Predicting Reaction Outcomes and Designing Synthetic Routes

10th and 11th of September 2014

Weetwood Hall Conference Centre and Hotel in Leeds.

The meeting brought together people from a range of disciplines to generate ideas and collaborations to tackle the challenges associated with

- Predicting unknown reaction outcomes
- Designing synthetic routes
- Applications driving the development of new chemistry.

The challenges require a cross-disciplinary approach, with particular relevence to Computer Science, Mathematics, Engineering and many parts of Chemistry. Researchers from both industry and academia are most welcome to attend.

The programme will consist of plenary talks to present some of the best current approaches to the problems, short presentations by participants on their potential contributions and interests, alongside brain-storming sessions to suggest ways forward and develop collaborations.

Attendees:

Dr Ben Andrews (GSK), Dr Robin Attrill (GSK), Dr Mike Bodkin (Evotec), Mr Anthony Cook (University of Leeds), Prof Ian Fairlamb (University of York), Dr Natalie Fey (University of Bristol), Dr David Flanagan (Wiley), Prof Val Gillet (University of Sheffield), Dr Jonathan Goodman (University of Cambridge, Organising Committee), Ms Jessica Gould (Croda Europe Limited), Prof Ron Grigg (University of Leeds), Mr Chris Hone (University of Leeds), Dr Fernando Huerta (ChemNotia), Dr Mike Hutchings (InfoChem), Prof Peter Johnson (University of Leeds, Organising Committee), Dr Chris Jones (TSB), Dr Mikhail Kabeshov (University of Cambridge), Dr Kelly Kilpin (University of Southampton/DaM Coordinator), Dr Frank Langbein (Cardiff University), Prof Alexei Lapkin (University of Cambridge), Dr Andrew Leach (Liverpool John Moores University), Dr Mark Leach (meta-synthesis), Dr Stuart Little (Croda Europe Ltd), Dr Daniel Lowe (NextMove Software Limited), Dr Jason Lynam (University of York), Prof Andrei Malkov (Loughborough University), Mr James McManus (University of Leeds), Ms Nicole McSweeney (LHASA), Dr Bao Nguyen (University of Leeds), Dr Stephanie North (InfoChem), Mr Alan Pettman (Pfizer), Ms Natasha Richardson (EPSRC), Dr Roger Sayle (NextMove Software Limited), Dr John Slattery (University of York), Mrs Gill Smith (Gillian Smith Associates), Prof Nguyen TK Thanh (UCL), Mr James Wallace (Lilly/University of Sheffield), Prof Richard Whitby (University of Southampton, Organising Committee) Prof David Woods (University of Southampton)

Program Day 1
Wednesday 10th September:

10:00 – 10:30	*Registration with Tea/Coffee and Refreshments*
10:30 - 10:45	Welcome and Introduction (**Prof Richard Whitby** - *University of Southampton & Dial-a-Molecule PI*)
	SESSION ONE: Chair – **Prof Richard Whitby** (*University of Southampton*)
10:45 – 11:15	**Dr Natalie Fey** (*University of Bristol*) Computational Tools for the Discovery and Optimisation of Organometallic Catalysts
11:15 – 11:45	**Prof Ian Fairlamb** (*University of York*) Intelligent Synthesis
11:45 - 12:15	**Prof Alexei Lapkin** (*University of Cambridge*) Non-Linear Dimensionality Reduction for Automated Reaction Optimisation and Discovery
12:15 – 12:45	**Prof Andrei Malkov** (*Loughborough University*) Catalytic Asymmetric Crotylation: Method Development
12:45 – 13:45	*Conference Lunch*
	SESSION TWO: Chair – **Dr Jonathan Goodman** (*University of Cambridge*)
13:45 - 14:15	**Dr John Slattery** (*University of York*) Understanding Organometallic Catalysis using Experiment and Theory: Could a similar, energy-landscape approach contribute to catalytic reaction prediction?
14:15 – 14:45	**Prof Nguyen TK Thanh** (*UCL*) Can we Dial Nanoparticles?
14:45 – 15:15	**Nicole McSweeney** (*LHASA*) Mutagenic Impurity Risk Assessment Purge Tool
15:15 – 16:00	**Breakout Session 1:** Predicting Reaction Outcomes
16:00 – 16:30	*Afternoon Refreshments*
	SESSION THREE: Chair – **Mrs Gill Smith** (*Gillian Smith Associates)*
16:30 – 17:00	**Prof David Woods** (*University of Southampton*) Statistical Learning Through Designed Experiments
17:00 – 17:30	**Dr David Flanagan** (*Wiley*) Synthetic Route Design with ARChem and RxnFinder from Wiley Science Solutions
17:30 – 18:00	**Dr Mike Hutchings** (*InfoChem*) ICSynth as an Idea Generator in Synthesis Planning
19:00 – 21:00	*Conference Dinner*

Program Day 2
Thursday 11th September:

	SESSION FOUR: Chair – **Prof Ian Fairlamb** (*University of York*)
09:00 – 09:40	**Prof Peter Johnson** (*University of Leeds)* Automated Synthesis Planning – Advances over the past 40 years
09:40 – 10:10	**Tony Cook** (*University of Leeds*) A Treatment of Stereochemistry in Computer Aided Organic Synthesis
10:10 – 10:40	**Dr Fernando Huerta** (*ChemNotia*) ICSynth: Forward Reaction Prediction Applications
10:40 – 11:00	*Morning Refreshments*
	SESSION FIVE: Chair – **Prof Peter Johnson** (*University of Leeds*)
11:00 – 11:30	**Dr Mike Bodkin** (*Evotec*) From Vectors to Sequence to Networks in Molecular Design
11:30 – 12:30	**Breakout Session 2**: Designing Synthetic Routes
12:30 – 13:30	*Conference Lunch*
	SESSION SIX: Chair – **Dr Andrew Leach** (*Liverpool John Moores University*)
13:30 – 14:00	**Dr Jonathan Goodman** (*University of Cambridge*) Do We Know Enough Chemistry?
14:00 – 14:30	**Dr Roger Sayle** (*NextMove Software Limited*) "Big Data" Reaction Yield Analysis from Pharmaceutical ELN's and Text Mining of Patent Applications
14:30 – 15:00	**Dr Mark Leach** (*meta-synthesis*) The Chemical Thesaurus: A Reaction Chemistry Database
15:00 – 15:20	*Afternoon Refreshments*
15:20 – 16:15	**Breakout Session 3**: Design a Machine to Make Many Different Compounds
16:15 – 16:30	Conclusions and Formal Closing (**Dr Jonathan Goodman** – *University of Cambridge*)

Session One: Predicting Reaction Outcomes:

a) For what reactions would you trust the answers?
Amide/ester formation
Suzuki
Click reaction (Azide-Alkyne Huisgen Cycloaddition)
Hydrogenation

Reactions which have been done oneself / simple / lots of data
Substrates which are free from troublemakers
Can we extrapolate beyond know data?
Can we predict behaviour of novel substrates?

b) How can this be expanded?
Need Reliability? Predictable? High yield?
Need complete mechanistic understanding - needs unlimited resource
Mine large reaction databases
Need reactions with broad substrate scope and good functional group tolerance
Broad and robust substrate scope? Remain selective over range of conditions and scale up?
Lack of data - people (particularly in the USA) do not report poorer results
Tackle this through publication process
- need criteria before a reaction can be a named reaction or a reliable reaction
Building up a more detailed picture of a reaction - understand it better - reliable?
Need a prize for good reactions conditional on reporting good and bad data on scope

c) Which specific reactions are most in need of analysis?
Reactions to scale up?
Atom economical?
Sustainable reactions?
Most used reactions need to be made more predictable
Non-precious metal catalysis (enzymes?)
Reduction of amides to amines using hydrogen

d) Is any software available now? What would a useful tool do?
ARChem and *ICSynth* might be helpful?

One sentence note on each lecture:

Natalie Fey: maps of chemical space work well for phosphorus donor ligands and can be used for both interpretation and prediction of suitable data

Ian Fairlamb: detail of the Stille reaction is very complicated although work from this group has lead to some understanding of the mechanism

Alexei Lapkin: however well the components are understood, complex processes are unpredictable

Andrei Malkov: catalytic asymmetric crotylation reactions can be understood in great detail

John Slattery: combined experimental and computational approches explain complex processes

Nguyen T K Thanh: It is still a huge challenge to predict the nanoparticle formation

