O				
4. NaOH				
5. CH_3COOH				
6. CH_3I				
7. CH_3Br				
8. CH_3Cl				
9. CH_3F				
10. CH_3OH				
11. CH_3CN				
12. CH_3NO_2				
13. CH_3I				
14. CH_3Br				
15. CH_3Cl				
16. CH_3F				
17. CH_3OH				
18. CH_3CN				
19. CH_3NO_2				
20. CH_3I				
21. CH_3Br				
22. CH_3Cl				
23. CH_3F				
24. CH_3OH				
25. CH_3CN				
26. CH_3NO_2				
27. CH_3I				
28. CH_3Br				
29. CH_3Cl				
30. CH_3F				
31. CH_3OH				
32. CH_3CN				
33. CH_3NO_2				
34. CH_3I				
35. CH_3Br				
36. CH_3Cl				
37. CH_3F				
38. CH_3OH				
39. CH_3CN				
40. CH_3NO_2				
41. CH_3I				
42. CH_3Br				
43. CH_3Cl				
44. CH_3F				
45. CH_3OH				
46. CH_3CN				
47. CH_3NO_2				
48. CH_3I				
49. CH_3Br				
50. CH_3Cl				
51. CH_3F				
52. CH_3OH				
53. CH_3CN				
54. CH_3NO_2				
55. CH_3I				
56. CH_3Br				
57. CH_3Cl				
58. CH_3F				
59. CH_3OH				
60. CH_3CN				
61. CH_3NO_2				
62. CH_3I				
63. CH_3Br				
64. CH_3Cl				
65. CH_3F				
66. CH_3OH				
67. CH_3CN				
68. CH_3NO_2				
69. CH_3I				
70. CH_3Br				



Weigh-Station #1 Sep 19, 2003 2:51:28 PM		
mrg	mrg/3828/4	
<h2 style="margin: 0;">Experiment Details</h2>		
Name	Planned	Actual
Europium acetate	0.2500 g	0.2518 g
4chloro dipicolinic acid monoh...	0.5000 g	0.4950 g
Triethylamine	0.6 ml	0.6 ml
Methanol	100.0 ml	110.0 ml
Acetone	30.0 ml	0.0 ml

<table border="1" style="margin: auto; width: 80%;"> <tr> <td style="width: 33%; text-align: center; font-size: 1.5em;">7</td> <td style="width: 33%; text-align: center; font-size: 1.5em;">8</td> <td style="width: 33%; text-align: center; font-size: 1.5em;">9</td> </tr> <tr> <td style="text-align: center; font-size: 1.5em;">4</td> <td style="text-align: center; font-size: 1.5em;">5</td> <td style="text-align: center; font-size: 1.5em;">6</td> </tr> <tr> <td style="text-align: center; font-size: 1.5em;">1</td> <td style="text-align: center; font-size: 1.5em;">2</td> <td style="text-align: center; font-size: 1.5em;">3</td> </tr> <tr> <td style="text-align: center; font-size: 1.5em;">0</td> <td style="text-align: center; font-size: 1.5em;">.</td> <td></td> </tr> </table>	7	8	9	4	5	6	1	2	3	0	.		<table border="1" style="margin: auto; width: 90%;"> <tr> <td style="width: 50%; font-size: 1.5em; padding: 10px;">Enter</td> <td style="width: 50%; font-size: 1.5em; padding: 10px;">Del</td> </tr> </table>	Enter	Del
7	8	9													
4	5	6													
1	2	3													
0	.														
Enter	Del														

All measurements completed.

123 - 1111 - Water	222 - 2222 - Tea	333 - 3333 - Sugar	444 - 4444 - Coffee
555 - 1233 - Miso Bro			

Quit
Weigh
Liquid-Measure
Bench
Store
Escape

Bench-Station #1

22-Sep-2003 16:01:39

dj
djbj3403

Experiment Details

Stage	Instructions	Done
1	Dissolve 4-fluorinated biphenyl in butanone	<input checked="" type="checkbox"/>
2	Add K2CO3 powder	<input checked="" type="checkbox"/>
3	Heat at reflux for 1.5 hours	<input type="checkbox"/>
4	Cool and add BrI10CB	<input type="checkbox"/>
5	Heat at reflux until completion	<input type="checkbox"/>
6	Cool and add water (30ml)	<input type="checkbox"/>
7	Extract with DCM (3x40ml)	<input type="checkbox"/>
8	Combine organics, dry over MgSO4 & filter	<input type="checkbox"/>
9	Remove solvent in vacuo	<input type="checkbox"/>

2 of 10 tasks completed.

