The use of rapid 2D design methods within a Design-to-Delivery software suite

Susan Boyd, Michaela Lenz & Mike Snarey

3rd Joint Sheffield Conference on Chemoinformatics

April 2004

SCYNEXIS™

Research Triangle Park, NC, USA

Fyfield Research Park, Essex, UK

Chemistry focused drug discovery partnering company

CGMP Kg Production Plant is in USA

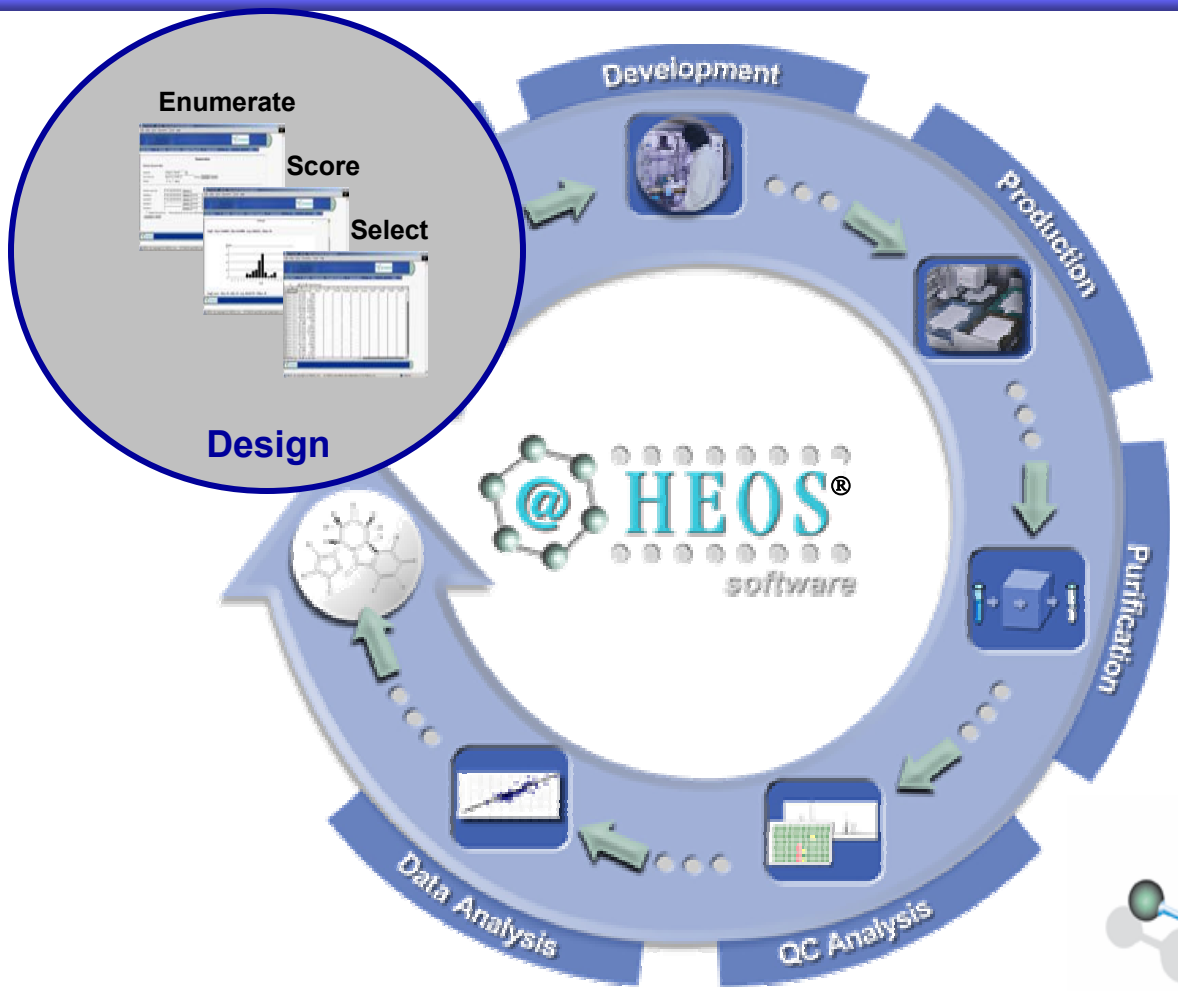
Radiochemicals Production is in UK

- **Founded by former Aventis Researchers in July 2000**
- **120+ employees Worldwide**
  - **73 US, 47 UK**