Nicole McSweeney: automated risk assessment for reactions is possible

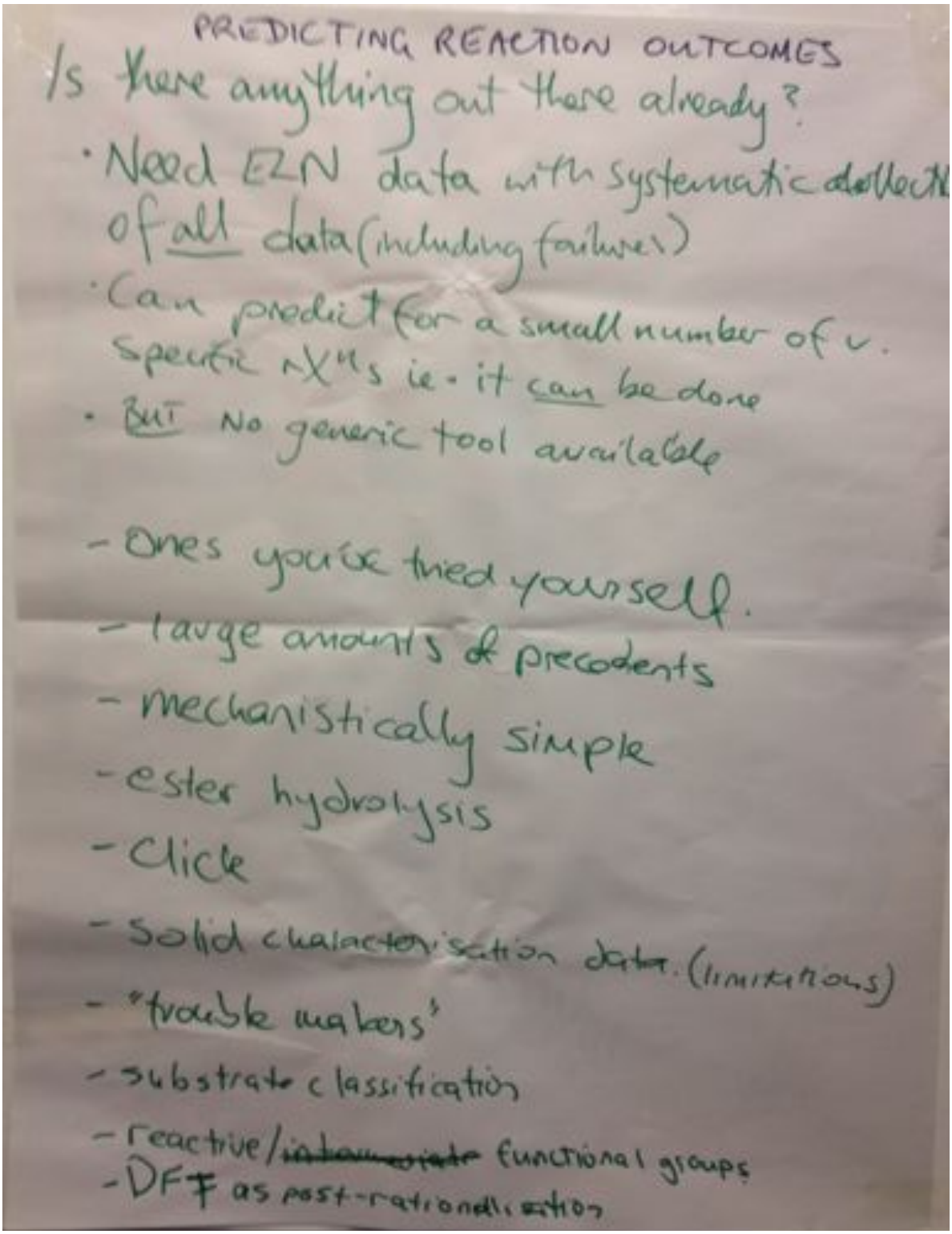
Some References:

Risk Assessment of Genotoxic Impurities in New Chemical Entities: Strategies To Demonstrate Control
Andrew Teasdale, David Elder, Sou-Jen Chang, Sophie Wang, Richard Thompson, Nancy Benz, and Ignacio H. Sanchez Flores
Org. Process Res. Dev., 2013, 17 (2), pp 221–230
DOI: 10.1021/op300268u

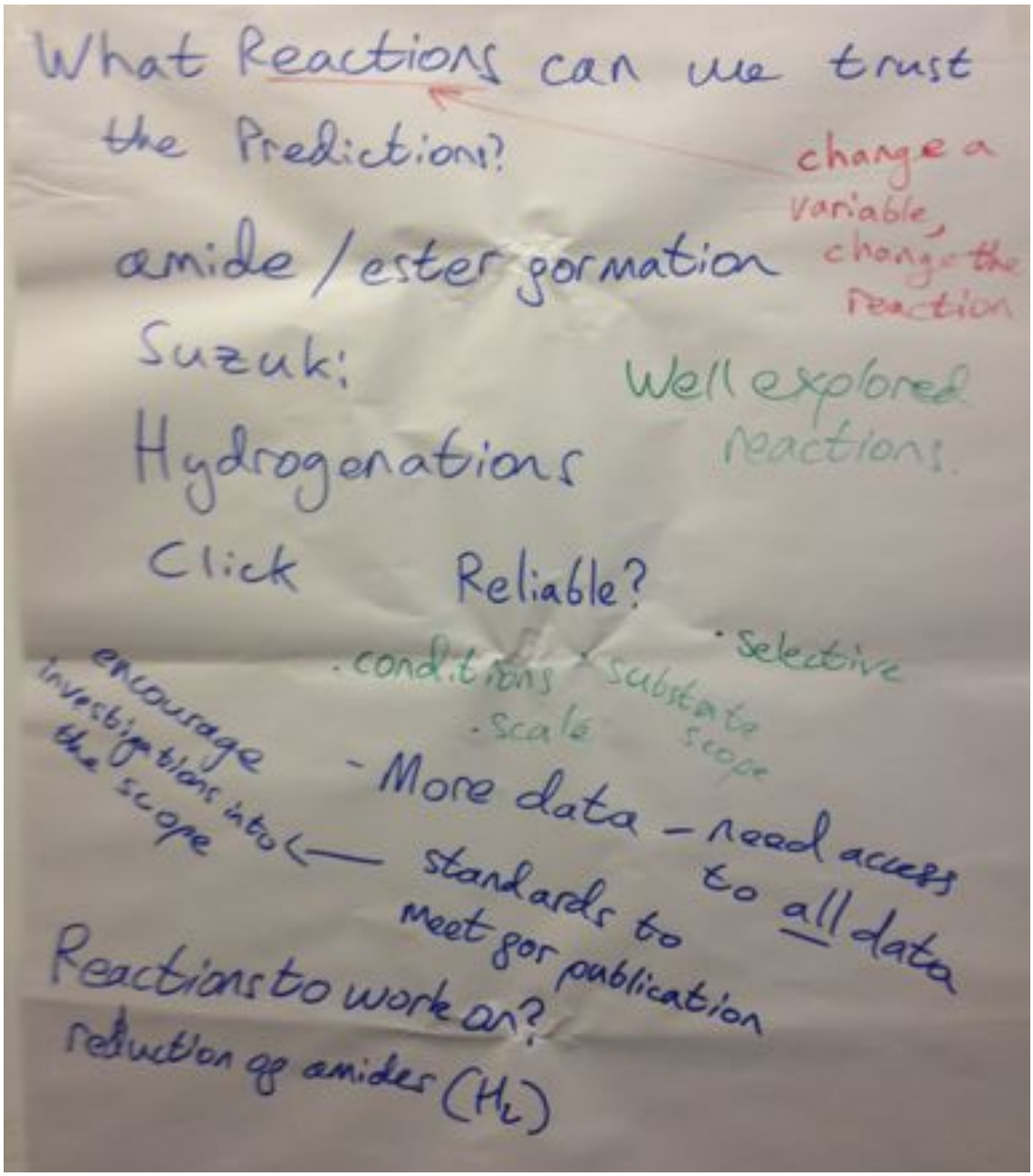
Systematic Exploration of the Mechanism of Chemical Reactions: The Global Reaction Route Mapping (GRRM) Strategy using the ADDF and AFIR Methods
Satoshi Maeda, Koichi Ohno, Keiji Morokuma
Phys. Chem. Chem. Phys., 2013, 15, pp 3683-3701
DOI: 10.1039/C3CP44063J

Chemical Space as a Source for New Drugs
Jean-Louis Reymond, Ruud van Deursen, Lorenz C Blum and Lars Ruddigkeit
MedChemComm., 2010, 1, pp 30-38
DOI: 10.1039/C0MD00020E

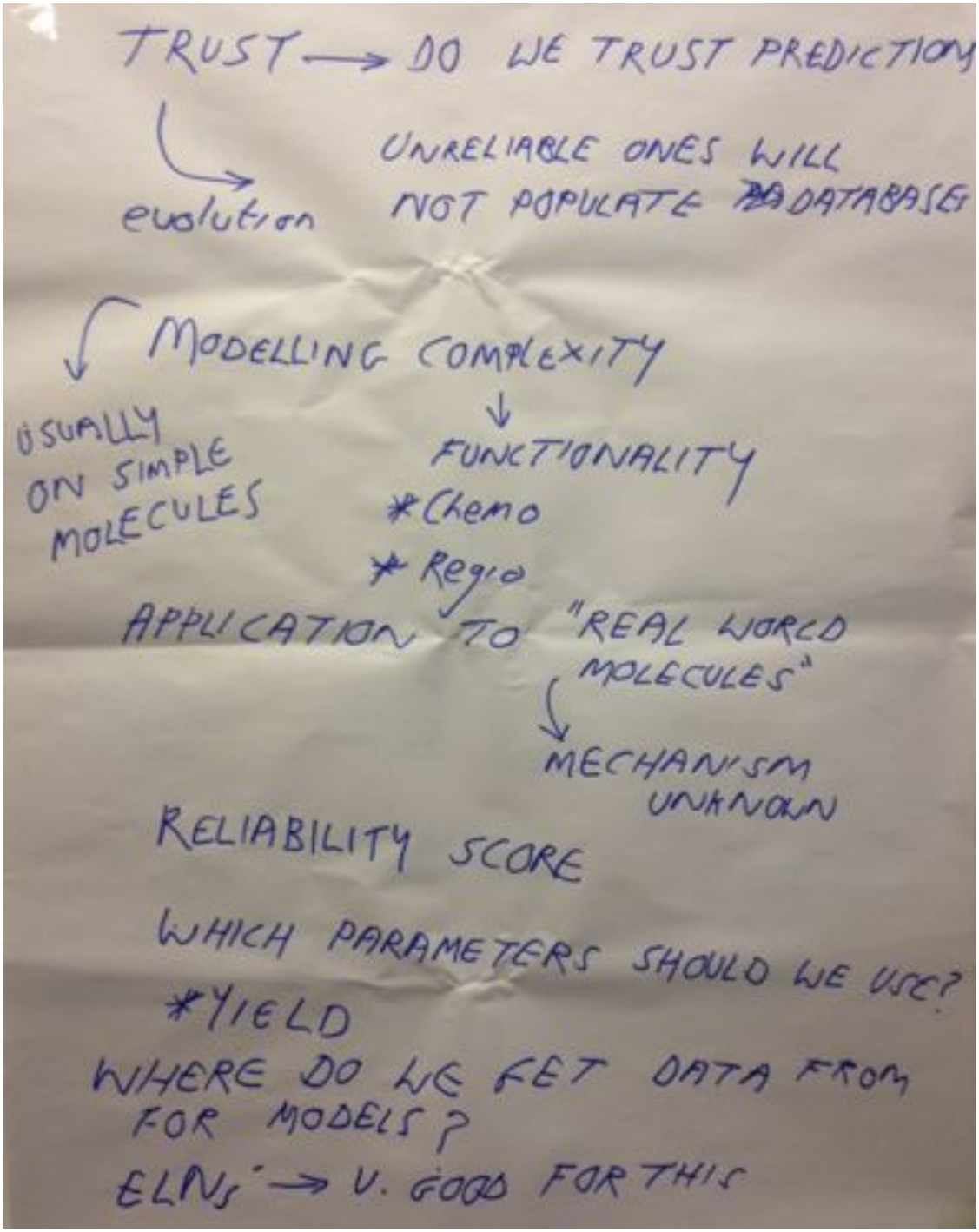
Stochastic Voyages into Uncharted Chemical Space Produce a Representative Library of All Possible Drug-Like Compounds
Aaron M Virshup, Julia Contreras-García, Peter Wipf, Weitao Yang, David N Beratan
JACS, 2013, 135(19), pp 7296-7303
DOI: 10.1021/ja401184g