Run
Weigh
Liquid-Measure
Bench
Store

A General Chemistry Laboratory Repository

Create new compound
(parent record)

The screenshot shows the 'Create new compound' form in the R4L interface. It includes fields for 'Name' (with examples like 2-(methylsulfonyl)benzoic acid and aspirin), 'Identifier' (with example TLT/1142), and a 'Sketch' section with a MarvinSketch applet showing a chemical structure of a substituted benzene ring. There is also a 'Collaborators' section with a table for entering names and email addresses.

Family Name	Given Name / Initials	Email address
1. Coles	Simon J	sjc5@ex.ac.uk
2. Hursthouse	Michael B	

Add new experiment type

The screenshot shows the 'Add new experiment type' form. It features a list of experiment types including Simple Crystal Diffraction, Powder X-Ray Diffraction, IR, UV-Vis, Mass Spec, Raman, Optical Microscopy, DSC, TGA, NMR, Solid State NMR, SHG Laser Spectroscopy, and Elemental Analysis. A dropdown menu is set to 'Add Experiment'.

Add metadata and
upload data files

The screenshot shows the 'Add metadata' form, which includes sections for 'Reference Material', 'Pan Bottom', 'Pan Lid', and 'Purge Gas'. Each section has a text area for describing the material and a dropdown menu for selecting a specific material or gas.

The screenshot shows the 'Add metadata' form, which includes sections for 'Experiment Metadata', 'Compound', 'Date', 'Time', and 'Location'. Each section has a text area for entering the relevant information.

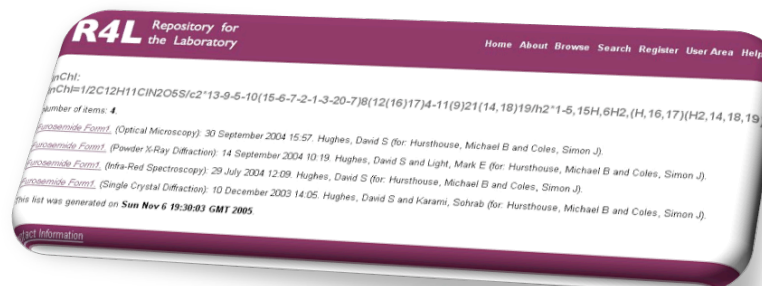
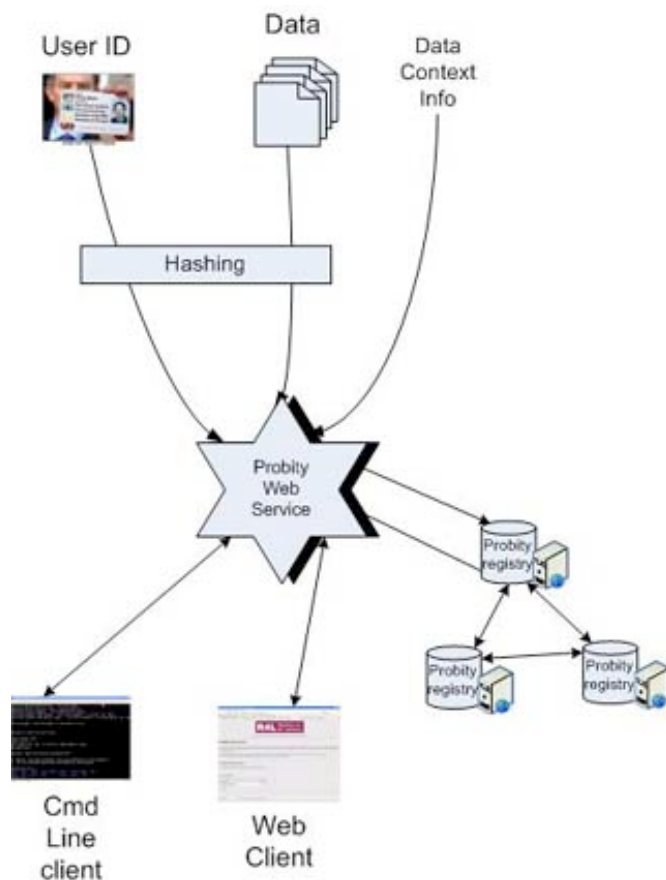
A General Chemistry Laboratory Repository

