Synthetic Chemical Literature

Matthew Clark

The Institute for Scientific Information, 3501 Market Street, Philadelphia, Pennsylvania 19104

Received January 6, 1999

Several databases track new and improved synthetic methods in organic chemistry. An analysis of the ISI Reaction Center, CASREACT, and ChemInformRX shows a consensus that nearly all new synthetic techniques are found in a core of about 100 sources. In addition, patents were found to be a significant, yet under-used, source of novel chemistry.

INTRODUCTION

It has often been observed that scientific knowledge in a given discipline is concentrated in a few journals. Bradford's law of scattering[1] is the name commonly given to this observed behavior. This effect and its origins have been studied extensively in a variety of contexts. Bradford's law indicates that (1) the main body of an area of study is contained in a relatively small subset of the journals in a field and (2) there is an exponentially decreasing rate of new information added as the number of journals is increased beyond this core set. This appears to be true for scientific literature when journal citation rates were measured: Of the journals covered in the Journal Citation Reports, Science Edition, 80% of all citations are from 16% of journals.[2]

Garfield examined the synthetic chemistry literature in 1973.[3] This work provides an updated view based on synthetic methods. In this study, we confirm that 25 years later, despite the increase in the number of journals published, Bradford's law still describes the distribution of novel synthetic techniques in organic literature. As was reported in 1973, 95% of new synthetic organic techniques are reported in a core of 100 journals and patent organizations.

METHODS

Since the determination that an article contains a new synthetic method is somewhat subjective and dependent on editorial policy, the study was performed on the ISI Reaction Center,[4] the CASREACT database,[5,6] and the ChemInformRX[7,8] database.

The data in the ISI Reaction Center was analyzed via the ISI Chemistry Server Web interface. The data for CASREACT and ChemInformRX were collected through the STN online system and the "smartselect" command to give lists of the sources of abstracted articles and the number of articles abstracted from each source. These results were then sorted by the number of articles abstracted per journal to produce graphs of the percent of articles covered vs the number of journals that the articles were drawn from.

RESULTS AND DISCUSSION

Figure 1 shows a log graph of the cumulative percentage of the total articles covered in three major databases in 1996 vs the number of journal titles considered. The distribution of articles among the journals follows the familiar shape described by Bradford's law.

Table 1. Top 20 Sources Contributing New Synthetic Techniques in ISI Reaction Center in 1996

source	articles selected	% of all articles selected
Tetrahedron Lett.	1700	18
U.S. Pat.	1427	15
J. Org. Chem.	901	9
Tetrahedron	512	5
Chem. Commun.	498	5
Synth. Commun.	432	5
J. Am. Chem. Soc.	398	4
Synlett	363	4
Tetrahedron: Asymmetry	307	3
Organometallics	195	2
Synthesis-Stuttgart	193	2
J. Chem. Soc., Perkin Trans. 1	181	2
Chem. Lett.	175	2
Heterocycles	175	2
Angew. Chem., Int. Ed. Engl.	161	2
J. Organomet. Chem.	139	1
Indian J. Chem., Sect. B	101	1
Liebigs Ann.	92	1
Chem. Pharm. Bull. Tokyo	90	1
J. Heterocycl. Chem.	75	1

Table 2. Top 20 Sources Contributing New Synthetic Techniques in CASREACT in 1996

source	articles selected	% of all articles selected
Tetrahedron Lett.	1271	13
Jpn. Kokai Tokkyo Koho	801	9
J. Org. Chem.	774	8
Tetrahedron	526	6
PCT Int. Appl.	363	4
Synth. Commun.	337	3
J. Am. Chem. Soc.	314	3
Eur. Pat. Appl.	310	3
Synlett	309	3
Organometallics	233	2
U.S. Pat.	220	2
Tetrahedron: Asymmetry	216	2
J. Chem. Soc., Perkin Trans. 1	200	2
J. Organomet. Chem.	182	2
Chem. Commun.	173	2
Synthesis	172	2
Heterocycles	154	2
Ger. Offen.	132	1
Angew. Chem., Int. Ed. Engl.	121	1
Zh. Obshch. Khim.	108	1

The initial curve, the "nuclear zone", is comprised of the first 10 journal titles; the linear portion encompasses from 10 to 100 titles, where the incremental number of articles

10.1021/ci990001e CCC: $18.00 © 1999 American Chemical Society
Published on Web 06/05/1999

Table 3. Top 20 Journals Contributing New Synthetic Techniques in ChemInformRX in 1996

journal	articles selected	% of all articles selected
Tetrahedron Lett.	1535	15
Tetrahedron	1105	11
J. Org. Chem.	960	9
Synth. Commun.	520	5
Chem. Commun.	391	4
Synlett	369	4
Tetrahedron: Asymmetry	361	3
J. Heterocycl. Chem.	264	3
Heterocycles	263	3
J. Am. Chem. Soc.	246	2
Bioorg. Med. Chem. Lett.	244	2
Izv. Akad. Nauk, Ser. Khim.	232	2
Zh. Org. Khim.	228	2
Synthesis	221	2
Indian J. Chem., Sect. B	216	2
J. Chem. Soc., Perkin Trans. 1	200	2
Khim. Geterotsikl. Soedin.	195	2
J. Med. Chem.	193	2
J. Chem. Res., Synop.	166	2
Chem. Pharm. Bull.	156	2