SCYNEXIS™

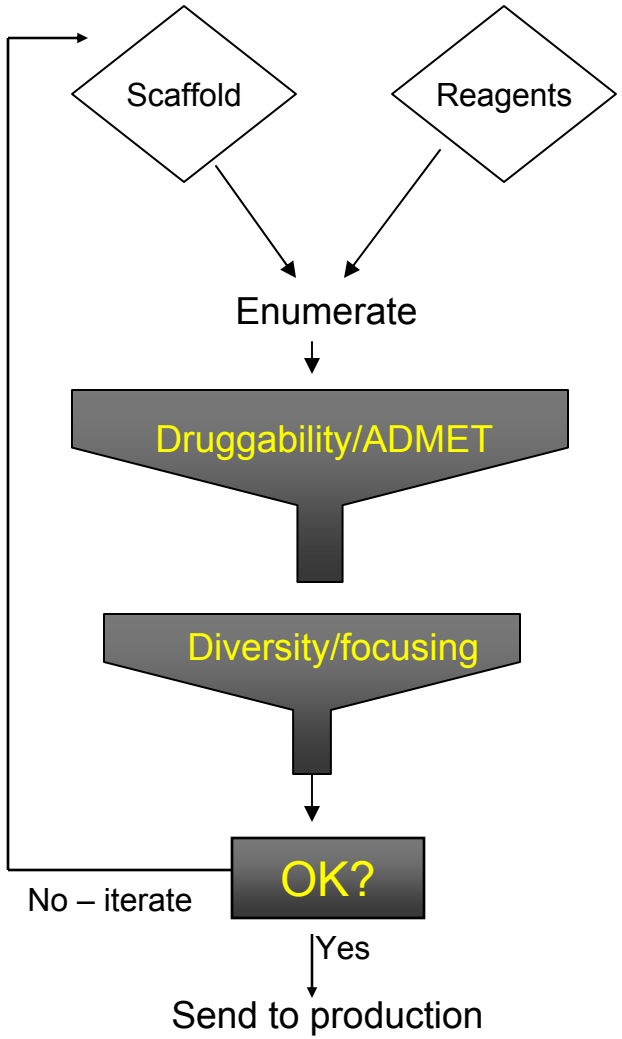
The HEOS® "Electronic Loop"

