PERSPECTIVE

The *Journal of Chemical Documentation* and the *Journal of Chemical Information and Computer Sciences*: Publication and Citation Statistics

Rabab Al Jishi and Peter Willett*

Information School, University of Sheffield, 211 Portobello Street, Sheffield S1 4DP

Received September 27, 2010

INTRODUCTION

As Agrafiotis et al. note, "Chemoinformatics is a vast discipline, standing on the interface of chemistry, biology, and computer science,"[1] and chemoinformatics research is hence reported in many different fora. Of these, one particular publication stands out as being the core journal for the discipline:[2] this is the *Journal of Chemical Information and Modeling* (hereafter normally abbreviated to JCIM), a peer-reviewed, English-language journal published by the American Chemical Society (ACS). JCIM started life in 1961 as the *Journal of Chemical Documentation* (hereafter JCD), changed its name to the *Journal of Chemical Information and Computer Sciences* (hereafter JCICS) in 1975, and adopted its present title in 2005.

The journal is currently celebrating the 50th anniversary of its founding in 1961: it is hence timely to review its development over the intervening years, and this paper reports a study of the Journal using the methods of bibliometrics.[3–6] We focus here not just on the articles that have appeared in the Journal but also, and arguably more importantly, on the citations to those articles, since citations are widely regarded as reflecting the impact of scientific research, whether of a specific article, of a specific individual or, as here, of a specific journal.[7–10]

Bibliometric analyses of specific journals are not uncommon, with a recent review by Anyi et al. discussing no less than 82 such studies that had been published in the period 1998−2008 across a wide range of disciplines.[11] Willett has previously reported bibliometric analyses of two chemoinformatics journals—*Journal of Molecular Graphics and Modeling*[12] and *QSAR & Combinatorial Science*[13]—and there have also been two publications that focus on JCICS. In 2001, Onodera discussed the changes that had taken place in the *Chemical Abstracts* indexing terms assigned to articles from the Journal, noting that an initial focus on the representation and searching of chemical substances had broadened to include subjects such as property prediction and molecular modeling.[14] Then, in 2005, Warr discussed the most important papers in JCICS (where "important" was defined as attracting at least 100 citations in the period 1997−2005) as part of a review of the historical development of chemoinformatics.[15] In this review, she highlighted many individual articles that had contributed significantly to the discipline, with the largest number of these having appeared in either JCD or JCICS; analogous historical reviews and lists of significant papers have since been published by Chen[16] and by Willett.[17]

In this paper, we present a publication and citation analysis of the articles that appeared in the Journal in the period 1961−2004 (i.e., in JCD and JCICS), the publications representing the content of the Journal (and hence of the discipline that the Journal is seeking to describe), and the citations representing the impact of that content on the broader academic community. We also discuss JCIM, but only very briefly since its constituent articles have had only a few years in which they have been able to accrue citations; however, JCIM is referred to frequently in the text. The data presented and discussed below were obtained using the Thomson Reuters Web of Science (WoS) system, this comprising the *Science Citation Index - Expanded*, the *Social Sciences Citation Index*, the *Arts & Humanities Citation Index*, and the *Conference Proceedings Citation Index - Science* files. Use was also made of the subject categories in the Web of Knowledge (WoK) system, of which WoS is an important part. The database searches were carried out in the period March−July 2010, and the outputs analyzed using the Analyze Results and Citation Reports tools in WoS.

DEVELOPMENT OF THE JOURNAL

The ACS Division of Chemical Literature was established in 1948 to provide a medium for exchanging news and views on the various aspects of chemical documentation via technical meetings and papers.[18] The latter appeared in a range of journals such as *Journal of Chemical Education*, *Chemical & Engineering News*, and *American Documentation* (now the *Journal of the American Society for Information Science and Technology*). The idea that the ACS should launch a new journal was first suggested as early as 1952, and the ACS Board of Directors approved the publication of the new *Journal of Chemical Documentation* in 1961, with Herman Skolnik as its founding editor.[19] On assuming this post,

* To whom all correspondence should be addressed: email p.willett@sheffield.ac.uk.

10.1021/ci100338j
Published on Web 10/14/2010

Skolnik listed three requirements for Division members: to carry out high quality investigations in the field of chemical documentation, to produce carefully written articles for the Journal, and to support it through personal subscription.[18] It was initially planned that two issues be published per year, but the Journal soon moved to quarterly publication (and now publishes monthly), with the first volume containing a total of 59 articles, 40 of which originated in the Division's technical programs. This pattern continued for over a decade, but by the mid-1970s the majority of the published articles had been submitted independently of the Division. Other changes over this period were a move from industrial to academic submissions, and an increasing number of articles from outside the United States.

In the early days, the principal focus of the Journal was the (increasingly computerized) documentation of the chemical literature, both textual and structural. By the mid-1970s, however, technologies such as structure and substructure searching and the use of artificial intelligence methods for synthesis design and structure elucidation had become well-established, and the Journal hence changed its name to JCICS in 1975 (the same year that the Division became the Division of Chemical Information). The reason for this was highlighted by Skolnik in his first editorial after the change: "The change of name in no way is meant to exclude any part of the scope of papers that we have been publishing over the past 14 years. Rather it is to encourage the submittal of papers covering a broader scope, most importantly to include more papers oriented to computer science."[20] Skolnik was succeeded as editor by Thomas Isenhour in 1982, whose August 1982 editorial noted that he was "assuming this position at a time of exponential development in computers, which should result in similar exciting changes in information and computer science." This was indeed the case in chemoinformatics, with the 1980s seeing, e.g., the first operational systems for 3D substructure searching, ligand−protein docking, and graphical, end-user access to the CAS Registry System. None of these developments would have been possible without the enhanced computing facilities that were then becoming widely available. Isenhour was succeeded in 1989 by George ("Bill") Milne, who oversaw a steady growth in the size and range of the Journal. Starting with the second issue of volume 33, JCICS was divided into three sections: "Chemical Information", "Chemical Computation", and "Molecular Modeling", with the journal name now having as a subtitle "Includes Chemical Computation and Molecular Modeling".[21] A further section, "Biopharmaceutical Computing", was added in the fifth issue of volume 41 (by which time publication was bimonthly). In 2004, after 15 very successful years, Milne handed over to the current editor, William L. Jorgensen, who oversaw the second change of name to the current appellation, noting in his first editorial that, "Again, the name change is intended to better reflect the evolving contents of the Journal, while still emphasizing the Journal's premium position in chemical information."[22]

