Journal of
Applied Crystallography
ISSN 0021-8898

Received 3 November 2000
Accepted 14 February 2001

A citation analysis of the Cambridge Crystallographic Data Centre

Jane Redman,[a] Peter Willett,[a]* Frank H. Allen[b] and Robin Taylor[b]

[a]Department of Information Studies, University of Sheffield, Western Bank, Sheffield S10 2TN, England, and [b]Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge CB2 1EZ, England. Correspondence e-mail: p.willett@sheffield.ac.uk

Citation analysis has been widely used to quantify the influence of research articles on the development of science. This paper reports a citation analysis of ten highly cited papers associated with the Cambridge Crystallographic Data Centre (CCDC), covering the variation of citation with time, the journals in which citations occur, and the types of organization and the geographic regions that use the Cambridge Structural Database. The ten most highly cited papers, comprising four database descriptions (CSD), two geometrical tabulations (TAB) and four basic research papers (RES), received a total of 8494 citations over the period 1981–1998, with more than half of these citations occurring in the literature published from 1995 onwards. The high citation rates of the database descriptions (3573 of 8494) indicate the value of crystallographic data. However, the large number of citations of the geometrical tables (3172) and the research papers (1767) indicate that this value resides not just in the raw data held in the Cambridge Structural Database, but also in the structural knowledge that can be derived from it. In the most recent years covered by the analysis (1995–1998), these ten CCDC publications have received more than 1000 citations per annum (CSD 507, TAB 398 and RES 153 citations per annum) and the detailed analysis shows that these papers, and the data that they discuss, are used not only by crystallographers but also by researchers across the entire range of the chemical sciences.

1. Introduction

The Cambridge Crystallographic Data Centre was established in 1965 to record the primary numerical, chemical and bibliographic data for published X-ray and neutron diffraction studies of small organic and metallo-organic molecules. Its principal product, the Cambridge Structural Database (Allen *et al.*, 1991; Allen & Kennard, 1993), now contains details of more than 225 000 structures, and is used by academic and industrial scientists in 46 countries worldwide. Allied to database creation, the CCDC has developed software for search, retrieval and display of CSD information, and has always carried out a programme of basic research and development based on information derived from the CSD. A key research interest has been to contribute to the development of novel methodologies, and related computer programs, for the analysis of retrieved data (see *e.g.* Taylor & Allen, 1994). More recently, the CCDC has begun to construct and distribute knowledge-bases of structural information derived from CSD data (see *e.g.* Bruno *et al.*, 1997; Allen, 1998), and to develop applications software that makes use of these knowledge-bases to address specific problems in structural chemistry and rational drug design (Jones *et al.*, 1995, 1997; Verdonk *et al.*, 1999).

Over the years, the CCDC has published several hundred papers which may be classified into five broad categories: (*a*) descriptions of the developing CSD and its associated software (the CSD System), (*b*) research methodologies, (*c*) scientific research applications of CSD data, (*d*) tabulations of structural information derived from the CSD, and (*e*) crystal structure analyses related to the research interests of the CCDC. Further, as the use of the CSD has grown over time, the CCDC has also compiled a database of published research applications: papers for which systematic use of the CSD has made a very substantial contribution to the published results. This citations database, DBUse, has been distributed as part of the CSD System since 1995, and currently contains nearly 800 references. DBUse is now also freely available *via* the CCDC's Web site (http://www.ccdc.cam.ac.uk).

Now that the CCDC has been established for nearly 40 years, it is of interest to analyse the impact on scientific research of the wide availability of crystal structure information in a readily accessible and organized form. Simple statistics on the published research output of the CCDC are of somewhat parochial, if not introverted interest, while analysis of the contents of the DBUse database should wait until this compilation is somewhat larger. However, over the past few years, citation analysis (Cronin, 1984; Garfield, 1979; Liu, 1993) has been widely used to quantify the influence of research articles on scientific development, and although the CSD is widely used by both academic and industrial organizations, we are unaware of any attempt, thus far, to quantify the extent of its use to support scientific research.