Hit Explorer Operating System



SCYNEXIS™

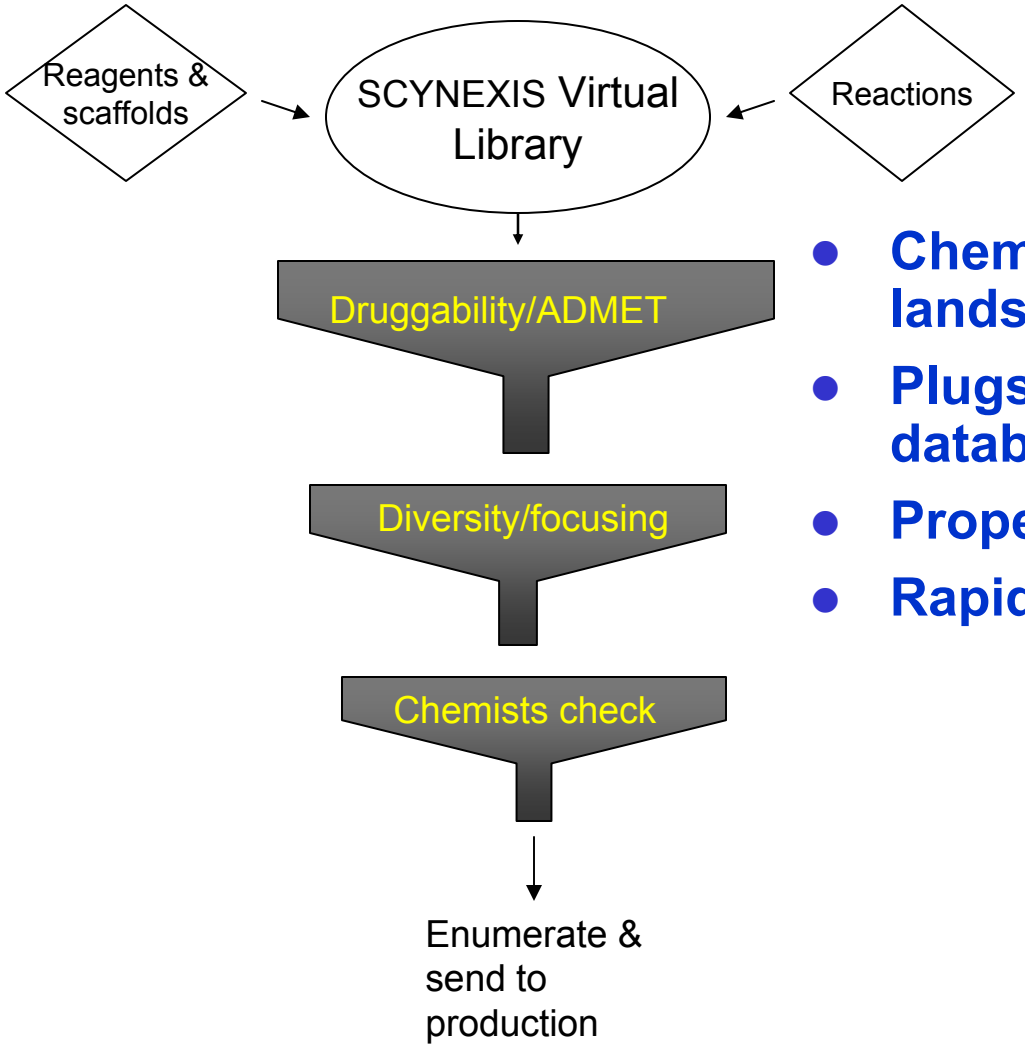
HEOS® Design Process Flow



- Chemists make decision which scaffold(s) to use up-front
- Time-consuming enumeration first step in process
- Re-enumeration needed to create final SD file for production, minus those monomers which scored unfavourably

SCYNEXIS™

Alternative Design Process Flow



- **Chemist can "survey the chemical landscape" ahead of decision-making**
- **Plugs into reagent & reaction databases (Oracle)**
- **Property calculation early in process**
- **Rapid enumeration**

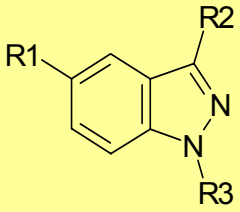
SCYNEXIS

BCI Markush Toolkit

- **Enumeration speeds of 50,000 compounds (with properties) per second !**
- **Computation of product properties from "Markush object"**
  - **Lipinski, rotatable bond count, molar refractivity**
  - **Fingerprints**
  - **Topological indices, atom counts etc…..**

R1 = alky, aryl (derived from benzoic acid reagents)

C
CC(C)C
Cc1ccccc1
etc



R2 = alkyl, aryl (derived from Grignard reagents)

CC
C(C)C
CCN(C)C
etc

R3 = alkyl, aryl (derived from hydrazine reagents)

