Aslib Proceedings

A bibliometric analysis of the literature of chemoinformatics
Peter Willett,

Article information:

To cite this document:
Peter Willett, (2008) "A bibliometric analysis of the literature of chemoinformatics", Aslib Proceedings, Vol. 60 Issue: 1, pp.4-17, https://doi.org/10.1108/00012530810847335
Permanent link to this document:
https://doi.org/10.1108/00012530810847335

Downloaded on: 10 May 2018, At: 02:09 (PT)
References: this document contains references to 47 other documents.
To copy this document: permissions@emeraldinsight.com
The fulltext of this document has been downloaded 1524 times since 2008*

Users who downloaded this article also downloaded:

(2011),"Journal bibliometric analysis: a case study on the Journal of Documentation", Journal of Documentation, Vol. 67 Iss 5 pp. 806-822 <a href="https://doi.org/10.1108/00220411111164682">https://doi.org/10.1108/00220411111164682</a>

(2011),"Computer science research in Malaysia: a bibliometric analysis", Aslib Proceedings, Vol. 63 Iss 2/3 pp. 321-335 <a href="https://doi.org/10.1108/00012531111135727">https://doi.org/10.1108/00012531111135727</a>

Access to this document was granted through an Emerald subscription provided by emerald-srm:395687 []

For Authors

If you would like to write for this, or any other Emerald publication, then please use our Emerald for Authors service information about how to choose which publication to write for and submission guidelines are available for all. Please visit www.emeraldinsight.com/authors for more information.

About Emerald www.emeraldinsight.com

Emerald is a global publisher linking research and practice to the benefit of society. The company manages a portfolio of more than 290 journals and over 2,350 books and book series volumes, as well as providing an extensive range of online products and additional customer resources and services.

Emerald is both COUNTER 4 and TRANSFER compliant. The organization is a partner of the Committee on Publication Ethics (COPE) and also works with Portico and the LOCKSS initiative for digital archive preservation.

*Related content and download information correct at time of download.

Downloaded by Indian Institute of Technology Kharagpur At 02:09 10 May 2018 (PT)

The current issue and full text archive of this journal is available at
www.emeraldinsight.com/0001-253X.htm

A bibliometric analysis of the literature of chemoinformatics

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

Received 10 February 2007
Revised 22 February 2007
Accepted 26 February 2007

Downloaded by Indian Institute of Technology Kharagpur At 02:09 10 May 2018 (PT)

Abstract

Purpose – The purpose of this article is to analyse the literature of chemoinformatics, a subject that has arisen over the last few years and that draws on techniques from a range of disciplines, most notably chemistry (particularly computational and medicinal chemistry), computer science and information science.

Design/methodology/approach – Discusses subject, author and citation searches of (principally) the web of knowledge database.

Findings – The *Journal of Chemical Information and Modeling* (previously the *Journal of Chemical Information and Computer Sciences*) is the core journal for the subject, but with many significant papers being published in journals whose principal focus is molecular modelling, quantitative structure-activity relationships or more general aspects of chemistry. The discipline is international in scope, and many of the most cited papers describe software packages that play a key role in modern chemoinformatics research.

Originality/value – This is the first bibliometric study of chemoinformatics, and one of only a very few that consider the bibliometrics of computational chemistry more generally.

Keywords Information retrieval, Chemistry, Research

Paper type Research paper

Introduction

Chemical information has been processed and exploited for many years, first in printed (Cooke, 2004) and then in computer form (Gasteiger, 2003; Hann and Green, 1999). It is now, under the name of "chemoinformatics", a key component of modern chemical research (Gasteiger and Engel, 2003; Leach and Gillet, 2003). Chemoinformatics' enhanced role has come about principally from the vast increase that has occurred in the volumes of data that need to be stored, searched and mined in research programmes for the discovery of biologically active molecules, most obviously but not exclusively in the pharmaceutical and agrochemical industries. These programmes involve the synthesis of large numbers of chemical compounds, followed by testing to identify those (normally very few) molecules that exhibit the biological activity of interest, e.g. lowering a person's blood pressure. The explosion in research data has been occasioned by technological developments that have enabled both chemical synthesis and biological testing to move from an inherently sequential to a massively parallel mode of processing: combinatorial synthesis enables large numbers – hundreds or even thousands – of structurally related molecules to be synthesised

The author gives thanks to Dr Val Gillet, Dr Alex Tropsha and Dr Wendy Warr for comments on this paper, and to Chemical Abstracts Service for providing the country-of-origin data for papers in the 2006 issues of *Journal of Chemical Information and Modeling*.

Aslib Proceedings: New Information Perspectives
Vol. 60 No. 1, 2008
pp. 4-17
© Emerald Group Publishing Limited
0001-253X
DOI 10.1108/00012530810847335

Downloaded by Indian Institute of Technology Kharagpur At 02:09 10 May 2018 (PT)

simultaneously, and high-throughput screening (HTS) enables these molecules to undergo testing for (normally) *in vitro* biological activity simultaneously.

The first formal definition of chemoinformatics was that of Brown (1998) who stated that "the use of information technology and management has become a critical part of the drug discovery process. Chemoinformatics is the mixing of those information resources to transform data into information and information into knowledge for the intended purpose of making better decisions faster in the area of drug lead identification and optimization", a definition that ties the subject very closely to the pharmaceutical industry where many of the key developments have taken place. A more general definition is that of Paris, as cited by Warr (1999):

> Chem(o)informatics is a generic term that encompasses the design, creation, organization, management, retrieval, analysis, dissemination, visualization and use of chemical information.

Most recently, Gasteiger (2006) has referred to it as "the application of informatics methods to the solution of chemical problems".