It must be emphasized that this is only a brief account of the Journal's history; more detailed accounts of its early days are provided by Metanomski[18] and by Skolnik.[23]

PUBLICATIONS

Journal of Chemical Documentation. JCD published a total of 55 issues across 14 volumes between January 1961 and November 1974, these issues containing a total of 900 items in WoS. Of these, 723 are described as articles, with the others including meeting abstracts, book reviews, editorial material, etc. Inspection of the individual items shows some variation in the way that these very early publications were categorized, and we have hence treated all 900 items as a whole in what follows, where the items will be described generically as articles. This variability is just one example of the fact that the database system that we now call WoS has changed considerably over the years: for example, there was an increase of ca. 39% in 1991 in the number of journals covered by the database, and many of the early records are missing institutional and address information.

Table 1. Most Productive Authors in JCD

author	articles
Skolnik, H.	41
Lynch, M. F.	15
Bernier, C. L.	12
Garfield, E.	12
Maizell, R. E.	12
Frome, J.	11
Granito, C. E.	11
Weil, B. H.	10
Starker, L. N.	9
Kuney, J. H.	8

A total of 901 authors published in JCD. The 10 most productive authors are listed in Table 1, and this list will not surprise anybody who was familiar with the field and/or the Journal in its early days. The largest number of articles was by Skolnik, with 16 of these being editorials. Next came Lynch, who moved in 1965 from being head of the Basic Research Department at CAS to the Postgraduate School of Librarianship and Information Science (now the Information School) at the University of Sheffield, which was the host institution for all but one of his 15 articles and which has subsequently been the source of many other papers in the Journal over the years. His location in such a department was typical of a time when the library and information science (LIS) community published extensively in JCD (as we discuss further below when considering citations). Thus, looking at other authors in Table 1, Bernier had senior information roles at the Defense Documentation Centre and the National Library of Medicine, Garfield was of course the founder of the Institute for Scientific Information and has been a life-long proponent of citation indexing, Weil was an information and copyright specialist at Exxon, and Maizell was the author of what was, for many years, the standard book about the chemical literature.[24]

Use of the Analyze Results tool to investigate the WoK subject areas in which items had been categorized showed that all but 23 of them had been categorized as belonging to computer science, with mathematical and computational biology (12 items) being the only other category appearing 10 or more times. Turning to location data, all but 12% of the items lack institutional information in the WoS database, and this is also the case when searching on the "Country/Territory" field. However, manual inspection of the items that did have geographical data showed that the overwhelming majority (88%) came from the U.S.A., with only the U.K., Canada, and Japan being the source for 10 or more items.

Table 2. Most Productive Authors in JCICS

author	articles
Randic, M.	67
Willett, P.	61
Jurs, P.C.	55
Trinajstić, N.	44
Heller, S.R.	40
Balaban, A.T.	39
Gutman, I.	36
Katritzky, A.R.	33
Basak, S.C.	32
Lynch, M.F.	32

Table 3. Most Used WoK Subject Areas for Indexing JCICS Articles

subject area	articles
computer science	3412
chemistry	3411
biochemistry and molecular biology	568
pharmacy and pharmacology	471
mathematics	295
medical informatics	253
genetics and hereditary	98
toxicology	80
information science and library science	73
life sciences and biomedicine—other topics	58

Table 4. Most Productive Institutions in JCICS

institution	articles
University of Sheffield	98
Chemical Abstracts Service	82
Ruđer Bošković Institute	66
National Institute of Chemistry of Slovenia	64
Pennsylvania State University	57
Texas A&M University	45
Université Paris 7—Denis Diderot	43
Drake University	38
University of Florida	35
University of Minnesota	33

Journal of Chemical Information and Computer Sciences. The 30 volumes of JCICS comprised 146 issues and contained a total of 3432 articles. These articles were the work of 4408 authors, the most productive of whom are listed in Table 2. Lynch is the only author common to both Table 1 and Table 2, reflecting his long and distinguished contribution to the field; indeed, JCICS had a festschrift for him in 1991.[25] The presence in this table of four leading researchers in the field of topological indices—Randic, Trinajstíc, Balaban and Gutman—highlights the key role that the Journal played in the development of this subject during the period under review.

Study of the WoK subject headings assigned to the 3432 articles showed that the most used are those listed in Table 3. The list makes clear the increasing range of topics included in the Journal as compared to JCD. In all, 76 subject areas are included, with singletons including criminology and penology, marine and freshwater biology, and ophthalmology, *inter alia*. While locational information is still not complete, there is far more of it for the JCICS publications than for the JCD publications. Thus, Table 4 lists the 10 institutions publishing most frequently in JCICS. Many of these reflect the presence of particularly productive authors, e.g., Lynch and Willett at the University of Sheffield, Randic and Trinajstić at the Ruđer Bošković Institute, and Katritzky at the University of Florida. All of the authors here work in areas that would now be considered typical of chemoinformatics, rather than LIS as for the authors in Table 1. Turning to the authors' nationalities, Table 5 shows that the United States was still by far the largest contributor to the Journal; however, JCICS is clearly more international in scope than was JCD, with significant contributions from a range of countries in both Western Europe and elsewhere.

Table 5. Most Productive Nations in JCICS

nation	articles
United States	1614
United Kingdom	345
Germany	246
Peoples Republic of China	157
Japan	142
France	135
Slovenia	91
Spain	90
Russia	80
Canada	79

CITATIONS

Journal of Chemical Documentation. The 900 JCD articles attracted 4724 citations over the period 1961−2009, as shown in Figure 1. It can be seen, hardly surprisingly, that the rate of citation has dropped off over time as (most of) the material in the JCD articles has changed from being current, to being obsolescent, and then to being obsolete. Residual citations continue to the present day, although these are principally due to a single, highly important paper that is discussed in more detail later in this section. There is a sharp spike in the citations to JCD in 1985; this is due in part to the 25th anniversary, special issue of the Journal (part 3 of volume 25), which contained a series of review articles (several of them historical in tone) that made extensive reference to previous JCD material. As is common, the citation counts for individual articles follow a highly skewed distribution with a few being heavily cited and the majority attracting few, if any, citations. The 20 most cited JCD papers are listed in Table 6, along with the total number of citations they have attracted to date. These 20 articles, comprising just 2.2% of the total, attracted 28.2% of the citations.