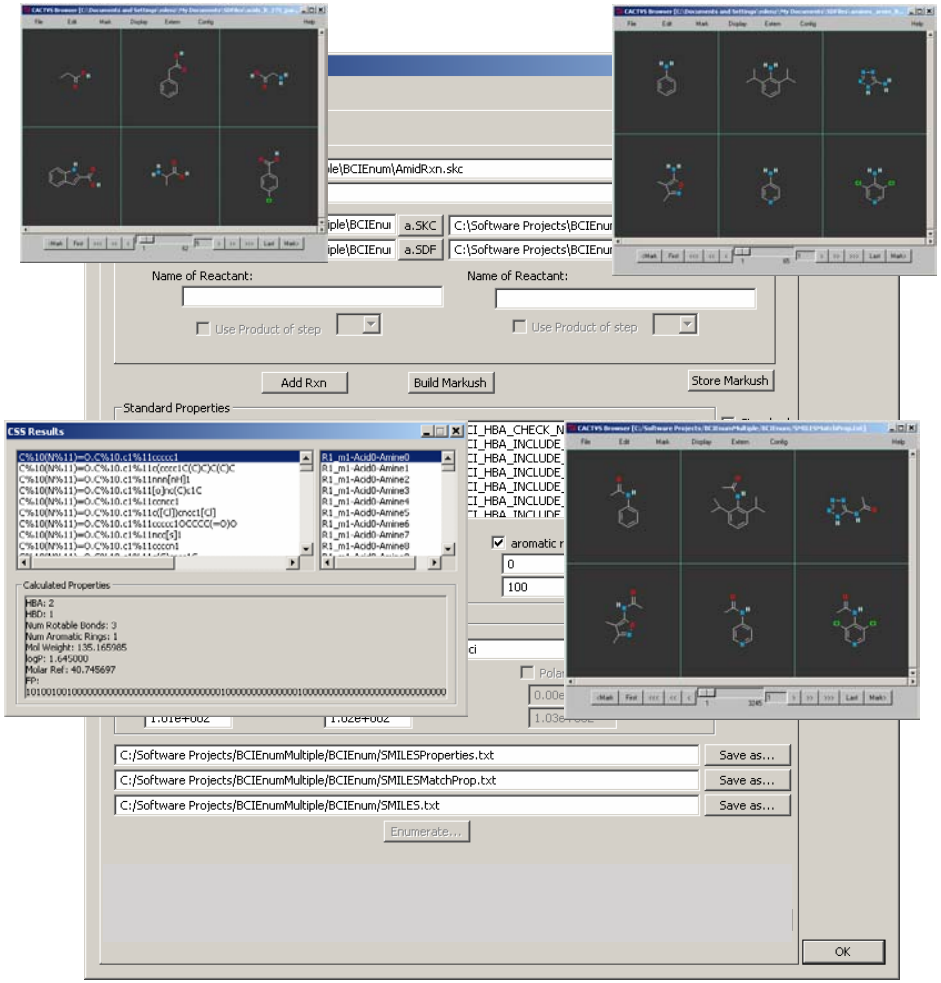
C(C)C
c1ccccc1
CN(C)CC
etc

Creation of Markush Objects with BCI Markush Toolkit

- **RG file**
- **Combinatorial Synthesis Scheme**
- **Editable Markush Object**

SCYNEXIS

Markush Enumeration System

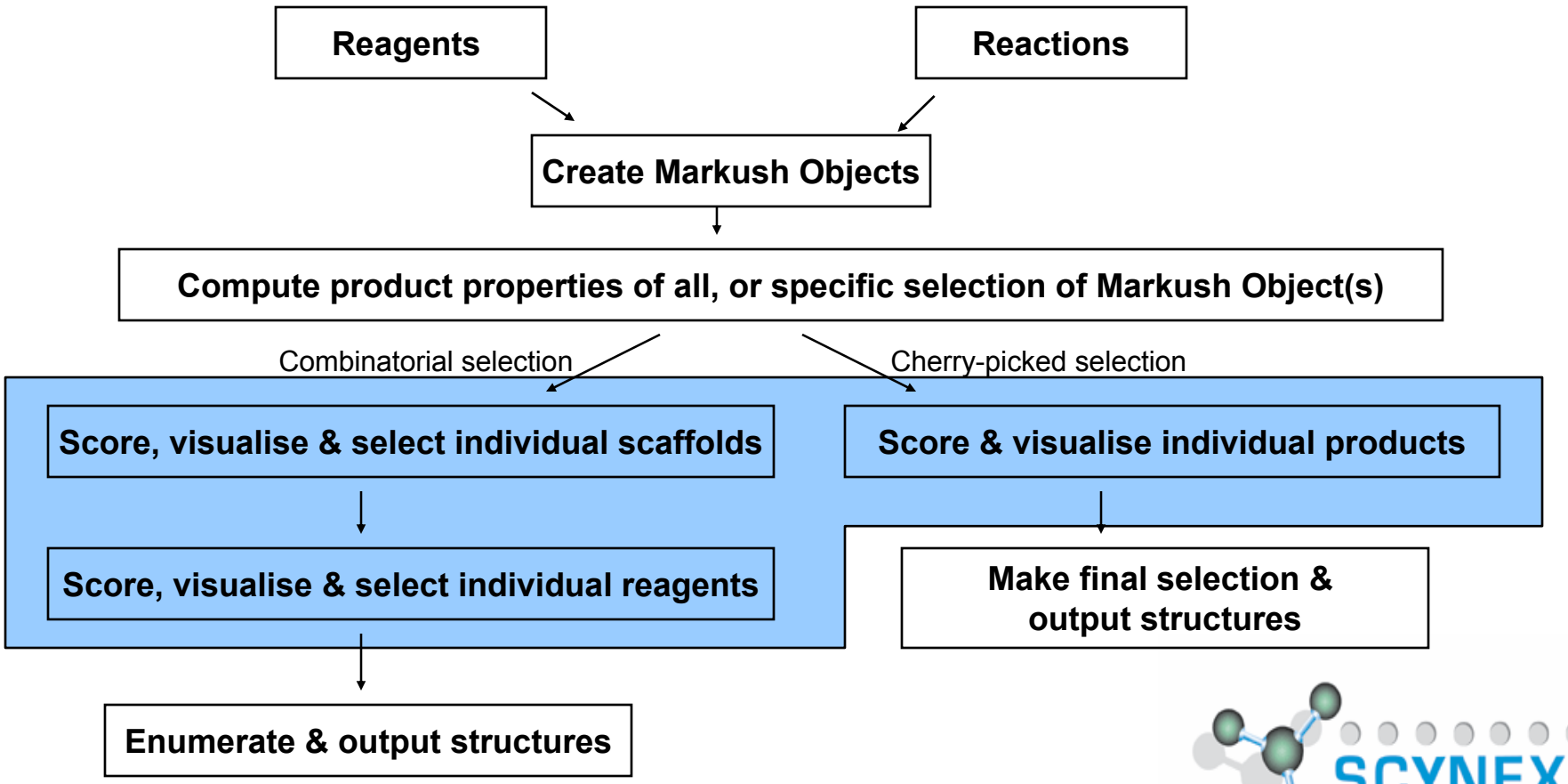


- Easy to use
- Fast

- No stereochemistry (CSS)
- All reactive sites react!

SCYNEXIS™

Product & Monomer Scoring

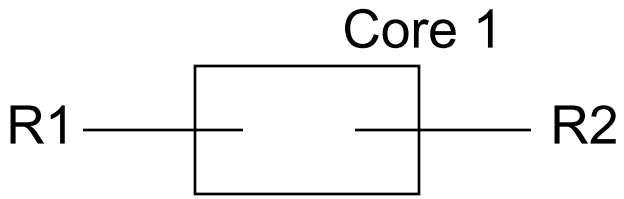


SCYNEXIS

Scoring on simple properties

Libraries (scaffolds) or monomers can be scored on simple property basis…..

- **Library basis**
  - **What % of products from a given Markush object (ie scaffold + suitable reagents) satisfy desired ro5 constraints?**
- **Monomer basis**
  - **What % of products built using this single reagent – from a given Markush Object – satisfy desired ro5 constraints?**



SCYNEXIS

Scoring using models

- Again, applicable to either whole libraries or to individual monomers
- Run virtual compounds against desired models, and score accordingly
- QSAR models, pharmacophore models, docking scores

…….2D virtual screening techniques

SCYNEXIS™

Virtual Screening

- **Two methods explored, for scoring purposes**
  - **Recursive partitioning decision trees**
    - **c4.5 : See http://www.cse.unsw.edu.au/~quinlan/ for c4.5 classification program and http://www2.cs.uregina.ca/~hamilton/courses/831/notes/ml/dtrees/c4.5/tutorial.html for a tutorial**
    - **Cactvs fingerprints: See http://www2.chemie.uni-erlangen.de/software/cactvs/index.html**
  - **Nearest-neighbours similarity searching**
    - **BCI fingerprint Tanimoto**