- Probity: A process to assert originality of a data record

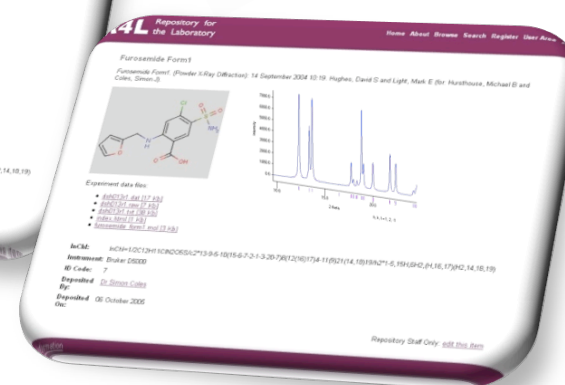
Register a new claim

File for user ID Hash	<input type="text"/>	Browse...
File data Hash	<input type="text"/>	Browse...
Data Context Information	<input type="text"/>	

Register



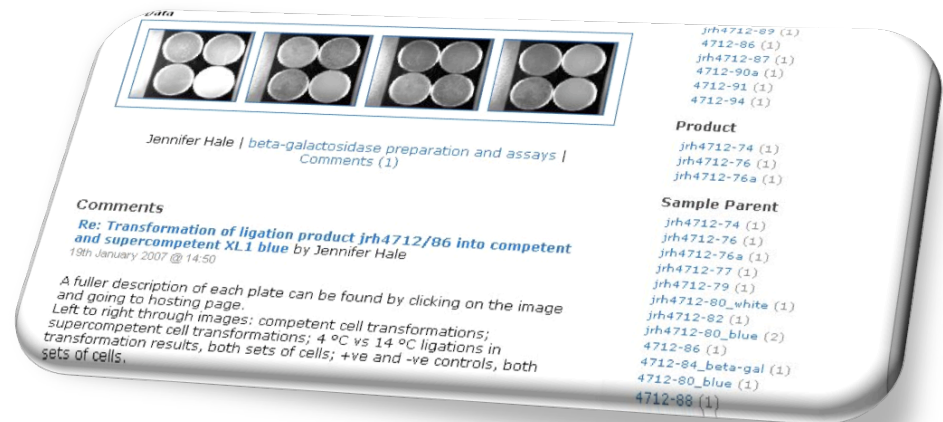
Search / Browse



Analysis & Discussion: Blogging Experiments

A repository can...

- Allow one to put, store and get
- Provide search and browse functionality
- **NOT** provide the presentation and discussion functions essential to working up a scientific study



Investigations into neutral drift

Test PCR (different primer, dNTP and enzyme stocks)
 18th May 2008 @ 14:29

Post Type: PCR
 Risk assessment: Mutagenesis/PCR risk assessment

Reaction	DNA template	µL Water	µL Buffer	µL Primer 1	µL Primer 2	µL dNTPS	µL Miscellaneous	µL Enzyme	µL	Product
1	Purified plasmid from the X-glu positive colonies in 5025/46 (experiment round 4)	1 Sterile filtered 13.05.08	8.8 GoTaq Buffer (04/08)	4 Beta-glu fwd	1 Beta-glu rev	1 Normal made 04/08	2 MgCl ₂	1.2 GoTaq (04/08)	1 (diluted 1in first)	4 Test (different primer, dNTP and enzyme stocks) product 1
2	Purified plasmid from the X-glu positive colonies in 5025/46 (experiment round 4)	1 Sterile filtered 13.05.08	8.8 GoTaq Buffer (04/08)	4 Beta-glu fwd 04/08	1 Beta-glu rev 04/08	1 Normal made 04/08	2 MgCl ₂	1.2 GoTaq (04/08)	1 (diluted 1in first)	4 Test (different primer, dNTP and enzyme stocks) product 2
3	Purified plasmid from the X-glu positive colonies in 5025/46 (experiment round 4)	1 Sterile filtered 13.05.08	8.8 GoTaq Buffer (04/08)	4 β-glu fwd (13.05.08)	1 β-glu rev (13.05.08)	1 Normal made 04/08	2 MgCl ₂	1.2 GoTaq (04/08)	1 (diluted 1in first)	4 Test (different primer, dNTP and enzyme stocks) product 3
4	None	1 Sterile filtered 13.05.08	9.8 GoTaq Buffer (04/08)	4 β-glu fwd (13.05.08)	1 β-glu rev (13.05.08)	1 Normal made 04/08	2 MgCl ₂	1.2 GoTaq (04/08)	1 (diluted 1in first)	4 Test (different primer, dNTP and enzyme stocks) product 4
5	Purified plasmid from the X-glu positive colonies in 5025/46 (experiment round 4)	1 Sterile filtered 13.05.08	8.8 GoTaq Buffer (04/08)	4 Beta-glu fwd	1 Beta-glu fwd	1 Normal made 13.05.08	2 MgCl ₂	1.2 GoTaq (04/08)	1 (diluted 1in first)	4 Test (different primer, dNTP and enzyme stocks)