The 4724 citations came from 2132 citing articles. These articles appeared in 459 different journals, with the 10 that contributed most citations listed in Table 7 (where it should be noted that *American Documentation* was the former name of the *Journal of the American Society for Information Science*). The citations are dominated by JCICS and by JCD itself, with only one other journal—the *Journal of Computational Chemistry*—having a chemical focus; instead, the seven remaining journals all have an LIS focus, a fact that might appear rather surprising at first sight. However, a recent bibliometric study of the links between chemoinformatics and information retrieval suggests a reason for this observation.[26] Chemistry was the first discipline to develop comprehensive information systems and services, and it has thus long been at the forefront of attempts to apply developments in technology to information processing. For example, work at CAS in the 1960s pioneered the use of computers for the

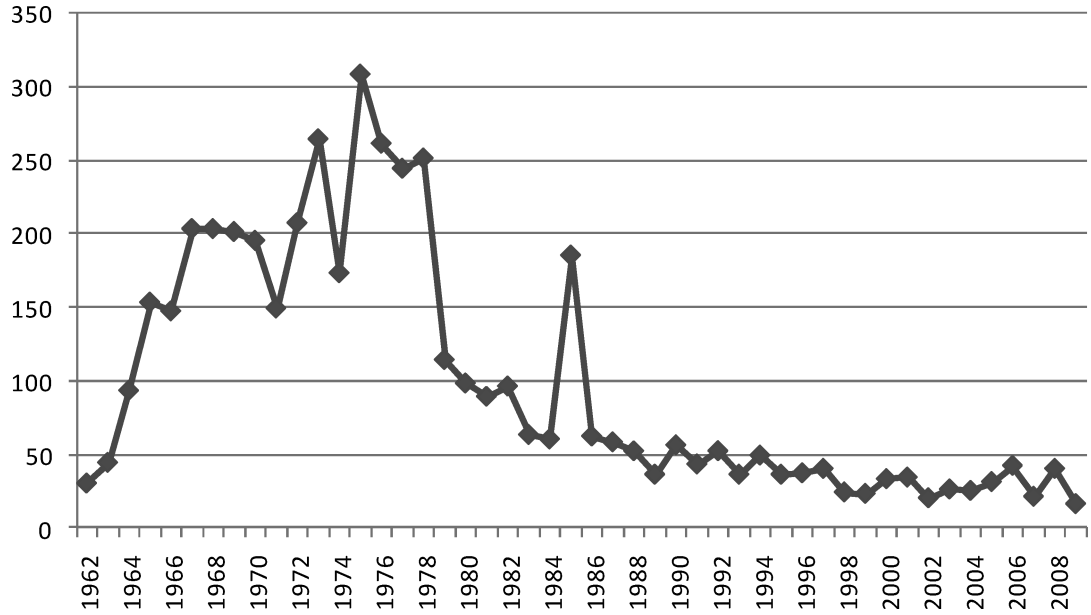

Figure 1. Citations over time to articles published in JCD.

Table 6. The 20 Most Cited Papers in JCD

article	citations	
	total	annual
Morgan, H. L. (1965). The Generation of a Unique Machine Description for Chemical Structures - A Technique Developed at Chemical Abstracts Service.	381	8.28
Sussenguth, E. H. (1965). A Graph-Theoretic Algorithm for Matching Chemical Structures.	134	2.91
Spialter, L. (1964). The Atom Connectivity Matrix (ACM) and Its Characteristic Polynomial (ACMCP).	76	1.62
Hosoya, H. (1972). Topological Index as a Sorting Device for Coding Chemical Structures.	68	1.74
Balaban, A. T.; Harary, F. (1971). The Characteristic Polynomial Does Not Uniquely Determine the Topology of a Molecule.	67	1.67
Adamson, G. W. et al. (1973). Strategic Considerations in the Design of a Screening System for Substructure Searches of Chemical Structure Files.	63	1.66
Gluck, D. J. (1965). A Chemical Structure Storage and Search System Developed at Du Pont.	54	1.17
Spialter, L. (1964). The Atom Connectivity Matrix Characteristic Polynomial (ACMCP) and Its Physico-Geometric (Topological) Significance.	52	1.11
Leiter, D. P. (1965). Installation and Operation of a Registry for Chemical Compounds.	50	1.09
Hyde, E. et al. (1967). Conversion of Wiswesser Notation to a Connectivity Matrix for Organic Compounds.	41	0.93
Plotkin, M. (1971). Mathematical Basis of Ring-Finding Algorithms in CIDS.	40	1.00
Figueras, J. (1972). Substructure Search by Set Reduction.	39	1.03
Thomson, L. H. et al. (1967). Organic Search and Display using a Connectivity Matrix Derived from Wiswesser Notation.	39	0.89
Blackwood, J. E. et al. (1968). Unique and Unambiguous Specification of Stereoisomerism about a Double Bond in Nomenclature and other Notation Systems.	38	0.88
Kennard, O. et al. (1972). Cambridge Crystallographic Data Centre. I. Bibliographic File.	36	0.92
Allen, F. H. et al. (1973). Cambridge Crystallographic Data Centre. II. Structural Data File.	32	0.84
Kudo, Y.; Sasaki, S. (1974). The Connectivity Stack, a New Format for Representation of Organic Chemical Structures.	31	0.86
Petrarca, A. E. et al. (1967). A Method for Generating Unique Computer Structural Representations of Stereoisomers.	31	0.70
Feldman, A. et al. (1963). The Automatic Encoding of Chemical Structures.	31	0.65
Rossler, S.; Kolb, A. (1970). The GREMAS System, an Integral Part of the IDC System for Chemical Documentation.	29	0.71

