Comparison of Beilstein CrossFire*Plus*Reactions and the Selective Reaction Databases under ISIS

Faiz A. Parkar[‡] and Don Parkin*[,†]

SmithKline Beecham, New Frontiers Science Park North, Third Avenue, Harlow, Essex CM19 5AW, and Chemical Database Service, CCLRC Daresbury Laboratory, Warrington, Cheshire WA4 4AD, UK

Received September 5, 1998

To answer the question "are small selective databases now redundant", a comparison was made between the smaller, more selective reaction databases [under ISIS (Integrated Scientific Information System)] and a larger, more comprehensive database (Beilstein, via CrossFire) using references, exact reactions, named reactions, and transformations. The results show a large overlap of reference data with a smaller overlap of reaction data. Depending on the search, the overall view is that the databases are complementary.

INTRODUCTION

A comparison study of the reaction retrieval systems REACCS,[1] SYNLIB,[2] and ORAC[3] was published about 12 years ago,[4] which concentrated on the functionality of the systems. Two years later a comparison of the chemical content of these systems was published by choosing functional group transformations, constructing three queries, and comparing references.[5] There was very little overlap in the three systems, despite the fact that they abstracted similar journals from a similar period. The emphasis was on careful selections of useful reactions and did not claim to be comprehensive.

Since then, the number and size of reaction databases have increased. Three years ago, a study using automated overlap analysis showed an expected, but acceptable, increase in overlap of bibliographic data for these selective databases.[6] Furthermore, it was noted that, as a rule of thumb, "only half of the entries with identical literature references will have the same reaction centers."

Although the selective databases can be said to have a small overlap, this is obviously not true for the large databases such as CASREACT[7] and Beilstein[8] (see Appendix 1), which offer more complete literature coverage within certain boundaries. With the easier access to these databases via desktop software programs SciFinder and CrossFire, numerous companies and academics now have access to both selective and comprehensive reaction databases. The use of such databases is expected to increase as more chemists become aware of the potential and ease of use of these databases. A recent article by Zass[9] indicates the move away from online databases to in-house databases. This, however, also means increased time and cost of maintenance as the databases increase in size and number and new desktop software versions appear.

If most of the data is available from one large database, the question regarding the future of the smaller databases is raised. Therefore, this study was undertaken to see if the smaller, selective databases are now redundant.

At present, the availability of Beilstein CrossFire*plus*Reactions within the academic community in the United Kingdom (UK) has not caused the number of users or number of accesses to the smaller selective databases under the Integrated Scientific Information System (ISIS)[10] to decrease; if anything, this figure has increased (see Figure 1). The Chemical Database Service (CDS)[11] is a national service, funded by the Chemistry Programme of the EPSRC (Engineering and Physical Sciences Research Council). It provides access for UK academics to a range of chemistry databases in crystallography, spectroscopy, and synthetic organic chemistry. A Beilstein Database Trial was run for 6 months and then for another 10 months before being transferred successfully to its permanent site at MIDAS[12] after securing funding.

METHOD

A full comparison can only be performed if factors such as ease of use are taken into account as well as the data. This could include query drawing (tools, templates, query atoms, and bonds), entering nonstructural data (references, yields, keywords, etc.), type of search (e.g., exact, same transformation), hit list size, viewing, and manipulation. Some of the different roles and functions of ISIS, Beilstein, and CASREACT were discussed recently in the Zass study.[9] However, preference for one drawing package over another is often personal. Furthermore, MDL Information Systems, Inc. (MDL), who distribute ISIS, now market, sell, and support the entire Beilstein electronic product range, including the CrossFire System, together with Beilstein Informationsysteme GmbH (BIS).

It is possible, therefore, that in the future the graphical user interface to these databases may be very similar. ISIS/Draw already can be used as the structure editor for Beilstein, with a few restrictions, and an MDL Beilstein/ISIS gateway has become available. Therefore, comparison of the ISIS and Beilstein databases was restricted to the differences and overlap of the data.[13] The following categories were used for the comparison:

[†] Chemical Database Service.
[‡] SmithKline Beecham.

