A Bibliometric Analysis of the *Journal of Molecular Graphics and Modelling*

Peter Willett[1], Krebs Institute for Biomolecular Research and Department of Information Studies, University of Sheffield, UK

Abstract. This paper reviews the articles published in Volumes 2-24 of the *Journal of Molecular Graphics and Modelling* (formerly the *Journal of Molecular Graphics*), focusing on the changes that have occurred in the subject over the years, and on the most productive and most cited authors and institutions. The most cited papers are those describing systems or algorithms, but the proportion of these types of article is decreasing as more applications of molecular graphics and molecular modelling are reported.

INTRODUCTION

The Molecular Graphics and Modelling Society (then the Molecular Graphics Society) was established in 1982, and 2007 is thus the Society's 25th anniversary. Its mission is to support and to encourage research into the use of novel computer techniques for the design of bioactive molecules, focusing on molecular modelling and related aspects of computational chemistry. One of the most important ways that it achieves its aims is the publication of one of the key academic journals in the field, the *Journal of Molecular Graphics and Modelling*. The journal was started in 1983, changing its original name (the *Journal of Molecular Graphics*) to its present title in 1997 with Volume 15, following its merger with *Chemical Design Automation News*. In what follows, we shall normally refer to the journal as *JMGM*, irrespective of which of the two names was used at any particular point in time. Graham Richards was the editor-in-chief of *JMGM* from 1984 to 1996, during which time it grew considerably (from 18 full articles and 132 pages in Volume 2 to 42 full articles and 393 pages in Volume 14) and the journal's current high standing owes much to his leadership during this period.

The *JMGM* webpage states that it "is devoted to the publication of papers on the uses of computers in theoretical investigations of molecular structure, function, interaction, and design. The scope of the journal includes all aspects of molecular modelling and computational chemistry, including, for instance, the study of molecular shape and properties, molecular simulations, protein and polymer engineering, drug design, materials design, structure-activity and structure-property relationships, database mining, and compound library

[1] Address: Department of Information Studies, University of Sheffield, 211 Portobello Street, Sheffield S1 4DP, UK; Telephone: 0044-114-2222633; Email: p.willett@sheffield.ac.uk

design". Thus, while molecular modelling continues to be the principal focus of the journal, its remit is quite broad, covering a range of topics in the general area of chemoinformatics and computer-aided molecular design. It is also now published in association not just with the Molecular Graphics and Modelling Society but also with the Computers in Chemistry Division of the American Chemical Society.

The increasing availability of databases containing publication and citation information has spurred the development of the subject of *bibliometrics*. This involves the analysis of a set of publications characterised by bibliographic variables such as the author(s), the place of publication, the associated subject keywords, and the citations. Spurred in large part by the pioneering work of Garfield [1], bibliometrics is now a well-established discipline with its own core journal, *Scientometrics*, and many papers appearing in other journals, in particular those in library and information science, e.g., the *Journal of the American Society for Information Science and Technology* and the *Journal of Information Science*. Bibliometric data can be used to analyse a range of phenomena, such as the development of a subject over time, the identification of the key researchers and key collaborations in the subject, the extent to which different journals carry highly-cited articles on the subject, and, increasingly, to provide performance indicators relating to the quality of the research of an individual or of an institution [2-6].

There have been few bibliometric analyses in the area of computers and chemistry; examples include studies by Onodera of the *Journal of Chemical Information and Computer Sciences* (as it was then named) [7], by Redman *et al*. of applications of the Cambridge Structural Database [8], by Behrens and Luksch of the contents of the Inorganic Crystal Structure Database [9] and, most recently, by Willett of the development of chemoinformatics [10]. There have, however, been several bibliometric analyses of specific journals, most notably in the field of library and information science where several of the key journals have been studied in some detail [11-14]. This short paper provides a bibliometric analysis of *JMGM*, focusing on the changes that have occurred in the subject over the years, the most productive and most cited authors and institutions, and a comparison of the results obtained with those for other journals that cover similar research areas.