Search

Archives
 May 2008 (58)
 April 2008 (77)
 March 2008 (72)
 February 2008 (113)
 January 2008 (67)
 December 2007 (61)
 November 2007 (181)
 October 2007 (323)
 September 2007 (181)
 August 2007 (67)
 July 2007 (89)
 June 2007 (123)
 May 2007 (128)
 April 2007 (98)
 March 2007 (116)

Sections
 Materials (71)
 Notes (31)
 Procedure (338)
 Product (1274)
 Safety (11)
 Templates (29)

Post Type
 Digestion (31)
 Enzyme (15)
 Cell_strain (28)
 Buffer (18)
 Primer (11)
 Thermocycler_program
 Template (27)

Facilitating Research

- Enables 'geographically distributed collaborative research'
- Can be open or private
- A useful approach for sharing 'failed' experiments?

PCR of beta-galactosidase third attempt by Jennifer Hale
14th December 2006 @ 11:10

Unfortunately the purification appears not to have gone well. Though I also can't get any consistency from the figures given by the nano-drop. These are the results I got:

	reading 1	reading 2	reading 3	reading 4	reading 5	reading 6	average
PCR product before*	282.3 ng/μL	283.4 ng/μL	281.1 ng/μL	N/A	N/A	N/A	282.3 ng/μL
PCR product after*	7.8 ng/μL	12.9 ng/μL	17.6 ng/μL	85.4 ng/μL	22.4 ng/μL	12.8 ng/μL	?

*Both reactions combined together after PCR

I'm going to do another PCR again. That step is working really well. I'm just not sure what to do about purifying it. The only other thing I can try is eluting in TE buffer rather than water (which it says you can also elute into)

In this purification I used preheated water and followed the instructions closely. Perhaps the DNA will elute into TE more effectively.

Re: PCR of beta-galactosidase third attempt by David Neylon
14th December 2006 @ 18:32

I would definitely compare these on a gel so as to see whether it is just the nanodrop that is the problem. It might help also if you are explicit about how much solution you are trying to purify and what the final volume is.

4712-86 (1)
4712-84_beta-ga
4712-80_blue (1)
4712-88 (1)

Sample Parent2
jrh4712-80_blue (1)
jrh4712-80_white (2)
4712-84_pBad (1)
4712-80_white (1)

Sample Parent 3
4712-74 (1)