The *Science Citation Index*, produced by the Institute for Scientific Information (ISI, http://www.isinet.com), contains all the end-of-article citations in papers published in more than 5000 high-impact scientific journals since 1974. The availability of this database, and

similar ones covering the social sciences, arts and humanities, has spurred interest in the use of such citation data for quantitative evaluations of the research productivity of groups of scientists, such as those working in an individual university, institution or department (see *e.g.* Bradley *et al.*, 1991; Kim & Kim, 2000; Martin *et al.*, 1987; Nederhof & Noyons, 1992). Although there are limitations in the use of such data for this purpose, owing to the many reasons for which a citation may be made, citation analyses have been shown to correlate well with peer reviews of the importance of scientific research activity (see *e.g.* Cronin & Overfeld, 1994; Oppenheim, 1997; Virgo, 1977). Similar methods are now being applied to the hypertext links between pages on the Web, on the basis that such 'sitations' are analogous to citations in the conventional printed literature (Ingwersen, 1998).

This paper therefore reports a detailed bibliometric study of key CCDC publications, citation of which may be taken as a measure of the scientific value of crystal structure analyses in general, and of crystallographic database resources in particular, to the wider scientific community.

2. Methodology

The citation data for this study, performed in 1999, came from the version of the Science Citation Index database that was available at that time to UK Higher Education institutions *via* the Bath Information Data Services (BIDS, at http://www.bids.ac.uk). The BIDS service covers the SCI database from 1981 to date (with abstracts included from 1991); the citation data used here were collected over the period 7–21 June 1999.

At that time, the CCDC's DBUse database recorded some 135 articles written by CCDC staff since 1973, either on their own or in collaboration with non-CCDC authors. Searching BIDS for all of these articles identified 10681 citations since 1981. As is often the case in bibliometric analyses, the great majority of these citations resulted from a small fraction of the total number of papers, and the decision was taken to select just ten of the papers for further detailed analysis. These ten papers, which between them accounted for 8494 citations, are representative of three of the five categories of publications identified above. These ten articles are summarized in Table 1 and are categorized as follows: CSD1–CSD4 are articles dealing with the CSD and its associated software and services (3573 citations); TAB1, TAB2 are compilations of tabulated structural data derived from the database (3172 citations); RES1–RES4 are papers describing basic research carried out using CSD data (1767 citations, out of a total of 3936 citations to all of CCDC's basic research output since 1981).

The full bibliographic data for each of the 8494 citations were downloaded from BIDS and analysed locally using Microsoft Word and Excel software, although most of the analyses involved the 4334 post-1994 citations (Redman, 1999). The analyses carried out on this large body of bibliographic data included: (*a*) the growth in citations over time, (*b*) the journals in which citations appear, (*c*) the types of organization that cite the selected articles, and (*d*) the geographical locations of the citing author(s).

Table 1
Articles selected for the detailed citation analysis.

Label	Title (reference)	Citations	
		Total	Post-1994
Database descriptions (CSD)		3573	2029
CSD1	Cambridge Crystallographic Data Centre: computer-based search, retrieval, analysis and display of information (Allen *et al.*, 1979)	1098	298
CSD2	Systematic analysis of structural data as a research tool in organic chemistry (Allen *et al.*, 1983)	974	362
CSD3	The development of versions 3 and 4 of the Cambridge Structural Database System (Allen *et al.*, 1991)	731	601
CSD4	3D search and research using the Cambridge Structural Database (Allen & Kennard, 1993)	770	768
Tables of standard geometry (TAB)		3172	1662
TAB1	Tables of bond lengths determined by X-ray and neutron diffraction. Part 1. Bond lengths in organic compounds (Allen *et al.*, 1987)	2386	1200
TAB2	Tables of bond lengths determined by X-ray and neutron diffraction. Part 2. Organometallic compounds and coordination complexes of *d*- and *f*-block metals (Orpen *et al.*, 1989)	786	462
Research articles (RES)		1767	643
RES1	The molecular structures of nucleosides and nucleotides. Part 1. The influence of protonation on the geometries of nucleic acid constituents (Taylor & Kennard, 1982*b*)	208	46
RES2	Crystallographic evidence for the existence of C—H···O, C—H···N and C—H···Cl hydrogen bonds (Taylor & Kennard, 1982*a*)	981	377
RES3	Geometry of the N—H···O═C hydrogen bond. 3. Hydrogen-bond distances and angles (Taylor *et al.*, 1984)	256	93
RES4	Hydrogen-bond geometry in organic crystals (Taylor & Kennard, 1984)	322	127
Totals		8494	4334