Table 7. Journals Most Frequently Citing JCD

journal	citations
Journal of Chemical Information and Computer Sciences	381
Journal of Chemical Documentation	301
Annual Review of Information Science and Technology	69
Journal of the American Society for Information Science	52
American Documentation	37
Journal of Documentation	34
Nauchno-Tekhnicheskaya Informatsiya Seriya 2: Informatsionnye Protsessy i Sistemy	27
Special Libraries	26
Journal of Computational Chemistry	22
Nachrichten für Dokumentation	21

creation and searching of text databases, and it was chemistry that saw some of the first attempts to provide end-user access to online bibliographic retrieval services when they started to become generally available in the 1970s. Accordingly, many of the citations to JCD reflect the fact that it was chemistry that had seen the initial development of such procedures, which were then adopted (with appropriate modification if necessary) by LIS specialists in other disciplines; i.e., these developments were of interest not only to chemical information specialists but also to the LIS community more generally. However, as chemoinformatics has developed, there have been far fewer articles in the Journal about chemical documentation (as evidenced by the change of name from JCD to JCICS) and far more articles about drug discovery and informatics (which is why LIS journals are noticeably absent from Table 10). The 274 journals that have cited JCD just once are an eclectic mix,

Table 8. WoK Subject Areas Most Frequently Citing JCD

subject area	citations
computer science, information systems	764
computer science, interdisciplinary applications	760
chemistry, multidisciplinary	740
information science and library science	463
chemistry, organic	78
multidisciplinary sciences	68
chemistry, analytical	66
chemistry, physical	65
biochemistry and molecular biology	63
mathematics, interdisciplinary applications	39

including the *Annual Journal of Public Health*, *Education for Information*, the *Journal of Insect Physiology*, and *Transactions of the American Geophysical Union*.

The 459 citing journals have been categorized as being in 42 different WoK subject areas. Table 8 lists the 10 areas that yielded most citations (note that pharmacology and pharmacy yielded the same number of citations as mathematics, interdisciplinary applications). The reader should note that there is a fair amount of duplication here as journals can be allocated to more than one subject area; nonetheless, the table makes clear that JCD has had the greatest impact in the chemical, computer, and information sciences. This is not surprising; more worthy of note is the fact that the 26 subject areas that provided just a single citation included fields as diverse as finance, geochemistry and geophysics, language and linguistics, meteorology and atmospheric science, and robotics.

As noted when discussing the publications, address data is often absent from early WoS records, and we can hence only make general statements about the extent to which different countries and different institutions have cited JCD. However, the three heaviest citing nations were the United States, the United Kingdom, and Germany, with 698, 230, and 139 citations, respectively, and the three heaviest citing

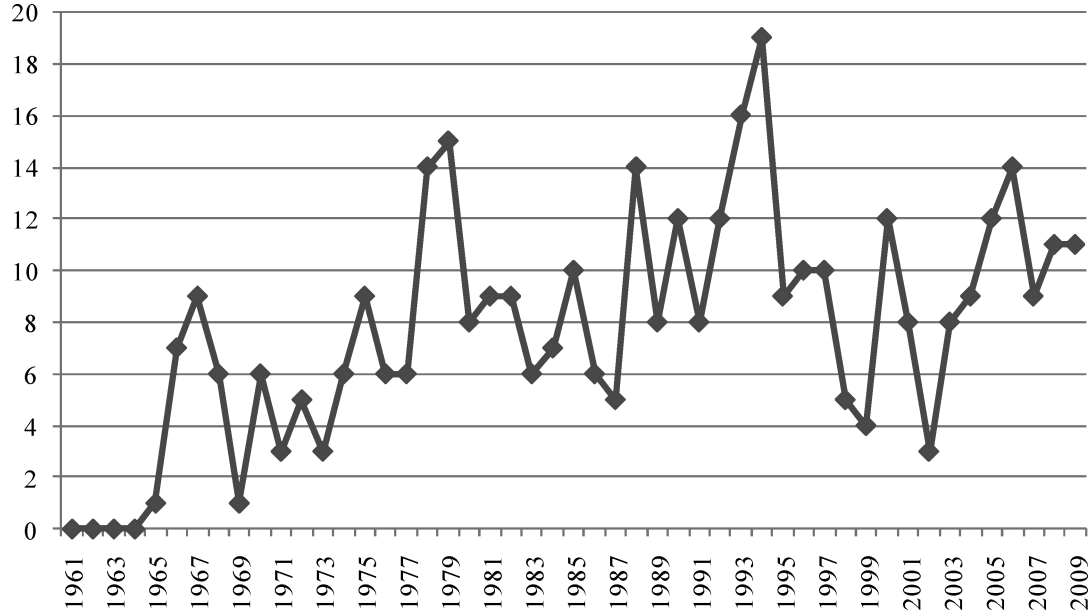

Figure 2. Citations over time to the article by Morgan.

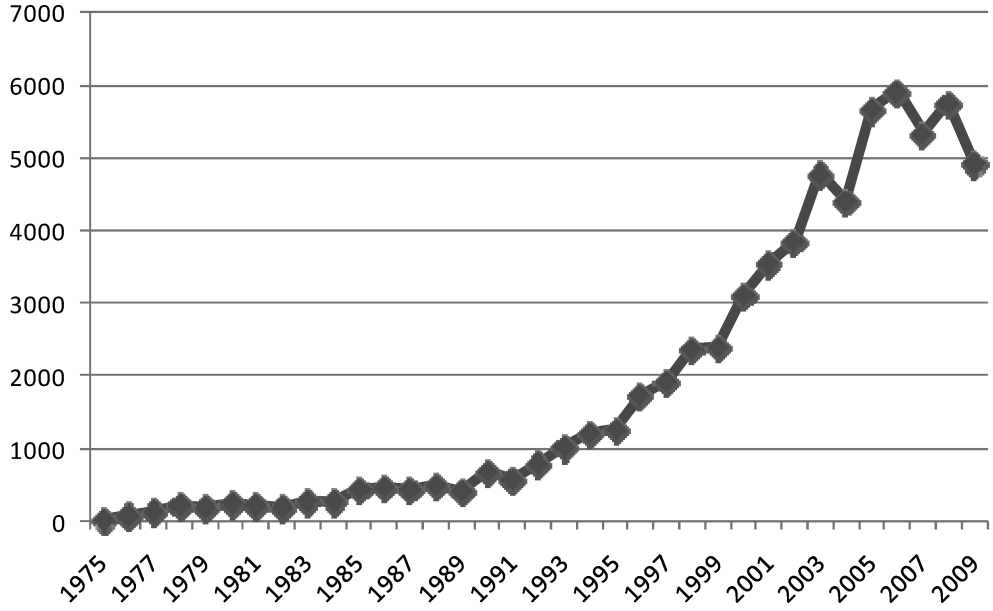

Figure 3. Citations over time to articles published in JCICS.