Search

Find

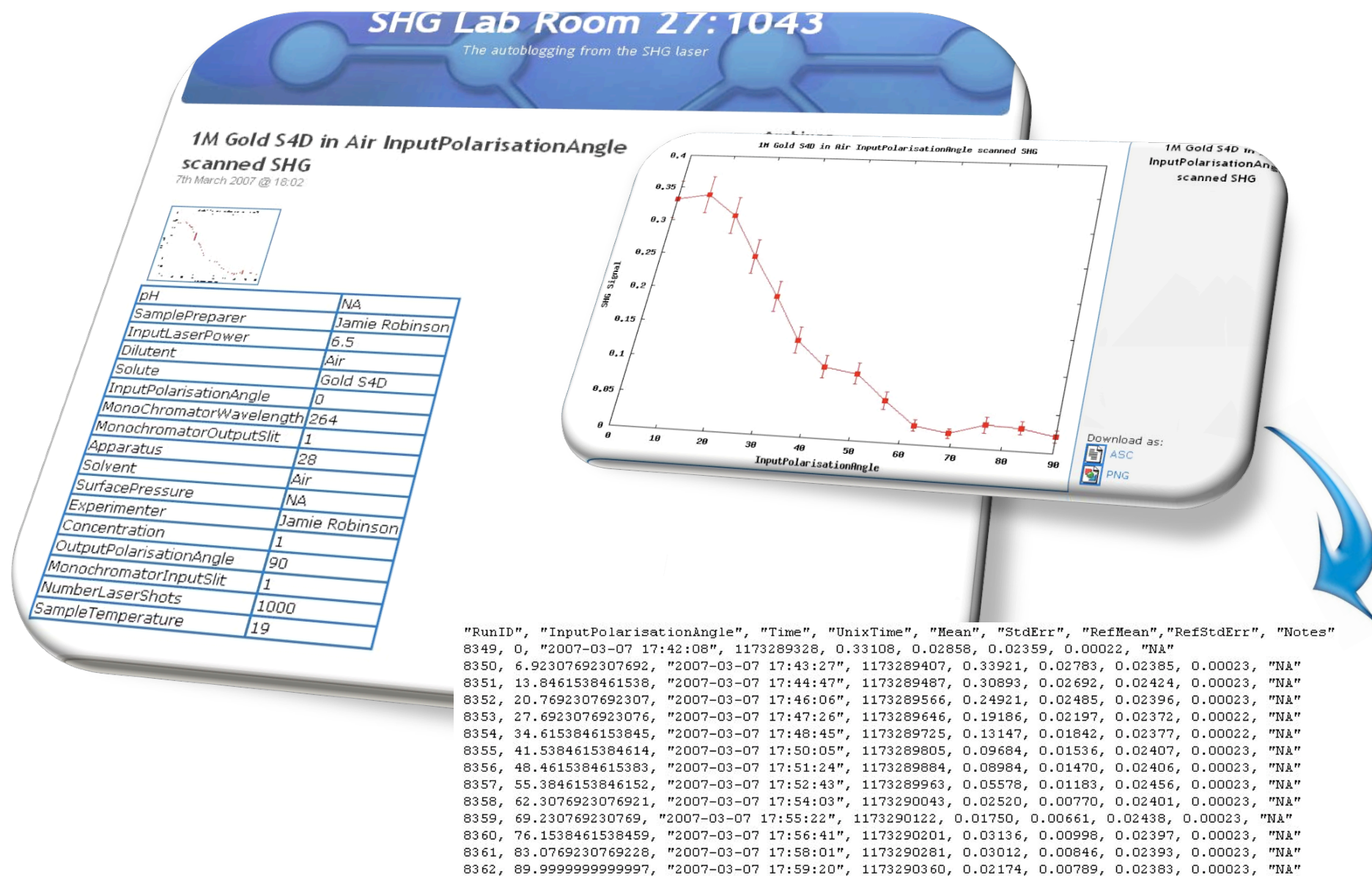
Links

Admin
New Post

Live Copy

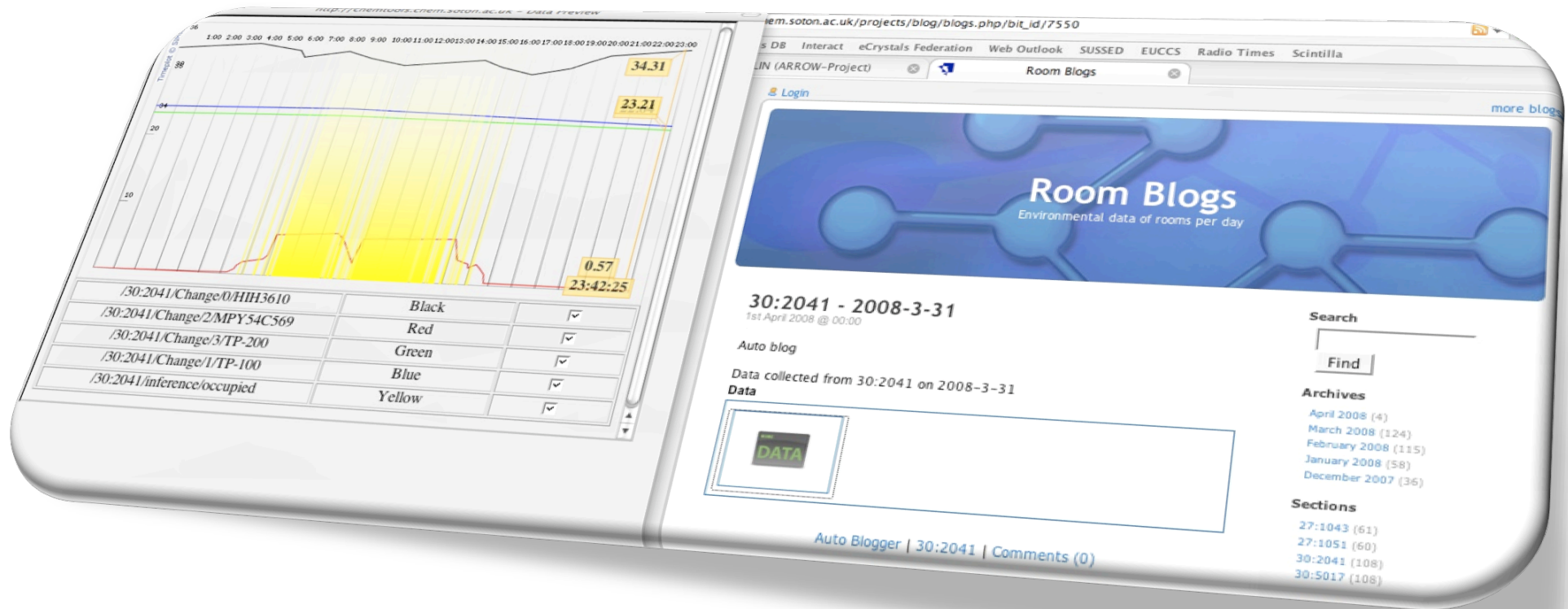