3. Results and discussion

3.1. Summary statistics

The 135 papers published by CCDC authors and considered here have together received 593 citations per year over the 18-year period covered by this analysis. Of these, 472 citations per year are accounted for by the ten most highly cited papers (Table 1), and the CSD and TAB papers, taken together, contribute 375 citations per year. Recent annual citation rates are, of course, much higher than this since many of these papers were not in print for the complete citation analysis period. Thus, while CSD1 received an average of 61 citations per year since 1981, CSD3 and CSD4 have received an average of 91 and 128 citations per year over their 8- and 6-year citation lifetimes, respectively. Similarly, TAB1 and TAB2 have received an average of 197 and 79 citations per year over their respective 12- and 10-year citation lifetimes. Average citation frequencies over the last 3 years of the analysis, 507 per annum for CSD1–CSD4 (taken together), 398 per annum for TAB1+TAB2 and 153 per annum for RES1–RES4 (taken together) are the best measure of current citation activity. The overall average citation rate for the ten papers of Table 1 in this most recent 3-year period is therefore 1058 per annum.

3.2. Growth of citations over time

The year of publication of each citing article was recorded so as to plot the frequency of citation over time. A peaked distribution (Fig. 1*a*) is seen for the earlier database descriptions CSD1–CSD3. Citations to CSD1 and CSD2 both peaked in 1993, some 14 years after publication of CSD1 and 10 years after publication of CSD2. By contrast, the 1991 update paper, CSD3, addressed a more established user base and reached a citation peak within 5 years of publication. CSD4, which is still relatively new in citation terms, has a citation frequency that is still increasing. Most