For many years, the principal source of data for bibliometric analyses was the *Web of Knowledge* (*WOK*). *WOK* draws on the publication and citation data from major academic journals contained in the *Web of Science* (comprising the *Science Citation Index*, the *Social Science Citation Index* and the *Arts and Humanities Citation Index*) produced by Thomson Corp. *WOK* has recently been joined by two other systems that provide access to such data:

Google Scholar produced by Google Inc. and *Scopus* produced by Elsevier B.V., and there is much interest as to the relative merits of the three systems for bibliometric research [15, 16]. Citation searching of the chemical literature can also be carried out using *SciFinder Scholar* produced by the American Chemical Society; however, its main power is in the area of text and structure searching and the citation-handling facilities are far less well-developed than for the other systems considered here.

The results presented in this paper are based solely on use of *WOK* for two reasons: *Google Scholar* does not, as yet, have the sophisticated post-processing facilities that *WOK* and *Scopus* provide to facilitate bibliometric analyses of the sort considered here; and *WOK* covers *JMGM* from Volume 2 onwards, whereas the Scopus coverage commences with Volume 6. Thus the analyses reported below are based on the 859 articles that appeared in *JMGM* from 1984 (Volume 2) to 2006 (Volume 24), with the counts and frequencies being generated using *WOK*'s Analyse Results and Citation Reports routines. Use was also made of the *Journal Citation Reports* (*JCR*) database (another part of the *WOK* and again produced by Thomson Corp.), which contains extensive bibliometric information relating to individual journals, rather than to publications and citations as in the case of the citation indices. All of the database searches were carried out in January 2007.

AUTHOR AND SUBJECT ANALYSIS

Bibliographic data is characterised by highly skewed inverse power distributions (sometimes referred to colloquially, but only approximately, as hyperbolic) [17]. Examples include: the occurrence of words in natural language texts, with a few words occurring very frequently and the great bulk of those occurring at all appearing but once (Zipf's Law); in carrying out a literature search for some specific topic, one or a small number of journals (the "core" journals) will produce a significant number of relevant articles, with many more journals being required to provide a corresponding further number of relevant articles (Bradford's Law); and a few authors will be highly productive in writing large numbers of articles on a subject, with most authors producing just a single article (Lotka's Law). These are all examples of a generalised inverse power law distribution of the form

$$f(k) = \frac{C}{k^{\beta}}$$

where $f(k)$ is the frequency of occurrence of some bibliometric item that is associated with each member of a population (k=1,2...) that is producing examples of these items, and where C and β are constants. In the case of authors writing academic articles, Lotka's original

k	1	2	3	4	5	6	7	8	≥9*
f(*k*)	1621	210	73	23	12	13	3	2	3*

Table 1. Number of authors, *f*(*k*), producing *k* articles in Volumes 2-24 of *JMGM*. *This column does not include 11 articles by that prolific author, Anon.

formulation posited an inverse square law, i.e., β=2, but this is only approximately correct. A mathematical analysis of the law by Egghe [18] suggested that β≤3 if a dataset follows exactly a success-breeds-success model, in which the more items that a source (such as an author) has produced already, the greater the probability that the source will produce another [19], but this upperbound is not observed in practice. For example, Pao analysed no less than 48 different datasets, covering a range of disciplines: she found that only two datasets closely fitted the inverse-square formulation, with the value of the exponent, β, ranging from 1.78 to 3.78 [20].

Author-productivity data for *JMGM* is shown in Table 1, which lists the numbers of authors, *f*(*k*), producing *k* articles over the years. Lotka analyses can be carried out in many ways [21]. Here, we have used the LOTKA program [22] (available at URL http://www.cindoc.csic.es/cybermetrics/articles/v4i1p4.html), which gives values for C and β of 0.834 and 3.02, respectively, with the latter thus lying well within the range of values observed by Pao. Inspection of the productive authors reveals many of the research leaders, past and present, in the field, with G.A. Arteca, J. Bajorath, R. Brasseur, A. Chatterjee, T.E. Ferrin, D.R. Flower, B.P. Gaber, D.S. Goodsell, R. Griffith, B. Maigret, N.H. Martin, J.P. Mornon, H. Nakamura, A.J. Olson, W.G. Richards, O. Tapia, K. Toma, H. Umeyama, W.J. Welsh, D.N.J. White and P. Willett all producing at least six articles in the journal. Several of these productive authors also publish extensively in other journals covering the same or related topics, e.g., *Journal of Chemical Information and Modeling*, *Journal of Computer-Aided Molecular Design*, *Journal of Molecular Modeling*, *QSAR and Combinatorial Science*, and *SAR and QSAR in Environmental Research* [10].