c4.5 Recursive Partitioning

- Top ten most plentiful “agonists”, “antagonists” & “inhibitors” pulled from MDDR database
  - 8696 compounds
- Cactvs fingerprints generated
- 2000 of these selected at random, as training set for c4.5
- Predicted the most likely of the ten classes for the remaining 6696 compounds

c4.5 MDDR dataset

Lipoxygenase Inhibitor	1141	(A)
PAF Antagonist	1002	(B)
gpIIb/IIIa Receptor Antagonist	987	(C)
Substance P Antagonist	915	(D)
Aldose Reductase Inhibitor	863	(E)
Phosphodiesterase IV Inhibitor	852	(F)
Muscarinic (M1) Agonist	850	(G)
Angiotensin II AT1 Antagonist	797	(H)
5 HT3 Antagonist	650	(I)
Thrombin Inhibitor	639	(J)

Results of c4.5 classification

Target	$A_{(actual)}$	$B_{(actual)}$	$C_{(actual)}$	$D_{(actual)}$	$E_{(actual)}$	$F_{(actual)}$	$G_{(actual)}$	$H_{(actual)}$	$I_{(actual)}$	$J_{(actual)}$
$A_{(pred)}$	74.8	3.8	2.2	2.7	7.7	4.6	1.3	1.2	1.1	0.7
$B_{(pred)}$	3.6	69.3	3.8	8.2	2.4	3.8	1.5	3.7	1.7	2.0
$C_{(pred)}$	1.2	3.2	82.5	1.5	2.6	2.1	0.3	2.3	0.9	3.5
$D_{(pred)}$	1.7	5.8	1.1	76.0	1.1	6.5	0.1	2.8	2.3	2.4
$E_{(pred)}$	9.7	1.2	5.7	0.3	76.4	2.4	1.1	1.7	0.3	1.2
$F_{(pred)}$	6.6	5.8	0.9	1.3	5.5	75.2	0.9	1.3	1.4	0.9
$G_{(pred)}$	3.3	3.3	2.2	0.4	1.3	1.2	81.0	0.7	3.3	3.3
$H_{(pred)}$	2.1	3.4	1.8	0.8	0.5	1.6	0.2	88.2	0.5	0.8
$I_{(pred)}$	4.1	6.5	1.2	2.0	1.0	2.0	5.3	1.8	73.6	2.6
$J_{(pred)}$	0.6	5.5	10.8	3.5	0.6	3.1	0.6	1.0	1.4	72.9

- **Classification 69-88% correct (average 77%)**
- **High off-diagonal values may not be "incorrect", but merely indicate some similarity between the ligands for these receptors**

SCYNEXIS™

Nearest-Neighbours

- 18,594 compounds with <action> entries extracted from MDDR
- Of these, 245 classified as kinase inhibitors
- 6 of these chosen randomly as probes
- Nearest-neighbours similarity (Tanimoto) run and 18,588 MDDR compounds ordered by similarity to the 6 probes
- Found
  - 15 hits in top 100 — 11 x random
  - 44 hits in top 1000 — 3.5 x random

SCYNEXIS

Translation into Affinity Scores

- **c4.5 RP**
  - Various gene family & target-specific RP models available
  - Virtual library is run through pre-computed decision tree
  - Scaffolds & monomers are assigned to a branch of the tree
  - The affinity score (gene family or target) for each scaffold or monomer is simply the % of the products from the scaffold or monomer assigned to the appropriate tree branch (eg kinase, PDE)
- **Nearest-neighbours**
  - Probes selected from known actives against target
  - The affinity score for a scaffold/library is the proportion of products of the library which score > user-defined similarity threshold
  - The affinity score of a monomer (in given library) is computed similarly

SCYNEXIS™

Future Work

- Substructural safety filters
- Allow virtual screening processing of input single structures, or input, pre-computed SD files (including external libraries for procurement)
- Additional virtual screening tools (2D & 3D)

SCYNEXIS™

Acknowledgements

- Michaela Lenz, SCYNEXIS
- Mike Snarey, Independent Consultant
- BCI
- Freeware providers!

susan@compchemsol.co.uk · www.compchemsol.co.uk

CompChem Solutions

SCYNEXIS