Table 4. Average Reactions per Article in the 1996 Reaction Center Database

source	average reactions/article
J. Org. Chem.	5.35
Tetrahedron	5.31
J. Am. Chem. Soc.	4.65
Tetrahedron Lett.	4.50
Synlett	4.47
Tetrahedron: Asymmetry	4.40
Synth. Commun.	3.36
Chem. Commun.	3.14
Organometallics	2.17
U.S. Pat.	2.09

decreases exponentially. After about 100 titles the exponential curve tails off, and the "Groos droop" is observed at about 100 titles.[9,10] This feature of the Bradford distribution, first described by Ole V. Groos, is found in many sets of data, such as the distribution of citations among journals in the Science Citation Index, and indicates that information in the field in question has been nearly exhausted within the parameters of the data selection.

All three databases indicate that nearly all articles with new synthetic methods are covered in about 100 sources, with the first 10 sources covering from 50 to 70% of all methods. The number of articles selected is remarkably similar among the three databases. The Reaction Center abstracted 9503 articles, CASREACT 9480, and ChemInform RX 10348.

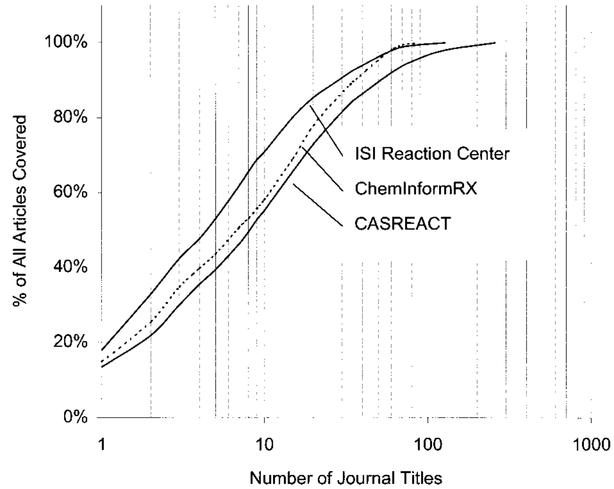

Figure 1. Articles reporting novel synthetic methods vs number of journal titles.

Tables 1, 2, and 3 report the top 20 sources of new synthetic techniques for the three databases studied. All three sources find *Tetrahedron Letters* to be the leading source for this information. The ISI Reaction Center and CASREACT find patents the second most prolific source, reflecting patents as an increasingly important source of information. The prominence of patents in these databases is intriguing given the scarcity of citations to patents in organic chemistry journals. CASREACT ranks the Japanese Kokai, PCT International Applications, and European patent applications as the second, fifth, and eighth most prominent sources of novel reactions. If all patent sources are combined, patents are the top sources of reactions in CASREACT.

Patents are an obviously important source of novel chemistry from their prominence in these databases. However, less than 1% of citations in the Science Citation Index are to patents. This suggests that patents are an under-used source of information for academic researchers who publish in scholarly journals. Corporate researchers, the source of most patents, do tend to cite other patents heavily in their patent filings.

Five of the top 10 journals are also in the top 10 reported in the 1973 *Nature* article. The similarity among the journals and the distribution of articles abstracted speaks to both the publishing habits of the field and the similarity of the abstracting policies of the databases. The curve in Figure 1 clearly shows a consensus among the three database producers that new synthetic techniques are reported in a subset of all chemistry journals.

While the ISI Reaction Center scans 350 journals for new synthetic techniques, we found that only 128 sources reported

Table 5. Most Cited Articles Reporting New Synthetic Techniques

citations 1985−1996	article
455	Ito, T.; Wang, J. X.; Lin, C. H.; Lunsford, J. H. Oxidative Dimerization Of Methane Over A Lithium-Promoted Magnesium-Oxide Catalyst. *J. Am. Chem. Soc.* **1985**, *107*, 5062−5068.
347	Sharpless, K. B.; et al. The Osmium-Catalyzed Asymmetric Dihydroxylation − A New Ligand Class And A Process Improvement. *J. Org. Chem.* **1992**, *57*, 2768−2771.
307	Ewen, J. A.; Jones, R. L.; Razavi, A.; Ferrara, J. D. Syndiospecific Propylene Polymerizations With Group-4 Metallocenes. *J. Am. Chem. Soc.* **1988**, *110*, 6255−6256.
298	Jeske, G. J.; et al. Highly Reactive Organolanthanides − Systematic Routes To And Olefin Chemistry Of Early and Late Bis(Pentamethylcyclopentadienyl) 4f Hydrocarbyl And Hydride Complexes. *J. Am. Chem. Soc.* **1985**, *107*, 8091−8103.
292	Corey, E. J.; Bakshi, R. K.; Shibata, S. Highly Enantioselective Borane Reduction Of Ketones Catalyzed By Chiral Oxazaborolidines − Mechanism And Synthetic Implications. *J. Am. Chem. Soc.* **1987**, *109*, 5551−5553.