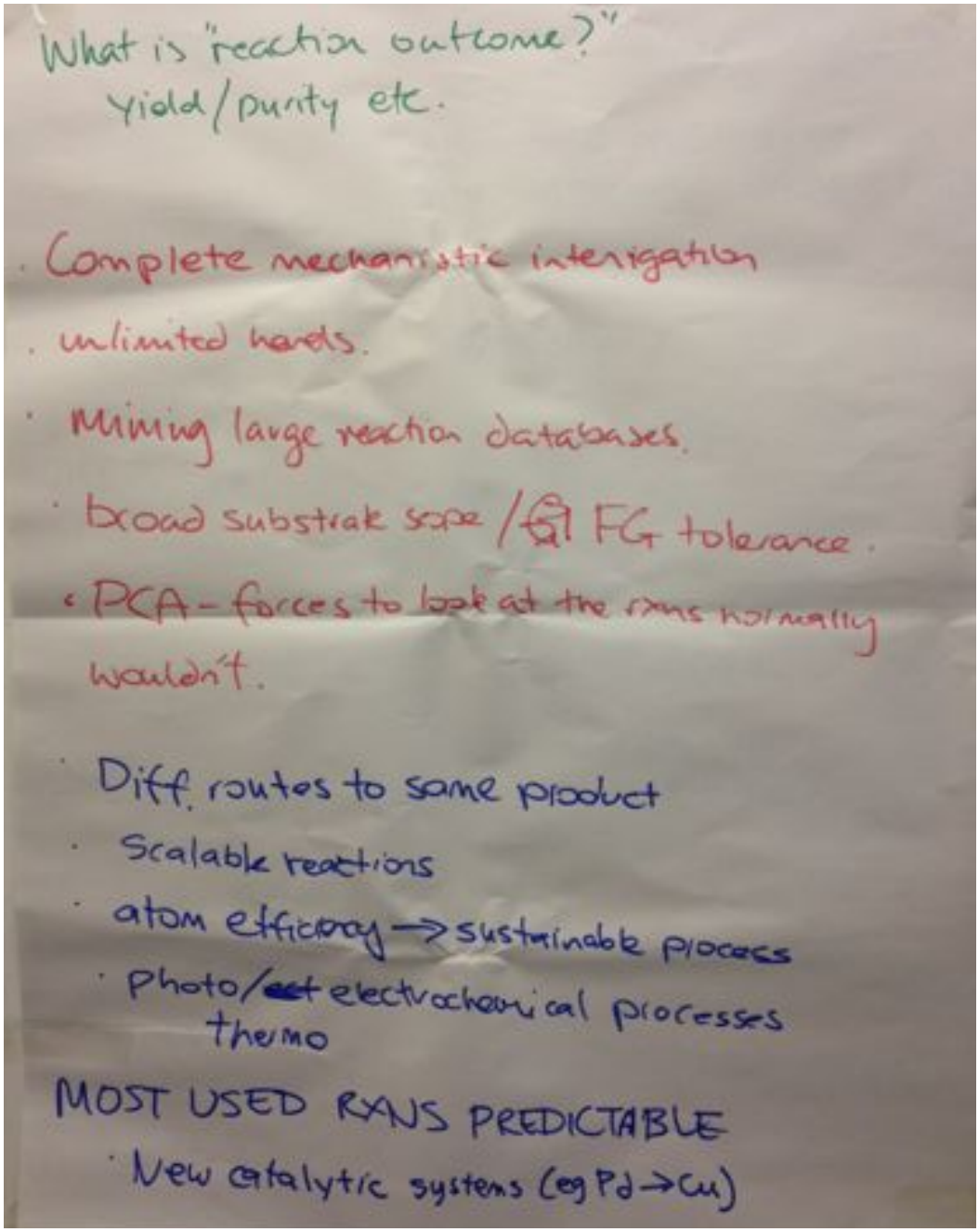

PREDICTING REACTION OUTCOMES
Is there anything out there already?
• Need ELN data with systematic collect
of all data (including failures)
• Can predict for a small number of v.
specific rxns ie - it can be done
• BUT No generic tool available
- Ones you've tried yourself.
- large amounts of precedents
- mechanistically simple
- ester hydrolysis
- Click
- solid characterisation data. (limitations)
- 'trouble makers'
- substrate classification
- reactive/ functional groups
- DFT as post-rationalisation

Session One; Group B



Session One; Group C




What is "reaction outcome?"
Yield / purity etc.
. Complete mechanistic interigation
. unlimited hands.
· Mining large reaction databases.
· broad substrate scope / FG tolerance.
· PCA - forces to look at the rxns normally wouldn't.
· Diff. routes to same product
· Scalable reactions
· atom efficiency → sustainable process
· Photo/ electrochemical processes
thermo
MOST USED RXNS PREDICTABLE
· New catalytic systems (eg Pd → Cu)

Session Two: Designing Synthetic Routes:

a) How to avoid a combinatorial explosion?
Pruning at each stage - need suitable criteria – which are somewhat subjective
Interactive pruning through suitable GUI?
Optimising scoring functions - learn from combichem and which reactions fail - analyse success of combichem results (data from GSK?)
Really big problem 10^19 after five deep synthetic analysis
- brute force computer power will never win

Avoid explosion by focus on a key transformation
- probably people do it this way
- any good references? Todd review? (DOI: 10.1039/b104620a)

Do we want to restrict explosion? Trade off between practicality and creativity
- capture and score depending on needs
- do not believe the first answer of a Google search

Training people to use tools is very important

b) What are useful families of targets?
3D fragments - poly aromatics - move away from flatland
Applying today's methods to broad range of transition metal compounds
Top 200 drug molecules:
http://cbc.arizona.edu/njardarson/group/top-pharmaceuticals-poster
Are drugs valuable enough?
"Did not come up with much"
Heterocycles, spirocycles, organofluorines
Key intermediates; fine chemicals; scaffolds; solar cells; … ?

c) What are good groups of reactions?
Good reactions - make reliable, tolerant, reproducible, automatable; transition metals
Telescoping reactions
- can we predict possibilities for telescoping from available data
- the necessary data may be available but un-analysed

Computers may not currently be good at capturing sensible groups of reactions
- but perhaps find common pairs?
- perhaps has been looked at by Grzybowksi

Green chemistry - what does this mean? low cost, low impact, etc
Enyzme catalysed processes

d) Do we need new reactions? If so, what should they do?
Yes ! Suzukiase and 3D Suzuki
Stereoselective regioselective CH activation
Simultaneous bond forming with control and stereochemistry
Designer enzymes
Reliable photochemistry
Avoiding protection/deprotection
Tandem reactions
Introducing fluorine
Both stoichiometric and catalytic
Enabling technologies - eg making phosgene *in situ* in flow
Re-examine old reactions and by-products of old reactions to develop new reactions.
Eg: improve the Schotten-Baumann reaction

One sentence note on each lecture:

David Woods: powerful statistical tools are available and underused by chemists

David Flanagan: ARChem now from Wiley (once Symbiosys) is a useful synthesis planning tool which can be used with many databases

Mike Hutchings: ICSynth: synthesis planning software restarted after a hiatus in the late 80s

Peter Johnson: how does the state of the art differ from Corey's work in the sixties? The principles laid down by Corey et al. in the 1970's are still relevant but modern tools like ARChem provide a practical aid for solution to everyday synthetic problems.

Tony Cook: it is possible to address stereochemical issues in automated synthesis planning

Fernando Huerta: ICSynth can go forwards as well as backwards but it does not take into account reagents or reaction conditions.