In this paper, we shall take 1998 as the starting point for our analysis, as this was when Brown's first formal definition of chemoinformatics appeared. That said, many of the basic techniques in chemoinformatics were developed prior to that date; indeed, the title of the paper by Hann and Green (1999) is "Chemoinformatics – a new name for an old problem". The 1998 starting point is thus rather arbitrary in nature and the interested reader is referred to several accounts (Chen, 2006; Engel, 2006; Willett, 2003) that describe the historical development of the subject and of its core technologies, e.g. the use of graph, statistical and expert-system methods for searching chemical structure databases, for predicting biological activity, and for designing synthetic pathways, respectively.

Bibliometrics involves the quantitative analysis of the literature of a subject domain, as represented by bibliographic entities such as keywords, classification codes, authors and citations. The newness of chemoinformatics – it is only recently that the first textbooks (Gasteiger and Engel, 2003; Leach and Gillet, 2003) and the first academic specialist courses (Wild and Wiggins, 2006) have appeared – means that there have been very few bibliometric analyses to date. Indeed, the only detailed study is that of Onodera (2001), which commenced with an analysis of the papers chosen for abstracting in subsection 20-5 of the *Chemical Abstracts* database. This subsection is entitled "Chemical information, documentation and data processing" and Onodera showed that the *Journal of Chemical Information and Computer Sciences* was by far the most frequently occurring journal in this subsection during the period 1972-2000 (i.e. mostly prior to the recognition of chemoinformatics as a distinct discipline). Onodera then analysed the indexing terms assigned to articles appearing in this core journal and demonstrated that there had been noticeable changes in content over the years, with the initial focus on information science and computer applications – particularly techniques for representing and searching databases of chemical structures – being broadened to encompass topics such as property prediction, simulation and modelling (which were referred to as the molecular information sciences). The change in focus has been reflected in changes in the name of the journal: it started life as the *Journal of Chemical Documentation* (1961-1974), then became the *Journal of Chemical Information and Computer Sciences* (1975-2004) and took its current title of the *Journal of Chemical*

Downloaded by Indian Institute of Technology Kharagpur At 02:09 10 May 2018 (PT)

Information and Modeling as recently as 2005; in what follows, we shall refer to this journal as *JCICS*, irrespective of the precise date of publication that is being considered. The move from traditional chemical information science to the broader molecular information sciences was also noted in a subsequent paper by Onodera (2003) that analysed the papers presented over 25 years at the Japanese "Symposia on Chemical Information and Computer Science"; this paper also considered the distribution of author affiliations and the relative importance of academic and of industrial contributions to the symposia. Finally, the most important papers in *JCICS*, defined as those attracting at least 100 citations since 1997, are briefly discussed in a review by Warr (2005) of the historical development of the field.

Journal coverage of cheminformatics

The very recent appearance of chemoinformatics as a distinct discipline is clearly indicated by the fact that there is still some disagreement as to its name, with two closely related names being used to describe the field: cheminformatics and chemoinformatics (and a third, chemiinformatics, that is arguably more correct from a linguistic point but far less mellifluous when spoken). A constantly updated analysis of Google postings (www.molinspiration.com/chemoinformatics.html) suggests that cheminformatics is used noticeably more frequently than chemoinformatics. Table I lists the postings frequencies for searches for the three chem?informatics variants and for four related phrases that occur in the literature; these searches involved Google, Google Scholar, the Web of Knowledge (WOK) and Scopus[1]. In this table the Google occurrence-frequencies are all described by the database as "about", the WOK occurrences are based on the title, keywords and abstract for each document in the *Science Citation Index*, the *Social Science Citation Index* and the *Arts and Humanities Citation Index*, and the Scopus occurrences are based on all fields; the Google Scholar, WOK and Scopus occurrences are from 1998 onwards. Of the three chem?informatics variants, cheminformatics is clearly the most used in common parlance, but chemoinformatics would appear to be the most used in the academic literature: in this respect, www.amazon.com lists six books with chemoinformatics in the title (Bajorath, 2004; Gasteiger, 2003; Gasteiger and Engel, 2003; Lavine, 2005; Leach and Gillet, 2003; Oprea, 2005), as against just one with cheminformatics (Noordik, 2004); there is also one entitled *Chemical Information Management* (Suhr and Warr, 1992). We shall generally use chemoinformatics in this paper.

Articles on chemoinformatics may not, of course, contain that particular word (or a variant); but articles that do contain it may be assumed (with a fair degree of

Table I.
Occurrences of search terms in Google, Google Scholar, the Web of Knowledge and Scopus

Term	Google	Google Scholar	Web of Knowledge	Scopus
Chemical documentation	695,000	66	1	34
Chemical informatics	50,400	129	20	39
Chemical information management	978	42	4	28
Chemical information science	779	17	2	5
Chemiinformatics	2,230	2	2	2
Cheminformatics	320,000	447	83	250
Chemoinformatics	191,000	5,636	99	473

Downloaded by Indian Institute of Technology Kharagpur At 02:09 10 May 2018 (PT)

