Reaction Databases at the Crossroad

Marudai Balasubramanian Ph.D.
Research Informatics
Pfizer Global Research and Development
Ann Arbor, MI, USA

ICIC 2006 Conference
10/22/2006 – 10/25/2006
Nimes, France

Disclaimer

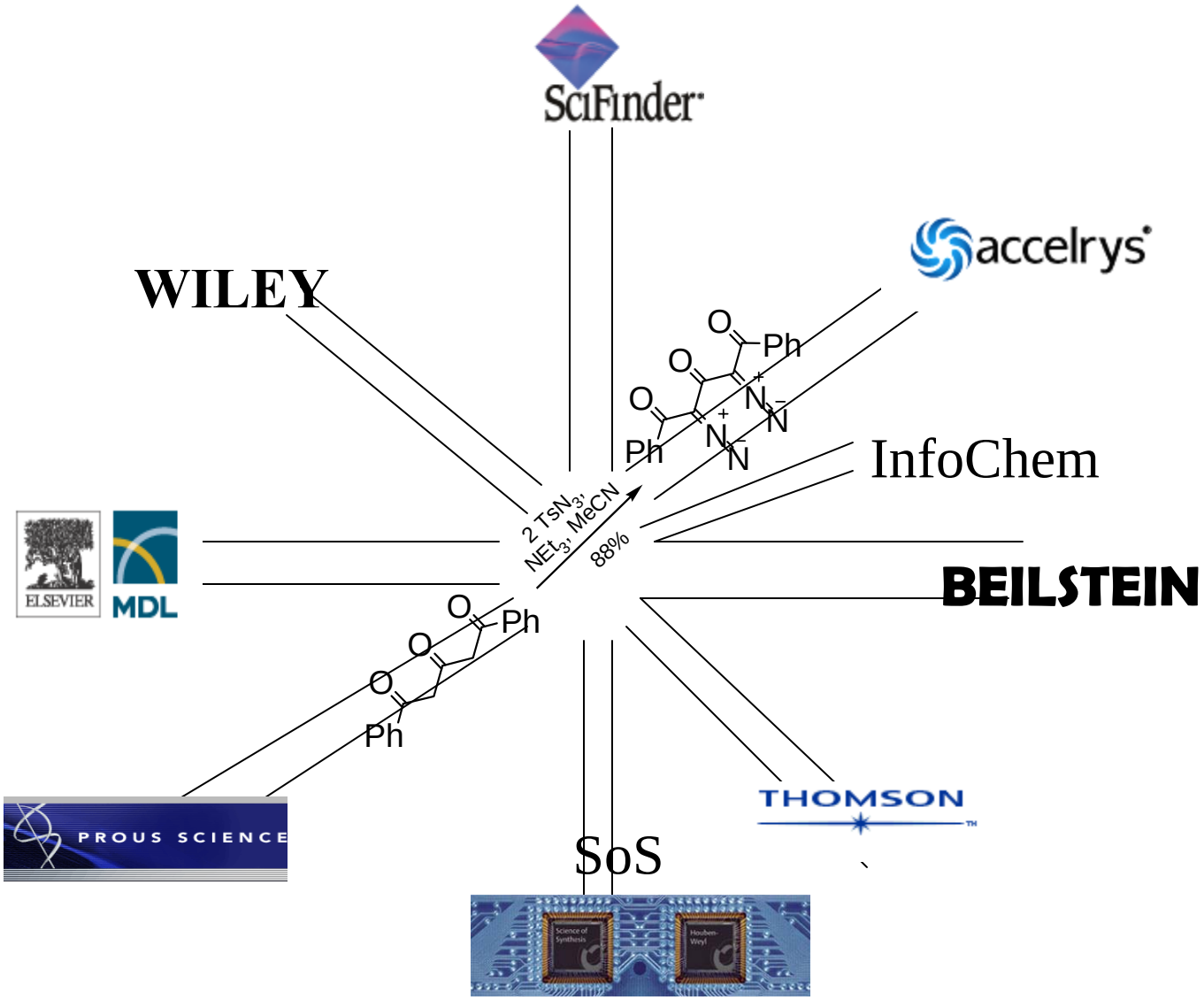
- **The data summary provided here on various RXn Databases/Resources is available in the public domain**

- **This presentation is for informational purposes only. This is not an endorsement of databases in any form.**

Why Search Reaction Databases?