10.1021/ci980334w CCC: $18.00
Published on Web 02/12/1999

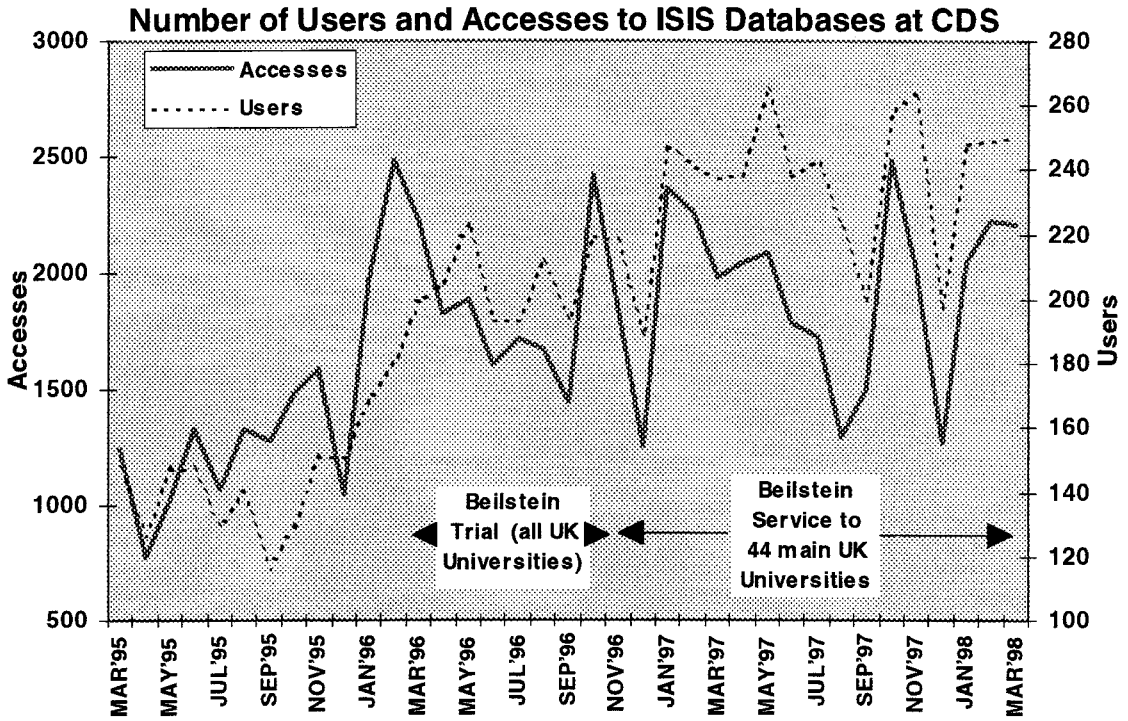

Figure 1. Number of users and accesses to ISIS databases at CDS.

A. References

- From selected reaction queries.
- From selected journals and years.

B. Reactions plus References

- Beilstein search for selected ISIS reactions and associated references by ISIS database.
- Beilstein search for selected ISIS reactions and associated references by Journal and Year.

C. Reactions.

- Named Reactions.
- Transformations.

A. References. Two approaches were used: comparison of references from selected reactions and comparison of references from journals chosen within selected periods to access any time dependence.[14,15]

A.1. From Selected Reaction Queries. A previous reaction database comparison[5] used various selected reaction queries to create a hit set, and the references analyzed. Queries 1 and 2 (see Figure 2) were repeated for Beilstein, CIRX[17] (92−97.2) and the rest of the evaluated ISIS databases REFLIB (CLF, CORE, METAL, CHIRAS, and Theilheimer), CHC, REACCS-JSM, and ORGSYN (see Appendix 1 for database descriptions).

Both reactions were fully mapped, and bond make/break or change attributes were added. For query 1, the two fragment structures of both the reactants and the products were grouped together to signify that they were part of the same molecule, and so limit the search.

ISIS searches were conducted as reaction substructure searches (RSS), and Beilstein searches were conducted with Implicit Free Sites.

A.2. From Selected Journals and Years. Because of the large number of references it was decided to look at specific journals rather than all journals found in both databases. Two sets of data were taken covering different periods, because ChemInform does not cover years before 1990. Therefore,

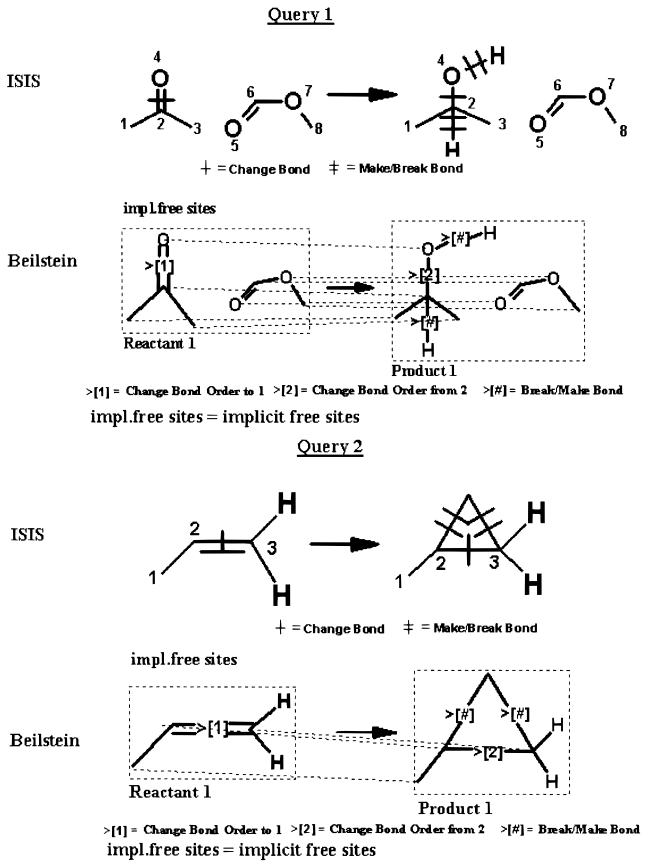

Figure 2. Reaction queries.

set A was taken from 1990 onward and includes the ChemInform databases, whereas set B was taken before 1990 and so excludes ChemInform.

