A User's View of Chemical Reaction Information Sources[†,‡]

ENGELBERT ZASS
Laboratorium für organische Chemie, ETH Zürich, CH-8092 Zürich, Switzerland

Received July 30, 1990

From a user's point of view, content and search facilities of publicly accessible databases such as Chemical Abstracts ONLINE, CASREACT, CRDS (Chemical Reactions Documentation Service), Beilstein Online, and of the in-house reaction database systems REACCS, ORAC, and SYNLIB are compared. Examples of reaction searches and their underlying search strategies and tactics are analyzed.

Although chemical reactions can be seen as the focal point of chemistry, filling most of the primary sources of chemical information, this is not clearly reflected in secondary sources which are dominated by *compound-oriented* abstract publications such as *Chemical Abstracts* (CA), or handbooks such as *Beilstein* or *Gmelin*. Sources specially devoted to chemical reactions, for example, *Theilheimer* or *Houben-Weyl*, are much smaller in volume and much less used by chemists. The appearance of online databases as a new medium and search tool did not at first change this circumstance, as the majority of these are based on printed secondary sources. We have had computerized keyword searching since the mid-seventies, substructure searching since the early eighties, but real reaction[1-3] searching only has been with us since the late eighties—with the notable exception of Derwent's Chemical Reactions Documentation Service (CRDS), available exclusively from the host ORBIT (now Maxwell Online) since 1975.[4]

An analysis of online searches in the chemistry library of the ETH Zürich seemed to indicate that the demand reflects this situation on the supply side: of the almost 1400 searches in 1989, only 2.3% were for specific reactions, and 2.1% for preparations of compounds, compared with about 20% each for author and keyword searches, and almost the same percentage for (sub)structure searching (5% via name, 14% via structure). However, this is only one side of the picture, as in a 3-month period last year, we had 650 accesses to the in-house system REACCS (Reaction Access System)[5-8] alone, with somewhat lesser usage for the comparable systems ORAC (Organic Reactions Accessed by Computer)[7-11] and SYNLIB (Synthesis Library)[7,8,12] also available at ETH. This gives a better impression of the need for reaction information in an organic chemistry laboratory and reflects several underlying factors:

- First, our usage policy; problems are submitted to the in-house system which incur no usage-dependent cost before they are tried in the more expensive online sources. Universities can get the in-house systems via academic programs where abstracting primary literature for the database is done in lieu of paying software and database license charges. For the public reaction databases CRDS and CASREACT, there is currently no academic program comparable to the one on the host STN for the Chemical Abstracts structure and literature files.

Second, ease of use (graphics-dominated, menu-driven systems vs basically text-oriented, command-driven systems) and ease of access for the individual chemist; personal computers connecting to the freely accessible in-house systems at the ETH computer center via the local area network already exist in many of our research laboratories and even in some undergraduate teaching laboratories. In contrast, online searching in public databases is normally done in the chemistry library under some kind of supervision.[13]

REACTION INFORMATION SOURCES

1.1. Printed Handbooks. In an "information scene" where most discussions are dominated by databases with their superior access, there are still the "classical" printed compendia such as *Theilheimer*; *Houben-Weyl*; *Organic Reactions*; *The Chemistry of Functional Groups*, edited by S. Patai; *Methodicum Chimicum*, by F. Korte; *Reaktionen der organischen Chemie*, by Krauch and Kunz about name reactions; *Compendium of Organic Synthetic Methods*, by the Harrisons and later editors; or R. C. Larock's recent *Comprehensive Organic Transformations*, to name only a few. We have no usage statistics for these, but we know from experience that they continue to be important, particularly for broad questions and orientational searches. The following discussion, however, will concentrate on reaction *databases*.

For a qualitative comparison and to get an impression of the indexing policies, it can be instructive to look at the different ways reaction information from the primary literature gets abstracted in several secondary sources. For obvious reasons the example given here cannot claim to be representative, but the conclusions drawn in the following are corroborated by several other examples we analyzed. A communication by Itô et al., Diastereoselective Synthesis of *N*-Acetyl-D,L-acosamine and *N*-Benzoyl-D,L-ristosamine,[14] in *Tetrahedron Letters* reports 14–16 reactions steps (the number depends on whether one counts intermediate reactions) in two half-page schemes, with no experimental details given.

1.2. Sources for Current Awareness. In *ChemInform* [presently only a printed publication produced by Fachinformationszentrum Chemie (Berlin) in cooperation with BAYER AG, but planned to be made available as a reaction database in the near future[15]], the essence of the primary publication is fairly reproduced in the form of an easily comprehensible reaction scheme plus bibliographic data and abstract (Figure 1).

The corresponding entry in another publication of the same type, the Institute for Scientific Information's *Current Chemical Reactions* (CCR), is shown in Figure 2. It gives a somewhat reduced view of the same paper, concentrating mainly on the key step, the intramolecular Michael addition of a urethane.

1.3. In-House Database Systems. CCR has been available since 1987 also as a database[2,16] (with more than 100 000 reactions during 1986–1989) in the REACCS system,[5,6] making it a useful tool for retrospective searches as well. The REACCS display of the same key step is reproduced in Figure 3. Besides the key step, three other reactions for each sequence from the two schemes in the original paper, plus the

[†] Based on a presentation at the Beilstein Institut Workshop on Computer Reaction Management in Organic Chemistry, Schloss Korb, Missian/Eppan nr. Bozen/Italy, May 29, 1990.

[‡] Dedicated to Prof. Albert Eschenmoser on the occasion of his 65th birthday.

0095-2338/90/1630-0360$02.50/0

Naturstoffe 1986

Kohlenhydrate
U 0500

8609-320

Diastereoselective Synthesis of N-Acetyl-D,L-acosamine and N-Benzoyl-D,L-ristosamine. — Das aus Ethylsorbat zugängliche Monoepoxid (I) dient als Ausgangsmaterial für die diastereoselektive Herstellung von N-Acetyl-D,L-acosamin (VI) und N-Benzoyl-D,L-ristosamin (X). (Zwischenstufen, IR-, NMR-Daten). — (HIRAMA*, M.; SHIGEMOTO, T.; YAMAZAKI, Y.; ITO, S.; Tetrahedron Lett. 26 (1985) 34, 4133—4136; Dep. Chem., Tohoku Univ., Sendai 980, Japan; engl.) — Ment

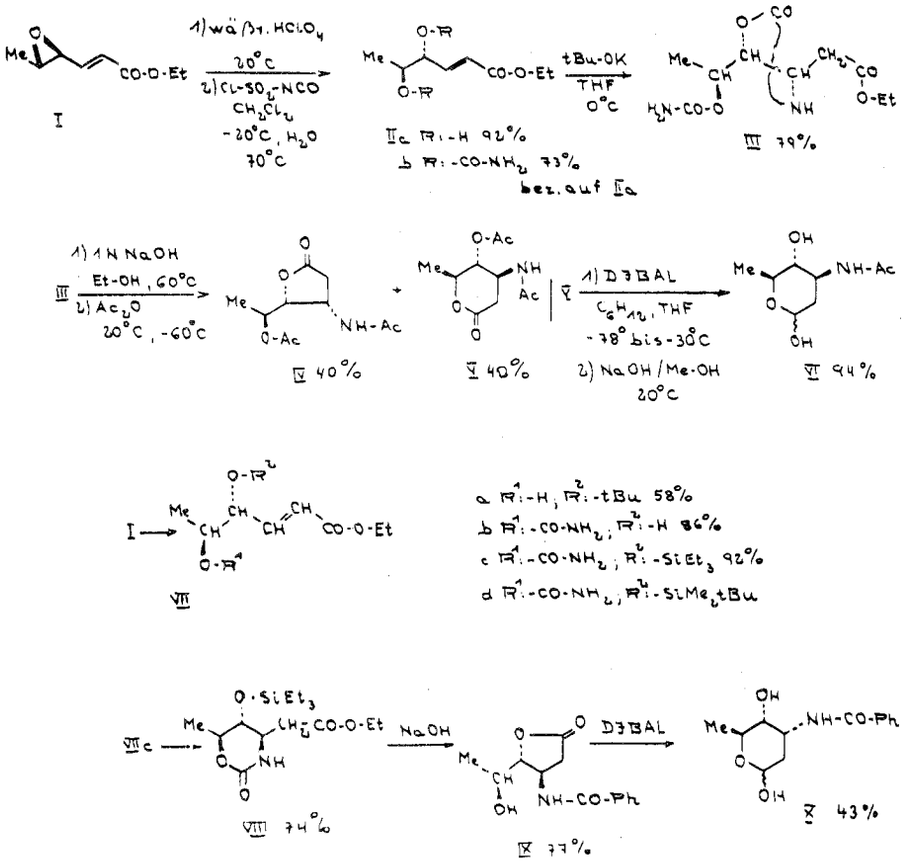


Figure 1. Abstract for the publication by Itô et al.[14] in *ChemInform* **1986**, *17*, 9. Reproduced by permission of Fachinformationszentrum Chemie GmbH. (Since 1987, abstracts are provided in English.)

030429

AMINO SUGARS

DIASTEREOSELECTIVE SYNTHESIS OF N-ACETYL-D,L-ACOSAMINE AND N-BENZOYL-D,L-RISTOSAMINE.
HIRAMA*M, SHIGEMOTO T, YAMAZAKI Y, ITO*S.
TOHOKU UNIV, DEPT CHEM, SENDAI 980, JAPAN.
TETRAHEDRON LETT 26(34),4133-6(1985).

N-Acyl derivatives of D,L-acosamine and D,L-ristosamine were synthesized with high stereoselectivity utilizing intramolecular Michael addition of γ- and δ-carbamoyloxy-α,β-unsaturated esters.