- **New synthetic methodology for target molecule is a challenging task for the organic chemist**
- **Help to identify the best reagent/condition for chemical transformation**
- **Speed up the process development**

Reaction databases at the crossroads



Commercial Rxn Db's/Resources

- CAS/STN: CASREACT; SCIFINDER
- BEILSTEIN (STN & via XF)
- MDL/Synthetic Methodology - Discovery Gate
- WILEY INTERSCIENCE (OR, OS & e-EROS)
- ACCELRYS (MOS, SPS, FA, BIOCAT & BIOSTER..)
- THOMSON/DERWENT (DJSM)
- THOMSON – CCR/IC - Web of Science
- SCIENCE OF SYNTHESIS (Houben-Weyl)
- PROUS SCIENCE INTEGRITY, PDDR & Synthline

So Many Commercial Rxn DB's - Why?

- Explosive growth of publications
- Reaction data
- New research areas & specialization
- Technology to build large DB's
- Web based products
- Faster and better content delivery
- Platform for sharing third party databases

Why search more than one reaction database?

- There is no “one size fits all”
- Customer’s need varies
- Content, focus, coverage period varies
- Abstracting guidelines are different
- Multiple sources of reaction data
- Refine, import & export tools
- Integration of internal and external data

There is a need for One-Stop-Shop

Is there one resource?

Not yet!

Vendors - bundling resources on one platform

Organic Chemists - Expectations?

- One-stop-shop
- Fewer number of relevant & useful reactions
- Focus on synthetic methodology
- Structure or reaction searchable
- Combination of key words & numerical data
- Refine/analyze, import and export tools
- Link to literature citations
- Integrate with internal data

STN Platform

- **Reaction Searchable files**
  - **CASREACT - > 10.6 M rxns; 1840-; J/P; Weekly update**
    - **CAS – Journals, 1985 –; Patents 1991–**
    - **Infochem – rxns from VINITI (USSR) & ZIC (Berlin)**
      - **Journals 1974 – 1991; Patents 1982 – 1991**
    - **INPI – Core reactions: 1840 – 1985**
    - **Biotransformation database (1971 – 1998)**
  - **DJSMONLINE**
  - **CHEMINFORMRX (1991 – present)**
  - **PS-Pharmaceutical Substances**

- **SCIFINDER**
  - **Rxn info, ~ 10.6 M**

STN Platform (Cont'd)

Structure searchable but not directly reaction searchable

- **BEILSTEIN (Note: XF reaction searchable)**
- **New and useful preparative methods; Chem & Biochem**
- **> 10 M RXNS; 1779 – present; Quarterly update**
- **176 + journals, Patents**
- **Patent coverage ceased in 1979 (see MDL PC DB)**
- **Experimental values: physical, chemical, spectral, pharmacological, ecological properties**
- **Not all compounds in Beilstein have CAS RN's**
- **Structure-based searching offers a more powerful and flexible method**
- **CAS Registry/CAPLUS (use react/prep roles)**

Text searchable (STN cont'd)

PROUSDDR

- Drugs in development/marketed (13,596 records)
- Information on CN, GN, BN, CAS#, Phase #, originator, licensee, MoA, structure, patent ref.
- 1980 – present; updated monthly

SYNTHLINE

- >9,600 drug products
- >15,300 syntheses; > 68,900 reactants
- *Drugs of the Future* & patents
- 1984 to the present; updated monthly

MDL Synthetic Methodology

- Access *via* MDL DiscoveryGate
- CHEMINFORMRX Library (CIRX, > 1M rxns)
  - Current Synthetic Methodology (CSM)
  - MDL Reference Library (RefLib, > 210,000)
    - THEILHEIMER, CHIRAS, METALYSIS & CHC
- MDL Solid-Phase Organic Reactions (SPORE)
- Third party Db's (Derwent, Wiley, Thomson)*
- Integrated major reference works*
- BEILSTEIN*
- MDL Patent Chemistry Database (PCD)*

* Platform MDL DG

MDL Patent Chemistry Database on DG

- Is it a patent database or reaction database?
- It is “ONE-STOP-SHOP” for Reaction info, bioactivity data, patent info & chemical sources
- Integrated with other MDL DB’s on DiscoveryGate
- Retro-synthetic reaction pathways *via* MDL DG
- Synthesis drill-down & planning
- Reaction display (>12/2003), > 2.4 M reactions
- 15,000 Markush reactions
- View specific examples of a Markush family
- Links to & from similar reactions in major reference works
- Link from reaction to any MDL data source
- Integration of internal & third party reaction databases
- DG Platform allows to access > 17 M reactions.