The choice of journals was then confined to those most likely to be found in libraries rather than obscure or specialist journals such as Magy.Kem.Lapja or Biocatalysis. This was

restricted further, in the journals containing a large number of references, to a specific year so that a manageable number of hits were obtained. The journals (and years) chosen for all the databases were

Set A

Acta Chim. Hung. (1990−1993); *J. Chem. Soc., Chem. Commun.* (1993); *J. Chem. Soc., Perkin Trans. 1* (1992); *J. Org. Chem*. (1994); *Synthesis* (1995); *Synth. Commun*. (1990, p 1579 onwards); and *Tetrahedron Lett.* (1996).

Set B

Chem. Ber. (1950); *Helv. Chim. Acta* (1960); *J. Chem. Soc., Chem. Commun.* (1975); *J. Chem. Soc., Perkin Trans. 1* (1980); *J. Med. Chem.* (1965); *Justus Leibigs Ann. Chem*. (1975); and *Synthesis* (1980).

B. Reactions plus References. B.1. By ISIS Database. Ten reactions, along with one reference associated with each reaction, were selected randomly from each of the 11 ISIS databases (REFLIB, CHC, REACCS-JSM, ORGSYN, CIRX92−CIRX98). Searches were conducted in Beilstein for the exact reaction, the reference, and, if the reaction was not found, for the product.

B.2. By Journal and Year. Ten to twelve reactions were selected randomly from each of the eight journals that were searched in set B for the years before 1990 to determine what effect, if any, the time period has on data. After removing all mapping, searches were conducted in Beilstein for the exact reaction and for the product, which was then matched with the required ISIS reference; the product search was conducted to see if it was in the database as a nongraphical reaction.

C. Reactions. Reactions were restricted to semispecific queries to avoid long search times and large numbers of hits (e.g., had to be aromatic or only ethyl esters or a methyl substituent). ISIS searches were conducted as RSS with explicit hydrogens where appropriate and were mapped automatically. Beilstein searches were conducted either "as drawn", except for some positions where R groups were allowed which were marked as MAX Free Sites, or as Implicit Free Sites. In most cases mapping was not applied, except where rearrangement, ring formation, or expansion occurred. In some cases searches were conducted with and without mapping to explore any differences.

C.1. Named Reactions. Representative reaction searches were conducted in both ISIS and Beilstein for named reactions (shown in Table 6). In most cases the reagent/catalyst was not specified except for the Reformatskii and Simmons−Smith reactions, in which searches were conducted with and without zinc and CH_2I_2, respectively. Formulas of the reagent/catalyst were added to the structure box in ISIS or stated as the reagent (rx.rgt) in the Fact Editor for Beilstein. Only one reactant was specified except for the Reformatskii and Diels−Alder reactions where two reactants were present, any ketone and ethyl bromoacetate being entered as reactant structures for the former and $CH_2{=}CH{-}CH{=}CRR' + CH_2{=}CHR$ for the latter.

C.2. Transformations. Representative reaction searches were conducted in both ISIS and Beilstein for various transformations (shown in Table 7). A search with and without absolute stereochemistry was conducted in Beilstein for *cis*-hydroxylation. Both reactants were drawn for allylic bromination, *N*-bromosuccinimide being entered in the structure box for ISIS and by either drawing it as a second reagent in Beilstein or by including the field rx.rgt or rxb.rgt using the Fact Editor.

The "ozonolysis and ring opening" search was conducted with and without ozone, which was added either by including it in the structure box for ISIS or by using the field rx.rgt or rxb.rgt for Beilstein.

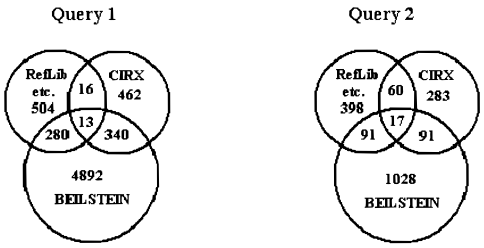

Figure 3. Number of unique references for the two reaction queries.

RESULTS AND DISCUSSION

A. References. The Fiz−Chemie databases ChemInform (CIRX) contain data from 145 different publications, although 43 are no longer abstracted. Thirty of these 145 publications are not found in Beilstein. This amounts to about 21 000 reactions or 6.2% of the reactions found in CIRX, or 2300 (3.8%) of the references found in CIRX.