Mike Bodkin: Reaction vectors are a powerful tool in reaction analysis

Some References:

Computer-aided synthesis design: 40 years on
Anthony Cook, A. Peter Johnson, James Law, Lahdi Mirzazadeh, Orr Ravitz and Aniko Simon
Wiley Interdisciplinary Reviews: Computational Molecular Science, 2012, 2 (1) pp 79-107

Analysis of the reactions used for the preparation of drug candidate molecules
John S. Carey, David Laffan, Colin Thomson and Mike T. Williams
Org. Biomol. Chem., 2006,4, 2337-2347
10.1039/B602413K

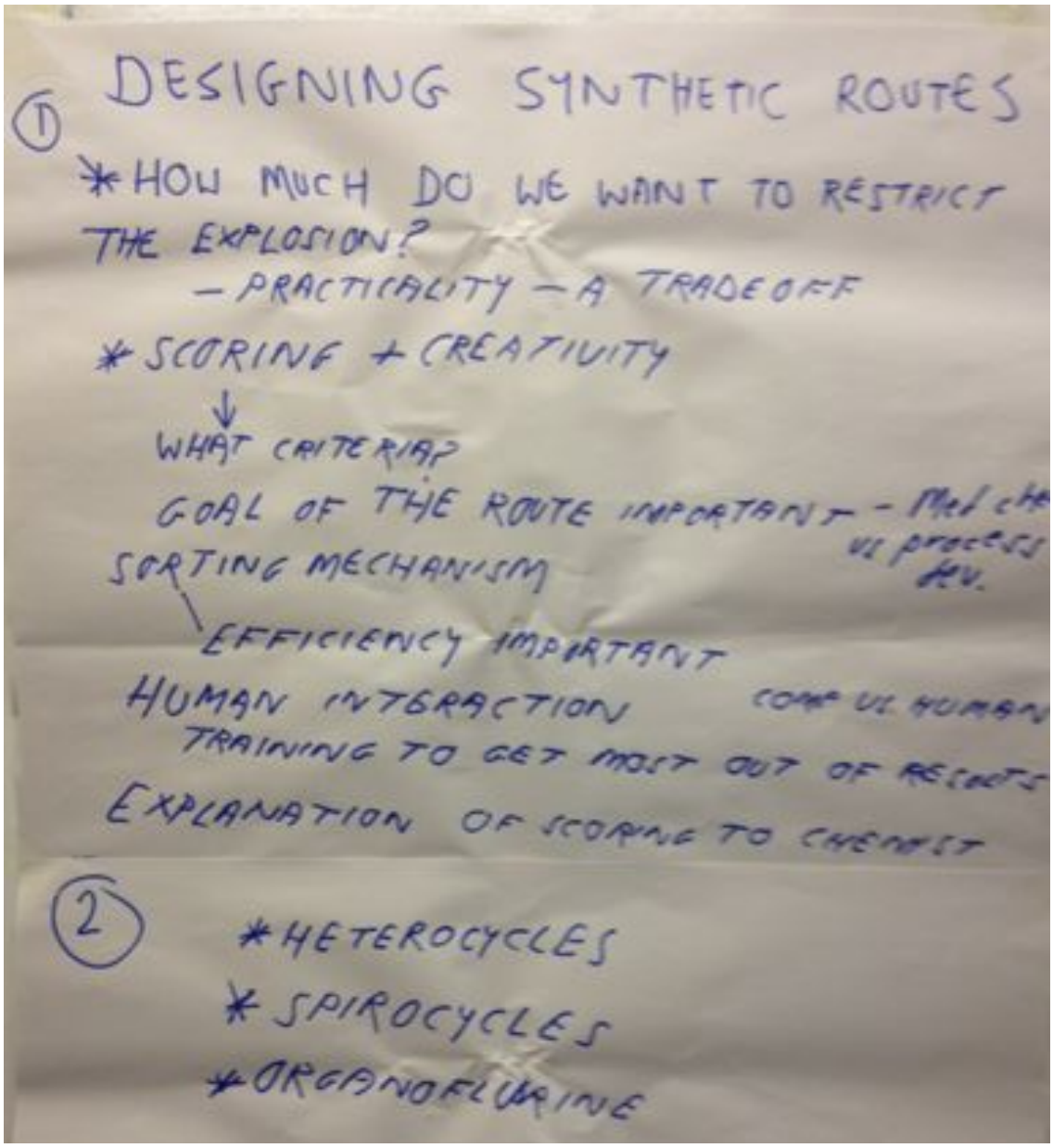
The Medicinal Chemist's Toolbox: An Analysis of Reactions Used in the Pursuit of Drug Candidates
Stephen D. Roughley and Allan M. Jordan
J. Med. Chem., 2011, 54 (10), pp 3451–3479
DOI: 10.1021/jm200187y

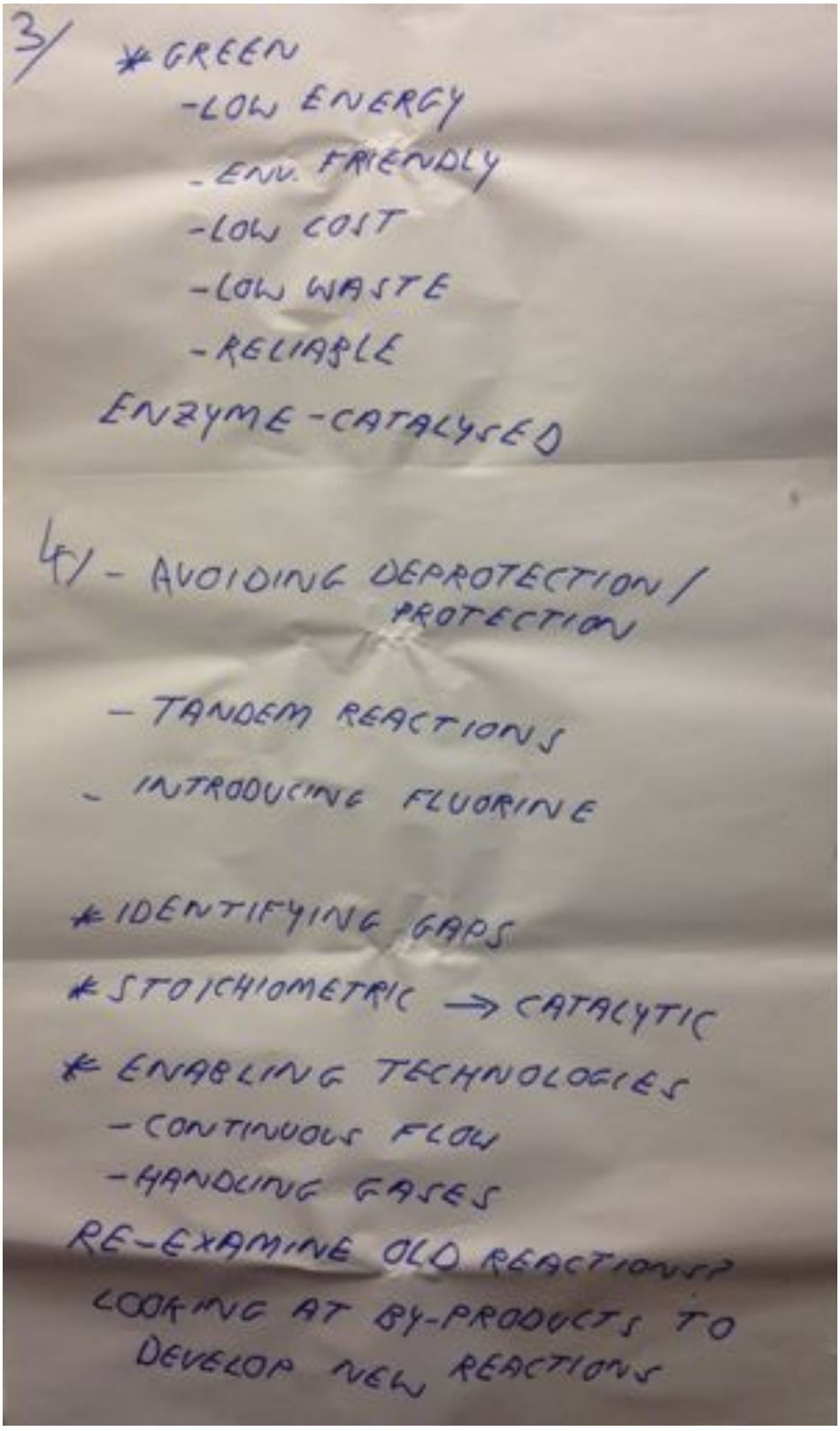
Advancing the Drug Discovery and Development Process
Prof. K. C. Nicolaou
DOI: 10.1002/ange.201404761