Automatic Blogging by Machines



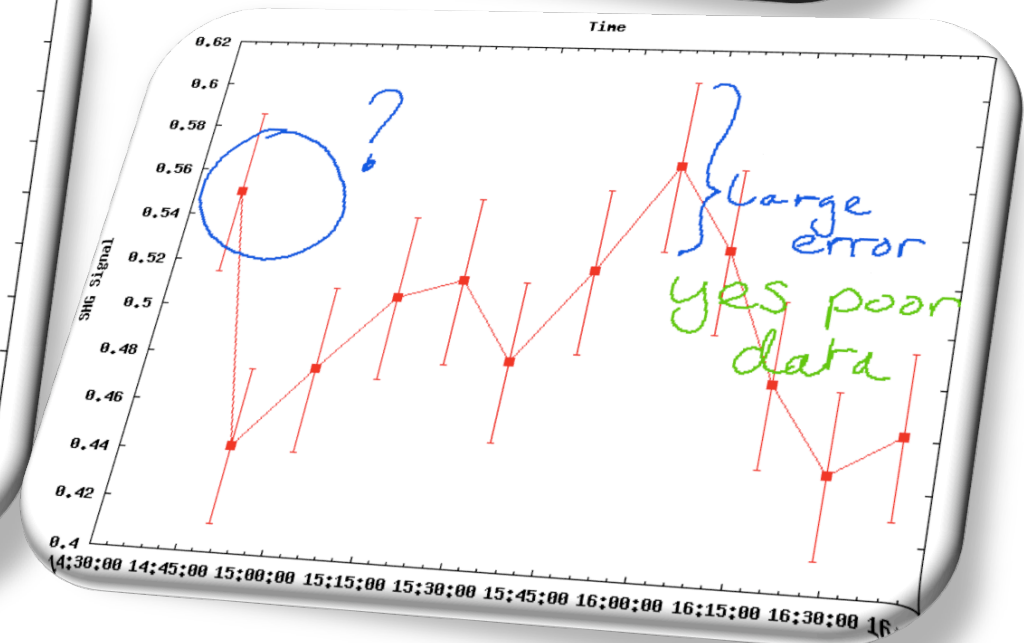
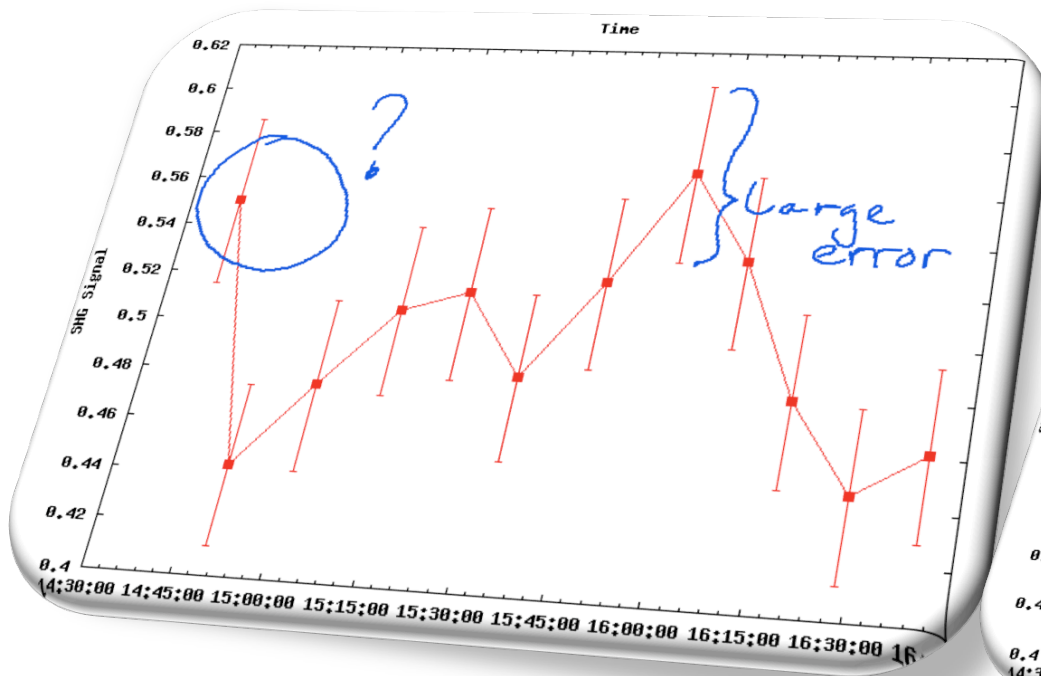
Automatic Blogging by Sensors