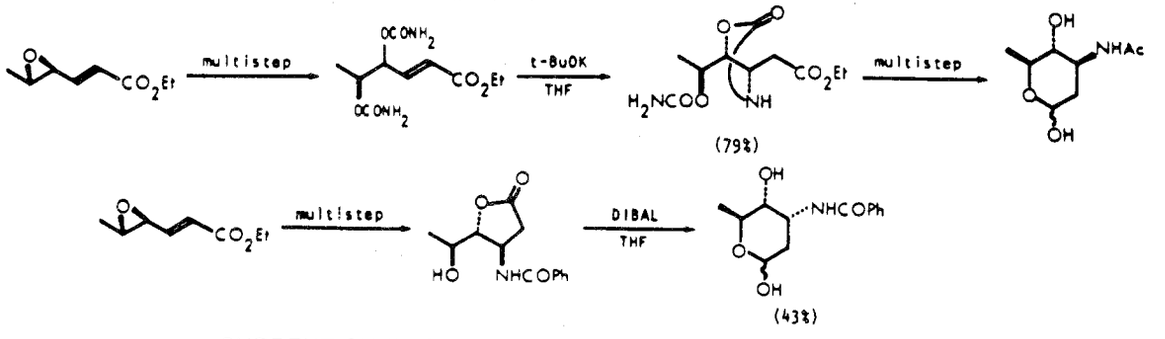


Figure 2. Abstract for the publication by Itô et al.[14] in *Curr. Chem. React.* **1986**, *8*, 87. Reproduced by permission of the Institute for Scientific Information.

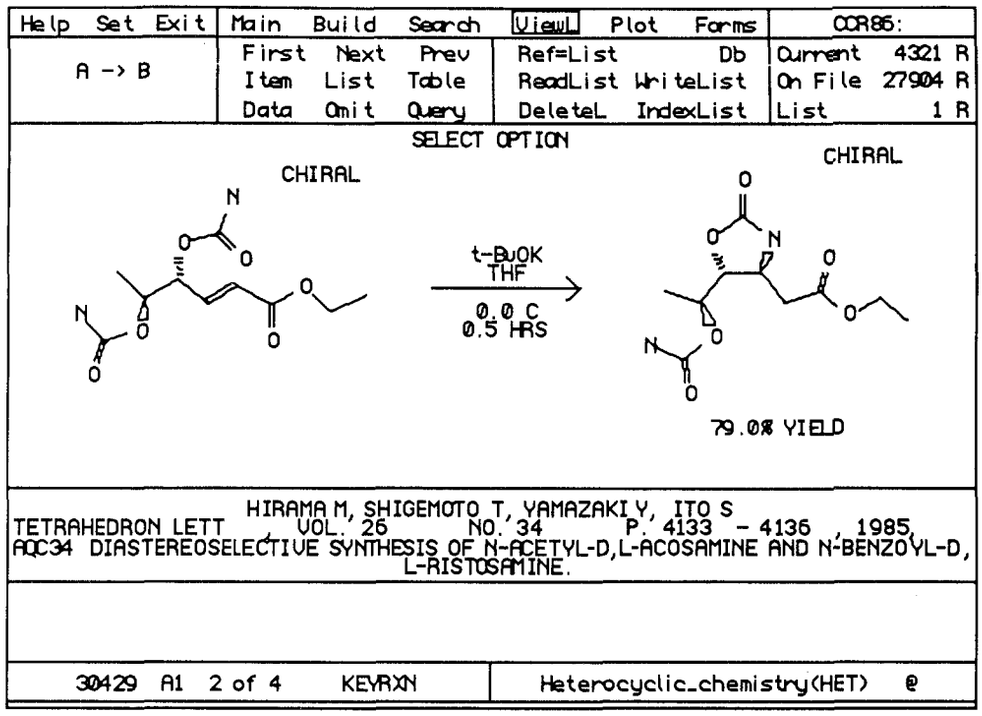


Figure 3. Display of key reaction step (KEYRXN, step 2 of 4 in sequence A1) from the publication by Itô et al.[14] in the REACCS CCR86 database.[21] Reproduced by permission of the Institute for Scientific Information and Molecular Design Limited.

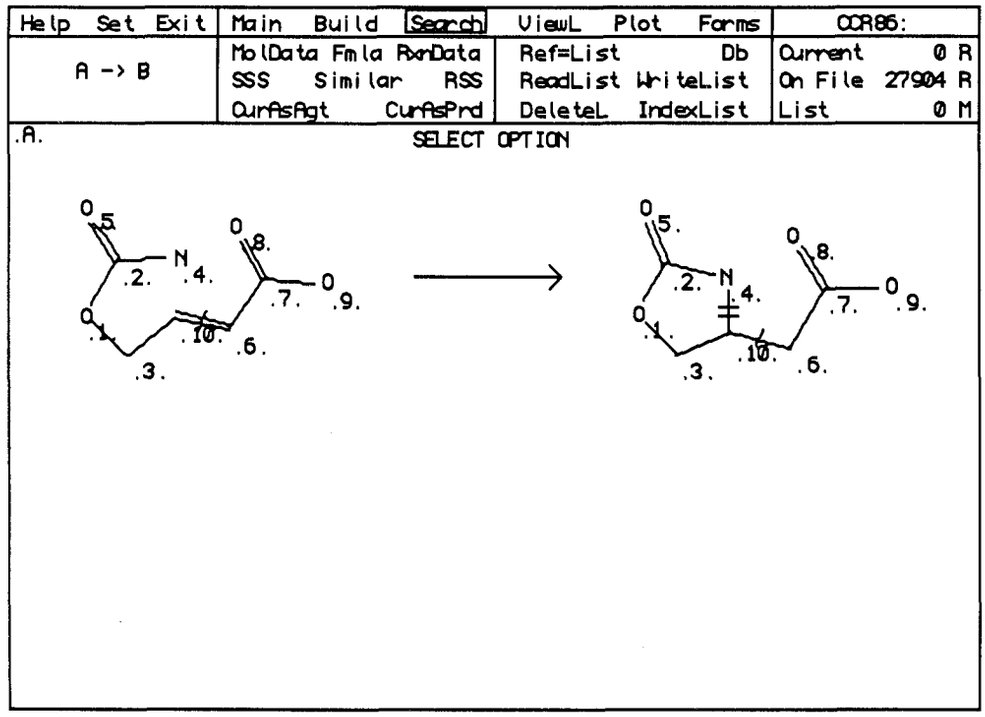


Figure 4. REACCS reaction substructure query (with mapping and bonds marked as reaction centers) for intramolecular Michael addition of a urethane.[21] Reproduced by permission of Molecular Design Limited.

two corresponding overall reactions are searchable and displayable in REACCS.[17]

Of the other in-house systems, ORAC (version 7.6 with 55 000 reactions)[9,10] was found to contain only the key step, and version 2.2 of SYNLIB (ca. 58 000 reactions)[12] did not abstract this or a similar reaction. The display of the key step in ORAC is conceptually similar in layout and content to that of REACCS in Figure 3 (cf. Figure 11); useful additional information about the presence of experimental details is given.[20]

Figure 4 shows one of several possible queries in REACCS for retrieving the key step of our example. Query input for all in-house systems is very similar, using a pointing device, usually a mouse, and menus for drawing. Reactions are searched mainly via a combination of (sub)structures with their roles in the reaction. For additional precision, corresponding atoms in reaction partners are mapped and reaction centers recognized automatically, with user intervention/correction possible. SYNLIB[7,8,12] is quite different in this respect; it is a "browsing" rather than a precise retrieval system. Product substructures are searched as such; any starting material substructure presently is only used to determine the qualitative

a

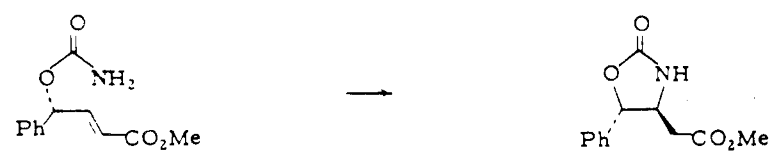
NC∩CC 77826A

H Potassium tert-butoxide
X

Intramolecular Michael addition
of ethyleneurethans
with asym. induction

2-Oxazolidones. Startg. 2-ethyleneurethan stirred with K-tert-butoxide in tetrahydrofuran for 30 min. at 0° under argon ⟶ product. Y 85%. (trans:cis 12:1). F.e. (66-79%) and bases (e.g. NaH), also tetrahydro-1,3-oxazin-2-ones, s.

M. Hirama, S. Itô et al.,
J.Am.Chem.Soc. 107, No.6, 1797-8 (1985); s.a.
Tetrahedron Letters 26, No.34, 4133-6; 4137-40 (1985).



b

```
AN  - 77826A HX
TI  - /INTRAMOLECULAR MICHAEL ADDITION OF ETHYLENEURETHANS WITH ASYM.
      INDUCTION .2. 2-OXAZOLIDONES . TETRAHYDRO-1,3-OXAZIN-2-ONES/
      NC-R-CC / /
AU  - HIRAMA M; ITO S ET AL
CI  - J.AM.CHEM.SOC. 107, NO.6, 1797-8 (1985); S.A. TETRAHEDRON LETTERS
      26, NO.34, 4133-6; 4137-40 (1985).
KW  - /URETHAN/ /ALLYL-O/ /OR/ /HOMOALLYL-O/ /ETHYLENE-C/ GIVE *ISOL*
      *DIHYDRO* *O-FUNCT* *URETHAN,CYCLIC* *OXAZOLE* OR *TETRAHYDRO*
      *OXAZINE-1,3* *AZACARBONYL-3* *(C-ESTER)* INTRAMOLECULAR
      (MICHAEL) 1,4-ADDN RING-CLOSURE,NC ALKOXIDE BASE=5
      (POTASSIUM-TERT-BUTOXIDE) SOLV=4 TEMP=3 OR BASE=6 HYDRIDE,ALKALI
MP  - *01* 019 02& 03& 032 111 122 126 13& 134 14- 150 153 23- 234 237
      242 243 247 26& 29& 29- 30& 302 33& 336 340 341 345 37- 373 375
      376 38- 392 45- 450 504 507 525 527 551 57- 589 599 601 624 641
      649 651 664 69& 697 704 74- 353 358 385
MP  - *01* 019 02& 03& 032 111 122 126 13& 134 14- 150 153 23- 234 237
      242 243 247 26& 29& 29- 30& 302 33& 336 340 341 345 37- 373 375
      376 38- 392 45- 450 504 507 525 527 551 57- 589 599 601 624 641
      649 651 664 69& 697 704 74- 353 358 385
MP  - *01* 019 02& 03& 032 111 122 126 13& 134 14- 150 153 23- 234 237
      242 243 247 26& 29& 29- 30& 302 33& 336 340 341 345 37- 373 375
      376 38- 392 45- 450 504 507 525 527 551 57- 589 599 601 624 641
      649 651 664 69& 697 704 74- 353 358 385
```

Figure 5. (a) Key reaction step from the publication by Itô et al.[14] in *J. Synth. Methods* **1985**, *11*, 12. Reproduced by permission of Derwent Publications Ltd. (b) Full display of CRDS database record for the key reaction step from the publication by Itô et al.[14] Reproduced by permission of Derwent Publications Ltd.