Hardly surprisingly, the subjects that these authors write about show some evidence of change over the lifetime of the journal. The first index for the journal appeared with Volume 4 in 1986, and we have hence compared the title entries in this with those in the index in Volume 14 (1996) and in the most recent index (Volume 24, 2006). Given the novelty of the technology, it is not surprising that the majority of the articles in the 1986 index (58% on the basis of inspecting the article titles) were concerned principally with the details of specific

programs or computer systems, with typical titles including "Easy projection of stereo movies" "Display algorithm for space-filling molecular models using a video array processor" "Protein secondary structural representations using real-time interactive computer graphics" and "Program for the visualization and interactive study of molecules on a calligraphic display system". Even so, there were already papers describing the applications of the technology to practical problems in computational chemistry and drug discovery, e.g., "Analysis of the pharmacological properties of clozapine analogues using molecular electrostatic potential surfaces", "Computational conformational analysis of cyclohexaglycyl" and "Computer-aided structural comparisons of clonidine and guanfacine with cyclazocine". Methods and software papers continue to provide the majority (64%) of the articles in the 1996 index, with the other articles including new applications (and present-day staples) such as molecular diversity analysis ("BOOMSLANG: a program for combinatorial structure generation"), bioinformatics ("STRUCTURELAB: a heterogeneous bioinformatics system for RNA structure analysis") and the Web ("The World Wide Web as a graphical user interface to program macros for molecular graphics, molecular modeling, and structure-based drug design"). By 2006, the journal's coverage had extended still further so that applications now provide the majority (64%) of the articles, with subjects such as machine learning, structure-based virtual screening, and QSAR *inter alia* complementing the more traditional focus on graphics and modelling.

THE CORE LITERATURE

The importance of a paper to the development of a field is often assessed by the number of citations that it has attracted, on the basis that subsequent workers in that field will have cited that paper if, and only if, they found it to be of value in their own research. All of the papers published in Volumes 2-24 were searched in *WOK*; the 859 papers attracted a total of 20,259 citations, and the 20 with the greatest numbers of citations are listed in Table 2. Many of these papers will undoubtedly be familiar to regular readers of *JMGM*, whatever their particular research interests.

Inspection of Table 2 makes very clear that the bulk of the articles here focus on methods rather than applications, with many of the former being what are, in effect, the "standard" references that are cited whenever anybody subsequently uses some particular software package. This is often clear from the title (e.g., the papers describing MIDAS, MOLMOL and What-If), with the articles by Dunker *et al*. (on protein structures) and by Golbraikh and Tropsha (on the analysis of QSAR models) being the most highly-ranked application papers. There is an obvious source of bias in the data, which is that older papers have a greater chance