Table 9. The 30 Most Cited Articles in JCICS

article	citations total	citations annual
Allen, F. H. et al. (1991). The Development of Versions 3 and 4 of the Cambridge Structural Database System.	1,325	69.74
Fletcher, D. A. et al. (1996). The United Kingdom Chemical Database Service.	891	63.64
Weininger, D. (1988). SMILES, A Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules.	616	28.00
Viswanadhan, V. N. et al. (1989). Atomic Physicochemical Parameters for Three Dimensional Structure Directed Quantitative Structure−Activity Relationships. 4. Additional Parameters for Hydrophobic and Dispersive Interactions and their Application for an Automated Superposition of Certain Naturally Occurring Nucleoside Antibiotics.	550	26.19
Willett, P. et al. (1998). Chemical Similarity Searching.	531	44.25
Rogers, D.; Hopfinger, A. J. (1994). Application of Genetic Function Approximation to Quantitative Structure−Activity Relationships and Quantitative Structure−Property Relationships.	448	28.00
Brown, R. D.; Martin, Y. C. (1996). Use of Structure−Activity Data to Compare Structure-Based Clustering Methods and Descriptors for Use in Compound Selection.	395	28.21
Krygowski, T. M. (1993). Crystallographic Studies of Intermolecular and Intramolecular Interactions Reflected in Aromatic Character of Pi-Electron Systems.	299	17.59
Carhart, R. E. et al. (1985). Atom Pairs as Molecular-Features in Structure−Activity Studies - Definition and Applications.	294	11.76
Hann, M. M. et al. (2001). Molecular Complexity and its Impact on the Probability of Finding Leads for Drug Discovery.	256	28.44
Ghose, A. K.; Crippen, G. (1987). Atomic Physicochemical Parameters for Three Dimensional Structure Directed Quantitative Structure−Activity Relationships. 2. Modeling Dispersive and Hydrophobic Interactions.	248	10.78
Oprea, T. I. et al. (2001). Is There a Difference Between Leads and Drugs? A Historical Perspective.	238	26.44
Randic, M. (1991). Resolution of Ambiguities in Structure−Property Studies by Use of Orthogonal Descriptors.	206	10.84
Platts, J. A. et al. (1999). Estimation of Molecular Linear Free Energy Relation Descriptors Using a Group Contribution Approach.	205	18.64
Hall, L. H.; Kier, L. B. (1995). Electrotopological State Indexes for Atom Types - A Novel Combination of Electronic, Topological, and Valence State Information.	201	13.40
Wessel M. D. et al. (1998). Prediction of Human Intestinal Absorption of Drug Compounds from Molecular Structure.	195	16.25
Brown, R. D.; Martin, Y.C. (1997). The Information Content of 2D and 3D Structural Descriptors Relevant to Ligand−Receptor Binding.	191	14.69
Randic, M. (1984). On Molecular-Identification Numbers.	186	7.15
Hawkins, D. M. (2004). The Problem of Overfitting.	184	30.67
Balaban, A. T. (1985). Applications of Graph-Theory in Chemistry.	183	7.32
Weininger, D. et al. (1989). SMILES. 2. Algorithm For Generation of Unique SMILES Notation.	182	8.67
Sadowski, J. et al. (1994). Comparison of Automatic 3-Dimentional Model Builders Using 639 X-ray Structures.	181	11.31
Wildman, S. A.; Crippen, G. M. (1999). Prediction of Physicochemical Parameters by Atomic Contributions.	175	15.91
Bergerhoff, G. et al. (1983). The Inorganic Crystal Structure Database.	175	6.48
Gemperline, P. J. (1984). A Priori Estimates of the Elution Profiles of the Pure Components in Overlapped Liquid-Chromatography Peaks Using Target Factor-Analysis.	175	6.73
Galvez, J. et al. (1994). Charge Indexes - New Topological Descriptors.	173	10.81
Schultz, H. P. (1989). Topological Organic Chemistry. 1. Graph-Theory and Topological Indexes of Alkanes.	173	8.29
Hall, L. B. et al. (1991). The Electrotopological State - Structure Information at the Atomic Level for Molecular Graphs.	166	8.74
Feher, M.; Schmidt, J.M. (2003). Property Distributions: Differences Between Drugs, Natural Products, and Molecules from Combinatorial Chemistry.	157	22.43
Randic, M. et al. (2000). On 3-D Graphical Representation of DNA Primary Sequences and their Numerical Characterization.	149	14.90

Table 10. Journals Citing the 30 Most Cited Articles in JCICS

journal	citations
Journal of Chemical Information and Computer Sciences	593
Journal of Chemical Information and Modeling	373
Journal of Medicinal Chemistry	283
Acta Crystallographica Section C- Crystal Structure Communications	196
Journal of Computer-Aided Molecular Design	175
QSAR & Combinatorial Science	120
Bioorganic & Medicinal Chemistry	118
Journal of Molecular Graphics & Modeling	100
Acta Crystallographica Section E - Structure Reports Online	98
Acta Crystallographica Section B - Structural Science	95

institutions were the University of Sheffield, Chemical Abstracts Service, and the Institute for Scientific Information, with 83, 33, and 31 citations, respectively.

Inspection of Table 6 suggests that the highly cited articles are of two main types: those that describe algorithms and those that describe systems. It can be seen that the former tend to be more highly cited; this is hardly surprising in that an algorithm can continue to be relevant for many years, whereas systems articles can date very rapidly as technology develops. The citations in the table are dominated by those to one of the algorithmic articles, specifically that by Morgan,[27] which has attracted almost 3 times as many citations as the next most-cited article (also an algorithmic paper); indeed, this single article has attracted 8.1% of all the citations to JCD.