2009 Hina Patel PhD Thesis (Gillett/Bodkin, Sheffield)
- no relationship between similarity and yield

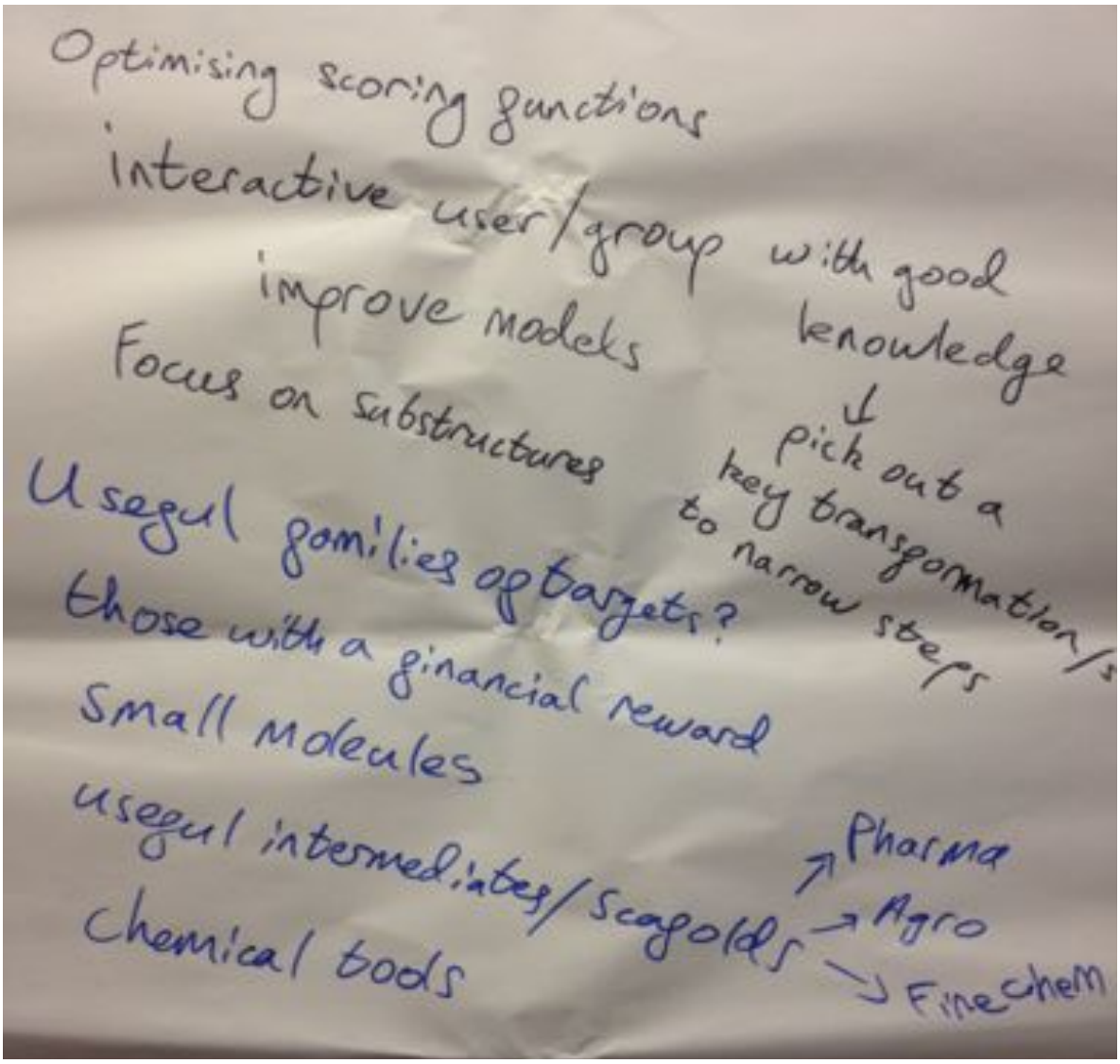
Knowledge-Based Approach to de Novo Design Using Reaction Vectors
Patel, H; Bodkin, MJ; Chen, BN; Gillet, VJ
Journal of Chemical Information and Modeling Volume: 49 Issue: 5 Pages: 1163-1184
DOI: 10.1021/ci800413m