Highly cited paper	Citations
R. Koradi, M. Billeter, K. Wuthrich, MOLMOL: A program for display and analysis of macromolecular structures, J. Mol. Graph. Modell. 14 (1996) 51-55.	3298
W. Humphrey, A. Dalke, K. Schulten, VMD: Visual molecular dynamics, J. Mol. Graph. 14 (1996) 33-38.	1732
G. Vriend, What-If – a molecular modelling and drug design program, J. Mol. Graph. 8 (1990) 52-56.	1505
R.M. Esnouf, An extensively modified version of MolScript that includes greatly enhanced coloring capabilities, J. Mol. Graph. Modell. 15, (1997) pp. 132-134.	1316
S.V. Evans, SETOR – hardware-lighted 3-dimensional solid model representations of macromolecules, J. Mol. Graph. 11 (1993) 134-138.	1151
T.E. Ferrin et al., The MIDAS display system, J. Mol. Graph. 6 (1988) 13-27.	982
M. Carson, Ribbon models of macromolecules, J. Mol. Graph. 5 (1987) 103-106.	514
W. Smith, T.R. Forester, DL_POLY_2.0: A general-purpose parallel molecular dynamics simulation package, J. Mol. Graph. 14 (1996) 136-141.	314
M.L. Connolly, The molecular-surface package, J. Mol. Graph. 11 (1993) 139-141.	243
A.K. Dunker et al., Intrinsically disordered protein, J. Mol. Graph. Modell. 19 (2001) 26-59.	242
A. Golbraikh, A. Tropsha, Beware of q²!, J. Mol. Graph. Modell. 20 (2002) 269-276.	172
H.E. Dayringer et al., Interactive program for visualization and modeling of proteins, nucleic-acids and small molecules, J. Mol. Graph. 4 (1986) 82-87.	145
C.C. Huang et al., CONIC – a fast rendered for space-filling molecules with shadows, J. Mol. Graph. 9 (1991) 230-236.	134
L. Laaksonen, A graphics program for the analysis and display of molecular-dynamics trajectories, J. Mol. Graph., 10 (1992) 33-34.	131
Kokalj A, XCrySDen - a new program for displaying crystalline structures and electron densities, J. Mol. Graph. Model. 17 (1999) 176-179.	125
M. Carson, C.E. Bugg, Algorithm for ribbon models of proteins, J. Mol. Graph., 4 (1986) 121-122.	110
D.L. Bergman, L. Laaksonen, A. Laaksonen, Visualization of solvation structures in liquid mixtures, J. Mol. Graph. Modell. 15 (1997) 301-306.	106
R.D. Clark et al., Consensus scoring for ligand/protein interactions, J. Mol. Graph. Modell. 20 (2002) 281-295.	105
B.M. Bode, M.S. Gordon, MacMolPlt: A graphical user interface for GAMESS, J. Mol. Graph. Modell. 16 (1998) 133-138.	100
P. Furet, A. Sele, N.C. Cohen, 3D molecular lipophilicity potential profiles – a new tool in molecular modeling, J. Mol. Graph. 6 (1988) 182-189.	90

Table 2. The most cited papers published in *JMGM*

of being cited, and it will be seen that many of the papers in the table date from early volumes when, as we have noted above, fewer applications papers were published; even so, a comparable bias towards system descriptions is observed if, e.g., one considers the most-cited papers published since 1996. This behaviour is not restricted to *JMGM*: amongst the most cited papers published in *Journal of Computer-Aided Molecular Design* are the standard

Journal	Relatedness ($\times 10^6$) of *JMGM* to *J*	Relatedness ($\times 10^6$) of *J* to *JMGM*
Journal of Computer-Aided Molecular Design	250.35	256.16
Journal of Chemical Information and Modeling	62.95	186.84
Journal of Computational Chemistry	162.85	66.96
Structure	30.00	141.65
Proteins	55.99	116.33
Acta Crystallographica D	15.04	111.91
SAR and QSAR in Environmental Research	31.48	98.87
Journal of Molecular Modeling	22.36	96.27
Current Opinion in Structural Biology	84.66	41.70
Protein Science	26.56	79.73

Table 3. Journals that are closely related to *JMGM*, based on the *Journal Citation Reports* database

references for packages such as AutoDock, CAVEAT, DISCO, GASP, HINT, IsoStar, LUDI, Molden and MOPAC; and this is also true of chemoinformatics journals in general [10].

There are several other journals that cover similar ground to *JMGM*. The subject relationships between journals can be explored in some detail using the Relatedness (*R*) data in the *Journal Citation Reports* database. *R* is determined for some period of time using a calculation that is based on three factors: the number of citations in that period from *JMGM* (in this particular case) to another journal, *J*; the total number of citations from *JMGM* to all journals; and the total number of articles in *J* [23]. Let *A* and *B* be two journals publishing P_A and P_B articles, respectively. Then let C_{AB} (or C_{BA}) be the number of times that *A* cites *B* (or that *B* cites *A*), and let CT_A (or CT_B) be the total number of citations in *A* (or *B*). Then the relatedness of *A* to *B*, R_{AB}, is defined as

$$\frac{C_{AB}}{P_B \times CT_A},$$

and similarly for R_{BA}. Table 3 lists the ten most closely related journals, ranked in descending order of the larger of the two possible *R* values (using citations from *JMGM* to *J;* and using citations from *J* to *JMGM*) in each case, as suggested by Pudovkin and Garfield [23]. Tables analogous to Table 3 can be created for *Journal of Chemical Information and Modeling* and for *Journal of Computer-Aided Molecular Design*, the two top-ranked journals here. This further emphasises the degree of commonality of subject: *JMGM* is the fifth "closest" journal for *Journal of Chemical Information and Modeling* and the fourth for *Journal of Computer-Aided Molecular Design*; and the latter two journals are mutual nearest neighbours.