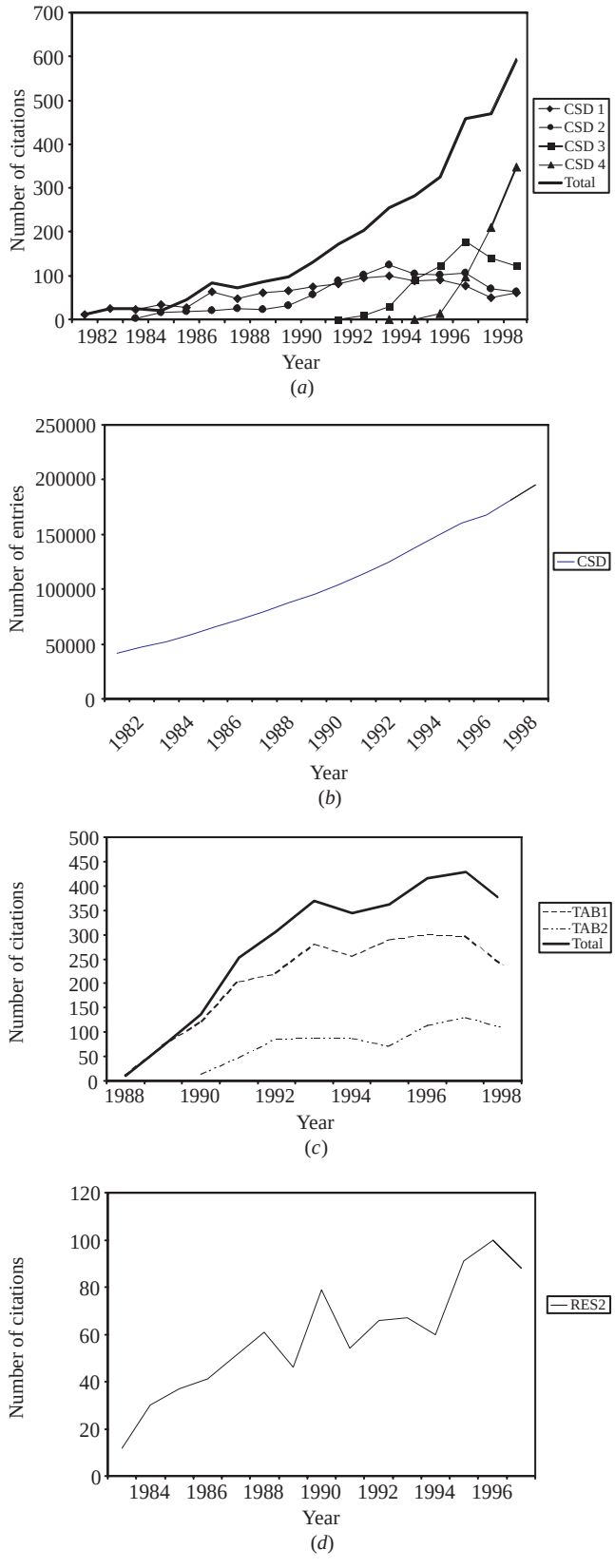

Figure 1
(*a*) Citation frequencies for the period 1981–1998 for the four database-description papers CSD1–CSD4. (*b*) Growth of the Cambridge Structural Database 1981–1998. (*c*) Citation frequencies for the period 1988–1998 for the geometrical-tabulation papers TAB1 and TAB2, and for their citation sums over this period. (*d*) Citation frequencies for the research paper RES2 over the period 1983–1998.

revealing is the plot of citation over time that uses the summed frequencies of CSD1–CSD4, also shown on Fig. 1(*a*). This plot shows a steady increase from 12 citations in 1981 to 132 citations in 1990, followed by a very rapid increase through the 1990s to the nearly 600 citations recorded in 1998.

At least three factors have contributed to the steady rise in the total citation frequency identified in Fig. 1(*a*): (i) the increasing international distribution of the CSD, particularly during the 1980s and early 1990s, (ii) the general growth in the rate of scientific publications, and (iii) the increasing productivity of crystallographers, who tend to provide the majority of citations to the CSD in their structural papers. Indeed, there is a clear similarity in the increase in total citation frequency of CSD1–CSD4 and the overall growth pattern of the CSD itself during the period 1981–1998, shown in Fig. 1(*b*).

The two tabulations of standard bond lengths, TAB1 and TAB2, show rather similar citation behaviour to each other, as shown in Fig. 1(*c*): a rapid rise over the first 3–4 post-publication years followed by a levelling off since then at 200–300 citations per year for TAB1 and 90–130 citations per year for TAB2. The plot of summed frequencies (Fig. 1*c*) shows that total citations to TAB1 and TAB2 have exceeded 300 per annum since 1992, just 5 years after publication of TAB1. There are no signs of a long-term decrease in citation levels over the past 5 years in either case, although the slight reduction in citation levels in 1998 is not readily explicable.

All of the four RES papers (Table 1) were published in the early 1980s, but their citation history is remarkably constant over time, albeit at rather different levels. Thus RES1, dealing with intramolecular geometry, has received an average of 13 citations per year over the 16-year period 1983–1998, and within the range from 7 to 21 citations annually. The three other papers, RES2–RES4, all deal with hydrogen bonding and show the citation behaviour exemplified by Fig. 1(*d*) for RES2. There is an initial rise to a plateau that spanned the years 1987–1994, followed by a significant further rise to a peak in the late 1990s. Two factors account for this behaviour: (i) the rise in importance of supramolecular chemistry, the modelling of protein–ligand docking modes, crystal engineering and related topics, and (ii) the inclusion of improved and systematic intermolecular nonbonded search procedures within the released CSD software from 1993 onwards. This increased interest in hydrogen bonding of all types, particularly the weaker hydrogen bonds and, indeed, in intermolecular interactions not mediated by hydrogen, is exemplified by the publication of two major monographs during the 1990s (Jeffrey & Saenger, 1991; Desiraju & Steiner, 1999).

3.3. Analysis by journal titles

Data relating to the number and titles of journals that have cited at least one of the papers of Table 1 since 1995 were collated. A total of 322 journals generated 4334 citations, with the top 13 journals (Table 2) generating more than 50% (2203) of these citations. Only three of these journals are purely crystallographic. The remaining ten are chemistry journals, with six of these specializing in inorganic and metallo-organic chemistry. We note here that the crystallographic databases are cited in two different ways: by use of an appropriate literature paper describing that database, or simply by citing the database itself as *e.g.* Cambridge Structural Database, Version 5.19 of April 2000, usually followed by the address of the CCDC. References of the latter form, which are frequent, were not picked up in this analysis.

To obtain further insight into the impact of the CSD in different areas of science, the 322 citing journals were grouped into four broad

Table 2
Top-ranked citing journals for the ten selected articles.

