Reacting to Chemists’ Needs

Engelbert Zass
zass@chem.ethz.ch
Reacting to Chemists’ Needs

Reaction Information Sources, their Providers and Users

- Historical Developments
- State of the Art & Future Needs
Reaction Information Sources, their Providers and Users

A CHC as a Tribute to Günther Grethe
Neue Möglichkeiten zur Recherche von organisch-chemischen Reaktionen: Ein Vergleich der «in-house»-Datenbanksysteme REACCS, SYNLIB und ORAC

Engelbert Zass und Stefan Müller


1985:
in-house Reaction Databases at ETH

The parties hereby have caused their duly authorized representatives to execute this Agreement. This Agreement shall be effective as of the last date of signing below.

MOLECULAR DESIGN LIMITED

By: Stephen F. Mendel  
Executive V.P., its duly authorized representative

Date: 5/29/85

EIDENKÖSISCHEN TECHNISCHE HOCHSCHULE (ETH-ZENTRUM)

By: C. W. Full  
Principal Investigator

Date: 4/14/86

AND

By: G. B. C. Full  
authorized officer, its duly

Date: 4/14/86

Comparison of Reaction Searching in REACCS, SYNLIB and ORAC

<table>
<thead>
<tr>
<th>Version</th>
<th>REACCS</th>
<th>SYNLIB</th>
<th>ORAC</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1</td>
<td>2.2</td>
<td>6.3</td>
<td></td>
</tr>
<tr>
<td>Reactions</td>
<td>815</td>
<td>ca. 55</td>
<td>20</td>
</tr>
</tbody>
</table>

via Input of Reactions (Reactant, Product)

Identical reaction ("match") FIND CURRENT - REACTANT AND PRODUCT

Rection of same type RSS SUBS/RXN SUBSTRUCUTURE (SUB/CONS) (REACTION)

via Input of Substructures ("half reactions")

Reactant RSS SUBSTRUCUTURE

Product RSS SUBS SUBSTRUCUTURE

Reagent/Solvent etc. (RSS or .) - -

via Input of (exact) Structures

Reactant CURASAGT - REACTANT (current or reactant)

Product CURASPRD - PRODUCT

Reagent/Solvent etc. CURASAGT (current or catalyst/.as solvent)
Skolnik Award: Reactions

1981: B.H. Weil
1982: R. Fugmann
1984: M. Hyams
1986: D.B. Baker
1987: W. Theilheimer
1989: M.F. Lynch
1990: E. Meyer
1991: W.T. Wipke
1992: J.-E. Dubois
1993: P. Willett
1994: A.T. Balaban
1995: R. Luckenbach
1996: M. Randic
1997: J. Gasteiger
1998: G. Wiggins
1999: S. Kaback
2000: S.R. Heller
2001: G. Grethe
Historical Developments

- Traditional printed sources
- 70’s: Early databases
- 80’s: In-house systems
Traditional Printed Sources

- Reaktionen der organischen Synthese
- Comprehensive Organic Transformations
- Houben-Weyl Methoden der organischen Chemie
- Formation of C-C Bonds
- Compendium of Organic Synthetic Methods

1984: Reaction Searching at ETHZ
70’s: Early Databases

- Chemical Abstracts
  1972
- CRDS
  Chemical Reaction Documentation Service 1975
1975: CRDS

AN - 76167U RQ
KW - /ALCOHOL/ GIVES *ALCOHOL* *SIMPLE-REACTION* SYNTH-IC SOLV=10
      TEMP=5 PRESSURE=5 (CARBON-NITRIDE) RE (RHENIUM) CARBONYLATION
      HYDROGENATION,CAT CO COMPLEX,CARBONYL (COBALT-CARBONYL)
TI - /HOMOLOGIZATION OF ALCOHOLS/ CC-A-OC /
CI - US-411837 287308-E.

Homologation of alcohols

\[
\text{CH}_3\text{-CH}_2\text{-CH}_2\text{-OH} \rightarrow \text{CH}_3\text{CH}_2\text{CH}_2\text{-CH}_3
\]

p-Methoxybenzyl alcohol and dicobalt octacarbonyl in benzene heated
under an initial pressure of 295 atm of synthesis gas (2 H₂ : 1 CO),
whereby gas absorption starts at ca. 70°C and is essentially complete 25
min. later at 180°C \(\rightarrow\) 9-(p-methoxyphenyl)ethanol (X: 44%) and p-
methoxytoluene (X: 16%). F.e.s. I. Wender et al., Am.Soc.74, 4079
(1952).
80’s: In-house Databases

- SynLib

- REACSS

- ORAC
1974 / 1981: SynLib

ACS Award for Computers in Chemistry

W. CLARK STILL has been a dominating force in the perfection of many of the computational tools used today in solving problems in synthesizing organic compounds. In addition, points out a colleague, “his own experimental work has dramatically demonstrated the value of molecular modeling” in this area. Still is professor of chemistry at Columbia University.