"transformation statistics" (bonds broken/formed, atoms added/eliminated).[40] With *similarity search* now available[10,22] in both REACCS and ORAC, we have a kind of browsing facility in these systems too.

1.4. CRDS. The *Journal of Synthetic Methods*[4] again abstracts only the key step (Figure 5a). This journal constitutes, together with *Theilheimer* Volumes 1–30, the CRDS reaction database. The corresponding record from this database is shown in Figure 5b, including the characteristic keyword and code indexing in the data fields KW and MP, respectively;[4] for the graphic information, the printed journal is indispensable, as CRDS has neither graphic input nor output. To retrieve such a reaction, it was encoded by four keywords and a (nonexhaustive) total of 5 reaction and 14 compound codes. Having been the only publicly available reaction database for more than 10 years,[23] CRDS' user interface (but definitely not its content) is outdated and not acceptable for end-users.

The *Journal of Synthetic Methods* is available as a REACCS database as well (REACCS–JSM, ca. 32 000 reactions since 1982), and *Theilheimer* is offered both by REACCS[19] and ORAC. While *Theilheimer* is part of the above-mentioned REACCS academic program, the relatively high license fee for REACCS–JSM[24] makes it not easily affordable for academic chemistry departments.

1.5. *Beilstein*. *Beilstein* and *Chemical Abstracts*, two of the most important chemical information sources, do also contain reaction information, although in terms of content and access points they are compound-oriented. Because Beilstein Online[3] only covers the literature until 1979, the same reaction

Table I: Comparison of Reaction Databases (August 1990)

	CRDS	CASREACT	REACCS	SYNLIB	ORAC
no. of reactions	(×1000) 80	960	234	63	117
coverage since	1942	1985	1942[a]	?	ca. 1960[a]
available since	1975	1988	1982	1983	1984
unit record	reaction	document	reaction	reaction	reaction
search control	commands	commands	menu/commands	menu	menu
query input	code	(graphics)[b]	graphics	graphics	graphics
search output	ref. only	reaction	reaction	reaction	reaction
search for[c]					
exact compounds	–	++	+	–	+
substructures	+	(+)[b]	++	+	+
stereochemistry	–	–	+	–	++
reaction centers	+	–	+	(+)	++
reaction conditions	(+)	(–)	+	(+)	+
reaction keywords	++	–	+	(+)	++
multistep reactions	–	++	+	–	–

[a] Oldest reactions found: 1869 (REACCS), 1881 (ORAC). [b] In CAS Registry File, not in CASREACT itself. [c] Rating: ++, very good; +, available; (+), limited; (–), very limited; –, not available.

example can of course not be shown in *Beilstein*. A "reaction search"[25] for the Michael addition of a urethane via product substructure, limiting the result with "PRE*paration*/F*ield* A*vailability*"[26] to those compounds that had their preparation described in *Beilstein*, turned out 13 such compounds on the host STN (6/18/90); a combination of these with the reaction/reactant keywords "(urethan? or carbamoyl?)/PRE*paration*.ED*uc*T" gave a zero result (cf. discussion below; "?" = STN truncation symbol).

1.6. *Chemical Abstracts*. In the printed *Chemical Abstracts*, the publication by Itô et al.[14] appears with the author abstract reproduced, but no structural formula added: "*N*-Acyl derivatives of D,L-acosamine and D,L-ristosamine were synthesized with high stereoselectivity utilizing intramolecular Michael addition of γ- and δ-carbamoyloxy-α,β-unsaturated esters." The contents of this paper were, according to the database record, indexed by four (issue index) keyword phrases, two *General Subject Index* entries, and 15 *Chemical Substance Index* entries. All but one of the compounds from the original schemes (virtually identical with the *ChemInform* scheme in Figure 1) were indexed; the missing one was only mentioned in a footnote of the paper, without information about its actual preparation. In addition, the primary starting material ethyl sorbate (mentioned only in the text of the paper, and reproduced neither in *ChemInform* nor CCR) and the reagent chlorosulfonyl isocyanate for urethane preparation were also indexed. Of the 15 compounds, 13 were *reaction*-indexed with "prepn", and the two remaining ones with "epoxidn" and "reaction", respectively. Eight more type-of-reaction keywords, among them "intramol. Michael addn.", were used.

Thus, there is indeed a lot of reaction information[27] in the compound-centered *Chemical Abstracts* and the corresponding database, but it is sometimes rather hard to find. Furthermore, the CA indexing policy[28] makes it far from complete: Reactants and intermediates have only been indexed systematically since 1973.[29,30] The situation is worse with reagents, catalysts, or solvents, as these are not at all routinely indexed, but only if they are new or applied in a novel way.[28,31]

On a more generic level, types of reactions or compound classes are only indexed in CA if they constitute the essence of a publication, or if several compounds of a class are treated.[27-29] In our example only the subject index headings "Michael reaction" (one of the approximately 40 name reactions that get their own heading)[27] and "Stereochemistry" were used. Consequently, with the probable exception of the key step, individual steps in syntheses will not be retrievable via keywords.

1.7. CASREACT. With these deficiencies of CA in reaction indexing on the one hand, and the need for broad, if not comprehensive, coverage of reaction information on the other, it seems quite obvious why Chemical Abstracts Service (CAS) embarked upon the extensive and expensive CASREACT project.[30,32] CASREACT is by far the largest and fastest growing of the databases discussed here (cf. Table I);[2] it is available exclusively via the host STN International. Figure 6a displays the Michael addition step from Itô's publication[14] in CASREACT.

While the majority of the *individual* steps and the overall reaction from our example can be searched and displayed in REACCS CCR (they are all linked via the same "external registry number"), only CASREACT makes *any* total or partial *recombination* in the entire sequence searchable and displayable. Figure 6b gives an example of that, with the Michael addition at the end of a longer sequence. The price to pay for this enhanced flexibility and searchability with the current explicit reaction indexing is redundancy—the 14–16 individual steps from the original publication are stored in CASREACT as 17 single-step and no less than 72 multistep sequences.[33]

This is one of the reasons why CASREACT needs a rather large variety of display formats—they are useful tools for the information specialist, but really bewildering for the occasional user. An online reaction display can easily extend over several screen pages; in this example, from 1 (Figure 6a) to 3.5 screen pages (of 33 lines length) if one displays online all intermediate steps from the longest pathway in Figure 6b (which shows a "condensed" version). This makes the use of a terminal or emulation program with scrollable graphics a must. The other databases discussed here restrict the display of reactions to one page having all the necessary information,[34] but they of course show individual steps or "compressed" multistep reactions only. In order to save cost and not be overwhelmed with information, we often display only the literature references from CASREACT after a graphic spot-check of the retrieved reactions. But the user wants to have all the information necessary to judge the relevance of a reaction on the screen, and not in the library!

Two important, interrelated aspects distinguish CASREACT (cf. Table I): it is not based on reactions as the unit record as all other reaction databases here, but on *documents* as is the CA database. Relative to the latter, however, it offers the important additional features of enhanced reaction indexing and graphic display. As a consequence, one cannot search for (sub)structures in CASREACT itself, but has to go to the STN Registry File instead. This is a complicating and sometimes limiting factor—functional group transformation reactions with small, common substructures cannot be searched for in the Registry File because of system limits (cf. example in 3.4). Fortunately, in a search for the key step from our