For the same period (1990−1996), the rest of the ISIS databases (RefLib, JSM, CHC, and Orgsyn) contain references from 140 different journals/publications of which 57 are not found in CIRX. This amounts to 2614 references or 10.9% of the total ISIS references for this period. The majority (87%) of references for these other ISIS databases are from 1946 to 1990, so overlap with the CIRX databases should be small.

A.1. From Selected Reaction Queries. The results from a comparison of unique references obtained from the two query reactions first shown in the publication by Noordik are shown in Figure 3.

A.1.a. RefLib etc. vs CIRX. Both query reactions showed an acceptable overlap between the ISIS datasets, query 1 showing only about 3% overlap whereas query 2 showing that 15% of the RefLib etc. references also can be found in CIRX, or 21% of the CIRX references are also in the other ISIS databases.

A.1.b. ISIS vs Beilstein. As expected, query 1 showed a much larger overlap; 55% of the RefLib etc. references and 74% of the CIRX references were found in the Beilstein hit list.

All ISIS searches were complete within 5 min, although approximately 500 unique references means about 800 reactions, which are not very manageable, even with the help of Clustering. However, Beilstein took 2.5−5 h[18] and the number of hits (7692 reactions) is definitely not manageable. Query 2 showed, remarkably, only a slight, acceptable increase in reference overlap among all three database sets; only 23% of the RefLib etc. references and 32% of the CIRX references were found in the Beilstein set of references. The Beilstein search took 15−30[18] min and gave a 1750-reaction hit list, whereas ISIS gave 514 (Reflib etc) and 740 (CIRX) reaction hit lists within a few minutes, all of which are not very manageable. The fact that a large number of hits was obtained for both the Beilstein searches indicates that much

Table 1. Number of Unique References in Various Journals and Years from 1990

	journal						
database	*Acta Chim. Hung.*	*Synth. Commun.*	*Tetrahedron Lett.*	*Synthesis*	*J. Org. Chem.*	*J. Chem. Soc. Chem. Commun.*	*J. Chem. Soc. Perkin Trans. 1*
year	1990, p 171−end 1993	1990, 20 (11), 1579 onward	1996	1995	1994	1993	1992
Beilstein	296	2836	2319	294	1468	1043	523
CIRX	37	2618	1535	258	874	315	421
rest of ISIS databases	0	1331	491	132	243	201	81
all ISIS databases	37	2683	1597	264	900	345	426
CIRX refs *not* in Beilstein	0	80 (3%)	119 (7.8%)	1 (0.4%)	7 (0.8%)	0	7 (1.7%)
rest of ISIS database refs *not* in Beilstein	0	12 (0.9%)	41 (8.4%)	3 (2.3%)	3 (1.2%)	0	2 (2.5.%)
all ISIS refs *not* in Beilstein	0	87 (3.2%)	123 (7.7%)	3 (1.1%)	7 (0.8%)	0	8 (1.9%)

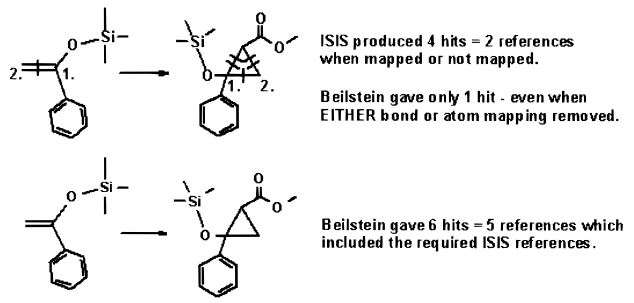

Figure 4. Effect of atom and bond mapping.

better processing tools are required if useful information is to be extracted from Beilstein.

These results may indicate the different abstracting philosophies of Beilstein and the selective reaction databases. Beilstein can be described as a molecule-oriented database; entries refer to a compound and its associated data, including some reactions, like the preparation or derivatization. The ISIS databases such as ChemInform focus on reactions and synthetic methods.

On searching Beilstein for some of the ISIS references that did not appear in the Beilstein list in query 2, a large number of these references were actually present in the database. Therefore these references were investigated a little further to see if, for example, the reaction appeared as nongraphical or just not abstracted. Although this was the case for one or two, on examining some of the references, reactions were found where a cyclopropane ring was indeed prepared from a double bond. These reactions may not have been picked up in the Beilstein search because of the MAPPING that was applied, as shown in Figure 4. (Note: the top reaction is shown as an ISIS sketch, but this can also be used in a Beilstein search.)

A.2. From Selected Journals and Years—Set A. The results of the search, displayed in Table 1, show a very large overlap for these journals. This is perhaps not surprising because most of these journals are among the most abstracted (see Table 2) and widely available journals. It is possible, however, that smaller overlaps in reactions exist (see Reactions plus References section).

A.3. From Selected Journals and Years—Set B. The results of the search, displayed in Table 3, again show a very large overlap in reference data for these journals. However, analysis of the reactions associated with these references was conducted (see Reactions plus References section), and the results for reaction data overlap are very different.