probability) to contain material about that subject. Journals that publish relevant material were hence sought using the query: “chemoinformatics OR cheminformatics OR ‘chemical informatics’”, the three most common search terms in academic usage in Table I. This search of the title, keyword and abstract fields retrieved 197 post-1997 documents in the WOK database, with 13 literature sources yielding a minimum of three documents as shown in Table II. Of these documents, the majority were journal articles with meeting abstracts the next most-common document type. With the exception of the top-ranked entry, which refers to papers presented at the twice-yearly national conferences of the American Chemical Society, it will be seen that the list is dominated by *JCICS*, hence confirming that it is the core journal for the subject. That apart the list contains several broadly-based chemical journals (*Drug Discovery Today*, *Current Opinion in Drug Discovery and Development*, *Chimia*, *Indian Journal of Chemistry Section A*, and *Molecules*), with the remainder being specialist journals covering topics that are very closely related to chemoinformatics such as bioinformatics, HTS and molecular diversity analysis, molecular modelling and quantitative structure-activity relationships (QSAR). These closely related subjects often appear in *JCICS*: for example, its 412 documents make it the largest single source represented in the 4,746 documents retrieved in a WOK search for “QSAR OR ‘quantitative structure-activity relationship*’”, ranking it higher than the specialist journals in the field, i.e. *Journal of Computer-Aided Molecular Design*, *Journal of Medicinal Chemistry*, *Journal of Molecular Graphics and Modelling* and *QSAR and Combinatorial Science* (previously entitled *Quantitative Structure-Activity Relationships*). In like manner, *JCICS*' 49 documents make it the largest source represented in the 1,308 citations retrieved in a WOK search for “molecular diversity”, with the other high-ranked journals here being from the fields of biology and genetics (where “diversity” has a rather different meaning). It is this increased scope (going beyond the traditional focus on chemical database searching noted by Onodera (2001)), that seems to have triggered the recent change of name from *Journal of Chemical Information and Computer Sciences* to *Journal of Chemical Information and Modeling* (Jorgensen, 2005).

Citation source	Number of documents
Abstracts of papers of the American Chemical Society	44
Journal of Chemical Information and Computer Sciences/Journal of Chemical Information and Modeling	22
Drug Discovery Today	11
Combinatorial Chemistry and High-Throughput Screening	5
Bioinformatics	4
Current Opinion in Drug Discovery and Development	4
Journal of Computer-Aided Molecular Design	4
Molecular Diversity	4
Quantitative Structure-Activity Relationships/QSAR and Combinatorial Science	4
Chimia	3
Indian Journal of Chemistry Section A	3
Journal of Biomolecular Screening	3
Molecules	3

Table II. Most frequently occurring literature sources in a search of the Web of Knowledge for cheminformatics, chemoinformatics or chemical informatics

Downloaded by Indian Institute of Technology Kharagpur At 02:09 10 May 2018 (PT)

It is clear that chemoinformatics (in its various linguistic forms) is perceived to be rather different from chemical information, since the addition of "OR 'chemical information'" to the WOK query in Table II yielded 1,024 documents in a wide-range of journals. Thus, the top three sources in Table II were joined at the head of the ranked list by physical and analytical chemistry journals (*Analytical Chemistry*, *Analytica Chimica Acta*, *Applied Surface Science* and *Applied Spectroscopy*), and large numbers of more general chemical journals appeared high in the rankings (e.g. *Journal of Chemical Education*, *Analyst*, *Journal of Chromatography A*, and *Proceedings of the National Academy of Sciences of the United States of America*). None of these are journals that researchers in chemoinformatics would regard as key sources for their discipline.

An alternative approach to the analysis of the core journals in a discipline has been described recently by Leydesdorff (2007). Drawing on an extensive analysis of WOK data, he has made available for each of over 7,000 journals those journals that were responsible in 2003-2004 for at least 1 per cent of the citations to a given journal. There are 15 such journals in the case of *JCICS*, these including *Combinatorial Chemistry and High Throughput Screening*, *Current Opinion in Drug Discovery and Development*, *Journal of Computer-Aided Molecular Design*, and *QSAR and Combinatorial Science* (and, of course, *JCICS* itself) from Table II, and *Journal of Molecular Graphics and Modelling,* and *SAR and QSAR in Environmental Research* from amongst those listed in Table III (as discussed in the next section).

Leydesdorff also provides comparable data for those journals providing at least 1 per cent of the citations from (rather than to) a chosen journal. In the case of *JCICS* there are just five such journals (apart from *JCICS* itself), of which there is one – *Journal of Computer-Aided Molecular Design* – from amongst those in Tables II and III. The much smaller number of "citations from", as against "citations to" journals shows that *JCICS* papers cite a range of journals, rather than focusing on just a small number covering the same subject domain. This may be due to the fact that chemoinformatics is still emerging as a topic in its own right and that it is inherently multi-disciplinary in nature, drawing on work in both more general subjects (chemistry, computing, and library and information science) and more specific subjects (databases, medicinal chemistry, molecular modelling, QSAR etc.), which would imply that only a few journals would meet the 1 per cent criterion. Journal data for citations to or from *JCICS* for the period 1981-1998 (i.e. before the emergence of chemoinformatics as a distinct discipline) are provided by Onodera (2001).

Bibliometric studies have traditionally used the WOK databases to obtain productivity and citation data, but the last few years have seen the introduction of several new sources of bibliometric information, most importantly the Google Scholar and Scopus databases. The relative merits of the various resources are being increasingly discussed (Jacso, 2005; Meho and Yang, 2007), and it has been suggested that multiple data sources need to be used if comprehensive statistics are to be obtained. In what follows, we have used just WOK data, but would not expect radically different conclusions were other sources to be used: for example, carrying out the search in Table II on Scopus gave a list of the 13 top-ranked journals that was headed by *JCICS* (the *Abstracts of papers of the American Chemical Society* does not appear in the *Scopus* database) and also contained *Combinatorial Chemistry and High-Throughput Screening*, *Current Opinion in Drug Discovery and Development*,

Downloaded by Indian Institute of Technology Kharagpur At 02:09 10 May 2018 (PT)

Table III. Most productive authors in the ten specialist journals identified in Table II