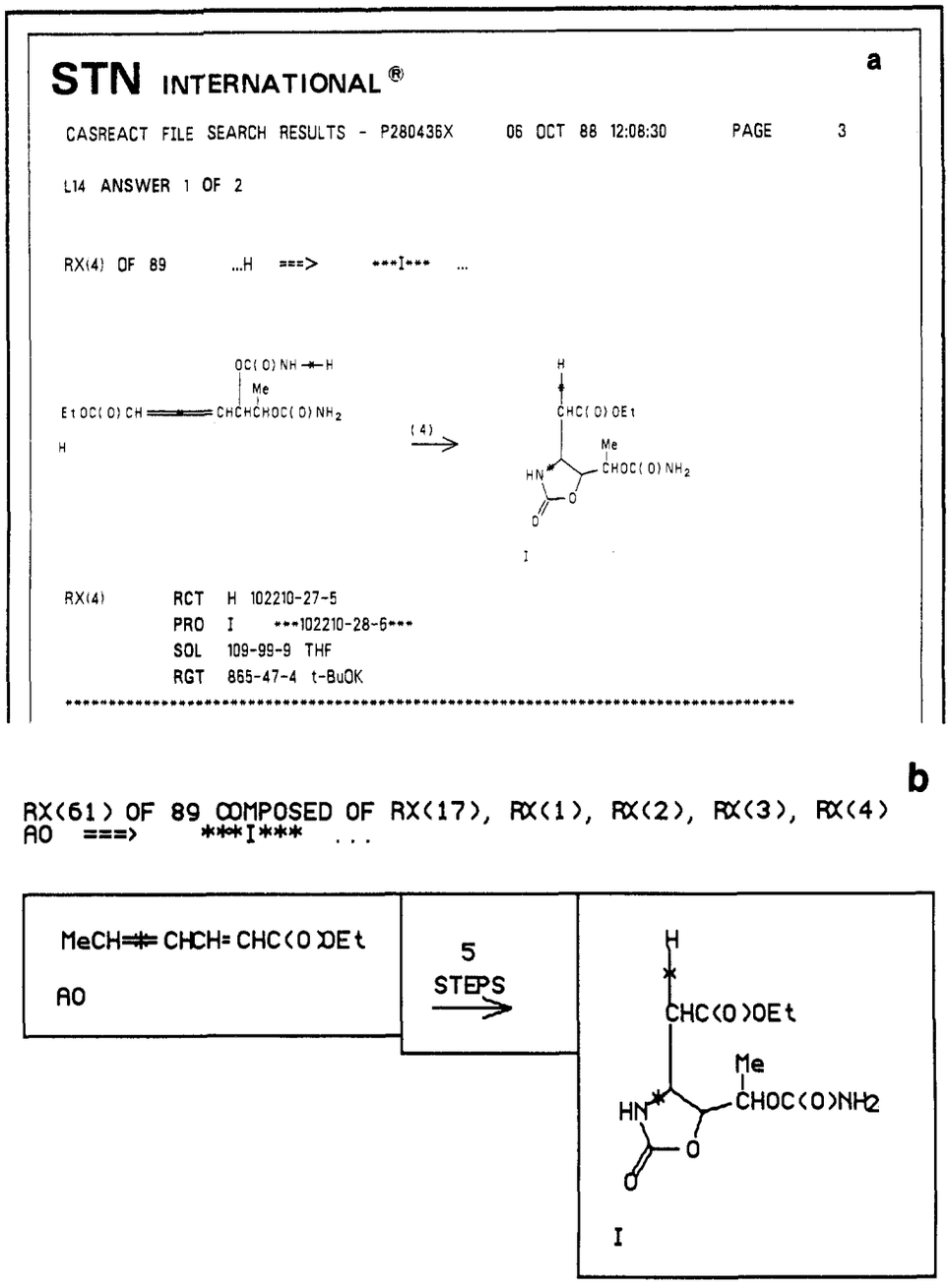


Figure 6. (a) *Offline* printout of key reaction step from the publication by Itô et al.[14] in the CASREACT database; display format FSPATH, i.e., "reaction map (letters representing reaction partners), diagram (structures), and summary (CAS Registry Numbers and roles); reaction centers marked with asterisks. Reproduced by permission of the American Chemical Society. (b) *Online* display of longest reaction sequence containing the key step from the publication by Itô et al.[14] in the CASREACT database (format PATH, i.e., reaction summaries for individual steps not included, intermediates omitted). Reproduced by permission of the American Chemical Society.

example, one substructure search was sufficient: following the tactics of *trying the least common substructure participating in a reaction first*, the 5-ring urethane product substructure gave only 8 compounds out of 9.3 million at that time in the STN Registry File; the five compounds of those present in CASREACT (identified by combining the primary result with the CASREACT locator field) gave two *documents* in CASREACT in a product role (⟨structure set no.⟩/PRO). The first publication was the one discussed here, and the other originated from the same research group. This is CASREACT reaction searching in a nutshell—using CAS Registry Numbers plus role indicators to retrieve reaction sequences with the defined participants of any length and regardless of intervening intermediates from the same document.

This presentation by example is supplemented in Table I with a summary of content and search facilities of reaction databases; the line is drawn here to exclude both CA and Beilstein as being mainly compound databases, albeit with reaction information.

USER NEEDS IN REACTION SEARCHING

After this sketch of the "supply side" of organic reaction information, let us now turn to user demands. Figure 7 shows a formalistic generalization of a reaction with the types of information involved. On the basis of that, the demand can be formulated in a simple fashion: we need to be able to search for any of the elements in Figure 7, alone or in any conceivable combination. The justification for this substantial demand is that in practice all kinds of questions occur, and the large amount of information already available must be matched by appropriate access points.

2.1. Database and System Requirements. More specifically, the following is required:

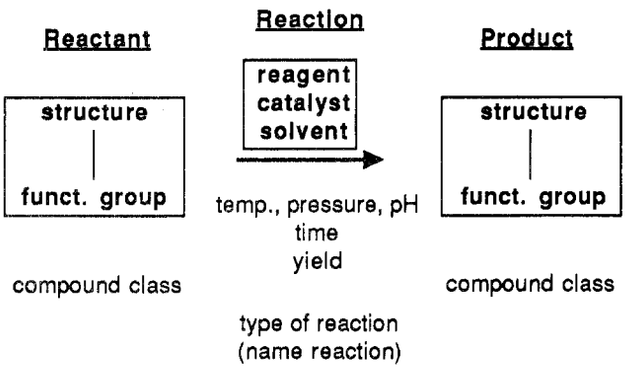


Figure 7. Generalized reaction description.

- systems must be highly interactive, with an appropriate user interface:
  menus for routine searches and occasional users, command shortcuts and enhanced search facilities for specialists
- graphical input (mouse) and output on standard terminals and emulation programs
- exact structure, substructure, and generic (compound class) searching for *all* reaction partners, including *stereochemistry* and *reaction centers*
  keyword searching for reaction types (thesaurus facility with cross-references/up-posting in hierarchy highly desirable)
  text/numeric (range) searching for data and reaction conditions
- entire database content searchable, individually or in any logical combination of queries and answer sets
- fast response times (subset searching for refinement, slow string search only acceptable for rarely used additional information, e.g., commentaries)
- easily comprehensible graphic display (one reaction/one page, reaction conditions on arrow between reactant and product), user-definable formats and tables for special purposes, export formats for text/graphics

In the author's opinion, REACCS and ORAC already fulfill most of these requirements; SYNLIB is inferior in some points to the other in-house systems;[40] the public online database CASREACT is still deficient in key issues such as specifying reaction centers (they are only marked with an asterisk in the display, see Figure 6a,b), stereochemistry, generic (keyword) search, and layout of display. CRDS and even more so *Beilstein* and *CA* cannot really be considered to be in the same class as the others, but nevertheless, users are not uncommonly forced to turn to them for reaction information, as shown in further examples below.

2.2. Reaction Coverage. CASREACT gives comprehensive coverage in a defined area, i.e., with predominantly *formal* selection criteria: virtually all reactions from the publications in about 110 important journals abstracted in the CA organic chemistry sections are included.[2,30] All the other reaction database producers do an *intellectual* selection within the primary sources covered to include representative reaction examples only, concentrating on key steps. Therefore, the number of reactions they contain is almost an order of magnitude smaller (cf. Table I). The difference is even larger with growth rates. For example, the *Journal of Synthetic Methods*, available as a database both via CRDS and REACCS, adds about 3000 reactions per year, while CASREACT grows by 200 000 reactions annually.

CASREACT and the information services (printed publications and databases) *ChemInform*, *Current Chemical Reactions* (CCR), and *Journal of Synthetic Methods*/CRDS cover together 306 (∝100%) periodicals; in the sequence given, individual totals (percentage of periodicals covered exclusively) are 110 (2.5%), 216 (26%, many chemical engineering journals), 126 (7%, mainly pharmaceutical journals), and 171 (13.5%, less-common journals and patents); 63 periodicals are covered by all four, 39 by any combination of three services (21 by *ChemInform*/CCR/*J. Synth. Methods*), 52 by two (23 by *ChemInform* and *J. Synth. Methods*), and 152 are exclusive to one service (80 to *ChemInform*).

In reality, both *comprehensive* and *representative/selective* sources are needed—one may want just a good example of the conversion of an oxime to an aziridine, or else all known examples of this transformation, probably in a restricted structural context. In the first instance, one would use the in-house systems or CRDS first, and only in case of unsatisfactory results refer to CASREACT. In the other case, one might still use the in-house systems as a starting point, but one would definitely need to turn to CASREACT and CA (for pre-1985 reactions, reactions from patents;[2] cf. Table III), perhaps even to Beilstein Online. A comparison of a keyword search in CA (host Data-Star) and CRDS for the oxime → aziridine reaction illustrates the principal difference in selection regardless of the difference in time coverage: for oxime + Grignard reagent CRDS 2/CA 17 references (1984), oxime + diazo compound 2/0, oxime + aluminum hydrides 3/18; of the 34 relevant references (out of a total of 44) in CA and the 7 (out of 12) in CRDS, only one was found in both databases. It is worth mentioning here that, with in-house system installations in the order of magnitude of a hundred worldwide, CRDS is—despite its unwieldy user interface—still the only public source of *selective* reaction information for a large number of chemists.

2.3. Generic Reaction Searches. The majority of reaction searches can best, or sometimes only, be expressed as (sub)structures; several examples of this category are discussed in the third part below. But relatively often one has also to answer questions like "electrochemical reduction of ketones", "[4+2] cycloaddition to enamines", or "conversion of esters into tertiary amines", which are more appropriately or more easily described by keywords then by substructures. At first glance, CA looks like the preferable source by the sheer wealth of information and precise indexing policy. Because of the limitations already mentioned, this is not always true. Keyword reaction searches in CA are only orientational in nature, with a good degree of precision if appropriate headings exist, but with a rather low recall and often quite demanding search profile construction.[35]

CASREACT at this time lacks generic search keys for compound classes or reaction types which would be desirable both for questions that can hardly be phrased otherwise and for functional group transformations that surpass the limits of the Registry File. The relatively scarce information in the *note field* (NTE; with the remarkable inclusion of failed reactions) or the useful feature to look for any catalyzed reaction simply are not enough.