Journal	Citations				
	CSD	TAB	RES	Total	%
Acta Crystallographica Section C: Crystal Structure Communications	351	264	53	668	15.4
Inorganic Chemistry	93	104	19	216	5.0
Organometallics	58	112	15	185	4.3
Journal of the American Chemical Society	86	57	39	182	4.2
Journal of the Chemical Society, Dalton Transactions	79	78	9	166	3.8
Acta Crystallographica Section B: Structural Science	79	38	27	144	3.3
Journal of Organometallic Chemistry	44	91	8	143	3.3
Inorganica Chimica Acta	36	55	14	105	2.4
Journal of Chemical Crystallography	29	51	11	91	2.1
Journal of Molecular Structure	48	16	18	82	1.9
Journal of the Chemical Society, Perkin Transactions 2	41	21	15	77	1.8
Polyhedron	41	24	9	74	1.7
Russian Chemical Bulletin	8	58	4	70	
Totals	993	969	241	2203	50.8

subject categories: biosciences (biology, biochemistry and biophysics), chemistry, crystallography, and other. The numbers of journals and numbers of citations for the CSD, TAB and RES papers are summarized in Table 3(*a*), which is ordered on the number of citations per journal, *k*, in each category. Inspection of Table 3(*a*) reveals that the 14 specifically crystallographic journals provide no less than 24.2% of the total number of citations, and have a *k* value of 75.0. These data reflect the greater awareness and use of the crystallographic databases amongst crystallographers, and a higher tendency for these journals to encourage or require a literature citation for the appropriate database when it is used.

The impact of the CSD in chemistry can be considered as broad rather than deep: a total of 223 journals contribute nearly 70% of all citations, but with a *k* value of 13.6 citations per journal. By contrast, uptake of the CSD in the biosciences is, as yet, rather sporadic: 238 citations, 5.5% of the total, are spread across 72 (22.4%) of all citing journals (k = 3.3). This latter result may well represent a false minimum: many protein crystal structures are refined using geometrical constraints derived (Engh & Huber, 1990) from the atomic resolution data in the CSD, but authors are most likely to cite the software package used for refinement, or the original non-CCDC compilers of the geometrical constraint values. Also, in future, when increasing numbers of high-resolution biological structures become available, comparisons with existing small-molecule results will become increasingly relevant.

The largest citation category, chemistry, was further divided into five journal-type subcategories: general, inorganic and organometallic, organic, physical, and pharmaceutical chemistry, and the corresponding data for these 223 journals and 3023 citations are shown in Table 3(*b*), again ordered by *k* values. Table 3(*b*) shows that one-third of all chemistry citations arise from a small number (24) of inorganic and organometallic chemistry journals, and the *k* value of 41.9 citations per journal begins to rival that observed for the purely crystallographic journals. Citations from the 178 general, organic and physical chemistry journals are much more broadly based with *k* values ranging from 7.3 to 16.6, but citations in pharmaceutical chemistry are sparse.

A number of factors contribute to the apparently disproportionate use of the CSD in organometallic and metal complex chemistry. First, crystallography is the analytical tool of choice in this area, and the specialist journals publish very significant numbers of crystal structures that contribute to the growth of the CSD (which now comprises about 50% of metallo-organic structures). Second, this area of chemistry is still developing, and has not reached the maturity of structural knowledge that is available in organic chemistry. Thus to advance further, the synthesis of very large numbers of novel metallo-organic molecules requires that their structures be compared carefully with existing benchmarks recorded in the CSD.

So far, journals in the biosciences and pharmaceutical chemistry areas belong to that long tail of journals in which citations to the ten selected CCDC papers occur only rarely. Indeed, there were no less than 116 journals that provided just a single citation in the period under review, this list including such diverse publications as the *American Journal of Physiology*, *Experimental Parasitology*, *Journal of the Society of Dyers and Colourists*, *Microporous Materials*, *Pesticide Science*, *Perspectives in Computing*, and *Solid State Communications*. The huge range of disciplines represented by journals such as these nevertheless demonstrates clearly the wide-

Table 3
Analysis of citations by subject area of the articles which have cited the ten papers of Table 1 since 1995.