- Continuous log of 'environmental' conditions in a laboratory
- Instant detection of erroneous events
- Correlate with inconsistencies in datasets



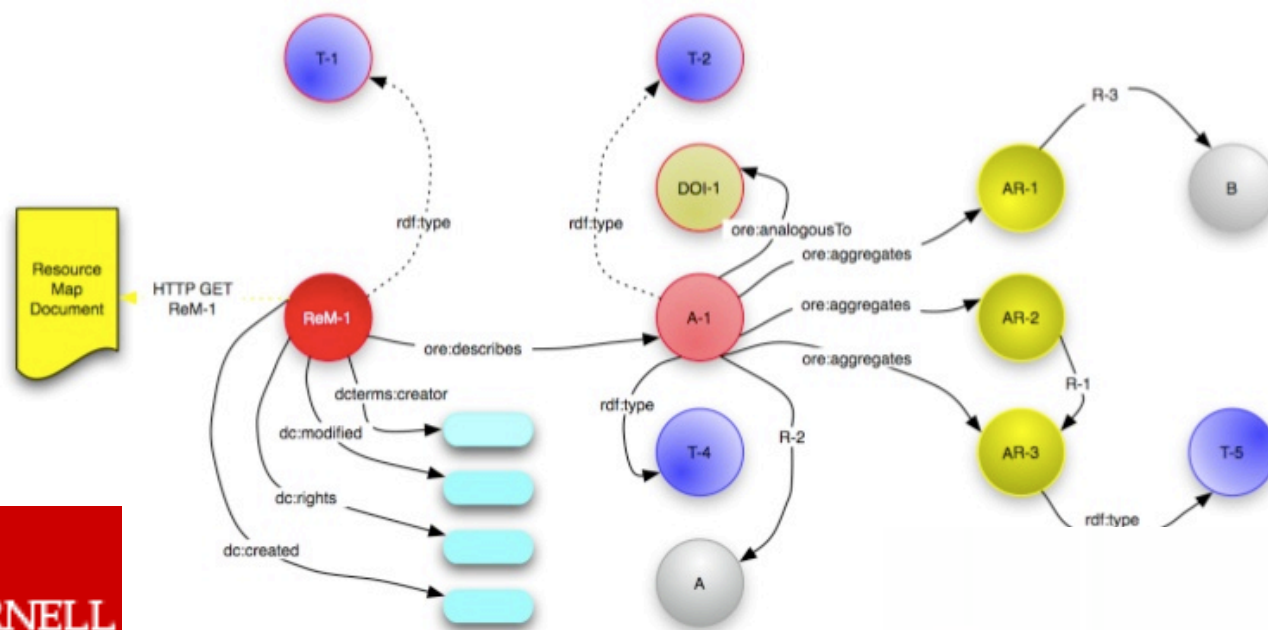
Comments and Collaborative Tools

- Annotation tools allow comments and foster collaboration and / or communication
- Need for more advanced Blog tools / technology around data

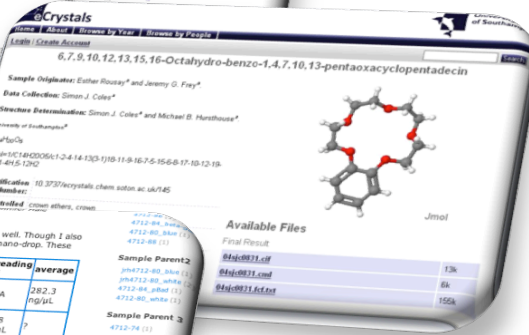
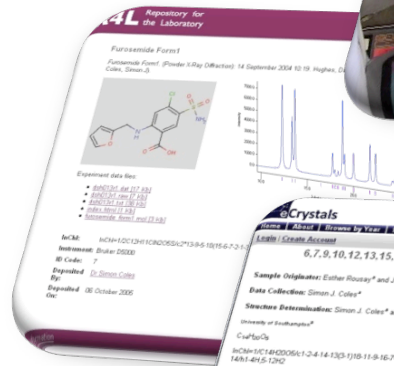
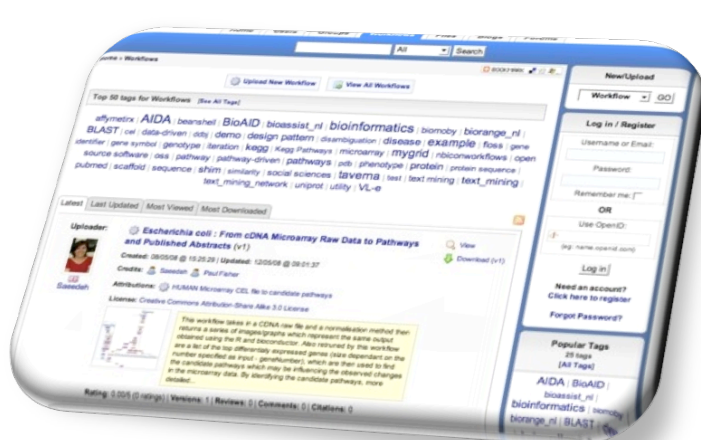