B. Reactions plus References. B.1. By ISIS Database. The results, displayed in Tables 4a and 4b, show that only

Table 2. Number of Unique References in Top 12 ISIS Journals from 1990 to 1996 Inclusive

journal	RefLib etc.	CIRX	Beilstein
Tetrahedron Lett.	5271	8946	15218
J. Org. Chem.	2504	5079	10349
Synth. Commun.	1514	2395	2901
J. Chem. Soc. Chem. Commun.	1393	2139	7829
J. Am. Chem. Soc.	1249	1911	12129
Synthesis	1085	1582	2120
Synlett.	1032	1788	0
J. Chem. Soc. Perkin Trans. 1	990	2268	3818
Tetrahedron	757	3917	6553
J. Heterocycl. Chem.	720	1630	2288
Chem. Lett.	654	1171	4210
Heterocycles	588	1589	2487

Table 3. Number of Unique References in Various Journals and Years before 1990

		references			ISIS refs *not* in Beilstein
journal	year	ISIS	Beilstein	combined	
Synthesis	1980	251	353	356	3 (1.2%)
J. Chem. Soc. Perkin Trans. 1	1980	189	514	520	6 (3.2%)
Justus Leibigs Ann. Chem.	1975	45	220	222	2 (4.4%)
J. Chem. Soc. Chem. Commun.	1975	198	571	579	8 (4%)
Synthesis	1970	68	73	85	12 (17.6%)
J. Med. Chem.	1965	51	242	242	0
Helv. Chim. Acta	1960	31	206	207	1 (3.2%)
Chem. Ber.	1950	42	112	113	1 (2.4%)

59% of the reactions selected randomly from the ISIS databases appear in Beilstein. Of these, 80% contained the required reference.

For the remaining 41% where no reaction was found, half contained the query reference AND product in Beilstein and a fourth contained either the reference or the product. Therefore, overall, the required reference was found for 74% of the queries, and a product was found for about 83% of the queries.

These figures were mainly due to the fact that either the reaction was nongraphical or that the reaction selected from ISIS was an overall reaction which Beilstein does not contain; the Beilstein reference did show the intermediates, however. In five cases, the required reference did not contain the required reaction, although the required reaction was found with other references. (Note: Similar figures were also found for Chemreact databases 1, 2, and 3. On average, 65% of the required reactions, 84% of the required references, and 81% of the products were found in Beilstein.)

Table 4

Results from Searching Beilstein for Specific Reaction/Reference/Product

	full reaction		reference		product	
database	yes	no	yes	no	yes	no
RefLib	6	4	8	2	9	1
CHC	4	6	10	0	7	3
JSM	8	2	10	0	10	0
ORGSYN	8	2	1	9	9	1
CIRX92	3	7	7	3	7	3
CIRX93	5	5	8	2	7	3
CIRX94	7	3	10	0	9	1
CIRX95	6	4	5	5	7	3
CIRX96	6	4	9	1	10	0
CIRX97	6	4	10	0	9	1
CIRX98	6	4	3	7	7	3
total average %	59.1	40.9	73.6	26.4	82.7	17.3

Summary of Results from Search Above

full reaction	required reference	required product	totals
no	no	no	12 (10.9%)
no	no	yes	4 (3.6%)
no	yes	no	7 (6.4%)
no	yes	yes	22 (20.0%)
yes	no	yes	13 (11.8%)
yes	yes	yes	52 (47.3%)

B.2. Reactions plus References by Journal and Year. Table 5 shows that, although there is a large overlap in reference data, there is a significant reduction in the overlap of reactions. On average only 56% of the required reactions were found, although the product was found in about 90% of cases.

The results seem to indicate that journal references with publication years before 1980 contain a significant number of nongraphical reactions (in this case, 39%). The other factor could well be that some of the ISIS reactions chosen were overall reactions and that intermediate reactions may be present in Beilstein for the required reference.

C. Reactions. C.1. Named Reactions. The results of searching for specific named reactions are shown in Table 6. The results show that, as expected, fewer reactions are found in ISIS than Beilstein (but the results are available quicker). No hits were obtained in only one case with ISIS. However, ISIS did produce two hits where Beilstein failed to produce any hits (mapped Strecker reaction), which is obviously due to mapping (see below).

An average of 31 hits for ISIS is a reasonable number to search through and a good base in which to start scanning literature references. In no case with ISIS did the search last for longer than 4 min. Beilstein, on the other hand, took up to 1 h 10 min,[13] with 30% of the searches taking longer than 20 min. Beilstein also often produced hit lists in the hundreds, and in one case, 1737 hits. This is too many hits to scan through effectively unless some form of clustering can be used, which is possible with ISIS.

C.2. Transformations. The results of searching for specific types of reactions are shown in Table 7. The results, as expected, are similar to the named reaction search. ISIS produced a small, manageable hit list after only a few minutes, extra bond mapping cutting down the hits even further (on average, about 20 hits). Beilstein gave a large range of results for both number of hits and time taken; some searches took longer than 1 h.[13]

Care had to be taken with Beilstein when defining reagents for a search. Beilstein does not allow for drawing solvents or catalysts, although solvents and catalysts can be included as reactants.