Bioinformatics (3,492)	*Combinatorial Chemistry and High-Throughput Screening* (532)	*Journal of Biomolecular Screening* (542)	*Journal of Chemical Information and Computer Sciences/Modeling* (1,903)	*Journal of Computer-Aided Molecular Design* (530)	*Molecular Diversity* (122)	*Quantitative Structure-Activity Relationships/QSAR and Combinatorial Science* (496)	*Journal of Molecular Graphics and Modelling* (582)	*Journal of Molecular Modeling* (550)	*SAR and QSAR in Environmental Research* (328)
Valencia, A. (27)	Crameri, R. (8)	Burns, D.J. (15)	*Randic, M. (38)*	*Willett, P. (9)*	Jung, G. (4)	Walker, J.D. (13)	Boyd, D.B. (12)	*Clark, T. (17)*	Devillers, J. (23)
Apweiler, R. (23)	Zucht, H.D. (8)	Warrior, U. (14)	*Willett, P. (33)*	Holtje, H.D. (8)	Yavari, I. (4)	*Cronin, M.T.D. (11)*	Goodsell, D.S. (10)	Capkova, P. (12)	*Cronin, M.T.D. (15)*
Dougherty, E.R. (22)	Schulz-Knappe, P. (7)	Pope, A.J. (12)	*Basak, S.C. (32)*	Liljefors, T. (7)	Dolle, R.E. (3)	Schneider, G. (11)	*Willett, P. (10)*	Forner, W. (11)	Schultz, T.W. (15)
Lengauer, T. (22)	Tammen, H. (7)	Oldenburg, K.R. (8)	Jurs, P.C. (32)	Filizola, M. (6)	Kim, S.W. (3)	*Fan, B.T. (10)*	*Bajorath, J. (9)*	Badawi, H.M. (10)	*Basak, S.C. (14)*
Baldi, P. (17)	Van Breemen, R.B. (7)	Sills, M.A. (7)	Katritzky, A.R. (30)	Oprea, T.I. (6)	Xu, J. (3)	Meldal, M. (10)	Gaber, B.P. (8)	Guseinov, I. (8)	*Fan, B.T. (11)*
Bork, P. (17)	Hess, R. (6)	Chung, T.D.Y. (6)	*Bajorath, J. (28)*	*Carbo-Dorca, R. (5)*	Afantitis, A. (2)	Roy, K. (10)	Griffith, R. (8)	Aviyente, V. (7)	*Dearden, J.C. (10)*
Vingron, M. (17)	Kellmann, M. (6)	Sewing, A. (6)	Karelson, M. (25)	Carotti, A. (5)	*Agrafiotis, D.K. (2)*	Schaper, K.J. (10)	Hubbard, R.E. (6)	Lai, L.H. (7)	*Mekenyan, O. (10)*
Kim, S. (14)	Actor, J.K. (5)	Wildey, M.J. (6)	*Carbo-Dorca, R. (21)*	*Clark, R.D. (5)*	Akerblom, E.B. (2)	Raevsky, O.A. (8)	Martin, N.H. (6)	Zakarya, D. (7)	*Doucet, J.P. (8)*
Ouzounis, C.A. (14)	Kyle, D.J. (5)	Arnold, F.H. (5)	Balaban, A.T. (20)	Dean, P.M. (5)	Andersson, PL (2)	Schuurmann, G. (8)	Welsh, W.J. (6)	*Bajorath, J. (6)*	Panaye, A. (8)
Stormo, G.D. (14)	Yao, S.Q. (5)	Ashman, S. (5)	Godden, J.W. (20)	*Fan, B.T. (5)*	*Bajorath, J. (2)*	Wiese, M. (8)	*Agrafiotis, D.K. (5)*	Jiao, H.J. (6)	Worth, A.P. (8)
Xu, Y. (14)	Appel, A. (4)	Banks, P. (5)	Gutman, I. (17)	*Gasteiger, J. (5)*	Baxter, A.D. (2)	Alberico, F. (7)	Chatterjee, A. (5)	Brickmann, J. (5)	Cunningham, A.R. (7)
Mewes, H.W. (13)	Bazylak, G. (4)	Beutel, B.A. (5)	Trinajstic, N. (17)	Karplus, M. (5)	Eriksson, L. (2)	Darvas, F. (6)	*Clark, R.D. (5)*	Hou, T.J. (5)	Dimitrov, S (7)
Noble, W.S. (13)	Heilker, R. (4)	Eglen, R.M. (5)	*Fan, B.T. (16)*	Ramos, M.J. (5)	Houghten, R.A. (2)	*Mekenyan, O. (6)*	*Clark, T. (5)*	Lanig, H. (5)	Gute, B.D. (7)
Zimmer, R. (13)	Hindsgaul, O. (4)	Fox, S. (5)	Xue, L. (16)	Sanz, F. (5)	Igglessi-Markopoulou, O. (2)	Vedani, A. (6)	Flower, D.R. (5)	Murray, J.S. (5)	Mills, D. (7)
Bateman, A. (12)	Jolley, M.E. (4)	Gribbon, P. (5)	Estrada, E. (15)	Verdonk, M.L. (5)	Jacchieri, S.G. (2)	Breton, R. (5)	Maigret, B. (5)	Selcuki, C. (5)	Netzeva, T.I. (7)
Bruniak S. (12)	Kassel, D.B. (4)	Kariv, I. (5)	*Gasteiger, J. (15)*	Carrieri, A. (4)	Johansson, E. (2)	Carreira, L.A. (5)	Reynolds, C.H. (5)	Wang, R.X. (5)	*Randic, M. (7)*
Kolchanov, N.A. (12)	Kay, B.K. (4)	Kofron, J.L. (5)	Rucker, C. (15)	Centeno, N.B. (4)	Koh, J.S. (2)	*Dearden, J.C. (5)*	Schulten, K. (5)	Weiss, Z. (5)	Rosenkranz, H.S. (7)
Sander, C. (12)	Kumar, A. (4)	Moore, K.J. (5)	Tropsha, A. (15)	Cruciani, G. (4)	Lee, E.J. (2)	Dorman, G. (5)	Waldman, M. (5)	Wu, G. (5)	Carlsen, L. (6)
Brass, A. (11)	Lehrach, H. (4)	Parker, C.N. (5)	*Gillet, V.J. (14)*	Gago, F. (4)	Melagraki, G. (2)	*Doucet, J.P. (5)*	Winkler, D.A. (5)	Xu, X.J. (5)	Carpy, A.J.M. (6)
Gerstein M. (11)	Mitscher, L.A. (4)	Prossnitz, E.R. (5)	Gomez-Nieto, M.A. (14)	*Gillet, V.J. (4)*	Perez-Paya, E. (2)	Fernandez-Forner, D. (5)	Burton, N.A. (4)	Yan, S.M. (5)	Marchand-Geneste, N. (6)

Notes: The bracketed number following each journal (or author) name is the number of documents published in that journal in the period 1998-2006; Italicised authors appear in more than one column of the table

Downloaded by Indian Institute of Technology Kharagpur At 02:09 10 May 2018 (PT)

Drug Discovery Today, *Journal of Computer-Aided Molecular Design*, *Molecular Diversity*, *Molecules* and *QSAR and Combinatorial Science*.