(*a*) Subject categorization by type of journal

Category	Number of journals making citation					Number of citations					
	CSD	TAB	RES	$n(J)$	$\%(J)$	CSD	TAB	RES	$n(C)$	$\%(C)$	$k = n(C)/n(J)$
Crystallography	14	12	9	14	4.3	523	399	128	1050	24.2	75.0
Chemistry	156	130	110	223	69.3	1352	1229	442	3023	69.8	13.6
Biosciences	53	20	26	72	22.4	139	31	68	238	5.5	3.3
Other	9	3	4	13	4.0	15	3	5	23	0.5	1.8
Totals	232	165	149	322	100.0	2029	1662	643	4334	100.0	

(*b*) Chemistry subcategorization by type of journal.

Subcategory	Number of journals making citation					Number of citations					
	CSD	TAB	RES	$n(J)$	$\%(J)$	CSD	TAB	RES	$n(C)$	$\%(C)$	$k = n(C)/n(J)$
Inorganic and organometallic	19	19	12	24	10.8	398	533	74	1005	33.2	41.9
Organic	17	16	10	21	9.4	153	142	53	348	11.5	16.6
General	68	68	57	108	48.4	542	464	214	1220	40.4	11.3
Physical	34	21	24	49	22.0	199	75	86	360	11.9	7.3
Pharmaceutical	18	6	7	21	9.4	60	15	15	90	3.0	4.3
Totals	156	130	110	223	100.0	1352	1229	442	3023	100.0	

spread impact of the CSD. The spread of journals is particularly noticeable if one considers that citations to the four RES papers, with their 643 post-1995 citations (out of a total of 4334, *i.e.* 14.8%) appear in no less than 149 distinct journals (out of a total of 322, *i.e.* 46.3%) identified in the analysis.

3.4. Analysis of corporate sources

The citation data relating to the corporate source of the citing articles are analysed by first classifying each corporate source into one of three major categories: non-commercial (academy, institute, university), commercial and 'other'. Only the first-named corporate source was used in this study, even though there were often two or more such sources, *e.g.* for a paper describing an academic–commercial collaboration. The non-commercial corporate sources clearly had one of the terms 'Academy', 'Institute' or 'University' appearing prominently in their name. Different departments within the same university or academy were isolated and treated as distinct corporate sources. 'National Research Centres' and 'National Laboratories' were classified as 'other'. Commercial organizations were generally recognized as such by the terms 'Co', 'Ltd', 'Corp.' or similar after their name. In ten cases, insufficient information was supplied by the citing authors, or abbreviations or locally accepted acronyms were cited in the corporate name. These few unknown sources were simply added to the 128 corporate sources in the 'other' category. The CCDC itself was treated as a singleton source.

A total of 1254 different corporate sources generated the set of 4334 post-1995 citations. The distribution across source categories is shown in Table 4, which is ordered by k_1, the number of citations generated by each category. On this basis, it is no surprise that the CCDC tops the list through self-citation of its own papers! However, the number of self-citations in this period (49, or 1.1% of the total) is insufficient to perturb the statistics for other corporate sources.

The vast bulk of the citations (89.2%, Table 4) comes from the non-commercial sector, with a mean number of citations per corporate source (k_1) of 3.7 that shows little variation (from 3.1 to 4.1) for the various types of non-commercial institutions. This value is 1.7 times higher than that for the 'other' sources and 2.3 times higher than the k_1 value for commercial sources. These differences are unsurprising: it is the norm for academic researchers to publish in the open literature, while their colleagues in the commercial sector are constrained by the patent-sensitivity and company confidentiality of their results. Nevertheless, the number of commercial sources may be underestimated by this analysis, since in an academic–commercial collaboration the first-named institution (used to compile the data used here) is most likely to be the academic laboratory, where the bulk of the hands-on work is most likely to have been carried out. Of the commercial corporate sources that were identified, pharmaceutical companies were the most prominent market sector, reflecting both the industry's need for detailed three-dimensional structural information to support projects in a wide variety of computational chemistry applications, e.g. protein–ligand docking, molecular modelling, *etc.*, and the CCDC's long-standing information provision to many companies in this sector.

A further classification of corporate sources according to the type of research associated with that source, *e.g.* biosciences, chemistry, crystallography, physics, *etc*., was attempted. Even though nearly all two-way source:subject combinations generated citations, the results merely confirmed those in Table 4: the combination of academic sources specializing in chemistry dominated the two-way classification.