Among the in-house reaction databases, REACCS offers searching by 28 controlled, partially computer-assigned reaction-type keywords, but SYNLIB offers nothing really useful of this kind.[40] The best possibilities with extensive controlled generic indexing[11] for reaction type, reagents, and solvents (the latter with a hierarchical thesaurus) are realized in ORAC. Thesaurus help is available, with cross-referencing of synonyms, e.g., HMPT/HMPA, and the entire indexing of individual reactions ("keys card") is displayable. For the Michael addition step from Itô's publication[14] discussed in the first part, the following indexing terms were used: nucleophilic, addition, 1,4-addition, intramolecular, stereoselective ("reaction keys"),

Michael addition ("actual reaction"), base, and alkoxide ("reagent keys").

2.4. Reaction Data Searches. ORAC and, to an even greater extent, REACCS allow searching for yield, temperature, and other reaction conditions. In these systems, such data are regularly indexed and stored in appropriate fields, while SYNLIB permits only a rather slow string search for them. Unfortunately, CASREACT lacks this kind of information almost completely,[2] with the exception of a few data in the *note field* (not standardized, searchable only in the *basic index*), and the yield, belatedly introduced in October 1986. Searches for reaction conditions are very useful, either to specify necessary constraints or to refine a search with a large number of answers.

Sometimes, one is looking for variations on a theme, i.e., the same structural or functional transformation with a variety of reagents or conditions. A user-definable *tabular display* with reagents, solvents, temperature, yield, reference, etc. is a standard feature in REACCS only. CASREACT would be by the very amount of information and the more comprehensive selection policy an ideal source for this kind of question, but it contains almost no reaction condition information besides yield, and the present formats are unsuitable for an acceptable display in such cases. A *role-specific* SELECT command would be very useful here. One could "extract" reagents and/or solvents from a large set of similar reactions and display a list of them (with graphics SET OFF) in the Registry File.

REACTION SEARCH EXAMPLES

3.1. Source Selection. On the *strategic level* ("where to search"), source selection is the most important decision, desired coverage the first criterion: representative examples, or as comprehensive a collection as possible. Generally speaking, selection is determined by availability (content, access) and feasibility (accessibility, cost). In a real world, of course, availability means the user's actual *knowledge* about content and access of reaction information sources and her/his *familiarity* with them. Feasibility is defined by the user's individual *perception* of accessibility and cost, strongly influenced by the usage policy of the institution he works with. At the ETH Zürich chemistry department, the in-house systems REACCS, ORAC, and SYNLIB are not only some of our most powerful sources of reaction information, but also definitely the cheapest. Although our students are therefore encouraged to try them routinely first, some problems are normally searched preferentially or even exclusively in CA because of its comprehensiveness. Cost considerations, potential problems with substructure searching in the Registry File that has to preceed most of the reaction searches, and display problems are the major reasons why we presently do not allow graduate students to access CASREACT by themselves—in contrast to the in-house systems, we regard this as a resource for advanced searchers, only to be used when those other sources have given no satisfactory result.

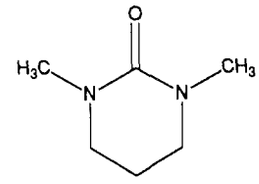
3.2. Preparation. A significant part of reaction searching is looking for information on the preparation of individual compounds or a substructure-defined group of them (cf. introduction). In terms of coverage *Chemical Abstracts*, with *Beilstein* for the older literature, is the source of choice for such problems, and the *tactics* ("how to search") seem to be straightforward. This is true within our limited experience with *Beilstein*, but not so with the *P* algorithm in CA. Table II shows disquieting discrepancies among the different hosts offering this feature; they all use their own (unpublished) algorithms with obviously quite different success. In the admittedly extreme case of the nitrate ion, the majority of "preparations" in STN dealt either with "formation" in the atmosphere, etc. or with "leaching" in soil. This problem is obviously not amenable to a simple algorithmic solution. Intellectual assignment would be definitely the best way to do it, although this is expensive. An expert system might be the second-best solution; cooperation of the hosts with CA in a joint development looks good from a user's point of view, somewhat naive as it may be with the known political and economic problems involved. The role definition in CASREACT is more precise, but since 1985 only the following preparations were found (6/18/90): Meldrum's acid, 2 (online display for one of them covered six screens in format FSPATH; cf. 1.7); adenine, 3; nitrate ion, 0.

Table II: The *P* Algorithm in Searching for Preparations (8/30/89)[a]

Data-Star	DIALOG	Questel	STN
	Meldrum's acid [2033-24-1]		
23 (13)	14 (13)	12 (11)	13 (12)
	Adenine [73-24-5]		
231	171	342	167
	Nitrate ion [14797-55-8]		
123	124	431	276

[a] References found with [Registry Number]*P*: total (relevant).

Table III: Reactions with DMPU as Solvent (3/2/89)



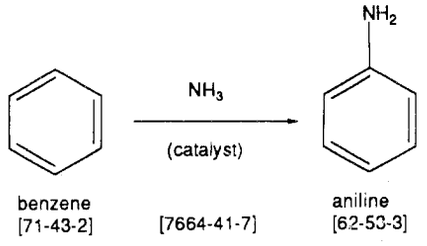
CASREACT	21 references total
	6/SOL*vent*[26]
	10/R*eaGen*T
	1/CAT*alyst*
	17 relevant
	15 exclusive in CASREACT
CA	49 references total
	(24 references before 1985)
	25 references after 1985
	11 relevant
	9 exclusive in CA
	1 very new (not yet in CASREACT)
	8 patents (not covered by CASREACT)

3.3. CA vs CASREACT. With respect to the indexing policy of CA (cf. 1.6), looking for reactions by reagent or even more so by solvent seems a clear-cut case for CASREACT. As the example in Table III illustrates, this is not exactly true. Despite all its deficiencies, CA is needed for *comprehensive* searches not only on all pre-1985 reactions—backward extension of CASREACT would imply complete reanalysis of documents by CAS which seems hardly feasible—but on reactions from patents as well, as these are presently not covered by CASREACT.[2] Inclusion of reaction examples from patents in CASREACT could be in line with current selection policies and would be a significant improvement.

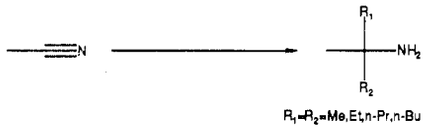
REACCS and ORAC were searched by structure for reactions with DMPU as solvent: 0 (out of 55000) in our present version of ORAC 7.6, 2 reactions in the REACCS Current Literature File.[18] For that particular problem, the REACCS Metalysis File (ca. 5000 metal-mediated reactions since 1974)[6] would be expected to give useful results, but is not available to us.

In Table IV, CA indexing for the amination of benzene is analyzed. All 14 relevant publications, among them 11 patents, were published before 1979, and no such reaction was found in ORAC, REACCS, Beilstein Online, or (obviously) CASREACT. This example again underlines the necessity to include CA as a reaction information source. From the

Table IV: Amination of Benzene (5/22/90)



Chemical Abstracts indexing (STN CA File)	refs
registry no. (product)/*P*	14
+reaction type: amination	4
+reactant trivial name: benzene	11
+reagent trivial name: ammonia	8
registry no. (reactant)	13
+reaction type: amination	4
+product trivial name: aniline	2
+reagent trivial name: ammonia	8
registry no. (reagent)	9
+reaction type: amination	1
+product trivial name: aniline	2
+reactant trivial name: benzene	7
	total 14



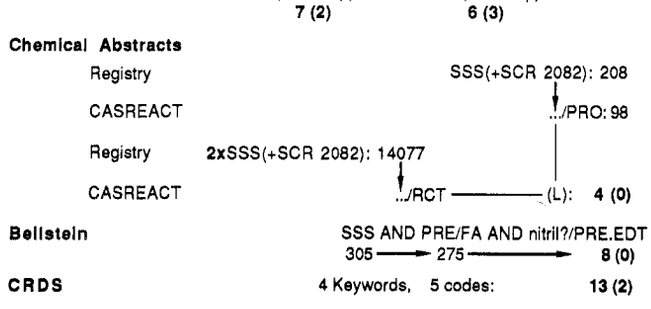


Figure 8. Reaction search for conversion of nitriles into tertiary amines (see text for explanations).

indexing, generally applicable search tactics using a combination of CAS Registry Numbers linked with trivial names and/or reaction-type keywords, depending on the desired precision and recall, can be deduced. Analyzed examples such as this and the following ones are used for teaching chemical information retrieval at the ETH Zürich.[13,36]

3.4. Substructure Reaction Searching. Figure 8 illustrates the search tactics for a typical functional group transformation. Such searches are straightforward in REACCS and ORAC: input of the query directly as given in Figure 8, then Reaction Substructure Search (RSS) and Reaction Search, respectively, both with automatic mapping and reaction center recognition. In Figures 8 and 13, the total number of reactions (for CASREACT and CA, documents!) retrieved are given, followed by the relevant ones in parentheses. The number of irrelevant reactions for the in-house system was higher than usual; this could have been improved by specifying substituents R^1,R^2 which were omitted in the interest of broad retrieval here. Individual database[2] results for REACCS were as follows: Current Literature File[18] total 1 (relevant 0); *Theilheimer*[19] 3 (1), REACCS–JSM (only Vols. 8–10, 1982–1984, 11 500 reactions) 2 (1), *Organic Syntheses* (Vols. 1–67, 1921–1988, 5000 reactions) 0, CCR86 (Vol. 7, 1986, 27 900 reactions) 1 (0). In CRDS, again a rather broadly coded query, used in order not to miss interesting reactions, found three highly relevant reactions. Only one of the seven relevant references found in all databases shown in Figure 8 appeared twice, namely in ORAC and CRDS.