Author productivity

The journals in Table II are those that have made most use of the term "chemoinformatics" (and its variants) and can hence be considered as having this topic as a focus of interest. However, an author analysis suggests that whilst researchers in bioinformatics and HTS (as exemplified by the journals *Bioinformatics*, *Combinatorial Chemistry and High-Throughput Screening*, and *Journal of Biomolecular Screening*) are aware of the importance of chemoinformatics, the most productive researchers do not publish frequently in the core chemoinformatics journals. The results of this analysis are shown in Table III, which summarises the outputs of WOK searches for 1998-2006 carried out on the specialist journals from Table II (i.e. *Bioinformatics*, *Combinatorial Chemistry and High-Throughput Screening*, *Journal of Biomolecular Screening*, *JCICS*, *Journal of Computer-Aided Molecular Design*, *Molecular Diversity* and *Quantitative Structure-Activity Relationships/QSAR and Combinatorial Science*) and on the three further specialist journals listed in the right-hand columns of the table (all of which carry articles on chemoinformatics but insufficient to appear in the top-ranked journals in Table II). These are: *SAR and QSAR in Environmental Research*, which covers QSAR-related topics analogous to those published in *QSAR and Combinatorial Science*; and *Journal of Molecular Graphics and Modelling* and *Journal of Molecular Modeling*. The latter two publications are, with *Journal of Computer-Aided Molecular Design*, the leading journals for the modelling of small chemical molecules (as against the modelling of biological macromolecules, which are covered in journals such as *Journal of Molecular Biology*, *Nucleic Acids Research* and *Proteins*).

For each journal in Table III, we have listed the 20 most productive authors in this period, using the analyse results and citation reports routines in WOK; similar, but more extended, facilities are available in the HISTCITE system (Garfield and Pudovkin, 2004). The reader should note that the use of a fixed cut-off (both here and elsewhere in the paper) means that there may well be other authors who published as many papers in a particular journal as the 20th-ranked author for that journal. Each column in the table represents one journal, with the number of papers published in the journal during 1998-2006 in brackets after the journal's name; each of the 20 elements of the column then contains an author name and the number of papers (bracketed) published in that journal during that period by that author. Authors appearing in more than one column, i.e. individuals who are productive in multiple journals, are listed in boldface italics.

Inspection of the extent to which individual authors publish across the range of journals listed here shows that none of the highly productive researchers in bioinformatics or HTS publish to any great extent in the chemoinformatics, molecular modelling and QSAR literatures (as represented by the other specialist journals in Table II); this comment applies to a lesser extent to *Molecular Diversity*, where the majority of the articles deal with combinatorial synthesis rather than the computational aspects of molecular diversity analysis. There is, however, a considerable degree of overlap between the other journals, and this is further emphasised if we include the three further specialist publications in the three right-hand columns of the table. Two of the authors in Table III, Bajorath and Fan,

Downloaded by Indian Institute of Technology Kharagpur At 02:09 10 May 2018 (PT)

publish extensively in four of the journals here, Willett publishes extensively in three, and there are 12 (Agrafiotis, Basak, Carbo-Dorca, Clark R.D. and Clark T., Cronin, Dearden, Doucet, Gasteiger, Gillet, Mekenyan, and Randic) who publish extensively in two of the journals. There is some degree of correlation between these highly productive authors: *QSAR and Combinatorial Science* and *SAR and QSAR in Environmental Science* share five highly productive authors, as do *JCICS* and *Journal of Computer-Aided Molecular Design*. The first pairing is hardly surprising given the titles and content of these two QSAR journals; the second pairing reflects the fact that *Journal of Computer-Aided Molecular Design*, one of the leading molecular modelling journals, publishes a fair number of database-related papers. *JCICS* has the greatest number (eight) of productive authors who are also productive authors in other journals, which again reflects the key role that this journal plays in chemoinformatics and its multi-disciplinary nature.

The core literature

One of the many uses of bibliometrics is the identification of the key publications in the development of a discipline, where the importance of a publication is assumed to be approximated by the number of citations to it, and we have hence sought the most cited papers in the core journal of *JCICS* (see also Warr, 2005) and in the associated specialist journals in the six right-hand columns of Table III. Searches were carried out for all documents in the chosen journals for the period 1998-2006, and the 4,411 resulting documents (of which over 90 per cent were articles) then ranked in decreasing order of the number of citations.