3.5. Analysis by geographical regions

The CSD system is distributed to academics in 46 countries worldwide and to more than 100 commercial organizations, principally pharmaceutical companies located in the USA, Western Europe and Japan. These sites (academic plus commercial) are sub-divided geographically in the first two columns of Table 5, with the remaining columns summarizing the post-1995 citation data for the ten key articles of Table 1 generated by authors from the various geographic regions, and publishing their work across all journals, both local and international. Table 5 is ordered on the number of citations generated per site, k_2, in each region.

Table 5 shows that authors from Western Europe plus the UK together generate more than half of the citations from only 34.7% of the sites. When Eastern Europe is also included, the complete European region generates 65.3% (2831) of citations from 42.5% (333) of all sites, giving a mean k_2 value of 8.5 citations per site. By

Table 4
Citation counts for different types of corporate source.

Type of corporate source	Number of sources $n(S)$	$\%(S)$	Number of citations $n(C)$	$\%(C)$	$k_1 = n(C)/n(S)$
CCDC	1	0	49	1.1	49.0
Non-commercial	1041	83.0	3868	89.2	3.7
Academy	60	4.8	248	5.7	4.1
University	853	68.0	3220	74.3	3.8
Institute	128	10.2	400	9.2	3.1
Other	138	11.0	296	6.8	2.1
Commercial	74	5.9	121	2.8	1.6
Totals	1254	99.9	4334	99.9	

Table 5
Analysis of citations by geographical location of the citing author(s).

	n(sites)	%	CSD	TAB	RES	$n(C)$	%	$k_2 = n(C)/n$(sites)
UK	53	6.8	338	174	52	564	13.0	10.6
Eastern Europe	77	9.8	231	334	75	640	14.8	8.3
Western Europe	203	25.9	806	565	256	1627	37.5	8.0
North America	212	27.0	430	274	148	852	19.7	4.0
Rest of the world	105	13.4	141	141	45	327	7.5	3.1
Far East	134	17.1	833	174	67	324	7.5	2.4
Totals	784	100.0	2029	1662	643	4334	100.0	

contrast, the region with the largest number of sites, North America, generates less than 20% of the citations from its 27% of sites, yielding a k_2 value of 4.0, less than half the European value. Comparative data from the Far East and the 'rest of the world' show even lower values of k_2: taken together, these regions generate just 15% (651) citations from 30.5% (239) of the sites, giving a k_2 value of 2.7 citations per site.

It is tempting to suggest that these activity differentials reflect increased proximity to the Cambridge Centre, and increased exposure to the CSD and its research applications at meetings within the UK and Europe. However, several other factors (at least) must also be taken into account, for example: (*a*) the actual number of users per site may differ from region to region, and is not known to the CCDC; (*b*) the relative levels of small-molecule crystallographic research activity, and indeed its chemical focus towards organic or metallo-organic chemistry, may have geographical variations; (*c*) levels of funding for scientific research, particularly at the student level, will also show global variations, with database analysis perhaps being seen as a cost-effective basis for research activities in some countries.

4. Conclusions

In absolute terms, the high citation rates of CSD1–CSD4 indicate the value of crystallographic data in general and of crystallographic databases in particular, and the longevity of several of the selected publications is a reflection of the definitive nature of databases such as the CSD, and of tabulations of structural metrics derived from them. Indeed, the large number of citations to the TAB and RES papers indicates that this value resides not just in the raw data, but also in structural knowledge that can be derived from that data. The success of TAB1 and TAB2, in particular, suggests that other similar geometry compilations would be popular, although many users of these compilations have suggested that updates should appear in a searchable electronic form. This is now being fulfilled by the availability of *IsoStar* (Bruno *et al.*, 1997), the CCDC's knowledge-base of intermolecular interactions, and by the ongoing development of Mogul, a knowledge-base of intramolecular geometry derived from the CSD and due for release as part of the distributed CSD System in 2001. Such computer-based products, updated on a regular basis as the CSD grows, have obvious and significant advantages over the static paper-based compilations of the 1980s.