MDL PCD on DG (cont'd)

- Not a patent database for prior-art search
- Patent coverage: US 1976 –; WO & EP 1978 – ; JP 2006 –
- Bridging the GAP - patent Lit - Beilstein vs MDL PC Db
- Org. Chem. & Life Sciences (pharma, agro, cosmetics)
- Experimental properties, bioactivity data (EC50 & LD50)
- Application, formulation, spectral & physical data
- Export to structure-activity-relationship tables
- Markush viewer - Markush structures (>12/2003)
- Claims text and Markush structure
- Compound location in the patent
- >1M prophetic compounds since 1976
- Substances >3 million
- >22,000 drug-target-binding constants
- 100 Million data – experimental values not calculated

MDL – Metabolites*

- Collection of xenobiotic transformations
- Comprised of drugs, agrochemicals, IC, EC
- Journals 60+; conferences (1990 - present)
- Metabolic schemes come from
  - *Pharmacokinetics* (1986-1990)
  - *Biotransformation von Arzneimitteln* (1977-83)
  - 1991 – Studies from New Drug Applications
- 78,000 transformations; 47,856 molecules
- 11,923 parent compounds; updated twice/year

* Not a reaction database

Wiley Interscience
Organic Reactions (OR)

- Rxns of current interest with prep. point of view
- Limiting factor and interfering influences
- Experimental conditions and reaction parameters
- Summary of all related published literature
- Coverage 1942 - present
- Currently consists of over 75,000 reactions
- Updated twice/year; 6,000 new reactions added/y

OR/OS/e-EROS: Searchable by reaction, structure, CAS registry number, reaction type, key words & numerical data

Wiley Interscience Organic Syntheses (OS)

- Authenticated preparative procedure
- Repeated, checked, re-checked
- Purification & Identification methods
- Spectral data & safety warning
- > 6000 reactions independently tested
- > 6400 molecules
- Coverage 1921 - present; Updated twice/year

Wiley Interscience e-EROS

- e-Encyclopedia of Reagents in Organic Synthesis
- Most suitable reagent for particular synthetic transformation
- > 50,000 reactions & > 3800 reagents
- Coverage 1900 - present
- Updated online twice a year.

"Product of the Chemists, by the Chemists, for the Chemists"

ACCELRYS
Methods in Organic Synthesis (MOS)

- Current awareness bulletin published by RSC
- Important new methods in organic synthesis
- Functional group interconversion and new reagents
- Abstracted from 100+ organic chemistry journals
- > 46,700 reactions
- 1991- present

ACCELRYS Solid Phase Synthesis (SPS)

- Focus on combinatorial and parallel synthesis
- Chemical rxns of substrate attached to solid supports (polystyrene, PEG, cellulose etc.)
- Covers primary literature, reviews and patents
- 1997 - present & quarterly update
- > 26,000 reactions

ACCELRYS - Failed Reactions (FA)

- Avoid repeating other's mistakes
- No reaction, did not proceed as expected
- Unexpected product (s) & side product (s)
- Suggests relevant ref. for right transformation
- Contains >13,500 reactions
- Semi - annual updates
- 2,000 reactions/year

ACCELRYS - InfoChem's (SPRESI)

- Academy of Science of the USSR (VINITI), Moscow & Central Information Processing for Chemistry (ZIC) in Berlin – since 1974
- InfoChem GmbH holds the rights on these data, 1990-
- Covers journals, books and patents
- Extracted from 600,000 references and 164,000 patents
- Provides direct access to over 3.7 million reactions 5.0 million structures, and 28 million factual data entries
- Property data includes boiling point, decomposition, density, dissociation, melting point, polymorphic transformation, refractive index, optical rotation, and sublimation.
- SPRESI - integrated into your corporate search database
- The data collection is offered as SDfile and RDfile.