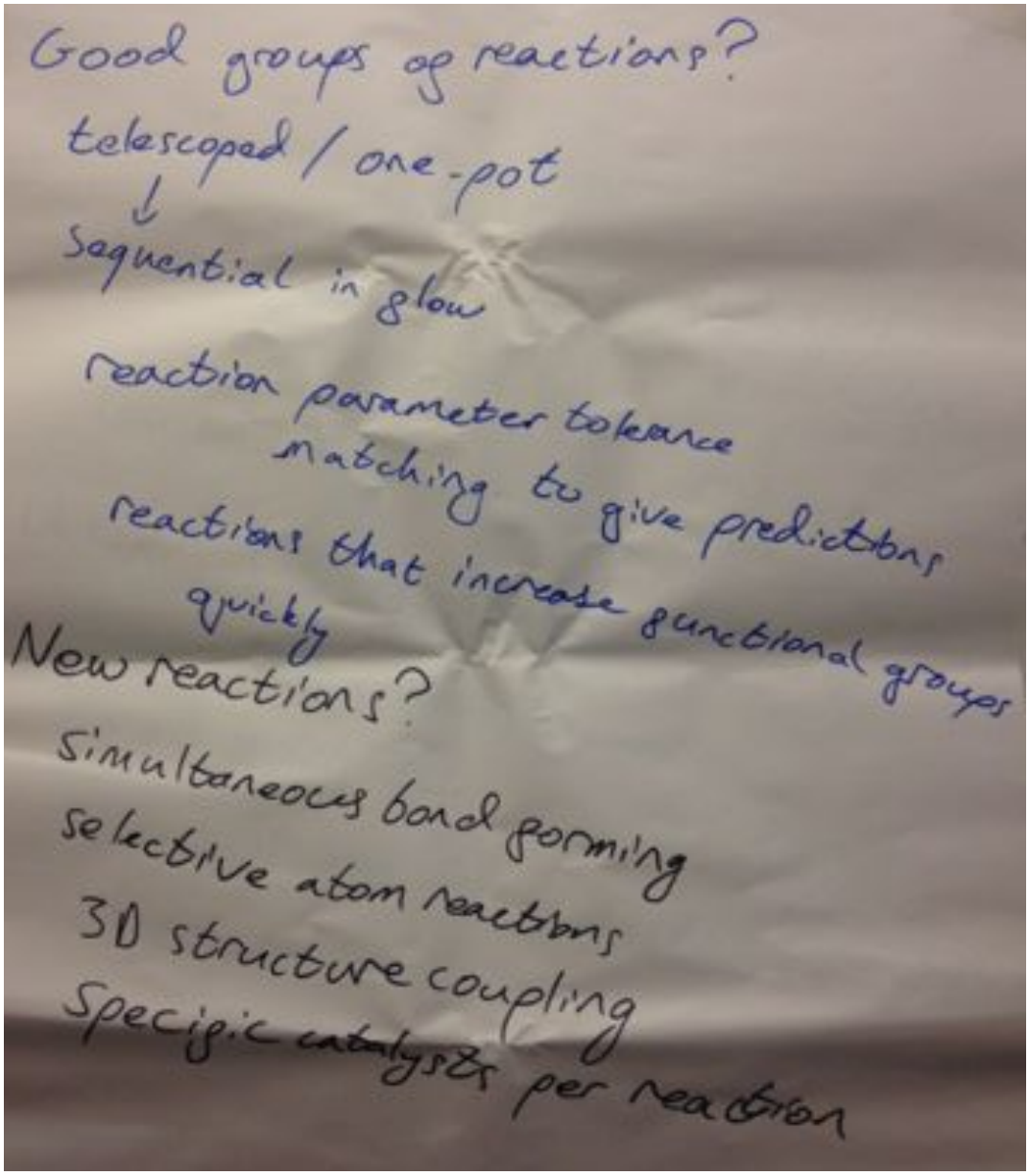
Session Two; Group D

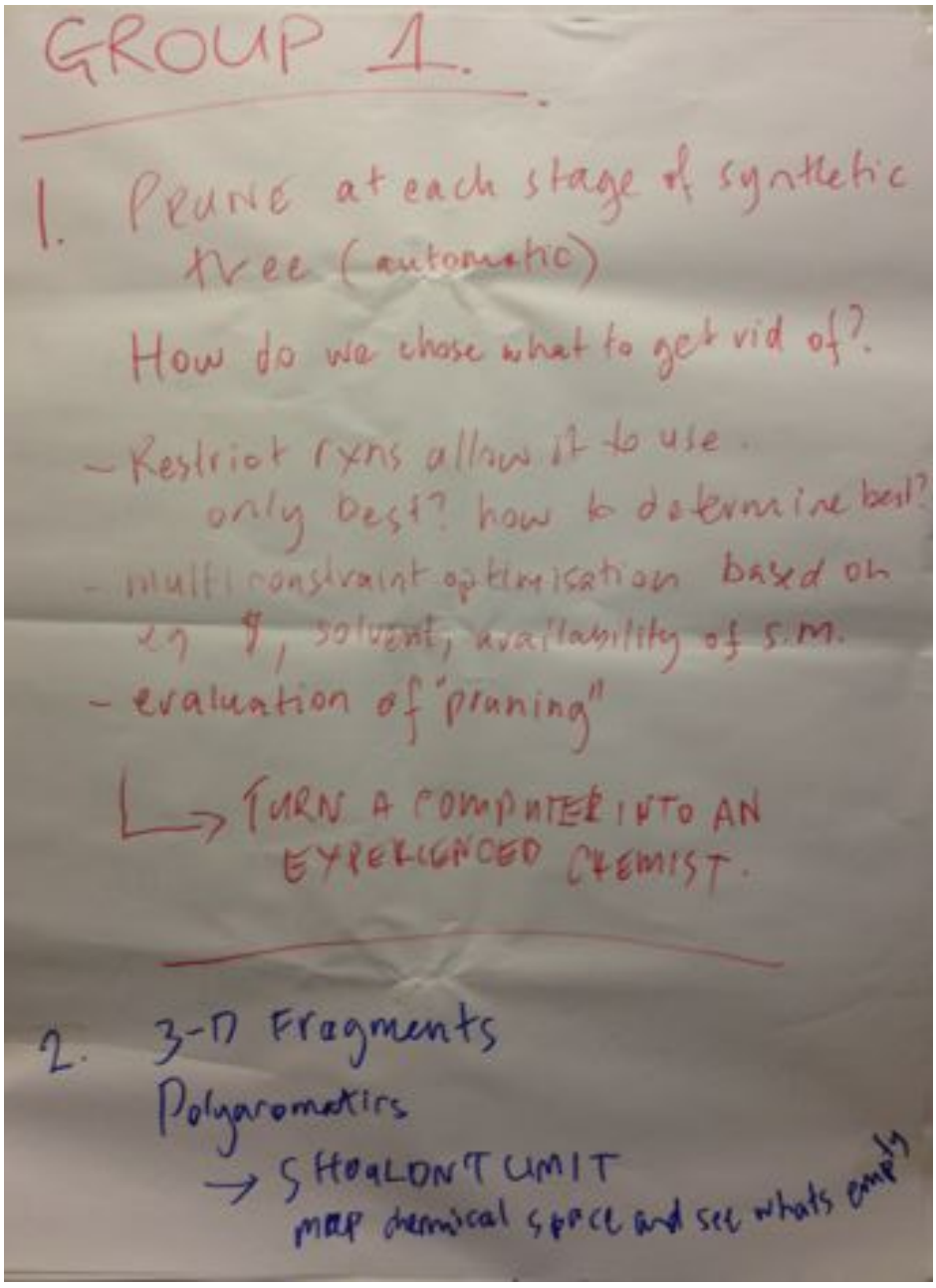


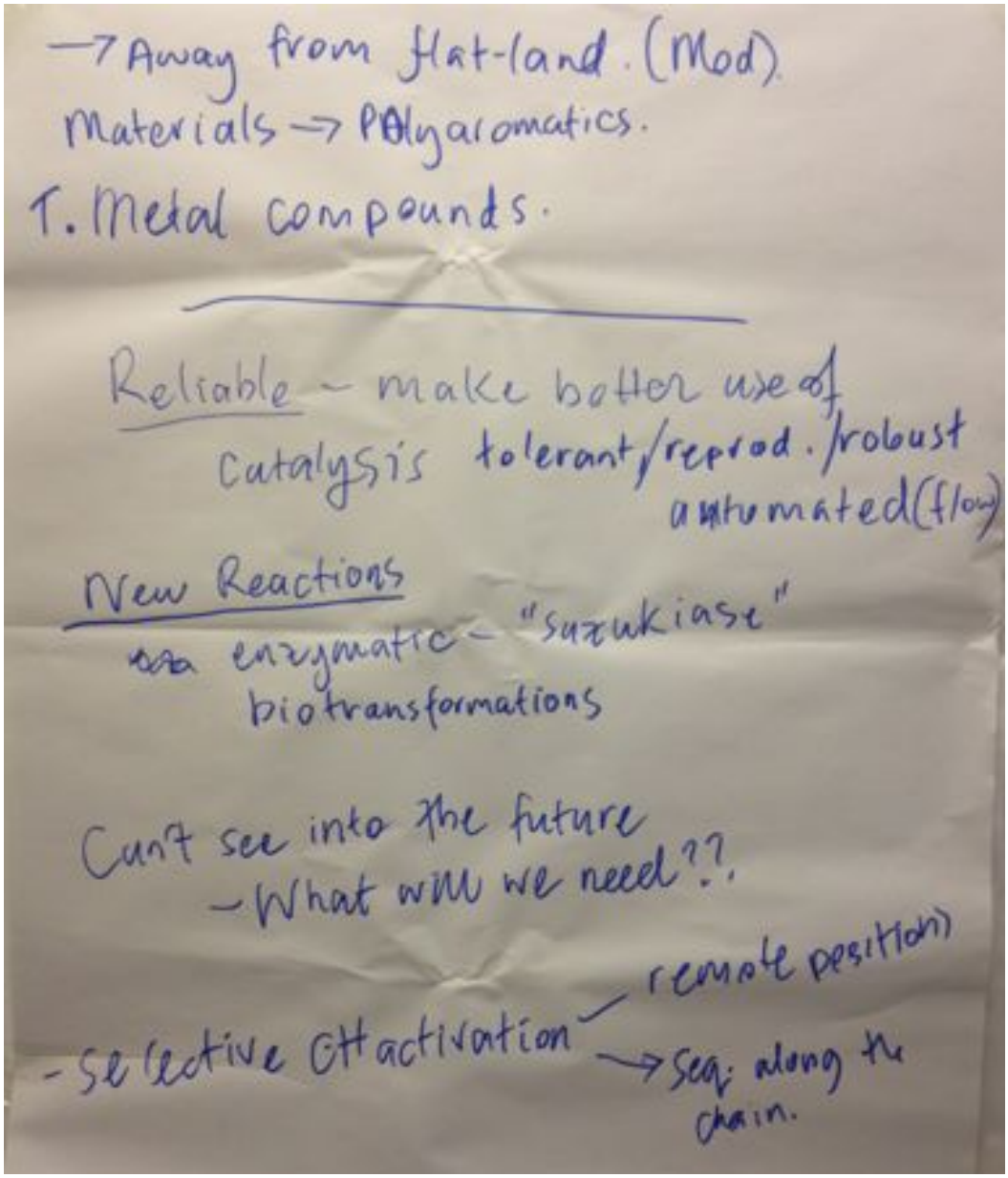

3/
* GREEN
- LOW ENERGY
- ENV. FRIENDLY
- LOW COST
- LOW WASTE
- RELIABLE
ENZYME-CATALYSED
4/ - AVOIDING DEPROTECTION / PROTECTION
- TANDEM REACTIONS
- INTRODUCING FLUORINE
* IDENTIFYING GAPS
* STOICHIOMETRIC → CATALYTIC
* ENABLING TECHNOLOGIES
- CONTINUOUS FLOW
- HANDLING GASES
RE-EXAMINE OLD REACTIONS?
LOOKING AT BY-PRODUCTS TO DEVELOP NEW REACTIONS

Session Two; Group E




Good groups of reactions?
telescoped / one-pot
Sequential in flow
reaction parameter tolerance
matching to give predictions
reactions that increase functional groups
quickly
New reactions?
simultaneous bond forming
selective atom reactions
3D structure coupling
Specific catalysts per reaction


GROUP 1.
1. Prune at each stage of synthetic tree (automatic)
How do we chose what to get rid of?.
- Restrict rxns allow it to use.
only best? how to determine best?
- multi constraint optimisation based on eg $, solvent, availability of s.m.
- evaluation of "pruning"
→ TURN A COMPUTER INTO AN EXPERIENCED CHEMIST.
2. 3-D Fragments
Polyaromatics
→ STHOLONT LIMIT
map chemical space and see whats empty


→ Away from flat-land. (Mod).
Materials → Polyaromatics.
T. Metal compounds.
Reliable - make better use of
catalysis tolerant/reprod./robust
automated (flow)
New Reactions
enzymatic - "suzukiase"
biotransformations
Can't see into the future
- What will we need??
- Selective CH activation → remote position
→ Seq. along the chain.

Session Three:
Design a Machine to make many different compounds:

a) How many different compounds could it make?
Peptide synthesisers and oligopeptide synthesisers and chemical plants
Precursors?
Synthesiser vs a screener
As many as we need
10^9 - 10^10
Milligram scale
Single step or sequential steps - rapid access to libraries? scale up?

b) How would it work?
Computer prediction - human prediction
Synthesis and purification and analysis instrument
Modular
Step - analyse - evolve - *repeat*
Do not do all the steps then purify and characterise
A lot of things already exist - 96 well plate; prep chromatography; etc;
Segmented flow for milligram quantities
No proprietory formats
After five years the machine should not be gathering dust
Need a robust reaction protocol - well defined - flow is good - controls everything carefully
In-line analytical methods needed
Modular is very important

c) How much would it cost?
Build out of lego..
£ 1-2 M is a feasible grant-funded budget
Less than developing libraries now
Must be sustainable
Compare H-cube and X-cube
General purpose - doing general reactions (eg Suzuki)
Need graduates and staff trained with a particular set of skills in data handling
Data capture important for future learning
Catalytic methodology

d) Could we get a grant to build it?
Pre-competitive - consortia of chemical companies

e) Would it produce useful molecules?
Choose:
Dull chemistry - new starting materials
New chemistry – challenging processes
Show-case methodology for academic groups

f) What would be the most difficult issues?
Worry about solubility issues
Need staff resource to optimise chemistry; rebuild; investment risk
Self cleaning reactors
Automation is not everything - needs good chemistry too.