Over the years, Still has developed a number of outstanding software systems. In 1974 he authored Synlib, the first computational retrieval system for reactions for synthetic organic chemistry. In an en-
1982: MDL REACCS

THE DATABASE
THEILHEIMER is the first computer-readable version of W. Theilheimer's Synthetic Methods of Organic Chemistry to display reactions using molecular structures. The database was designed to be read by Molecular Design's REACCS™ program. Using REACCS, you can search THEILHEIMER for reactions by simply drawing molecular structures on your terminal screen. Thus, you eliminate the need to learn new symbol schemes and to manually search through volumes of hard-copy indices.

THEILHEIMER contains nearly 35,000 reactions—all of those published from 1942 through 1981 and appearing in the original Theilheimer volumes 1 through 35. The database also includes full structures for approximately 70,000 molecules representing reactants, products, catalysts, and solvents.
Computer aids to synthesis planning

A. Peter Johnson

The rapid decrease in the cost of computer hardware allied to the development of user-friendly software has brought a number of useful tools within the reach of the synthetic organic chemist. Some currently available tools will help at the planning stage, while others will help in the detailed choice of procedure to accomplish particular conversions.
Example: Oximes -> Aziridines

\[
\begin{align*}
\text{Oximes:} & \quad \text{Aziridines} \\
\begin{array}{c}
\text{R} + \text{B} \rightarrow \text{C} \\
\text{NOROXIMINE} + \text{METHYL-METHYLAMINE} \rightarrow \text{1-aza-1,2,3,4-tetrahydrobenzene}
\end{array}
\end{align*}
\]

\[
\begin{align*}
\text{Procedure:} & \quad \begin{aligned}
\text{1 Reflux/THF/6C} \\
\text{2 reflux} \\
\text{3 H2O/E:20} \\
\end{aligned} \\
\text{Yield:} & \quad 67^\circ C, 45\%
\end{align*}
\]

Examples are given of the reduction of a variety of cyclohexenone oximes to give aziridines. The best yield is 56%.

Compiled by: S TURNER
Experimental: Given
State of the Art & Future Needs

- 90’s: Large databases
- 95’s: Usage-friendly systems
  Special databases
- 00’s: ???
1991: Reaction Searching at ETHZ

1) in-house reaction databases:
   REACCS, ORAC, SYNLIB substructures (roles)

2) public online databases:
   CA online (before 1967: CA CSI) CASRN + keywords
   Beilstein (up to 1960) BRN + role field + keywords
   CASREACT (since 1985) CASRN + role substructures + roles (mapping, sites)
   Registry / CA online substructure(s) + keywords/CASRN
Large Databases

- ISI CCR (1987)
- ChemReact (1991)
- ChemInform RX (1991)
- Beilstein
  - STN 1988
  - CrossFire plus Reactions (1996)
1991: ChemReact

---

**ChemReact**
- 370,000 reactions (REACCS)
- Display file of 1.8 million reactions
- Single-licence: 150,000
- Site-licence: 300,000

**ChemSynth**
- 80,000 reactions (REACCS)
- Display file of 1.8 million reactions
- Single-licence: 60,000
- Site-licence: 120,000

**ChemSelect**
- 10,000 reactions (REACCS)
- 10,000 reactions (CHEMBASE)
- Single-licence: 8,500
- Single-licence: 3,200

(50% discount for academic institutions)
(suggested prices in DM, excl. VAT, June 1992)

---

**Selective Reaction Database**

**Comprehensive Reaction Database**

**Reactions in 1st Literature**

- Intellectual
- Algorithmic

---

Springer New Media
October 5, 1991

Dr. Engelbert Zass
ETH Zürich
Laboratorium für Organische Chemie
ETH-Zentrum
CH-8092 Zürich
Switzerland

Dear Dr. Zass:

Recently Molecular Molecular Design Limited and Fachinformationszentrum Chemie GmbH (FIZ Chemie) have signed an agreement that gives MDL the exclusive rights for marketing electronic versions of ChemInform, a compendium of reaction information currently produced as a weekly journal by the German organization. This large (60,000 reactions/year), high-quality collection of reaction information will also serve as the source for a new 8000 reactions/year database. Current Synthetic Methodology (CSM) will replace the existing databases CLF, CHIRAS and METALYSIS. The first release of this semi-annual database will be in Fall 1992. To reflect the changes in your agreement with MDL I ask you to sign the enclosed copies of the Amendment to the Academic Software Agreement and return both copies to me. You will notice that we have also added a clause to the agreement requiring members of the Academic REACCS Project to provide MDL in December of each year with an annual report about the usage of the system at your site in research and teaching.

Please feel free to contact me if I can be of any assistance regarding the Academic REACCS Project. I am looking forward to a continued collaboration.

With best personal regards,

Quenter Grethe

cc.: Ms. Deborah Fournier
     Dr. Alex Kos
Usage-Friendly Systems

- MDL ISIS
- CrossFire plus Reactions
- CAS SciFinder (Scholar)
- ISI Reaction Center
- ......
Ease of Usage vs. Power of Search
Special Databases

Chemistry Databases

 Accord Database Explorer (ADE)
 BioCatalysis
 Bioster
 BioTransformations
 Failed Reactions
 Metabolism
 Methods in Organic Synthesis
 Protecting Groups
 Solid Phase Synthesis

Synthetic Methodology

ChemInform Reaction Library: Solution-phase synthetic methodology

Comprehensive Heterocyclic Chemistry (CHC)

CrossFire Beilstein: World's largest chemical facts database

Current Synthetic Methodology (CSM): New representative synthetic methodology

ORGSYN: Organic Syntheses

RX-JSM: Journal of Synthetic Methods

Reference Library of Synthetic Methodology (ReFLib): Collection of representative novel syntheses

Solid-Phase Organic Reactions (SPORE): Solid-phase synthetic methodology

THEILHEIMER: Synthetic Methods of Organic Chemistry

Future Needs: Current Deficits

- Strong **Bias** of current systems
  - Organic reactions
  - Preparative reactions
Organometallic & Inorganic Reactions
Future Needs: Current Deficits

- Insufficient Reaction “Qualifiers”
  - Reagents, Catalysts, Solvents
    - lack of standardization
    - lack of (sub)structure search
  - Reaction Conditions
    - insufficient data
    - insufficient search facilities
Future Needs: Current Deficits

- Multiple Reactions
  - Multi-Step Reactions
  - Reaction Sequences
CASREACT: Reaction Sequences

**STN INTERNATIONAL®**

CASREACT FILE SEARCH RESULTS - P280434V  06 OCT 88 12:38:18  PAGE  6

1/4 ANSWER 1 OF 2

RX(37) OF 89 COMPOSED OF RX(1), RX(2), RX(3), RX(4)
RX(1)  _A ====> B
RX(2)  E - B ====> UNKWN-F
RX(3)  _UNKWN-F ====> H
RX(4)  _H ====> ***1***

\[
\begin{align*}
\text{A} & \xrightarrow{(1)} \text{B} \\
& \xrightarrow{(2)} \text{UNKWN-F}
\end{align*}
\]

RX(1)  RCT A  84175-58-4
PRO B  '02147-67-1
SOL 7732-18-5 Water
RGT 7601-90-3 HOD4
RX(2)  RCT E  189-71-5, B 102147-67-1
PRO UNKWN-F
SOL 75-05-2 CH2D2
RX(3)  RCT UNKWN-F
PRO H 102210-27-5
SOL 7732-18-5 Water
RX(4)  RCT H 102210-27-5
PRO I  ***102210-24-6***
SOL 109-99-9 THF
RGT 665-47-4 t-BuOK
ChemInform Reaction Schemes
Where is Sandy Lawson’s Rabbit?