References

Allen, F. H. (1998). *Acta Cryst.* A**54**, 758–771.
Allen, F. H., Bellard, S., Brice, M. D., Cartwright, B. A., Doubleday, A., Higgs, H., Hummelink, T. W. A., Hummelink-Peters, B. G., Kennard, O., Motherwell, W. D. S., Rodgers, J. R. & Watson, D. G. (1979). *Acta Cryst.* B**35**, 2331–2339.
Allen, F. H., Davies, J. E., Galloy, J. J., Johnson, O., Kennard, O., Macrae, C. F., Mitchell, E. M., Mitchell, G. F., Smith, J. M. & Watson, D. G. (1991). *J. Chem. Inf. Comput. Sci.* **31**, 187–204.
Allen F. H. & Kennard, O. (1993). *Chem. Des. Autom. News*, **8**, 1, 31–37.
Allen, F. H., Kennard, O. & Taylor, R. (1983). *Acc. Chem. Res.* **16**, 146–153.
Allen, F. H., Kennard, O., Watson, D. G., Brammer, L., Orpen, A. G. & Taylor, R. (1987). *J. Chem. Soc. Perkin Trans. 2*, pp. S1–S19.
Bradley, S. J., Willett, P. & Wood, F. E. (1991). *J. Inf. Sci.* **18**, 225–232.
Bruno, I. J., Cole, J. C., Lommerse, J. P. M., Rowland, R. S., Taylor, R. & Verdonk, M. L. (1997). *J. Comput.-Aided Mol. Des.* **11**, 525–537.
Cronin, B. (1984). *The Citation Process: the Role and Significance of Citations in Scientific Communication.* London: Taylor Graham.
Cronin, B. & Overfeld, K. (1994). *J. Am. Soc. Inf. Sci.* **45**, 64–72.
Desiraju, G. R. & Steiner, T. A. (1999). *The Weak Hydrogen Bond*. Oxford University Press.
Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–398.
Garfield, E. (1979). *Citation Indexing: its Theory and Applications in Science, Technology and the Humanities*. New York: Wiley.
Ingwersen, P. (1998). *J. Docum.* **54**, 236–243.
Jeffrey, G. A. & Sänger, W. (1991). *Hydrogen Bonding in Biological Structures.* Berlin: Springer Verlag.
Jones, G., Willett, P. & Glen, R. C. (1995). *J. Mol. Biol.* **245**, 43–53.
Jones, G., Willett, P., Glen, R. C., Leach, A. R. & Taylor, R. (1997). *J. Mol. Biol.* **267**, 727–748.
Kim, M.-J. & Kim, B.-J. (2000). *J. Inf. Sci.* **26**, 111–119.
Liu, M. (1993). *J. Docum.* **49**, 370–408.
Martin, B. R., Irvine, J., Narin, F. & Sterrit, C. (1987). *Nature (London)*, **330**, 123–126.
Nederhof, A. J. & Noyons, E. C. (1992). *J. Am. Soc. Inf. Sci.* **43**, 249–256.
Oppenheim, C. (1997). *J. Docum.* **53**, 417–487.
Orpen, A. G., Brammer, L., Allen, F. H., Kennard, O., Watson, D. G. & Taylor, R. (1989). *J. Chem. Soc. Dalton Trans.* pp. S1–S83.
Redman, J. (1999). *A Citation Analysis Study of the Cambridge Crystallographic Data Centre (CCDC) to Determine the Extent and Range of Uses of their Data as Reported in the Scientific Literature*, MSc dissertation, University of Sheffield.
Taylor, R. & Allen, F. H. (1994). *Structure Correlation*, edited by H.-B. Bürgi & J. D. Dunitz, pp. 111–161. Weinhiem: VCH.
Taylor, R. & Kennard, O. (1982*a*). *J. Am. Chem. Soc.* **104**, 5063–5070
Taylor, R. & Kennard, O. (1982*b*). *J. Mol. Struct.* **78**, 1–28.
Taylor, R. & Kennard, O. (1984). *Acc. Chem. Res.* **17**, 320–326
Taylor, R., Kennard, O. & Versichel, W. (1984). *Acta Cryst.* B**40**, 280–288.
Verdonk, M. L., Cole, J. C. & Taylor, R. (1999). *J. Mol. Biol.* **289**, 1093–1108.
Virgo, J. (1977). *Lib. Quart.* **47**, 415–430.