The 4,411 documents attracted a total of 35,228 citations, with the 20 most highly cited documents (all articles) listed in Table IV: many of these articles will be familiar to workers in the field of chemoinformatics, whatever their particular specialism. A characteristic of chemoinformatics is the widespread use of certain software packages (often available via specialist software companies such as Accelrys Inc. or Tripos Inc., *inter alia*) for, e.g. displaying molecules or searching databases. This has the result that many of the articles listed in Table IV are the "standard" references that are cited whenever anybody subsequently uses these packages: such articles are denoted in the table by "(S)" after the citation count. Obvious examples are GROMACS and MOLDEN (the two top papers in Table IV), DOCK and XCrySDen, as well as two others in the table where this is not obvious from the title of the paper: those by Pearlman and Smith and by Clark *et al.* describe the Diverse Solutions and CScore software packages, respectively. Indeed, the two most cited articles in the history of *JCICS* (Warr, 2005) come into this category, these being the standard references for the database searching systems used by the Cambridge Crystallographic Data Centre (Allen *et al.*, 1991) and by the Daresbury Chemical Database Service (Fletcher *et al.*, 1996); both of these pre-date our 1998 starting point and are not included in Table IV only because they have insufficient citations in the period 1998-2006. Review articles – denoted by "(R)" in Table IV – often attract large numbers of citations, e.g. from the introductory sections of subsequent papers, and there are two reviews here – those by Willett *et al.* (1998) and by Taylor *et al.* (2002). Of the remaining 12 articles in the table, no less than four discuss the characteristics that differentiate drugs from other, non-drug molecules (Hann *et al.*, 2001; Oprea, 2000; Oprea *et al.*, 2001), and there are two on the calculation of binding energies (i.e. the strength with which a drug molecule attaches itself to a

Downloaded by Indian Institute of Technology Kharagpur At 02:09 10 May 2018 (PT)

Table IV.
Most cited articles in seven chemoinformatics journals (those heading the seven right-hand columns in Table III) in 1998-2006

Highly cited article	Citations
Lindahl, E. *et al.* (2001), “GROMACS 3.0: a package for molecular simulation and trajectory analysis”, *Journal of Molecular Modeling*, Vol. 7, pp. 306-317	854 (S)
Schaftenaar, G. and Noordik, J.H. (2000), “Molden: a pre- and post-processing program for molecular and electronic structures”, *Journal of Computer-Aided Molecular Design*, Vol. 14, pp. 123-134	701 (S)
Willett, P. *et al.* (1998), “Chemical similarity searching”, *Journal of Chemical Information and Computer Sciences*, Vol. 38, pp. 983-996	291(R)
Dunker, A.K. *et al.* (2001), “Intrinsically disordered protein”, *Journal of Molecular Graphics and Modelling*, Vol. 19, pp. 26-59	239
Ewing, T.J.A. *et al.* (2001), “DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases”, *Journal of Computer-Aided Molecular Design*, Vol. 15, pp. 411-428	181 (S)
Golbraikh. A. and Tropsha, A. (2002), “Beware of q^2!”, *Journal of Molecular Graphics and Modelling*, Vol. 20, pp. 269-276	167
Wessel, M.D. *et al.* (1998), “Prediction of human intestinal absorption of drug compounds from molecular structure”, *Journal of Chemical Information and Computer Sciences*, Vol. 38, pp. 726-735	157
Oprea, T.I. *et al.* (2001), “Is there a difference between leads and drugs? A historical perspective”, *Journal of Chemical Information and Computer Sciences*, Vol. 41, pp. 1308-1315	145
Bohm, H.-J. (1998), “Prediction of binding constants of protein ligands: A fast method for the prioritization of hits obtained from de novo design or 3D database search programs”, *Journal of Computer-Aided Molecular Design*, Vol. 12, pp. 309-323	143 (S)
Platts, J.A. *et al.* (1999), “Estimation of molecular linear free energy relation descriptors using a group contribution approach”, *Journal of Chemical Information and Computer Sciences*, Vol. 39, pp. 835-845	137
Hann, M.M. *et al.* (2001), “Molecular complexity and its impact on the probability of finding leads for drug discovery”, *Journal of Chemical Information and Computer Sciences*, Vol. 41, pp. 856-864	131
Taylor, R.D. *et al.* (2002), “A review of protein-small molecule docking methods”, *Journal of Computer-Aided Molecular Design*, Vol. 16, pp. 151-166	130 (R)
Kokalj, A. (1999), “XCrySDen – a new program for displaying crystalline structures and electron densities”, *Journal of Molecular Graphics and Modelling*, Vol. 17, pp. 176-179	122
Oprea, T.I. (2000), “Property distribution of drug-related chemical databases”, *Journal of Computer-Aided Molecular Design*, Vol. 14, 251-264	113
Pearlman, R.S. and Smith, K.M. (1999), “Metric validation and the receptor-relevant subspace concept”, *Journal of Chemical Information and Computer Sciences*, Vol. 39, pp. 28-35	112 (S)
Clark, R.D. *et al.* (2002), “Consensus scoring for ligand/protein interactions”, *Journal of Molecular Graphics and Modelling*, Vol. 20, pp. 281-295	103 (S)
Wang, R.X. *et al.* (2002), “Further development and validation of empirical scoring functions for structure-based binding affinity prediction”, *Journal of Computer-Aided Molecular Design*, Vol. 16, pp. 11-26	102
Katritzky, A.R. *et al.* (2000), “Structurally diverse quantitative structure-property relationship correlations of technologically relevant physical properties”, *Journal of Chemical Information and Computer Sciences*, Vol. 40, pp. 1-18	100
Rusinko, A. *et al.* (1999), “Analysis of a large structure/biological activity data set using recursive partitioning”, *Journal of Chemical Information and Computer Sciences*, Vol. 39, pp. 1017-1026	97
Gillet, V.J. *et al.* (1998), “Identification of biological activity profiles using substructural analysis and genetic algorithms”, *Journal of Chemical Information and Computer Sciences*, Vol. 38, pp. 165-179	97

Notes: R denotes a review and S denotes a software package biological receptor) (Bohm, 1998; Wang *et al.*, 2002). The trends noted here continue if one goes further down the list of highly-cited documents, with the next ten rank positions containing two further reviews, three further software descriptions, and two further articles on the calculation of binding energies.