The rather broad approach used for the smaller databases is not feasible in the CAS files. Following a standard tactic, the presumably least common reaction partner, in this case the tertiary amine, was tried first in a substructure search (SSS, Figure 8) in the STN Registry File. Limiting with the CASREACT *screen* (SCR) *2082*, which restricts the substructure search to only those compounds that occur in CASREACT, is essential here to get below system limits. The intermediate CASREACT result (98 documents) obviously needed further qualification by searching for the reactant side as well. For the common nitrile substructure, even use of *screen 2082* is not sufficient to get below system limits, so two online (or one batch) substructure searches are needed. This clearly illustrates the problems of having to use the Registry File prior to a reaction search proper in CASREACT. Apart from search limits, the cost of the preceeding substructure search(es)—often two per reaction—contributes significantly to the total search cost. This problem is somewhat alleviated by the reduced search charge introduced in January 1990 ($69/DM 133 as of July 1990) for a substructure search restricted by *screen 2082*. The STN (L) operator in the following steps, linking product and reactant role-qualified structure sets, assures that only reactions in the same sequence, and not anywhere else in the same document, are retrieved; none of the reactions in the four documents found were relevant (Figure 8; 7/19/89). This illustrates a second problem with CASREACT: lack of explicit or implicit reaction center specification is bound to lead to false hits like those above which contain the required functional groups, but not in the desired reactivity context.

A search of that transformation in CA was not deemed to be feasible because of lack of precision, either with a product substructure ("preparation") or with a keyword approach. In Beilstein Online,[3,25] however, a similar tactic worked in principle (Figure 8): product substructure search (potentially less problematic in that still significantly smaller structure database), restriction with "PRE*paration*/F*ield*A*vailability*"[26] and truncated reactant (PRE*paration*.ED*uc*T) name fragment search gave eight product records. These were printed out in the default format, which gives the substance identification information (structure, name, printed *Beilstein* citation, etc.) as well as the hit fields; inspection of the PRE field showed that none of them was relevant (7/19/89).

As *Beilstein* indexes preparations (under product name and structure) as well as reactions (under name and structure of the organic reactant), a reaction search could in principle start from either side. The majority of the reaction searches we have done so far in Beilstein Online have not been successful (except preparations). The real limitation lies in the fact that product and reactant records are not directly linked. In every one of the two approaches, the link exists only between a structure with an *implicit* role (structure with preparation reported = product, one with reaction reported = reactant) and a nonstandardized, sometimes incomplete name of the other reaction participant. In addition, this name information is missing for the period 1960–1979, i.e., the material not yet processed for the printed handbook. DIALOG announced recently (6/1/90) that 498 000 Beilstein Registry Numbers identifying starting materials were added to 642 000 records reporting preparations, thereby establishing a precise reaction link for 22% of the 2 939 000 Beilstein Online records that contain preparation information (total number of compounds on 6/18/90: 3 417 000). STN is following suit, and there are more ambitious plans to produce a special Beilstein reaction database.[37]

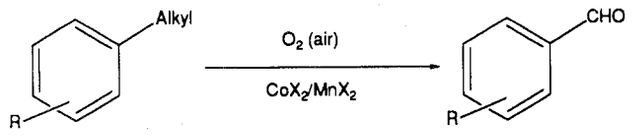


Figure 9.

In another search (2/5/89), examples for the conversion of a benzylic alkyl group to an aldehyde function with air in the presence of certain cobalt(II) and manganese(II) salts as co-catalysts were desired (Figure 9). Search tactics in the CAS databases should consider cost, too. This problem was therefore attacked with an approach that not only avoided the problematic, very common reactant/product substructures but also provided the presumably cheapest way to the most specific element in the query, the metal salts. After a STN Registry File "dictionary search" with name and molecular formula fragments, 409 Mn and 835 Co salts were crossed over into CASREACT and linked with the Registry Number for oxygen, using both catalyst and reagent roles for all three reaction participants to give three documents, all with relevant reactions.

In REACCS and ORAC, however, queries with reactant and product substructures were searched and combined with oxygen as reagent: "formula=H0O2 as agent" in REACCS, structure "O=O" as Actual Reagent in ORAC. There are several possibilities, particularly in REACCS, for the formulation of that query. As in the preceeding example, a rather general approach was used; Co/Mn salts were not specified. The browsing effect implicit in this tactic is desired; scanning the results or narrowing them down with a second search using the result subset from the first is fast and easy. We found that inexperienced users tend to overspecify, obtaining a more precise result, but probably often missing related information that can be of interest to them. In this example, ORAC gave no reactions even with the broad query, and the 10 found in REACCS were not relevant in a strict sense, as the desired structural transformation retrieved did not involve Co or Mn salts.

Figure 10 shows the ORAC query input (Reaction Search mode, before mapping) for selective reduction of cyclohexane β-ketoesters to the corresponding ketoalcohols, an example for a common class of problems, *selective* functional group transformation in a partially or completely defined structure environment. The only reaction retrieved in this search is displayed in Figure 11; the same one was also found in CA and CASREACT (see Figure 12a) as well as in CRDS. The other relevant reference from this latter database, and one each from the REACCS CLF[18] and SYNLIB (out of two retrieved) were not duplicated in any other database searched. Different reagents thus were found: KH or LDA/AlH_3, LDA/$LiAlH_4$, $NaBH_4$. A search in Beilstein Online was not successful.

Prior to a search (2/2/90) in CASREACT for this problem, both reactant and product substructure searches in the Registry File again were necessary. In contrast to the first example, *screen 2082* was not added, because there were no search limit problems here. Another reason for using this screen is the reduced search charge mentioned above, but with a search in CA assumed to be necessary (and later justified by the result!), this was not feasible here. By qualifying the compound sets with the respective roles and linking them, three (two relevant) documents were retrieved. The first part of a reaction display is reproduced in Figure 12a; the upper (OCC*urrence*)[26] part shows hit and display statistics, the lower part (FSPATH) the first of *five* relevant *hit reactions* from that *one document*. The nonrelevant document again had a reaction with correct structures but undesired reaction centers (Figure 12b). These are marked with asterisks, but are not searchable. "Hit highlighted" by three asterisks on each side are the letters in the "reaction map" (Figure 12a,b) and the CAS Registry Numbers in the "reaction summary" (Figure 12a only; cf also Figure 6a,b) representing the structures responsible for retrieval of a particular reaction.

In the STN Ca File, a first search profile "⟨product structure set⟩/P AND ⟨reactant structure set⟩" found 37 references, including among many nonrelevant ones all three retrieved in CASREACT before. Narrowing down in CA by linking [STN (L) operator;Registry Number and keyword in the same index entry] the reactant structure set with the reaction-type keywords "redn OR reducti?" produced five

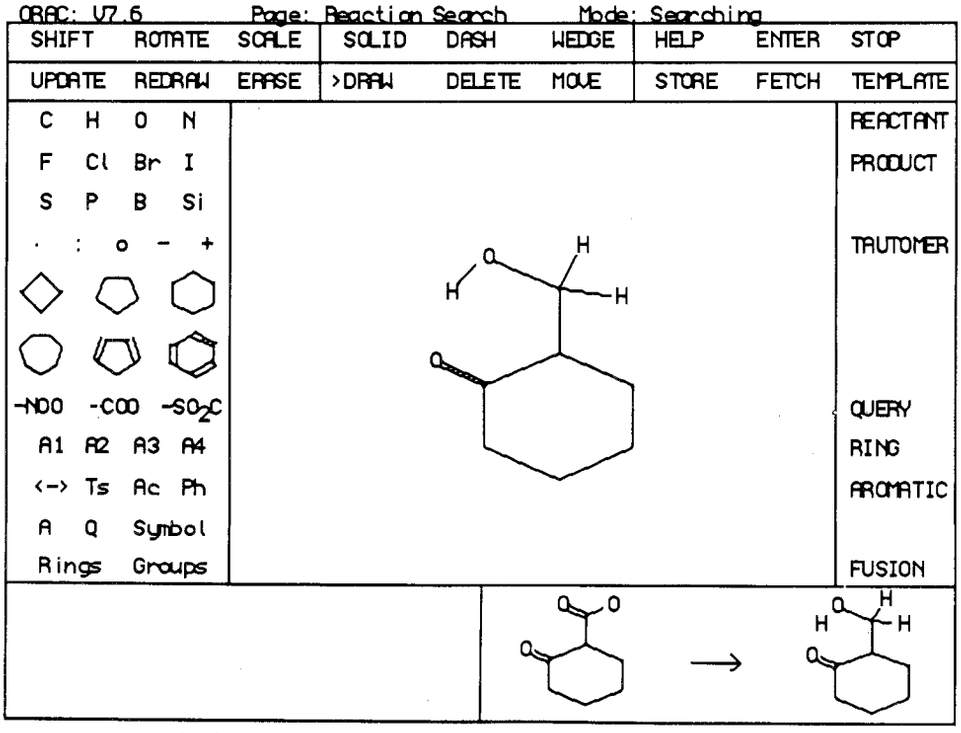


Figure 10. ORAC query input for selective reduction of β-ketoesters. Reproduced by permission of ORAC Ltd.[21]