Multistep Reactions: The RABBIT Approach

ALEXANDER J. LAWSON* and HARTMUT KALLIES

Beilstein Institute, Varrentrappstrasse 40-42, D-6000 Frankfurt/Main 90, FRG

Received August 10, 1990

A simple concept (RABBIT) for the coding of multistep reactions is presented. A RABBIT (Random Access Black Box Indexing Term) code contains purely structural information and may be represented as a vector in two-dimensional space. Applications of the concept on a PC basis allow automatic structural checking to maintain chemical relevance in the tree search of the multistep path at query time. An example is discussed from an experimental implementation of the concept.
Future Needs: Current Deficits

- Insufficient **Post-Processing**
  “The larger the database, the smaller the toolset”
  - Viewing Reactions
  - Exporting Reactions
  - Clustering/Sorting Reactions
MDL ISIS Reaction Browser

Summary

RefLib 95.1.2

Cluster #  Item #  Value:

Variation


Literature Reference

Variation

A high-yielding, general procedure for acetalisation of carbonyl compounds
Ref.

Comments
The use of polyisyl diphenylphosphine iodine complex (polyisyl
dishenyl phosphorane iodide) means that no water is formed,
reflux procedures. The initially-formed adduct
carbohydrate attack at the carbonyl C atom with a molecule

Protecting Groups

Cluster #  Item #  Value:

PO  Ketone (dialkyl)  PO  Ethylene acetal

Lability conditions

(L-C6H4)3N5bC6(1 eq.)/MeCH20 C/1 h

(COOH)2{[Ag, 10%]//H2O/Silica/CH2Cl2/22 C
(MeCN)2PdCl2(0.05 eq.)/Acetone//H2O/20 C/45 h
(MeCN)3(triphenyl)Ru(0/02CFC3)/2/Acetone/20 C/4 h
(Ph)3-(CH2)3Br(Bu2)/N(Bu2)/2(Cat.)/CH3COCl(1 eq.)/No solvent/20 C/4 h
[n-Bu]2SmNC][2O(1 mol%)/(MeOH2CH2H2)/20(2 Parts)/H2O(1 Part)/200 C
2,4,6-Triphenylpyrydium.BF4(2 eq.)/h/H2O/20 C/3 h

AcOH(70%)/H2O/90 C/1 h
Problem with a common Functional Group Transformation

\[ \text{O} \quad \text{O} \quad \text{N} \rightarrow \text{HN} \]

- CASREACT
- CrossFire Beilstein 8066
- MDL ChemInform Reaction Library 926
- Chemistry Server 151
MDL Reaction Browser

[Diagram of a reaction browser interface]
“Cluster by Reagent”
Future Needs: Current Deficits

- **Strong Integration with**
  - 1º Literature *(vertical)*
  - Printed handbooks *(vertical)*
  - Other reaction databases *(horizontal, homogeneous)*
  - Other databases *(horizontal, heterogeneous)*
Lack of Integration

CrossFire ↔ ACD ↔ CIRX
Homogeneous Integration: MDL iMRW

Integrated Major Reference Works (iMRW)

Integrated Chemical Information - Triangle of Electronic Information

Reaction/Molecule Structural Databases
(MDL, Beilstein, Third Party, Compound Warehouse etc.)

Tertiary Sources
(Major Reference Works, Reviews etc.)

Primary Information
(Primary journals from multiple publishers, patents, proprietary reports, etc.)

Links: molecule identifiers, reaction class codes, citations, keywords etc.

Global Search Hit Statistics

- Encyclopedia of Reagents for Organic Synthesis (Wiley)
  - 134 Hits

- Comprehensive Organic Functional Group Transformations (Elsevier)
  - 52 Hits

- Comprehensive Asymmetric Catalysis (Springer)
  - 6 Hits

To go back to your query or navigate in the hitlists please don’t use the browser’s ‘back’ and ‘next’-buttons but the following link:

Back to query

© InfoChem GmbH 2000
Heterogeneous Integration: ISI Reaction Center

Search Results – Reaction Details

Hay AM, Hobbs Dewitt S, MacDonald AA, Ramage R. Use of tetrahydrofuran as an anchor antibacterial, ciprofloxacin.

Reaction 9 of 23

REACTION CONDITIONS

Citing Articles – Summary

Use of tetrahydrofuran as an anchor group for the solid/solution phase synthesis of the quinolone antibacterial, ciprofloxacin

Hay AM, Hobbs Dewitt S, MacDonald AA, et al.
SYNTHESIS STUTTGART
(11) 1979-1985 NOV 1999

Related Records – Summary

These documents in the database are related to parent record:

Hay AM. Use of tetrahydrofuran as an anchor group for the solid/solution phase synthesis of the quinolone antibacterial, ciprofloxacin

Recent advances in the preparation of heterocyclic on solid support: A review of the literature J CYMB: CHEM 2. (3) 195-214 MAY-JUN 2000