A different set of hits were obtained for drawn reagents than for factual reagents, and this depended on which field was used. For instance, rx.rgt=nbs (NBS, *N*-bromosuccinimide) gave 0 hits as did defining it as a reagent in a nongraphical reaction (rxb.rgt=nbs). Several different solvents and co-reagents are also defined within Beilstein, so that NBS on its own (rx.rgt=nbs) gave only 358 hits, whereas rx.rgt=nbs* gave 674 hits. The same was true for the Ozonolysis reaction, which was complicated further by field value names of o3, ozon, or ozone, as well as with other solvents and reagents, as shown in the examples below for the amount of data available for just the reagent.

rx.rgt=ozon* = 2643 reactions
rx.rgt=o3* = 3893 reactions

	rx.rgt
26	ozon
2252	ozone
12	ozone (excess)
33	ozone−oxygen
10	ozone, conc.hcl
2951	o3
11	o3 (o2)

rx**b**.rgt=ozon* = 789 reactions
rx**b**.rgt=o3* = 226 reactions

	rxb.rgt
770	ozone
4	ozone−air mixture
155	o3
8	o3, air
16	o3, h2o

Searching for specific transformations in the presence of other potentially reactive groups is a task that is difficult in printed journals but is well suited to electronic databases. Two such transformations were included in the list, epoxidation of a cyclic double bond but not an acyclic double bond and oxidation of sulfur compounds but not double bonds. The results show that ISIS produced a workable list of hits (in less than 2 min), whereas Beilstein took much longer (in one case, approximately 1 h) and gave hit lists that were unmanageable in size (i.e., 3269 hits for selective

Table 5. Results from Searching Beilstein for Specific Years

journal	year	ISIS refs *not* in Beilstein	reaction + required ref[a]	reaction found[a]	product found[a]	nongraphical reaction[a]
Synthesis	1980	3 (1.2%)	80	90	100	0
J. Chem. Soc. Perkin Trans. 1	1980	6 (3.2%)	60	60	70	0
Justus Leibigs Ann. Chem.	1975	2 (4.4%)	10	20	80	70
J. Chem. Soc. Chem. Commun.	1975	8 (4%)	0	40	80	40
Synthesis	1970	12 (17.6%)	17	58	92	25
J. Med. Chem.	1965	0	10	40	100	50
Helv. Chim. Acta	1960	1 (3.2%)	17	75	100	42
Chem. Ber.	1950	1 (2.4%)	10	60	100	10

[a] Values are stated as percents.

Table 6. Results of Searching for Specific Named Reactions

named reaction	number of hits: ISIS	number of hits: Beilstein
Baeyer−Villiger	23	322
Biginelli reaction	25	46
Curtius rearrangement	77	16
Duff reaction	82	1737
Fischer indole	35	347
Knoevenagel condensation	44	504
Knorr pyrazole	26	62
Michael condensation	0	10
Reformatskii, without Zn	63	833
Reformatskii, with Zn	42	195
Simmons−Smith, without CH_2I_2	17	73
Simmons−Smith, with CH_2I_2	9	9
Robinson annelation	31	236
Strecker degradation	2	not mapped = 12 mapped = 0
Diels−Alder	30	not mapped = 151 mapped = 1
Blanc reaction/rule	4	not mapped = 12 mapped = 10
Wittig	21	not mapped = 227 mapped = 45
average	31	242

epoxidation of cyclic double bond and 1324 hits for selective oxidation of sulfur).

CONCLUSIONS

As with the study by Noordik 3 years ago,[6] there was an expected, but acceptable, increase in overlap of bibliographic data between the selective databases (CIRX compared with the other ISIS databases). A very much larger overlap (>90%) in reference data was shown for the collective ISIS databases compared with Beilstein. The overlap in reaction data, however, is much smaller, partially because a large percentage of reactions in Beilstein are nongraphical, at the moment, and partially because the ISIS databases contain overall reactions.

Beilstein CrossFire*plus*Reactions can be very fast when searching for precise factual, structural, or reaction data and is probably the recommended choice for specific data. However, nonspecific searching can take much longer, and this is particularly evident within the UK academic environment where a large number of users (up to 60 concurrent users) access the database. In general, Beilstein will take longer than ISIS on a nonspecific search. Also, with Beilstein, a large number of references can be obtained within an answer set (e.g., toluene has 4788 reactions and 3131 nongraphical reactions associated with it), therefore better viewing and hit list manipulation tools would help. Some specialist data are available under ISIS, such as the Protecting Groups[19] or Solid-Phase Synthesis[20] databases; such data are best accessed by this route.