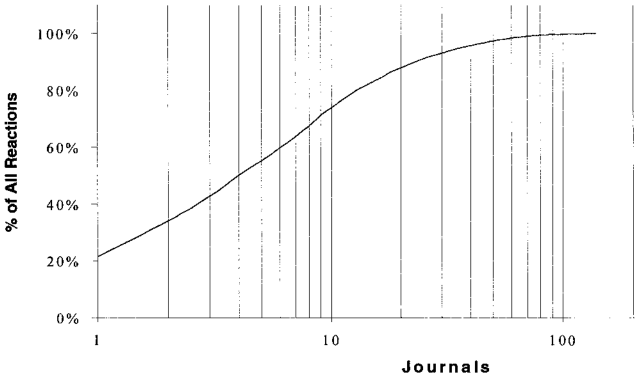

Figure 2. Number of reactions vs journals covered in reaction center.

one or more articles with new synthetic techniques in 1996. Of these, 43 had two or one article covered.

The CASREACT database, summarized in Table 2, gave results similar to those of the Reaction Center. Although CASREACT nominally covers over 9000 sources,[11] only 261 actually reported reactions covered by CASREACT in 1996. Of these, 90 reported only one reaction.

ChemInformRX covers fewer journals than the other two databases, 87, but extracts more articles per journal from them, resulting in about the same total number of articles abstracted. All but three journals covered more than one reaction.

The reactions abstracted per article were examined in detail for the Reaction Center. The journals focusing on synthetic chemistry tend to have the most articles contributed and the most novel reactions per article. Patents, although the number two source of reactions, have fewer reactions per patent. After the top 10 articles, both the articles per journal title and the reactions drop quickly to one reaction per article.

Figure 2 shows the concentration of reactions by journal. These data show the concentration to be similar to those of the articles themselves, even though the top journals tend to report more reactions per article as shown in Table 4.

An examination of the impact factors, the total number of citations to articles in a journal divided by the number of articles published in the journal, for organic chemistry journals shows that indeed the journals selected by all three databases are the highest impact journals in the field.

It is interesting to note the most cited articles in the ISI Reaction Center. The data in Table 5 is for cumulative citations 1985−1996. While most of the reactions are in the earlier years of the data, the 1992 Sharpless article stands out with a large number of citations in a short period of time and has continued to receive a large number, reaching over 500 by early 1999.

This table points out that the most-cited articles in organic synthesis tend to have a strong inorganic or organometallic component. The number of reactions reported involving inorganic or organometallic components has steadily risen to 17% of all reactions in the Reaction Center, indicating the increasing importance of these reagents in organic synthesis.

CONCLUSIONS

By the consensus of three major producers of reaction databases, over 95% of new synthetic methods are reported in a core of about 100 sources. The distribution of information among journals is consistent with those seen in other fields, where there is a diminishing return after the core set of sources is reached.

It is fascinating to note that with the explosion of journals in the past 25 years, seven of the top 10 journals of organic chemistry in 1973 are still in the top 10 list for 1996. The movements in the list reflect the long-standing trend toward rapid communication journals, moving *Chemische Berichte*, *Bulletin De La Societe Chimique de France*, and *Zhurnal Obshchei Khimii* from the list of top contributors.

REFERENCES AND NOTES

(1) Bradford, S. C. *Documentation*; Public Affairs Press: Washington, DC, 1950.
(2) *Journal Citation Reports*; Institute for Scientific Information: Philadelphia, 1996.
(3) Garfield, E.; Revesz, G.; Batzig, J. The Synthetic Chemical Literature from 1960 to 1969. *Nature* **1973**, *242*, 307−309.
(4) Institute for Scientific Information, 3501 Market Street, Philadelphia, PA 19104.
(5) Chemical Abstracts Service, 2540 Olentangy River Road, Columbus, OH 43210.
(6) Blake, J. E.; Dana, R. C. CASREACT − More Than A Million Reactions. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 394−399.
(7) FIZ Chemie Berlin GmbH, Postfach 12 03 37, D-10593 Berlin, Germany.
(8) Parlow, A.; Weiske, C.; Gasteiger, J. Cheminform − An Integrated Information-System On Chemical-Reactions. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 400−402.
(9) Groos, O. V. Bradford's law and the Keenan-Atherton Data. *Am. Doc.* **1967**, *18*, 46.
(10) Garfield, E. Bradford's Law and Related Statistical Patterns. *Curr. Contents* **1980**, *19*, 5−12.
(11) http://info.cas.org/ONLINE/DBSS/casreactss.html, revised 11/97.

CI990001E