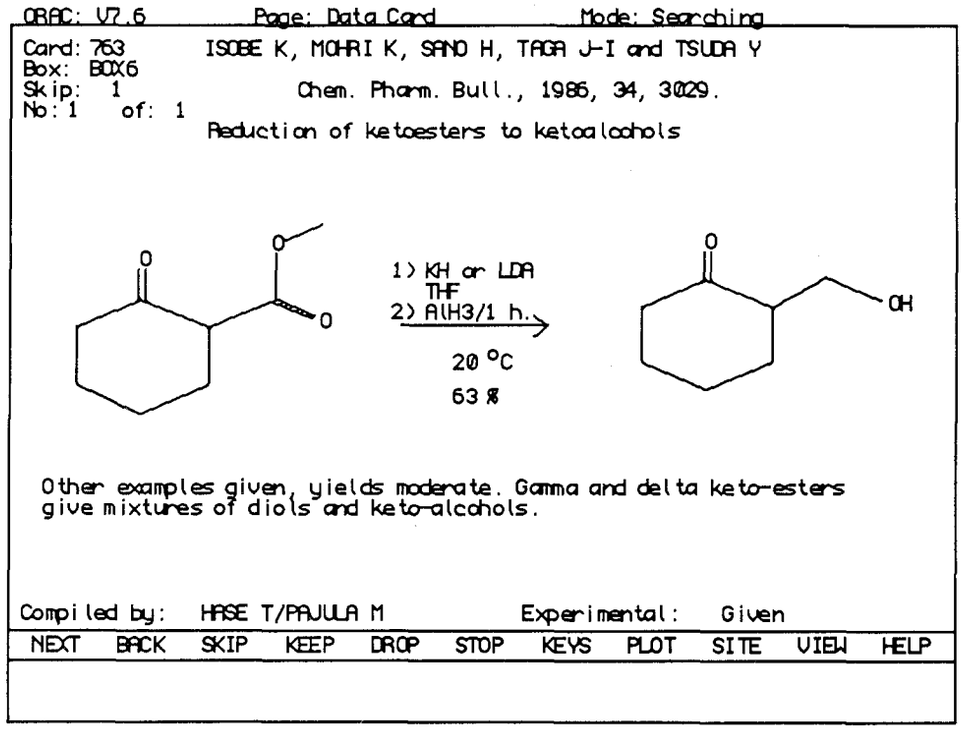


Figure 11. Selective reduction of β-ketoester retrieved in ORAC. Reproduced by permission of ORAC Ltd.[21]

references, two relevant (one also found in CASREACT). As already discussed in the first part, most of the reaction information concerning products and reactants is also in CA, but only CASREACT makes it comfortably (graphic display), precisely (roles), and reliably (easy query formulation) accessible.

In Figure 13, a search strategy for a ring contraction is shown. Because of expected problems with reaction center recognition, a Reaction Substructure Search in REACCS and a Substructure Search in ORAC for reactant and product were used, both *without* mapping. *With automatic mapping*, RSS in REACCS gave the same result, while a Reaction Search (mapping obligatory) in ORAC gave *no* answer. The mapping algorithms gave the following atom correspondences and reaction centers, given here with the numbering for reactant/product in Figure 13: 1/1, 2/2, 3/3, 4/4, reaction center *bonds* 2–7, 4–5 in reactant, 2–5, 4–5 in product for REACCS; in ORAC, 1/1, 2/2, 3/3, 4/4, 5/7, *atoms* 4, 5, 6, 7 in reactant and 4, 5 in product marked with an asterisk as reaction centers.

Searches in both CA and CASREACT used similar tactics[38] (Figure 13; 10/30/89, structure sets from the Registry File limited with CASREACT/*LoCator*[26] before crossover). Five references were found altogether: the first was common to REACCS, ORAC, and CRDS, but not found in CASREACT or CA. It was missed in CA because β-lactones as products, although mentioned in the title, were not indexed in any way for this 1969 publication. A second reference was common to CASREACT and CA. The other three were exclusively in one database only, a patent in CA, a publication indexed for a derivative (Registry Number with appended "D") in CASREACT, and a different patent in CRDS.

CONCLUSIONS

Several deficiencies and desired improvements have been discussed above in the content of search examples. A concluding qualitative, by necessity somewhat subjective, ranking of databases by both reaction content (coverage) and access (search facilities) is given in Figure 14. The arrows indicate the general direction in which the database producers should move in the author's opinion: for the in-house-systems, one would like to see more of same, that is, more reactions with slightly (more than that only in SYNLIB) improved access; for CASREACT, content and searchability should be about equally improved, as this database is (in our concept) assigned the role of the comprehensive reaction source. The need for reaction-center searching was illustrated by two of the examples. Accidentally, only the very first example included stereochemistry, although it was not necessary here to specify it in a search. Stereodifferentiation in searching will definitely become more important in the near future. This is a convenient feature to enhance precision even in small databases, particularly when one is looking at a file like REACCS' Chiras with presently ca. 10 000 asymmetric reactions.[6] In ORAC, one can search for both relative and absolute configuration of (sub)structures, while REACCS at present only allows the first possibility (the second is planned for version 8). Large, fast growing databases such as CASREACT are even more in need of such features to avoid large numbers of false hits that are both expensive and annoying to end-users. CA should of course keep covering the entire field of chemistry and will continue to be important for pre-1985 reactions and those from patents unless CASREACT will index them. Within this mission, it is probably not realistic to expect greatly enhanced reaction search facilities in CA; exhaustive reaction keyword indexing, for example, might destroy some of the very qualities we like about CA, as the amount of information would increase dramatically and unacceptably, at least for the printed indexes. But some improvements, such as better standardization of reaction keywords in the now virtually uncontrolled index modification, or role flags, perhaps even some reaction links between indexed compounds for documents not covered by CASREACT, might be both technically and economically feasible. Beilstein should preferentially improve access to the wealth of reaction information it contains—roles are already there, but the structure links are not (yet; cf. above). For CRDS, it is really the user interface that must be brought up to date.

Today, reaction databases are still comparatively less powerful and comprehensive than structure databases, and despite remarkable progress in the last few years, the search facilities

a

L9 ANSWER 3 OF 3

NUMBER OF HIT REACTIONS	5
NUMBER OF REACTIONS IN PATH	5
NUMBER OF REACTIONS IN SPATH	5

FIELD	COUNT
RX(24)	2
RX(25)	2
RX(26)	2
RX(29)	2
RX(34)	2

RX(24) OF 34 COMPOSED OF RX(1), RX(2)

A ===> ***E***

RX(1)	RCT	A ***1655-07-8***
	PRO	UNKWN-B
	SOL	109-99-9 THF
	RGT	7693-26-7 KH
RX(2)	RCT	UNKWN-B
	PRO	E ***5331-08-8***
	RGT	7784-21-6 Aluminum hydride (AlH_3)

RX(82) OF 300 COMPOSED OF RX(19), RX(20), RX(21), RX(8) **b**

... ***AX*** ===> ***Y*** ...

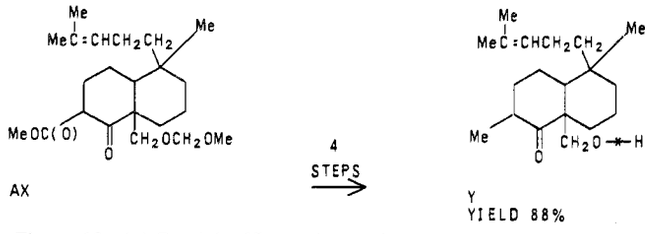


Figure 12. (a) Partial *offline* printout for the selective reduction of β-ketoesters to ketoalcohols from the CASREACT database (display format OCC FSPATH; see text). Reproduced by permission of the American Chemical Society. (b) Partial *offline* printout ("reaction map and diagram") of the *false hit* for the selective reduction of β-ketoesters to ketoalcohols from the CASREACT database (cf. text). Reproduced by permission of the American Chemical Society.

leave several important things to be desired. More completely standardized user interfaces[39] are a prerequisite for increased end-user searching. We do not yet have enough experience with similarity searching in REACCS and ORAC[10,22] to decide on the usefulness and general desirability of this feature. Across-document retrieval of reaction sequences is now being discussed;[37] this sounds promising indeed, but one would first like to see improved accessibility of the reaction information already present within documents.

ACKNOWLEDGMENT

We are indebted to Molecular Design Limited (REACCS), ORAC Ltd. (ORAC), Distributed Chemical Graphics (SYNLIB), and Chemical Abstracts Service (CA structure and literature files) for their respective academic programs, as without them this study would hardly have been possible. Financial support for online searching by the Eidgenössische Drucksachen-und Materialzentrale, Bern, reduced rates for *Journal of Synthetic Methods*/CRDS provided by Derwent Publications Ltd., and a test installation of CCR86 from the Institute for Scientific Information are gratefully acknowledged. I thank Prof. Claude Wintner (presently on sabbatical leave from Haverford College, PA) for his valuable advice on English style and grammar.

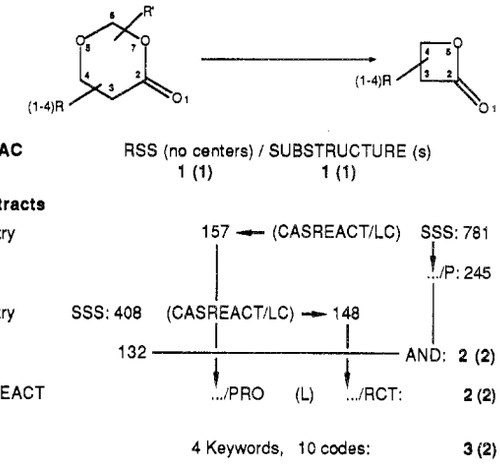


Figure 13. Reaction search for rearrangement of substituted dioxanones to β-lactones.