Packaging and Interoperability

- New moves in Digital Libraries community to enable distributed repositories to exchange content
- OAI-ORE (Open Archives Initiative – Object Reuse and Exchange)
- <http://www.openarchives.org/ore/>
- Describes an aggregation of objects in an exchangeable format
- Microsoft funded eChemistry testbed project



Towards a New Model for Chemical Information Exchange



A screenshot of a web page showing a table with PCR results and a text block. The table has columns for 'Reading', 'Reading', 'Reading', 'Reading', 'Reading', 'Reading', and 'Average'. The text block discusses the results of the PCR and mentions a 'PCR of beta-galactosidase third attempt'.

Reading	Reading	Reading	Reading	Reading	Reading	Average
282.3	283.4	281.1	N/A	N/A	282.3	282.3
7.8	12.9	17.6	85.4	22.4	12.8	?

Both reactions combined together after PCR

I'm going to do another PCR again. That step is working really well. I'm just not sure what to do about purifying it. The only other thing I can try is eluting in TE buffer rather than water (which it says you can also elute in).

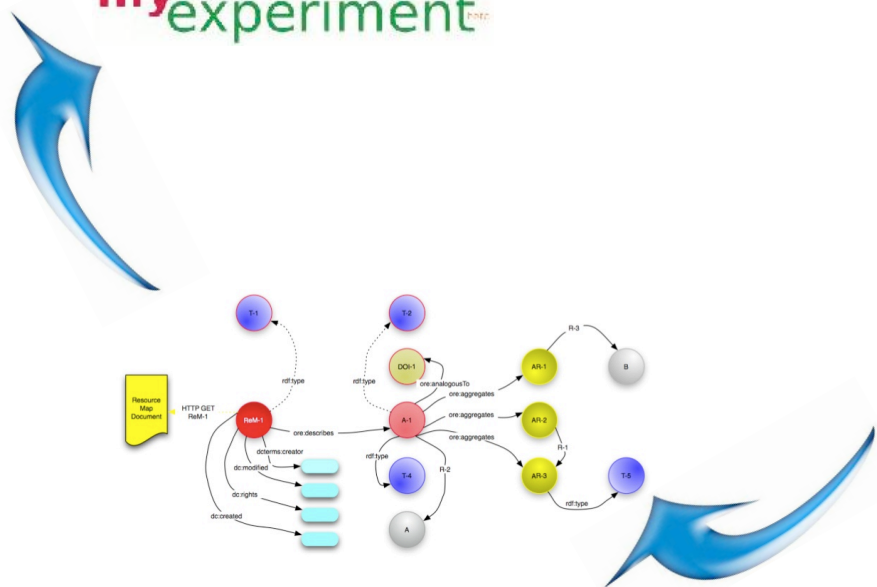
In this purification I used preheated water and followed the instructions closely. Perhaps the DNA will elute into TE more effectively.

Re: PCR of beta-galactosidase third attempt by David Neylon

14th December 2006 @ 19:32

I would definitely compare these on a gel so as to see whether it is just the phenotype that is the problem. It might help also if you are explicit about any much solution you are trying to purify and what the final volume is.

myexperiment



A solid foundation for Open/Self-Publishing of Chemistry Data???

Thanks to:

- **Jeremy Frey**, Andrew Milsted, Richard Stephenson, Cameron Neylon, Jamie Robinson, Steven Wilson



- **Dave DeRoure**, **Les Carr**, Chris Gutteridge, Tim Myles-Board, Arouna Woukei



- **Liz Lyon**, Rachel Heery, Monica Duke, Michael Day, Traugott Koch, Manjula Patel