Care must be taken with reaction searches in Beilstein, because it is not primarily a reaction database. No hits does not necessarily mean no data are available; it may simply be the way the query was constructed. For instance, adding mapping, drawing the reagent, or using the Fact Editor without comprehensive field names and values can reduce the hit size. The exact reaction may also not be found because it is present as two single-stage reactions or is nongraphical. The overall results for reaction searching indicate that it may be better to search ISIS for a lead first, because the hit list is obtained quickly and can be manageable. If no hits or very few hits are obtained, then Beilstein can be used.

The use of one database over the other would therefore appear to depend on the type of search query and the expected number of hits (i.e., for reference, fact, or specific

Table 7. Results of Searching for Specific Types of Reactions

reaction type	number of hits: ISIS	number of hits: Beilstein
epoxidation (of cyclic double bond, not acyclic)	100	3269
benzidine rearrangement	3	69
cis-hydroxylation (e.g., Millas)	2	71 absolute stereochemistry = 30
nitration (e.g., Zincke)	2	not mapped = 240 mapped = 91
allylic bromination (e.g., Wohl−Ziegler)	32	reagent drawn = 27 all nbs options in Fact Editor = 359 rxb.rgt=nbs = 0 rx.rgt=nbs = 358 rx.rgt=nbs,* = 674
reductive acylation	8	mapped = 5
rearrangement of tertiary acetylenic alcohols (e.g., Rupe reaction)	4	mapped = 12
ring expansion, C6−C7	atom mapped = 64 + bond mapped = 2	ring atoms = * mapped = 0 after 1 h not mapped = 376 069
ring expansion, any ring by CH_2	atom mapped = 5578 + bond mapped = 5	not correct ring expansion
cyclobutane formation from cyclic alkenes	8	mapped = 15
ozonolysis and ring opening	without o3 = 24, with = 9	unmapped = 106 with rx.rgt=ozone = 8 with rx.rgt=o3 = 3 mapped = 37 with rx.rgt=o3* or ozon* = 14 with rxb.rgt=o3* or ozon* = 2
oxidation of sulfur but not double bond	44	mapped = 1324
reductive amination	58	R & N mapped = 154

product search, choose Beilstein; for any other reaction search, choose ISIS first, then Beilstein). Overall, the two types of databases are complementary and the small selective databases still have a large role to play in reaction searching.

APPENDIX 1

REFLIB **REF**erence **LIB**rary of SyntheticMethodology, includes the following:

- CLF — The **C**urrent **L**iterature **F**ile. Synthetic methods database of new developments in organic chemistry; 36 601 reactions; 1983−1991.
- CORE — ORAC/CORE Database, includes synthesis of asymmetric compounds, heterocyclic compounds, and large rings and novel synthetic methods; 75 000 reactions; 1946−1991.
- METALYSIS — Transition metal-mediated methods. Metal-mediated transformations database covering organic applications of transition metal catalysts and reagents; 11 999 reactions; 1974−1991.
- CHIRAS — Asymmetric synthesis. Comprehensive collection of useful methods of asymmetric synthesis; 13 220 reactions; 1975−1991.
- Theilheimer — Synthetic methods of organic chemistry. A compendium of high-yield functional group transformations and synthetic methods selected from the literature; 46 784 reactions; 1946−1980.

CHC **C**omprehensive **H**eterocyclic **C**hemistry. Sourced from *Comprehensive Heterocyclic Chemistry* (Permagon Press); it contains collected reviews of heterocyclic chemistry, synthesis and reactions; 42 376 reactions; compendium up to 1984.

REACCS-JSM **J**ournal of **S**ynthetic **M**ethods. Sourced from the *Journal of Scientific Methods* (Derwent); information is abstracted from world-wide patent as well as primary chemical literature focusing on novel functional groups and ring chemistry, protecting group chemistry, and heterocyclic chemistry; 58 857 reactions; 1980 to the present.

ORGSYN **ORG**anic **SYN**thesis. Sourced from *Organic Synthesis* (Wiley & Sons); it contains contributed procedures useful to chemists in diverse areas of organic chemistry; 5201 reactions; 1921 to the present.

ChemInformRX Produced by FIZ Chemie; it covers the advances in synthetic research since 1991 and consists of separate annual releases starting from CIRX92, giving a total of >400 000 reactions.

SPG **S**ynopsys **P**rotecting **G**roups. Selected information on the full range of protecting group chemistry; 25 595 reactions.

SPS **S**olid-**P**hase **S**ynthesis. A database from Synopsys containing a selection of relevant material from primary literature, reviews, and patents (mainly 1997) from >1000 publications; 5445 reactions.

ACDRX Available Chemicals Directory (for MDL's Reaction Gateway). Essentially a database for finding chemicals and their suppliers. Contains around 0.25 million unique compounds from >500 supplier catalogues.

Beilstein Taken from the *Beilstein Handbook of Organic Chemistry*, it contains approximately 7 million structures and their associated chemical and physical properties, preparative methods, chemical behavior, and literature references, along with about 10 million reactions extracted from the literature since 1779.

CASREACT CASREACT contains information on 1.26 million single-step and 1.86 million multi-step organic reactions, drawn from Chemical Abstracts' organic sections since 1985. Information includes reactants, solvents, and catalysts.