ACCELRYS - Biocatalysts (BIOCAT)

- Developed in collaboration with Bryon Jones (U Toronto) and Herbert Holland (Brock U)
- Chemical synthesis using biocatalysts, enzymes and microorganisms
- Selectivity (chemo-, regio-, and enantio-)
- Environmental benefits
- Covers literature and patents from 1913 - present
- > 45,300 rxns; annual updates

ACCELRYS - BIOSTER*

- Dr. Istvan Ujvary, PPI, Hung. Acad. of Science
- Drugs, agrochemicals and enzyme inhibitors
- Search for bio-analog pairs, functional groups
- Larger structural fragment or moiety
- > 14,300 hypothetical transformations
- > 12,000 pairs of bioisosters
- > 36,000 biologically active molecules
- Journal articles, reviews and books
- Covers last 35 years with annual updates

* Not a reaction database

Thomson – Derwent (DJSM)
Derwent Journal of Synthetic Methods

- All new or improved synthetic methods
- Known processes effected by new reagents
- Novel functional group transformation
- Protecting groups
- Uses of organometallic and metal complexes in organic synthesis
- Expansion and adaptation of Theilheimer's
- > 180 journals & 41 patent authorities
- 1980 – present; > 87000 rxns; biannual updates

Thomson Scientific – CCR & IC

- Accessed through Web of Science
- New or improved preparation and rxn method
- 1985 – present; > 750,000 rxns; J/P
- Reactions specifically claimed from patents
- 1986 – 1996 only United States Patents
- Since 1997 - , from 40 major patent authorities
- DWPI examined for claimed rxn methods
- > 139,000 rxns from INPI; J/P; 1840 - 1985
- 45,000 rxns year added overall (J+P)
- 6,000 – 8,000 rxns/year from patents

Science of Synthesis – Houben-Weyl

- Synthetic methodology, scope, limitation & functional group compatibility
- Provides background information, history, nomenclature, structure, stability, reactivity
- Unique reference work – consisting of 48 volumes
- Interactive table of contents – extensively indexed
- > 1800 – present; covers journals, books, patents
- > 140,000 selected reactions
- Search fields: structure, role of structure in a reaction, catalyst, solvent, CAS #, bibliography, full text search, name reaction and R-group
- Crossover to other databases and e-journals
- Refine tools: temperature & yield (range, <, >, = etc.)
- Drawing plug-ins: Java applet, ISIS/draw & ChemDraw

Prous Science Integrity*

- Complete synthetic schemes for drugs in development and marketed drugs
  - > 14,354 schemes/end products
  - > 60,000 intermediates
  - 1990 - present
- Coverage: Journals >1500 & 11 patent authorities
- Literature and 300 conferences (0-2 weeks)
- US, EU & WO patents (24 hrs)
- Japanese patents (5 days)

* Not a reaction database but structure & text searchable

Some unmet needs in reaction searching

- **Batch mode - upload of structures, CAS RN#, names**
- **Recognize various representation of functional groups**
- **eg.,** N_3 **or –N=N+=N-; =N+=N- or** $=N_2$**;** NO_2 **or** SO_3**, N-O or S-O**
- **Multistep reactions A→B→C not searchable; A→B, B→C**
- **Text within tables and schemes are not searchable**
- **Use alpha or beta; can not use the symbol α or β**
- **Name reaction searching - by structure or name alone?**
  Hofmann reaction, Hofmann degradation, Hofmann-type, Hofmann rearrangement, Hofman elimination, Hofmann–Löffler–Freytag reaction
- **Author name index which covered all possibilities:**
  eg., Alan R Katritzky
  19 BEILSTEIN + 6 SCIFINDER + 5 MDL SM
- **Display all rxn conditions & reagents →**

Comparison of search results from SCIFINDER, BEILSTEIN & MDL SM

Transformation – Indole as reactant
Sub structure search (allowed substitutions & ring fusion)
Refine: Publication year = 2003
Author = Victor Snieckus
Name reaction Suzuki-Miyaura cross coupling

	MDL SM	BEILSTEIN	SCIFINDER
# References	2 [OL1899] [AGCI3399]	2 [OL1899] [OL3519]	1 [OL1899]
# Reactions	46	45	44

Found the unique reference in all three databases/resources

Comparison of hit answers from SCIFINDER, BEILSTEIN

Reactions of oxazole; Reactant = oxazole; sub structure search, free sites; Year 2003

	JHC	Hetero-cycles	Synthesis	TL	JOC	Total # of references
SCIFINDER	1	7	3	9	8	28
BEILSTEIN	0	5	5	10	7	27
# of ref. Both SF+B	-	5	2	9	5	-
# of add. ref.	-	2 SF	1 SF + 3 B	1 B	3SF+ 2 B	-

Take-home message

Novelty & Prior-art search

- Conduct searches in all available reaction resources & Db's
- Consult information professionals for additional resources

Comprehensive Search

BEILSTEIN & SCIFINDER complementary to each other

- **Combination of MDL SM + BEILSTEIN + PCD on DG**
- **Any specialized reaction database for specific subject area**