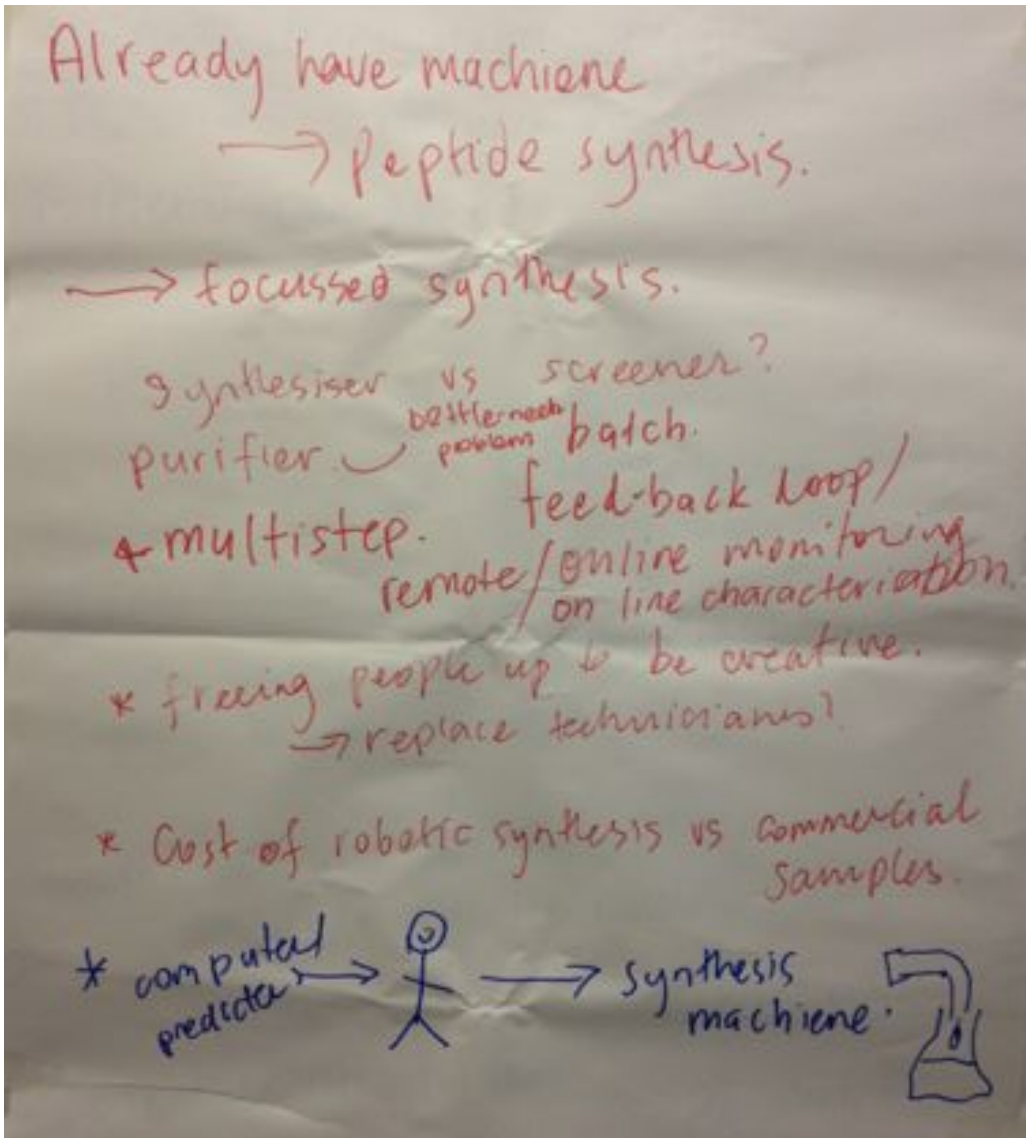
One sentence note on each lecture:

Jonathan Goodman: Reaction InChI and computational methods can help us order chemistry; we need to know more, but we can do more with what we have.

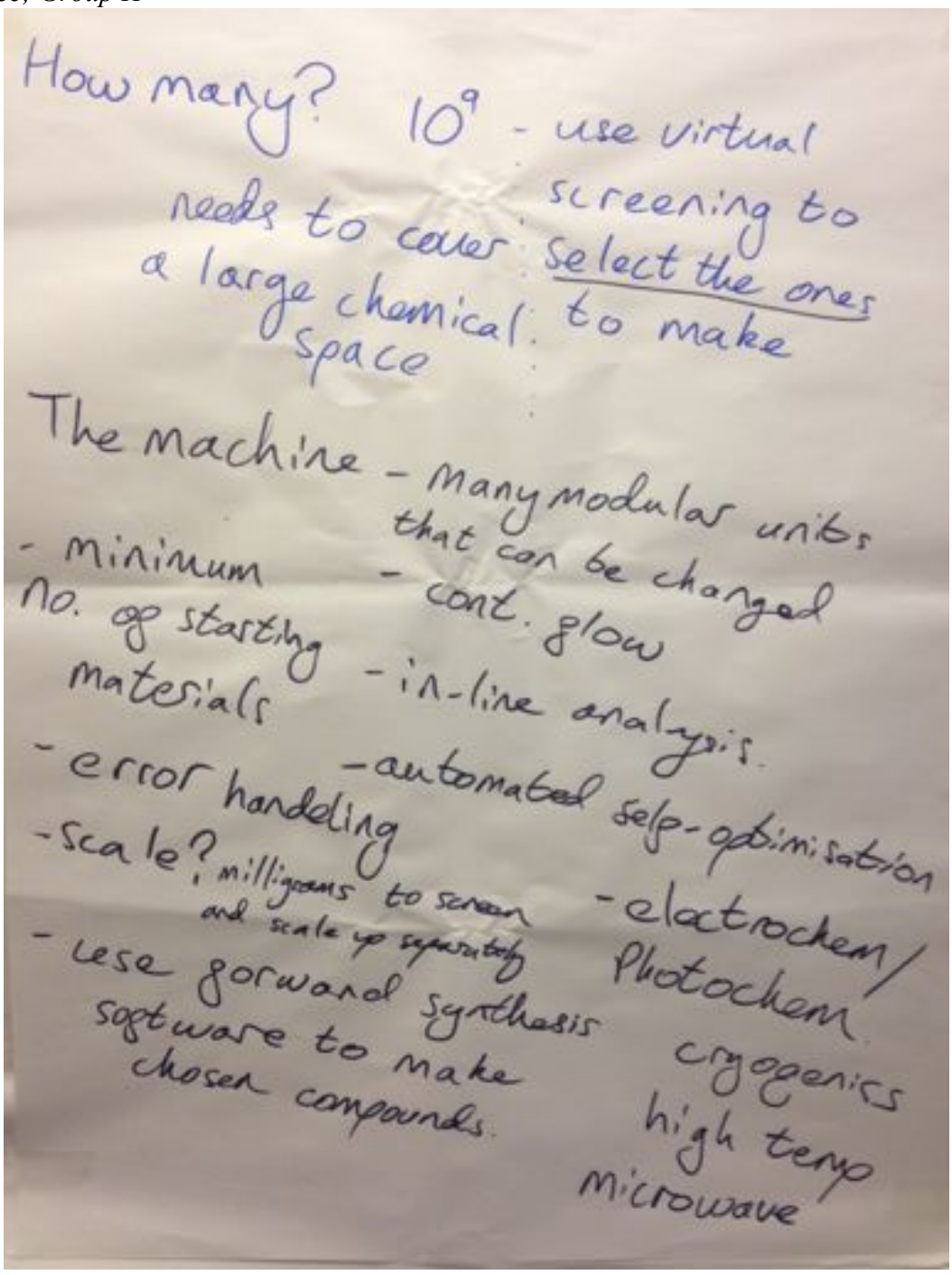
Roger Sayle: Big data allows us to discover well-known rules of medicinal chemistry, and many new insights