One may ask why the Morgan article continues to attract citations to the present day, some 45 years after its publication (see Figure 2). As with any "citation classic", some of these citations result from the so-called Matthew Effect ("unto he who hath shall it be given"), where authors cite a paper because it has been cited in papers that those authors have read. However, it seems unlikely that this alone could account for the paper's continuing impact, and we have hence inspected the 94 most recent citations, which cover the period 2000−2009. Of these, just 39, i.e., less than half, have

appeared in the mainstream chemoinformatics literature (with 25 of these 39 in JCICS or JCIM), and with almost the same number, 37, appearing in other chemistry journals. Some of these are fora where references to work in chemoinformatics would not be unexpected (e.g., *ChemMedChem* or *Journal of Medicinal Chemistry*), but others reflect a broad range of chemical specializations (e.g., *Atmospheric Environment*, *Cellular Chemistry and Technology*, and *Powder Diffraction*). The remaining articles comprise 13 from computer and information science (e.g., *Journal of Information Science* and *Lecture Notes in Computer Science*) and five from bioinformatics (e.g., *BMC Bioinformatics* and *Journal of Proteome Research*). Inspection of the titles and abstracts of the citing papers shows that several of them cite the Morgan paper in relation to the Pipeline Pilot circular substructures,[28] while other citations appear in review articles discussing the development of chemoinformatics. However, the article is most commonly cited in the context of graphs or graph descriptors, demonstrating the continuing relevance of this 1965 paper to current research on matching (normally molecular) graphs and chemoinformatics, respectively.

Journal of Chemical Information and Computer Sciences. The 3432 items published in the Journal attracted 65439 citations between 1975 and 2009, a mean value of 19.1 citations per item. Figure 3 shows the growth in citations over time: the annual citation rate appears to have peaked and to now be in decline as the JCICS items start to age rapidly. We have chosen to focus here on the 30 most cited articles listed in Table 9 (30 rather than 20 since the total number of citations to JCICS is so much larger than for JCD). These articles attracted 9348 citations, i.e., 0.9% of the total number of articles attracted 14.3% of the total citations, and most readers of this journal will be familiar with several of these highly cited items.

The two articles in Table 9 attracting most citations are discussions of databases, specifically of the Cambridge Structural Database of 3D crystallographic information and the Daresbury Chemical Database Service, which provides access to a range of chemical databases (including the Cambridge Structural Database) for the U.K. academic community. The large number of citations here is undoubtedly due in part to their being the "standard reference" that the databases recommend users to cite when reporting the use of system data in their research. Weininger's two SMILES papers can also be considered as standard references for users employing this very popular structure representation. It is interesting that the 1991 Allen et al. paper continues to be cited (11 citations in 2010 up to the start of August) even though it has long been superseded as the recommended standard reference for the Cambridge Structural Database.[29,30] In like vein, the 1998 review of molecular similarity in chemoinformatics by Willett et al. is by now arguably obsolescent; however, it has already attracted 41 citations in 2010, despite the ready availability of many subsequent, detailed reviews of the subject.[31–34] Reviews are often heavily cited, and other reviews in Table 9 include the crystallographic database description by Bergerhoff et al. (which is analogous to that by Allen et al.), and the articles by Balaban and by Oprea et al. There is no single article that dominates the citations, as was the case with the Morgan article in Table 6. All but five are from the past century; of the others, those by Hann et al. and by Hawkins have notably high annual citation rates and might thus be expected to rise further up the rankings in the future. The subject matter discussed in the articles in Table 9 is noticeably different from that in Table 6, with discussions of topics such as drug likeness, QSAR, topological indices, and molecular diversity *inter alia*.

Table 11. WoK Subject Areas Citing the 30 Most Cited Articles in JCICS

subject area	citations
chemistry, multidisciplinary	2478
computer science, interdisciplinary applications	1638
chemistry, medicinal	1095
computer science, information systems	1007
biochemistry and molecular biology	961
chemistry, organic	850
pharmacology and pharmacy	804
chemistry, physical	710
crystallography	706
chemistry, inorganic and nuclear	609

Table 12. Nations Citing the 30 Most Cited Articles in JCICS

nation	articles
United States	2079
United Kingdom	1876
Germany	725
Spain	609
China	465
Poland	335
France	320
India	314
Italy	301
Switzerland	237

Table 13. Institutions Citing the 30 Most Cited Articles in JCICS

institution	articles
Instituto de Catálisis y Petroleoquímica, CSIC	180
University of Cambridge	179
University of Warwick	149
University of Oxford	139
University of Sheffield	125
University of Valencia	109
University of Warsaw	104
GlaxoSmithKline	104
University College London	100
Central University of Las Villas	90

The top 10 journals citing the 30 papers are shown in Table 10, a list that is very different from the JCD one in Table 7 for reasons that we have discussed previously. The list is in no way surprising, containing all of the principal journals in chemoinformatics[2] and other, closely related journals such as the *Journal of Medicinal Chemistry* and *Bioorganic & Medicinal Chemistry*. The various sections of *Acta Crystallographica* appear in the list overwhelmingly as a result of citing the two crystallographic database articles mentioned above plus the database article by Fletcher et al. In all, 2783 publications have cited the Journal, with many of the 505 single citers being conference proceedings, e.g., the 2005 *IEEE Congress on Evolutionary Computation* or the 2008 *International Conference on Advanced Computer Theory and Engineering*.

Table 11 lists the top 10 subject areas citing the set of 30 most-cited JCICS items, these subjects being in the broad areas of chemistry and computer science. There were 20

Table 14. Institutions Citing Five of the Most Cited Articles in JCICS

institution	articles
University of Sheffield	66
University of Illinois	51
Merck	47
GlaxoSmithKline	35
Novartis	31
Pfizer	30
AstraZeneca	30
Jadavpur University	25
University of North Carolina	25
Russian Academy of Sciences	24

single citers, these including areas as diverse as anthropology, clinical neurology, remote sensing, and sociology. Table 12 lists the nations most frequently citing the 30 JCICS papers. The United States is the most frequent citing nation, with 28.6% of the total articles in the table. This is a much smaller fraction than the 54.2% of the entries in Table 5, which list the nations publishing most frequently in JCICS. Finally, Table 13 lists the most frequently citing organizations. Some of these might not be immediately thought of as centers of chemoinformatics research, but inspection of the data shows that institutions may occur here because of heavy citation of specific individual articles in Table 9. Thus, the great majority of the citations for the University of Warwick and the University of Oxford are to the 1996 article by Fletcher et al. describing the Daresbury Chemical Database Service (which is made available to all U.K. universities to support chemical, and chemically related, research), and the great majority of the citations for the Instituto de Catálisis y Petroleoquímica and for the University of Warsaw are to the crystallographic articles by Allen et al. and by Krygowski et al.