The relationships in Table 3 were further investigated for each of the highly cited papers in Table 2. Specifically, *WOK* cumulates the citations from a given journal to a given paper, and one can thus identify the journals with most citations to that paper. Taking the top-25 such journals for each of the 20 papers in Table 2, all but four of the 20 papers had at least three of the related journals in Table 3 citing it (those missing such high-ranked journals were the papers by Bergman *et al*., Bode and Gordon, Kokalj, and Smith and Forester), with the paper by Connolly having no less than eight such high-ranked journals.

These close subject relationships are reflected in other recent studies. Leydesdorff has made available for each of over 7000 journals those journals that were responsible in 2003-04 for at least 1% of the citations to a given journal [24]. There are 19 such journals in the case of *JMGM,* these including (of course) *JMGM* itself and most of those in Table 3 (the exceptions being *Current Opinion in Structural Biology, Journal of Molecular Modeling* and *SAR and QSAR in Environmental Research*). Willett's analysis of productive authors in chemoinformatics also highlights the relationships between *JMGM* and, in particular, *Journal of Chemical Information and Modeling* and *Journal of Computer-Aided Molecular Design* [10].

SOURCE OF ARTICLES

The Molecular Graphics and Modeling Society is international in scope and this is reflected in the publications in *JMGM*. Table 4 lists the geographical data for the most productive of the 55 countries from which articles came. The USA provided 38.1% of the articles published in the journal, but there are another 11 countries that provide at least 2% of the articles for which Country/Territory data are available in the *WOK* database. Note that Table 4 does not contain an entry for the United Kingdom as such, since England, Scotland, Ireland and Wales are entered separately in *WOK*. Note also the perhaps surprisingly low ranking of the People's Republic of China; however, it must be remembered that Table 4 relates to a period of over two decades and more recent articles provide a rather different picture. Specifically, considering articles from 2000 to date: the USA plays a still larger role, providing no less than 45.8% of the articles; England is still second but with just 8.5%; the People's Republic of China is now in fourth place; and Brazil and India (in tenth and eleventh positions) also now provide more than 2% of the articles.

Finally, Table 5 lists the most productive institutions from amongst the 687 institutions that contributed articles (in whole or in part), this table reflecting many of the key research groups

Country	%
USA	38.1
England	18.3.
Japan	7.2
France	6.8
Germany	4.3
Australia	3.5
Spain	3.5
Switzerland	3.0
Canada	2.7
Italy	2.7
Sweden	2.7
People's Republic of China	2.3

Table 4. Countries providing at least 2% of the articles published in *JMGM*.

Institution	%
University of Sheffield	2.2
University of Oxford	2.0
University of California at San Francisco	1.9
CNRS	1.8
University of Cambridge	1.5
University of North Carolina	1.5
Scripps Research Institute	1.3
University of Minnesota	1.2
Birkbeck College, University of London	1.0
Université de Paris 07	1.0
Naval Research Laboratory, Washington	1.0

Table 5. Research centres providing at least 1% of the articles published in *JMGM*.

in the field. All but five of the top 50 institutions are universities, governmental or not-for-profit organisations; the highest-ranked commercial organisation is IBM at position 26, and it is followed in the rankings by Accelrys, Asahi Chemical, Bristol Myers Squibb and Tripos. Commercial organisations do not normally figure in listings such as these, since they are focused on producing some commercial product rather than academic knowledge; the fact that several such organisations do appear here (albeit not at very high positions in the ranking) reflects the fact that much of the research and the majority of the applications work in this subject field is carried out in industry.

CONCLUSIONS

2007 sees the Molecular Graphics and Modelling Society celebrating its 25th anniversary, and this paper has discussed the publications in, and the citations to, articles in its associated

journal, the *Journal of Molecular Graphics and Modelling*. The most highly-cited papers are those describing systems or algorithms, but the proportion of these types of article is decreasing as more applications of molecular graphics and molecular modelling are reported. The journal is international in scope, receiving papers from around the world, but with the USA being by far the largest source of articles.