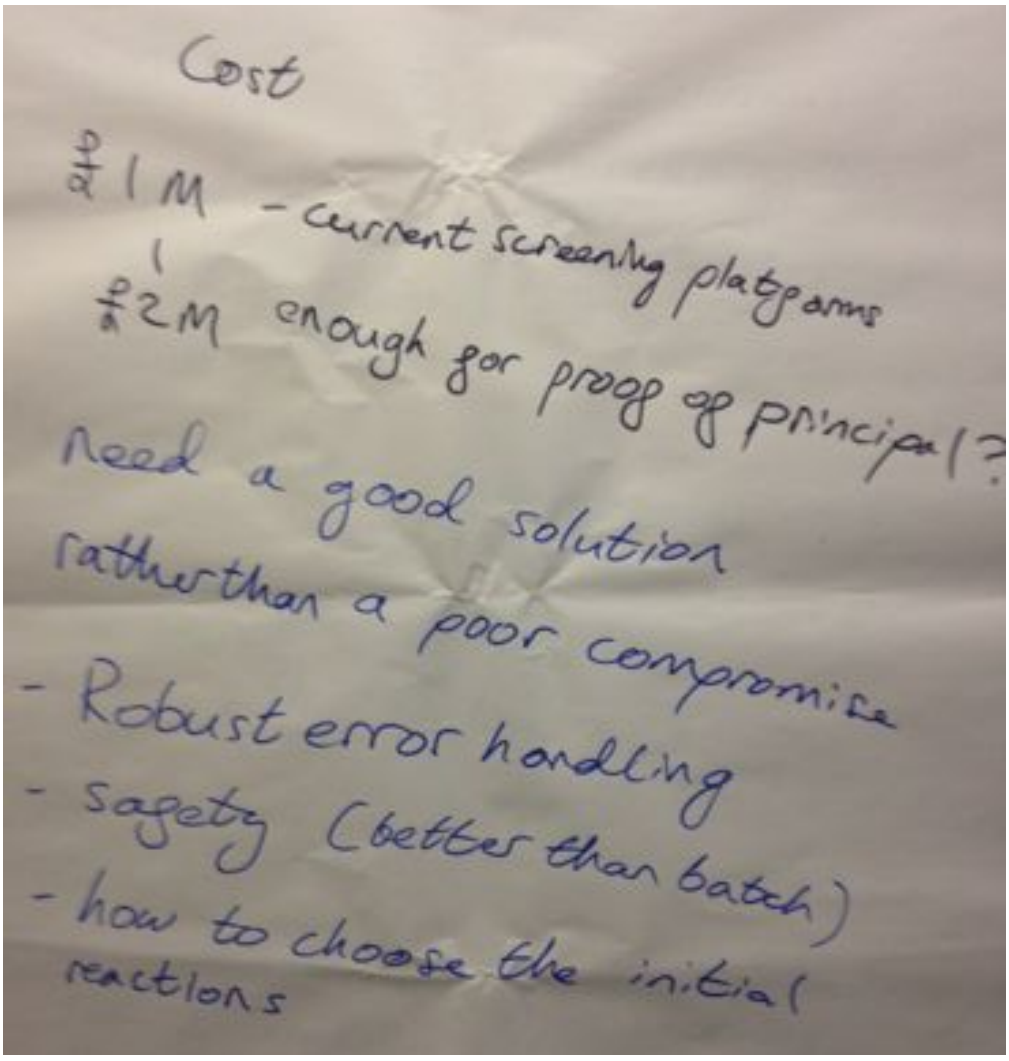
Mark Leach: www.chemthes.com

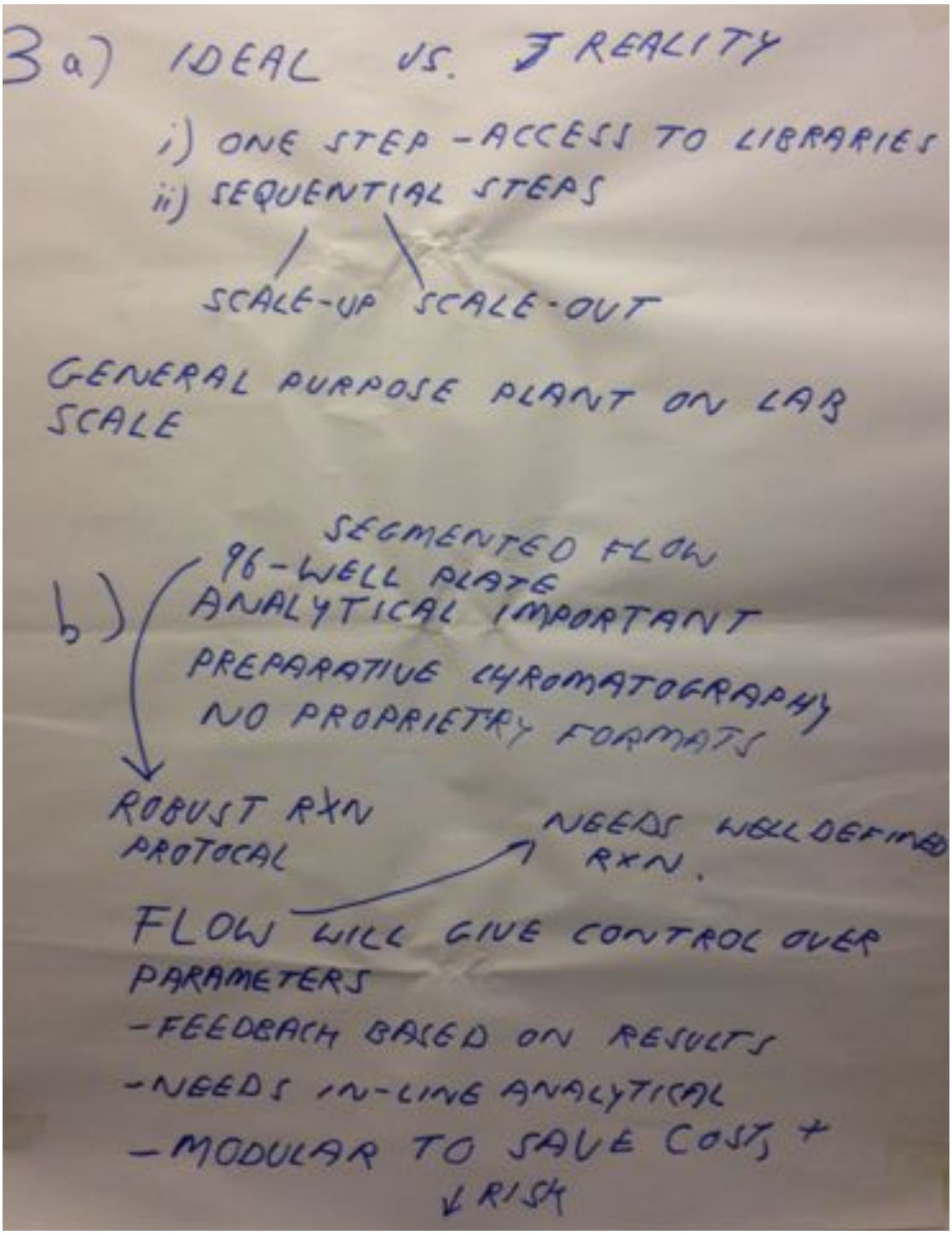
Session Three; Group G

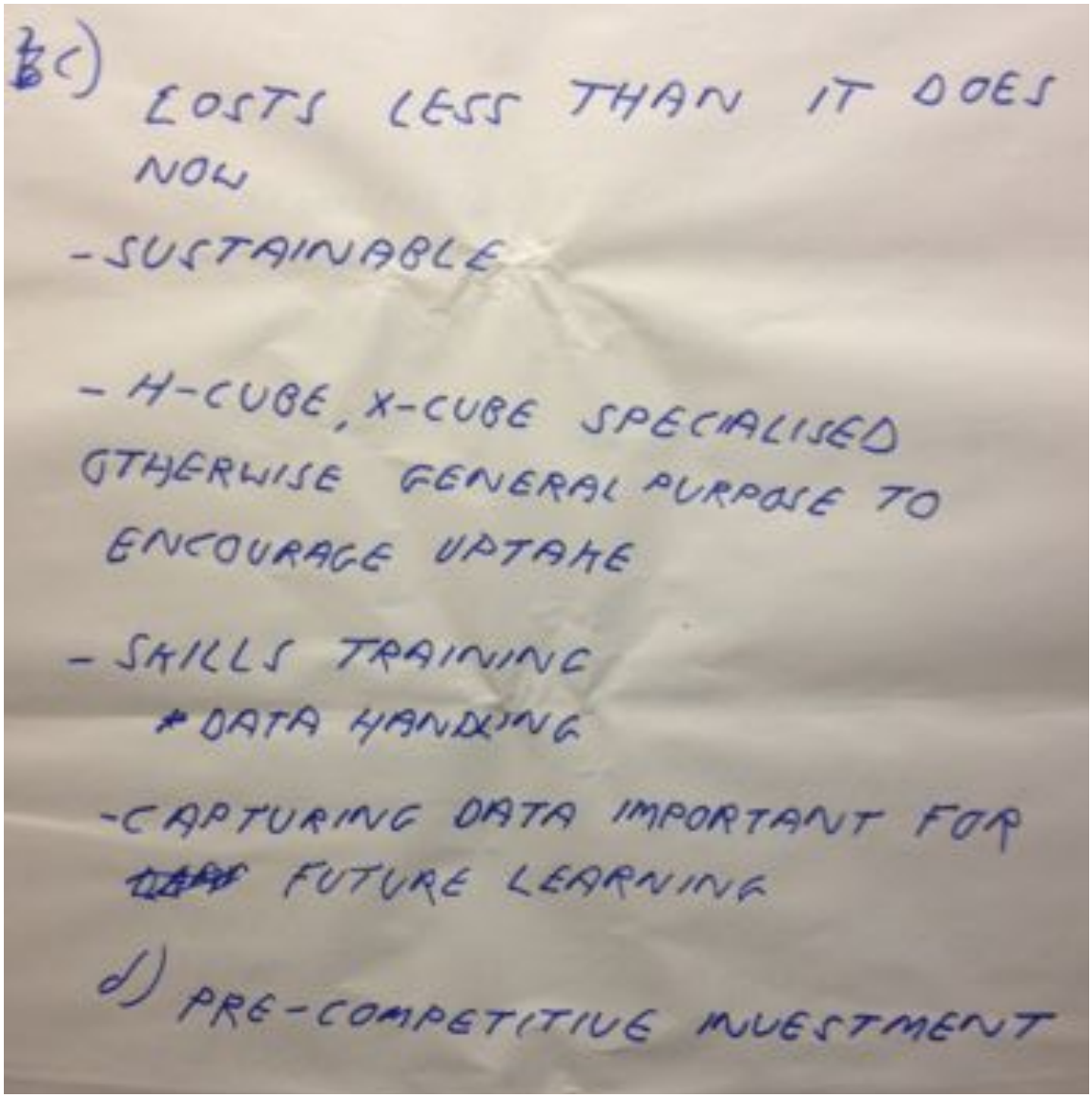


Session Three; Group H




Cost
£1M - current screening platforms
£2M enough for proof of principal?
need a good solution
rather than a poor compromise
- Robust error handling
- sagety (better than batch)
- how to choose the initial
reactions


3 a) IDEAL VS. REALITY
i) ONE STEP - ACCESS TO LIBRARIES
ii) SEQUENTIAL STEPS
SCALE-UP
SCALE-OUT
GENERAL PURPOSE PLANT ON LAB SCALE
b)
SEGMENTED FLOW
96-WELL PLATE
ANALYTICAL IMPORTANT
PREPARATIVE CHROMATOGRAPHY
NO PROPRIETARY FORMATS
ROBUST RXN PROTOCAL
NEEDS WELL DEFINED RXN.
FLOW WILL GIVE CONTROL OVER PARAMETERS
- FEEDBACK BASED ON RESULTS
- NEEDS IN-LINE ANALYTICAL
- MODULAR TO SAVE COST + ↓ RISK


2b c)
COSTS LESS THAN IT DOES NOW
- SUSTAINABLE
- H-CUBE, X-CUBE SPECIALISED OTHERWISE GENERAL PURPOSE TO ENCOURAGE UPTAKE
- SKILLS TRAINING
* DATA HANDLING
-CAPTURING DATA IMPORTANT FOR FUTURE LEARNING
d) PRE-COMPETITIVE INVESTMENT

e). "DULL" CHEMISTRY, NEW REAGENTS

- NEW CHEMISTRY, HARD TO ACCESS CHEMISTRY

RAPID ACCESS TO LIBRARIES TO SHOWCASE A METHODOLOGY

f) REAL LIFE IMPLEMENTATION

*SOLUBILITY ISSUES

*STILL NEED RESOURCE TO OPTIMISE CHEMISTRY SO IT IS GOOD ENOUGH

*COSTS FOR RE-BUILD, MANUAL INTERVENTION

* INVESTMENT RISK

~~*SE~~

- SELF-CLEANING REACTION
- MODULAR APPROACH
- AUTOMATION ISN'T EVERYTHING