It is clear that the statistics above are heavily influenced by the natures of specific articles, and we have hence carried out comparable analyses based on just the top five articles from Table 9, excluding those that are database or software descriptions, reviews, or crystallographic articles. The five articles are hence those by Viswanadhan et al., Rogers and Hopfinger, Brown and Martin, Carhart et al., and Hann et al., these covering mainstream topics for the Journal such as clustering, QSAR, and virtual screening. The results for citing journals differ from those in Table 10 only in that the three sections of *Acta Crystallographica* are replaced by *Drug Discovery Today*, the *European Journal of Medicinal Chemistry*, and *Combinatorial Chemistry & High Throughput Screening*. There are few differences from the results in Tables 11 and 12 for subjects and nations, respectively; however, inspection of the institutions citing these five articles, as shown in Table 14, shows marked differences from the corresponding data in Table 13. Specifically, the table demonstrates very clearly the importance of the Journal to the world's major pharmaceutical companies, with only one-half of the institutions here being academic in character.

Finally, and for completeness, there are some very brief comments about JCIM. The five volumes for the period 2005−09 carried 1256 articles, with the list of the 10 most productive institutions being headed by AstraZeneca and with two other companies—Novartis and Pfizer—also represented. This list also reflects the emergence of strong new academic groups such as those at the University of Cambridge, the

Table 15. The 20 Most Cited Articles (To the End of 2009) in JCIM

article	citations	
	total	annual
Irwin, J.; Shoichet, B. (2005). ZINC - A Free Database of Commercially Available Compounds for Virtual Screening.	336	84.0
Schuchardt, K. et al. (2007). Basis Set Exchange: A Community Database for Computational Sciences.	116	58.0
Chen, H. et al. (2006). Evaluating Molecular-Docking Methods for Pose Prediction and Enrichment Factors.	84	28.0
Wolber, G.; Langer, T. (2005). LigandScout: 3-D Pharmacophores Derived from Protein-Bound Ligands and their Use as Virtual Screening Filters.	82	20.5
Guha, R. et al. (2006). Blue Obelisk - Interoperability in Chemical Informatics.	76	25.3
McGaughey, G. et al. (2007). Comparison of Topological, Shape, and Docking Methods in Virtual Screening.	58	29.0
Jorissen, R.; Gilson, M. (2005). Virtual Screening of Molecular Databases Using a Support Vector Machine.	57	14.3
Hert, J. et al. (2006). New Methods for Ligand-Based Virtual Screening: Use of Data Fusion and Machine Learning to Enhance the Effectiveness of Similarity Searching.	56	18.7
Chou, K.; Cai, Y. (2005). Prediction of Membrane Protein Types by Incorporating Amphipathic Effects.	55	13.8
Maggiora, G. (2006). On Outliers and Activity Cliffs - Why QSAR Often Disappoints.	53	17.7
Yap, C.; Chen, Y. (2005). Prediction of Cytochrome P450 3A4, 2D6, and 2C9 Inhibitors and Substrates by Using Support Vector Machines.	53	13.3
Truchon, J.; Bayly, C. (2007). Evaluating Virtual Screening Methods: Good and Bad Metrics for the "Early Recognition" Problem.	48	24.0
Yang, J. et al. (2005). Consensus Scoring Criteria for Improving Enrichment in Virtual Screening.	48	12.0
Spalek, T. et al. (2005). Application of the Genetic Algorithm Joint with the Powell Method to Nonlinear Least-Squares Fitting of Powder EPR Spectra.	47	11.8
Dimitrov, S. et al. (2005). A Stepwise Approach for Defining the Applicability Domain of SAR and QSAR Models.	43	10.75
Kirchmair, J. et al. (2006). Comparative Performance Assessment of the Conformational Model Generators Omega and Catalyst: A Large-Scale Survey on the Retrieval of Protein-Bound Ligand Conformations.	42	14.0
Li, H. et al. (2005). Effect of Selection of Molecular Descriptors on the Prediction of Blood-Brain Barrier Penetrating and Nonpenetrating Agents by Statistical Learning Methods.	42	10.5
Fink, T.; Reymond, J. (2007). Virtual Exploration of the Chemical Universe up to 11 Atoms of C, N, O, F: Assembly of 26.4 Million Structures (110.9 Million Stereoisomers) and Analysis for New Ring Systems, Stereochemistry, Physicochemical Properties, Compound Classes, and Drug Discovery.	39	19.5
Papa, E. et al. (2005). Statistically Validated QSARs, Based on Theoretical Descriptors, for Modeling Aquatic Toxicity of Organic Chemicals in *Pimephales promelas* (Fathead Minnow).	39	9.8
Baroni, M. et al. (2007). A Common Reference Framework for Analyzing/Comparing Proteins and Ligands. Fingerprints for Ligands and Proteins (FLAP): Theory and Application.	28	14.0

Rheinische Friedrich-Wilhelms-Universität Bonn, and Indiana University. The articles attracted 8546 citations (to the end of 2009) from 4650 citing articles, with the 20 most cited articles listed in Table 15. Inspection of the table shows that the two most cited articles are again database descriptions, and that the fifth position is occupied by a software description. Analysis of the citations to these articles reveals few major differences from the general picture obtained from analysis of the citations to the JCICS articles. That said, it is worth noting the appearance of *Bioorganic & Medicinal Chemistry Letters*, *ChemMedChem* (a new journal), and *Journal of Physical Chemistry A* in the top 10 citing journals, and of two Chinese institutions—the Chinese Academy of Sciences (in fact a whole series of institutions) and Zhejiang University—in the top 10 citing institutions, this reflecting the dramatic growth in Chinese science in recent years.