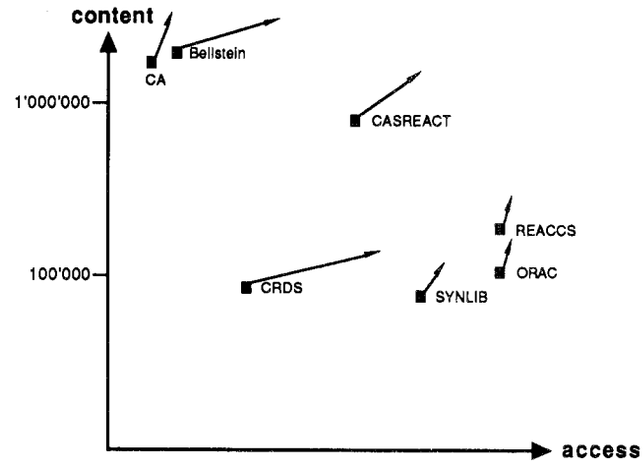


Figure 14. Qualitative comparison of reaction information sources.

REFERENCES AND NOTES

(1) Willett, P., Ed.; *Modern Approaches to Chemical Reaction Searching*; Gower: Aldershot, U.K., 1986.
(2) Dana, R. C. Where do they all come from? Appropriate Coverage of the Literature for a Chemical Reactions Database. *Online Inf.* [*Proc. Online Inf. Meet.*] **1989**, *13*, 129–138.
(3) Barth, A. Status and Future Developments of Reaction Databases and Online Retrieval Systems. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 000.
(4) Finch, A. F. The Chemical Reactions Documentation Service: Ten Years On. *J. Chem. Inf. Comput. Sci.* **1986**, *26*, 17–22.
(5) Developed by Molecular Design Limited, 2132 Farallon Dr., San Leandro, CA 94577. Kasparek, S. V. *Computer Graphics and Chemical Structures*; Wiley: New York, 1990; Part 4. Kos, A. J.; Grethe, G. Reaktionsdatenbanken-Werkzeuge für den Synthese-Chemiker. *Nachr. Chem., Tech. Lab.* **1987**, *35*, 586–594.
(6) McHale, P. J. Reaction Databases for Use with REACCS. *Online Inf.* [*Proc. Online Inf. Meet.*] **1989**, *13*, 155–160.
(7) Zass, E.; Müller, S. Neue Möglichkeiten zur Recherche von organisch-chemischen Reaktionen: Ein Vergleich der 'in-house'-Datenbanksysteme REACCS, SYNLIB und ORAC. *Chimia* **1986**, *40*, 38–50.
(8) Borkent, J. H.; Oukes, F.; Noordik, J. H. Chemical Reaction Searching Compared in REACCS, SYNLIB, and ORAC. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 148–150.

(9) Distributed by ORAC Ltd., 18 Blenheim Terrace, Woodhouse Ln., Leeds LS2 9HD, U.K. Johnson, A. P. Computer aids to synthesis planning. *Chem. Br.* **1985**, *21* (1), 59–67.
(10) Hopkinson, G. A. Recent Developments in Reaction Searching. *Online Inf.* [*Proc. Online Inf. Meet.*] **1989**, *13*, 139–145.
(11) Johnson, A. P.; Cook, A. P. In *Modern Approaches to Chemical Reaction Searching*; Willett, P., Ed.; Gower: Aldershot, U.K., 1986; pp 184–201.
(12) Distributed Chemical Graphics, 10788 Lockart Rd., Philadelphia, PA 19116. Chodosh, D. F. In *Modern Approaches to Chemical Reaction Searching*; Willett, P., Ed.; Gower: Aldershot, U.K., 1986; pp 118–145.
(13) Zass, E. Online-Datenbank-Recherchen an der Hochschule. Chimia **1987**, *41*, 96–103. Zass, E. Chemie-Information an der Hochschule. *Mitteilungsbl.—Ges. Dtsch. Chem., Fachgruppe Chem.-Inf.* **1989**, *16*, 41–54.
(14) Hirama, M.; Shigemoto, T.; Yamazaki, Y.; Itô, S. Diastereoselective Synthesis of *N*-Acetyl-D,L-acosamine and *N*-Benzoyl-D,L-ristosamine. *Tetrahedron Lett.* **1985**, *26*, 4133–4136.
(15) Gasteiger, J.; Weiske, C. ChemInform—An Integrated Information System on Chemical Reactions. *Online Inf.* [*Proc. Online Inf. Meet.*] **1989**, *13*, 147–154. Parlow, A.; Weiske, C.; Gasteiger, J. CHEMINFORM—An Integrated Information System on Chemical Reactions. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 000.
(16) Garfield, E. International Collaboration Helps Establish ISI Reactions Club: Current Chemical Reactions In-House Database Is Catalyst to Research Creativity. *Curr. Contents* **1987**, *13*, 3–7.
(17) Among the in-house databases, CCR has an unusually high percentage of *multistep* reaction sequences: of the reaction steps in the REACCS[6] CCR/Current Literature File[18]/Theilheimer[19] databases, for example, 34/45/62% are single step only, 1.1/0.4/<0.1% five-step sequences. ORAC and SYNLIB databases presently do not contain (explicit) multistep sequences.
(18) REACCS Current Literature File (CLF), produced by Molecular Design Limited; version 89.2 with 28 450 reactions from the primary literature 1983–1989.
(19) REACCS Theilheimer database, produced by Derwent Publications Ltd. and Molecular Design Limited;[6] corresponding to printed Vols. 1–35, 1946–1981, 46 800 reactions.
(20) There is an error in the ORAC display in the configuration of the asymmetric center formed in the Michael addition, probably due to conversion of the somewhat unusual drawing of the product in the original publication (which is reproduced without changes in both *ChemInform* and CCR, see Figures 1 and 2) to a more conventional representation.
(21) Apple LaserWriter printout of a graphics screen dump saved as MacPaint (bitmap) document in the Tektronix 4105 emulation mode of the VersaTerm PRO 2.2 terminal emulation program for Apple Macintosh.
(22) Grethe, G. Similarity Searching in REACCS. A New Tool for the Synthetic Chemist. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 000.
(23) The even older, powerful, but unfortunately not publicly available GREMAS system with ca. 4 million reactions cannot be discussed here; cf. Fugmann, R.; Kusemann, G.; Winter, J. H. The Supply of Information on Chemical Reactions in the IDC System. *Inf. Process. Manage.* **1979**, *15*, 303–323. Fricke, C.; Fugmann, R.; Kusemann, G.; Nickelsen, I.; Ploss, G.; Winter, J. H. In *Modern Approaches to Chemical Reaction Searching*; Willett, P., Ed.; Gower: Aldershot, U.K., 1986; pp 68–77. Fricke, C.; Nickelsen, I.; Fugmann, R.; Sander, J. GREDIA: A New Access to GREMAS Databases. *Tetrahedron Comput. Methodol.* **1989**, *2*, 167–175.
(24) Approximately Sfr. 70 000 for the REACCS-JSM license + ca. Sfr. 10 000 annually, minus a 30% reduction for universities; for comparison, online access to CRDS requires an annual subscription fee of £785 (keyword access to database and *J. Synth. Methods*; academic discount −60%, −75% for printed journal only), plus £445 for access to coding; for the actual online searches, connect time on the host ORBIT/Maxwell Online is $110/h, plus $0.30 per printed citation.
(25) Cf. *Online Searching on STN. Beilstein. How to Find Preparations and Reactions*; Springer: Berlin.
(26) Capital letters: Role/data field abbreviation, italics: explanation.
(27) Beach, A. J.; Dabek, H. F., Jr.; Hosansky, N. L. Chemical Reactions Information Retrieval from Chemical Abstracts Service Publications and Services. *J. Chem. Inf. Comput. Sci.* **1979**, *19*, 149–155.
(28) Cf. the *Chemical Abstracts Index Guides*, Appendix II: Indexes to Chemical Abstracts; Organization and Use, $14.
(29) Rowlett, R. J., Jr. An Interpretation of Chemical Abstracts Service Indexing Policies. *J. Chem. Inf. Comput. Sci.* **1984**, *24*, 152–154.
(30) Blake, J. E.; Dana, R. C. CASREACT: More than a Million Reactions. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 394–399.
(31) For problems with reaction searching in CA, see also examples in Zass, E. In *A Guide for the Perplexed Organic Experimentalist*, 2nd ed.; Loewenthal, E., Ed.; Wiley: London, 1990; Chapter 2.
(32) Blower, P. E.; Dana, R. C. In *Modern Approaches to Chemical Reaction Searching*; Willett, P., Ed.; Gower: Aldershot, U.K., 1986; pp 146–164.
(33) Blower, P. E., Jr.; Chapman, S. W.; Dana, R. C.; Erisman, H. J.; Hartzler, D. E. In *Chemical Structures*; Warr, W. A., Ed.; Springer: Berlin, 1988; p 399.
(34) Some supplementary information like additional literature references in REACCS or the keyword indexing in ORAC (cf. 2.3) needs additional display pages.
(35) Cf. search examples in refs. 13, 31, and Zass, E. In *Computer in der Chemie*, 2nd ed.; Ziegler, E., Ed.; Springer: Heidelberg, 1985; p 24.
(36) Zass, E. In *Chemical Information, Proceedings of the Montreux 1989 International Chemical Information Conference*; Collier, H. R., Ed.; Springer: Berlin, 1989; p 55. Zass, E. *J. Chem. Inf. Comput. Sci.*, to be submitted.
(37) Lawson, A. J.; Kallies, H. Multistep Reactions: The RABBIT Approach. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 426–430.
(38) The desired substitution pattern (Figure 13, 1–4 substituents on atoms α and β to the C=O group in both reactant and product) was not specified in REACCS, ORAC, or CRDS but defined in the substructure queries for the Registry File, using two substructures with variable substituents ("generic groups").
(39) Warr, W. A., Ed.; *Chemical Structure Information Systems. Interfaces, Communications, and Standards*; ACS Symposium Series 400; American Chemical Society: Washington, DC, 1989.
(40) Note Added in Proof: Version 3.0 of SYNLIB which has become available in the meantime shows significant improvements, e.g., reactant substructure search, more "constraints" for bond transformations, reaction conditions, and compound classes.