REFERENCES

[1] E. Garfield, Citation Indexing -- Its Theory and Application in Science,Technology, and Humanities, ISI Press, Philadelphia, PA 1979.

[2] B. Cronin, The Citation Process. The Role and Significance of Citations in Scientific Communication, Taylor Graham, London, 1984.

[3] C.S. Wilson, Informetrics, Ann. Rev. Inf. Sci. Tech. 34 (1999) 107-247.

[4] W.W. Hood, C.S. Wilson, The literature of bibliometrics, scientometrics, and informetrics, Scientometrics 52 (2001) 291-314.

[5] C.L. Borgman, J. Furner, Scholarly communication and bibliometrics, Ann. Rev. Inf. Sci. Tech. 36 (2002) 3-72.

[6] M. Thelwall, L. Vaughan, L. Björneborn, Webometrics, Ann. Rev. Inf. Sci. Tech. 39 (2005) 81-135.

[7] N. Onodera, A bibliometric study on chemical information and computer sciences focusing on literature of JCICS, J. Chem. Inf. Comput. Sci. 41 (2001) 878-888.

[8] J. Redman, P. Willett, F.H. Allen, R. Taylor, A citation analysis of the Cambridge Crystallographic Data Centre, J. Appl. Cryst. 34 (2001) 375-380.

[9] H. Behrens, P. Luksch, A bibliometric study in crystallography, Acta Cryst. B62 (2006) 993-1001.

[10] P. Willett, A bibliometric analysis of chemoinformatics, Aslib. Proc., in the press (2007).

[11] E. Bonnevie, A multifaceted portrait of a library and information science journal: the case of the Journal of Information Science, J. Inf. Sci. 29 (2003) 11-23.

[12] E. Nebelong-Bonnevie, T.F. Farndsen, Journal citation identity and journal citation image: a portrait of the Journal of Documentation, J. Docum. 62 (2006) 30-57.

[13] T.E. Nisonger, JASIS and library and information science journal rankings: a review and analysis of the last half-century, J. Amer. Soc. Inf. Sci. 50 (1999) 1004-1019.

[14] A.P. Young, Library Quarterly, 1956-2004: an exploratory bibliometric analysis, Lib. Quart. 76 (2006) 10-18.

[15] P. Jacso, As we may search - comparison of major features of the Web of Science, Scopus, and Google Scholar citation-based and citation-enhanced databases, Curr. Sci. 89 (2005) 1537-1547.

[16] L. Meho, K. Yang, A new era in citation and bibliometric analyses: Web of Science, Scopus and Google Scholar, J. Amer. Soc. Inf. Sci. Tech., in the press (2007).

[17] R.A. Fairthorne, Empirical hyperbolic distributions (Bradford-Zipf-Mandelbrot) for bibliometric description and prediction, J. Docum. 25 (1969) 319-343.

[18] L. Egghe, The duality of informetric systems with applications to the empirical laws, J. Inf. Sci. 16 (1990) 17-27.

[19] D. de Solla Price, A general theory of bibliometric and other cumulative advantage processes, J. Amer. Soc. Inf. Sci. 27 (1976) 292-306.

[20] M.L. Pao, An empirical examination of Lotka's Law, J. Amer. Soc. Inf. Sci. 37 (1986) 26-33.

[21] M.L. Pao, Lotka's Law: a testing procedure, Inform. Proc. Manag. 21 (1985) 305-320.

[22] B. Rousseau, R. Rousseau, LOTKA: a program to fit a power law distribution to observed frequency data, Cybermetrics 4 (2000) http://www.cindoc.csic.es/cybermetrics/articles/v4i1p4.html.
[23] A.I. Pudovkin, E. Garfield, Algorithmic procedure for finding semantically related journals, J. Amer. Soc. Inf. Sci. Tech. 53 (2002) 1113-1119.
[24] L. Leydesdorff, Visualization of the citation impact environment of scientific journals: An online mapping exercise, J. Amer. Soc. Inf. Sci. Tech. 58 (2007) 25-38.