CONCLUSIONS

The *Journal of Chemical Information and Modeling* is the premier international journal for research in chemoinformatics. Its reputation is built firmly on those of its previous manifestations, the *Journal of Chemical Documentation* and the *Journal of Chemical Information and Computer Sciences*. In this paper, we have reported data for the articles published in these two earlier manifestations and for citations to these published articles. The study has highlighted the impact of the research published in the Journal (as exemplified by the very wide range of publications and disciplines that cite it), the long-lived impact of some of the highly cited articles, and the changes in focus of the Journal that have occurred since its foundation in 1961. These changes are continuing,[22] and it is sincerely to be hoped that the Journal will be as successful over the next 50 years as it has been during its first half-century.

ACKNOWLEDGMENT

We thank Dr. Wendy Warr for helpful comments.

REFERENCES AND NOTES

(1) Agrafiotis, D. K.; Bandyopadhyay, D.; Wegner, J. K.; van Vlijmen, H. Recent advances in chemoinformatics. *J. Chem. Inf. Model.* **2007**, *47*, 1279–1293.
(2) Willett, P. A bibliometric analysis of chemoinformatics. *Aslib Proc.* **2008**, *60*, 4–17.
(3) Hood, W. W.; Wilson, C. S. The literature of bibliometrics, scientometrics, and informetrics. *Scientometrics* **2001**, *52*, 291–314.
(4) Borgman, C. L.; Furner, J. Scholarly communication and bibliometrics. *Ann. Rev. Inf. Sci. Technol.* **2002**, *36*, 3–72.
(5) Nicolaisen, J. Citation analysis. *Ann. Rev. Inf. Sci. Technol.* **2007**, *41*, 609–641.
(6) Bar-Ilan, J. Informetrics at the start of the 21st century - a review. *J. Informetrics* **2008**, *2*, 1–52.
(7) Garfield, E. *Citation Indexing -- Its Theory and Application in Science, Technology, and Humanities*; ISI Press: Philadelphia, PA, 1979.
(8) Cronin, B. *The Citation Process. The Role and Significance of Citations in Scientific Communication*; Taylor Graham: London, 1984.
(9) Bornmann, L.; Daniel, H. What do citation counts measure? A review of studies on citing behavior. *J. Doc.* **2006**, *64*, 45–80.
(10) Neuhaus, C.; Daniel, H.-D. Data sources for performing citation analysis: an overview. *J. Doc.* **2008**, *64*, 193–210.
(11) Anyi, K. W. U.; Zainab, A. N.; Anuar, N. B. Bibliometric studies on single journals: a review. *Malaysian J. Lib. Inf. Sci.* **2009**, *14* (1), 17–55.
(12) Willett, P. A bibliometric analysis of the *Journal of Molecular Graphics and Modelling*. *J. Mol. Graphics Modell.* **2007**, *26*, 602–606.
(13) Willett, P. A bibliometric analysis of *Quantitative Structure-Activity Relationships* and *QSAR & Combinatorial Science*. *QSAR Comb. Sci.* **2009**, *28*, 1231–1236.
(14) Onodera, N. A bibliometric study on chemical information and computer sciences focusing on literature of JCICS. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 878–888.
(15) Warr, W. A. Twenty five years of progress in cheminformatics. In *229th American Chemical Society National Meeting*, San Diego, CA, 2005. Available at http://www.warr.com/25years.html (accessed August 31, 2010).
(16) Chen, W. L. Chemoinformatics: past, present and future. *J. Chem. Inf. Model.* **2006**, *46*, 2230–2255.
(17) Willett, P. From chemical documentation to chemoinformatics: fifty years of chemical information science. *J. Inf. Sci.* **2008**, *34*, 477–499.
(18) Metanomski, W. V. *50 Years of Chemical Information in the American Chemical Society 1943−1993*; American Chemical Society: Washington, DC, 1993.
(19) Skolnik, H. Editorial. In retrospect. *J. Chem. Doc.* **1970**, *10*, 226.
(20) Skolnik, H. Editorial. Annual report for 1974. *J. Chem. Doc.* **1975**, *15*, 2.
(21) Milne, G. W. A. Editorial. Sections in the *Journal of Chemical Information and Computer Sciences*. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 183.
(22) Jorgensen, W. L. Editorial. *J. Chem. Inf. Model.* **2005**, *45*, 1.
(23) Skolnik, H. The journal for chemical information and computer scientists: a 25-year perspective. *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 137–140.
(24) Maizell, R. E. *How to Find Chemical Information*, 3rd ed.; Wiley: New York, 1998.
(25) Milne, G. W. A.; Willett, P. A tribute to Michael F. Lynch. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 175.
(26) Willett, P. Information retrieval and chemoinformatics: is there a bibliometric link? In *The Janus Faced Scholar - a Festschrift in Honour of Peter Ingwersen*; Larsen, B., Schneider, J. W., Åström, F., Schlemmer, B., Eds.; Royal School of Library and Information Science: Copenhagen, 2010; pp 107−115.
(27) Morgan, H. L. The generation of a unique machine description for chemical structures - a technique developed at Chemical Abstracts Service. *J. Chem. Doc.* **1965**, *5*, 107–113.
(28) Rogers, D.; Hahn, M. Extended-connectivity fingerprints. *J. Chem. Inf. Model.* **2010**, *50*, 742–754.
(29) Allen, F. H. The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Crystallogr.* **2002**, *B58*, 380–388.
(30) Wong, R.; Willett, P.; Allen, F. H. The scientific impact of the Cambridge Structural Database: a citation-based study. *J. Appl. Crystallogr.* **2010**, *43*, 811–824.
(31) Bender, A.; Glen, R. C. Molecular similarity: a key technique in molecular informatics. *Org. Biomol. Chem.* **2004**, *2*, 3204–3218.
(32) Eckert, H.; Bajorath, J. Molecular similarity analysis in virtual screening: foundations, limitation and novel approaches. *Drug Discovery Today* **2007**, *12*, 225–233.
(33) Sheridan, R. P. Chemical similarity searches: when is complexity justified. *Expert Opin. Drug Discovery* **2007**, *2*, 423–430.
(34) Willett, P. Similarity methods in chemoinformatics. *Ann. Rev. Inf. Sci. Technol.* **2009**, *43*, 3–71.

CI100338J