Onodera (2001) noted that a large fraction of *JCICS* articles originated from outside of the US, this fraction being greater than for any of the other journals published by the American Chemical Society, the world's largest publisher of chemical literature. This observation applies to the field of chemoinformatics more generally. Table V lists the geographical data for the ten most productive countries in the 1997-1998 issues of *JCICS* (Onodera, 2001) and in the set of 4,411 chemoinformatics documents described above. The US provided 34.1 per cent of the latter set of documents, but there were another 16 countries that provided at least 2 per cent of those for which country/territory data are available in the WOK database. Note that Table V does not contain an entry for the UK as such, since England, Scotland, Northern Ireland and Wales are entered separately in WOK; note also the perhaps surprisingly high *JCICS* rankings for Romania and Croatia, both of which have productive groups working in a very specific area of QSAR and publishing much of their research in this journal. The most obvious difference between the two parts of the table is the emergence of the People's Republic of China and India, both of which now compete strongly with the

JCICS 1997-1998		*JCICS* 2006		Chemoinformatics papers	
Country	%	Country	%	County	%
USA	40.6	USA	29.9	USA	34.1
England	10.4.	England	12.6	Germany	10.5
France	5.8	Germany	10.4	England	10.5
Germany	5.8	Japan	5.0	PR China	6.7
Slovenia	5.5	India	4.3	France	6.1
Japan	4.3	Italy	4.0	Spain	4.9
Romania	4.3	Canada	3.2	Italy	4.5
Croatia	4.0	France	3.2	Japan	3.5
Russia	3.7	Spain	3.2	India	3.1
PR China	3.2	Switzerland	3.2	Switzerland	2.8

Source: Onodera (2001)

Table V. Most productive countries for 347 papers in *JCICS* 1997-1998 for 278 papers in *JCICS* 2006, and for 4,411 documents in seven chemoinformatics journals in 1998-2006

Research centre	%
National Institute of Chemistry, Ljubljana	1.6
University of Erlangen-Nurnberg	1.6
University of Sheffield	1.6
University of Minnesota	1.5
Environmental Protection Agency	1.1
Russian Academy of Sciences	1.1
Liverpool John Moores University	1.0
Pennsylvania State University	1.0
Chinese Academy of Sciences	1.0
University of Cambridge	1.0

Table VI. Most productive research centres for 4,411 documents in seven chemoinformatics journals in 1998-2006

Downloaded by Indian Institute of Technology Kharagpur At 02:09 10 May 2018 (PT)

Downloaded by Indian Institute of Technology Kharagpur At 02:09 10 May 2018 (PT)

traditionally productive European research groups. For comparison with Onodera's figures, the table also contains data for the papers published in *JCICS* in 2006, which further demonstrate the broad spread of chemoinformatics research.

Finally, Table VI lists the most productive institutions in the set of 4,411 documents, this table reflecting many of the key research groups in chemoinformatics (e.g. those at the Universities of Erlangen-Nurnberg, Sheffield and Cambridge) and modelling or QSAR (e.g. the Environmental Protection Agency, the University of Minnesota, Liverpool John Moores University, and Pennsylvania State University). The most productive here is the National Institute of Chemistry in Ljubljana, Slovenia, which has conducted extensive research in various aspects of QSAR. All but two of the top 50 institutions are universities, governmental or not-for-profit organisations with just two – Tripos, one of the major chemoinformatics software companies, at rank-position 27 and Pfizer, the world's largest pharmaceutical research firm, at rank-position 36. 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 two such organisations do appear here reflects the fact that much of the leading-edge research in chemoinformatics is carried out in industry, principally by software companies that are developing chemoinformatics packages and by pharmaceutical companies who purchase and use these packages or develop their own in-house software.

Conclusions

Chemoinformatics first appeared as a distinct discipline in the late-1990s, since when it has generated a considerable literature. Analysis of data from, principally, the *Web of Knowledge* database shows that the *Journal of Chemical Information and Modeling* (previously the *Journal of Chemical Information and Computer Sciences*) is the core journal for the subject, but with many significant papers being published in journals whose principal focus is molecular modelling or QSAR, or more general aspects of chemistry. This paper highlights the most productive authors and institutions, noting the international nature of the discipline, and the most cited papers, many of which describe software packages that play a key role in modern chemoinformatics research.

Note

1. All the database searches in this paper were carried out in December 2006 and January 2007.

References

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. and Watson, D.G. (1991), "The development of versions 3 and 4 of the Cambridge Structural Database System", *Journal of Chemical Information and Computer Sciences*, Vol. 31, pp. 187-204.

Bajorath, J. (Ed.) (2004), *Chemoinformatics: Concepts, Methods and Tools for Drug Discovery*, Humana Press, Totowa, NJ.

Bohm, H.-J. (1998), "Prediction of binding constants of protein ligands: a fast method for the prioritization of hits obtained from de novo design or 3D database search programs", *Journal of Computer-Aided Molecular Design*, Vol. 12, pp. 309-23.

Brown, F.K. (1998), "Chemoinformatics: what is it and how does it impact drug discovery?", *Annual Reports in Medicinal Chemistry*, Vol. 33, pp. 375-84.

Downloaded by Indian Institute of Technology Kharagpur At 02:09 10 May 2018 (PT)

Chen, W.L. (2006), “Chemoinformatics: past, present and future”, *Journal of Chemical Information and Computer Sciences*, Vol. 46, pp. 2230-55.

Clark, R.D., Strizhev, A., Leonard, J.M., Blake, J.F. and Matthew, J.B. (2002), “Consensus scoring for ligand/protein interactions”, *Journal of Molecular Graphics and Modelling*, Vol. 20, pp. 281-95.

Cooke, H. (2004), “A historical study of structures for communication of organic chemistry information prior to 1950”, *Organic and Biomolecular Chemistry*, Vol. 2, pp. 3179-91.