REFERENCES AND NOTES

(1) *REACCS* (***REA****ction* ***ACC****ess* ***S****ystem*); MDL Information Systems Inc., San Leandro, CA.

(2) *SYNLIB* (***SYN****thesis* ***LIB****rary*); Distributed Chemical Graphics Inc., Philadelphia, PA.

(3) *ORAC* (***O****rganic* ***R****eactions* ***A****ccessed by* ***C****omputer*); MDL Information Systems Inc., San Leandro, CA.

(4) Zass, E.; Mueller, S. New Research Opportunities Concerning Organic-Chemical Reactions—A Comparison of the In−House Databank Reaccs, Synlib and Orac Systems. *Chimia* **1986**, *40*, 38−50.

(5) 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.

(6) Boiten, J.-W.; Ott, M. A.; Noordik, J. H., Automated Overlap Analysis of Reaction Databases. *J. Chem. Inf. Comput. Sci*. **1995**, *35*, 115−120.

(7) Blake, J. E.; Dana, R. C. CASREACT; More than a Million Reactions. *J. Chem. Inf. Comput. Sci*. **1990**, *30*, 394−399.

(8) Beilstein: Beilstein Informationsysteme GmbH, Frankfurt, Germany. See also: Jochum, C. The Beilstein Information System Is Not a Reaction Database, or Is It? *J. Chem. Inf. Comput. Sci*. **1994**, *34*, 71−73.

(9) Zass E. Using Beilstein Reaction Database in an Academic Environment. In *The Beilstein System. Strategies for Effective Searching*; Heller, S. R., Ed.; American Chemical Society: Washington, DC, 1998; p 99.

(10) *ISIS* (*Integrated Scientific Information System*); MDL Information Systems Inc., San Leandro, CA.

(11) *CDS* (*Chemical Database Service*). See: The United Kingdom Chemical Database Service. Fletcher, D. A.; McMeeking, R. F.; Parkin, D. *J. Chem. Inf. Comput. Sci*. **1996**, *36*, 746−749.

(12) *MIDAS* (*Manchester Information Datasets and Associated Data*); University of Manchester, Manchester, UK.

(13) ISIS searches were obtained via ISIS/Base (version 2.1.1) from the CDS and Beilstein searches via CrossFire (version 3.1, build 13) Commander (version 2.3, build 1) from MIDAS. All searches were conducted during normal working hours (9:00 a.m. to 5:00 p.m.). It should, however, be noted that any times quoted are for the UK academic environment, and times for other environments will vary depending on the quality and load of the server.

(14) ISIS references are often written as year, volume, and issue number then page number(s), whereas Beilstein references are usually volume and issue number, year, then page range. However, not only is the format different, but the information may be different, such as different author spelling and page numbers (e.g., ISIS = **Gibbs**, D. E.; Verkade, J. G. Synth Commun [SYNCAV] 1976, 6, 103. Beilstein = **Gibs**; Verkade; SYNCAV Synth Commun.; 6; 1976; **103**, **104**, **106**; and 6; 1976; **103**, **104**, **107**) and even page ranges (for example, CIRX = 1997, 27 (4), **691**−**699**; Beilstein = 27; 4; 1997; **691**−**700**). Therefore, literature comparisons were restricted to just journal CODEN, year, and starting page number.

(15) ISIS database references were obtained by viewing the hit list as a table of references, then exporting it as a table (from the RXN>VARIATION>LITREF>FULL_CITATION field) to a text file. Duplicates were removed using Microsoft Excel. Beilstein citation searches were displayed then printed to a PostScript file. Structure

query searches were displayed by using the Define User View, and choosing all the Citation (CIT) fields plus the Reaction Classification (RX.CL) field (a reaction details field must be chosen to associate the reference with a reaction). These PostScript files were then read in by Ghostscript via Gsview[16] and converted to text. Superfluous data were removed by manipulations within Microsoft Word and Excel.

(16) Gsview and Ghostscript: GSview is a graphical interface for Ghostscript under MS−Windows or OS/2. Ghostscript is an interpreter for the PostScript page description language used by laser printers and is owned by Aladdin Enterprises, Menlo Park, CA. For documents following the Adobe PostScript Document Structuring Conventions, GSview allows selected pages to be viewed or printed.

(17) *CIRX* (*Cheminform Reaction Databases*); FIZ CHEMIE, Berlin.

(18) Depends on time of day/number of users (see ref 13).

(19) SPG: Protecting Groups Database, Synopsys Scientific Systems Ltd., Headingley, Leeds, LS6 2EN, UK.

(20) Solid-Phase Synthesis databases, e.g., SPS from Synopsys Scientific Systems Ltd., Headingley, Leeds, LS6 2EN, UK, or SPORE (Solid-Phase Organic REactions) developed jointly by MDL Information Systems Inc., San Leandro, CA, and FIZ CHEMIE, Berlin.

CI980334W