Dunker, A.K., Lawson, J.D. and Brown, C.J. (2001), “Intrinsically disordered protein”, *Journal of Molecular Graphics and Modelling*, Vol. 19, pp. 26-59.

Engel, T. (2006), “Basic overview of chemoinformatics”, *Journal of Chemical Information and Modeling*, Vol. 46, pp. 2267-77.

Ewing, T.J.A., Makino, S., Skillman, A.G. and Kuntz, I.D. (2001), “DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases”, *Journal of Computer-Aided Molecular Design*, Vol. 15, pp. 411-28.

Fletcher, D.A., McMeeking, R.F. and Parkin, D. (1996), “The United Kingdom Chemical Database Service”, *Journal of Chemical Information and Modeling*, Vol. 36, pp. 746-9.

Garfield, E. and Pudovkin, A.I. (2004), “The HistCite system for mapping and bibliometric analysis of the output of searches using the ISI Web of Knowledge”, paper presented at the Annual ASIS&T Meeting, Newport, RI, available at: www.garfield.library.upenn.edu/papers/asistposter2004.pdf (accessed 5 October 2007).

Gasteiger, J. (Ed.) (2003), *Handbook of Chemoinformatics*, Wiley-VCH, Weinheim.

Gasteiger, J. (2006), “The central role of chemoinformatics”, *Chemometrics and Intelligent Laboratory Systems*, Vol. 82, pp. 200-9.

Gasteiger, J. and Engel, T. (Eds) (2003), *Chemoinformatics: A Textbook*, Wiley-VCH, Weinheim.

Gillet, V.J., Willett, P. and Bradshaw, J. (1998), “Identification of biological activity profiles using substructural analysis and genetic algorithms”, *Journal of Chemical Information and Computer Sciences*, Vol. 38, pp. 165-79.

Golbraikh, A. and Tropsha, A. (2002), “Beware of q^2!”, *Journal of Molecular Graphics and Modelling*, Vol. 20, pp. 269-76.

Hann, M. and Green, R. (1999), “Chemoinformatics: a new name for an old problem?”, *Current Opinion in Chemical Biology*, Vol. 3, pp. 379-83.

Hann, M.M., Leach, A.R. and Harper, G. (2001), “Molecular complexity and its impact on the probability of finding leads for drug discovery”, *Journal of Chemical Information and Computer Sciences*, Vol. 41, pp. 856-64.

Jacso, P. (2005), “As we may search: comparison of major features of the Web of Science, Scopus, and Google Scholar citation-based and citation-enhanced databases”, *Current Science*, Vol. 89, pp. 1537-47.

Jorgensen, W.L. (2005), “Editorial”, *Journal of Chemical Information and Modeling*, Vol. 45, p. 1.

Katritzky, A.R., Maran, U., Lobanov, V.S. and Karelson, M. (2000), “Structurally diverse quantitative structure-property relationship correlations of technologically relevant physical properties”, *Journal of Chemical Information and Computer Sciences*, Vol. 40, pp. 1-18.

Kokalj, A. (1999), “XCrySDen – a new program for displaying crystalline structures and electron densities”, *Journal of Molecular Graphics and Modelling*, Vol. 17, pp. 176-9.

Lavine, B.K. (Ed.) (2005), *Chemometrics and Chemoinformatics*, Chemical Society, Washington, DC.

Downloaded by Indian Institute of Technology Kharagpur At 02:09 10 May 2018 (PT)

Leach, A.R. and Gillet, V.J. (2003), *An Introduction to Chemoinformatics*, Kluwer, Dordrecht.

Leydesdorff, L. (2007), "Visualization of the citation impact environment of scientific journals: an online mapping exercise", *Journal of the American Society for Information Science and Technology*, Vol. 58, pp. 25-38.

Lindahl, E., Hess, B. and van der Spoel, D. (2001), "GROMACS 3.0: a package for molecular simulation and trajectory analysis", *Journal of Molecular Modeling*, Vol. 7, pp. 306-17.

Meho, L. and Yang, K. (2007), "Impact of data sources on citation counts and rankings of LIS faculty: Web of Science versus Scopus and Google Scholar", *Journal of the American Society for Information Science and Technology Early View*, available at: www3.interscience.wiley.com/cgi-bin/jissue/76504585 (accessed 5 October 2007).

Noordik, J.H. (Ed.) (2004), *Cheminformatics Developments: History, Reviews and Current Research*, IOS Press, Amsterdam.

Onodera, N. (2001), "A bibliometric study on chemical information and computer sciences focusing on literature of JCICS", *Journal of Chemical Information and Computer Sciences*, Vol. 41, pp. 878-88.

Onodera, N. (2003), "Research trends of chemical information and computer sciences in Japan observed from the publications at the 'Symposia on Chemical Information and Computer Science': a bibliometric analysis of the subjects and the authors' affiliated institutions", *Journal of Computer-Aided Chemistry*, Vol. 4, pp. 1-17.

Oprea, T.I. (2000), "Property distribution of drug-related chemical databases", *Journal of Computer-Aided Molecular Design*, Vol. 14, pp. 251-64.

Oprea, T.I. (Ed.) (2005), *Chemoinformatics in Drug Discovery*, Wiley-VCH, Weinheim.

Oprea, T.I., Davis, A.M., Teague, S.J. and Leeson, P.D. (2001), "Is there a difference between leads and drugs? A historical perspective", *Journal of Chemical Information and Computer Sciences*, Vol. 41, pp. 1308-15.

Pearlman, R.S. and Smith, K.M. (1999), "Metric validation and the receptor-relevant subspace concept", *Journal of Chemical Information and Computer Sciences*, Vol. 39, pp. 28-35.

Platts, J.A., Butina, D., Abraham, M.H. and Hersey, A. (1999), "Estimation of molecular linear free energy relation descriptors using a group contribution approach", *Journal of Chemical Information and Computer Sciences*, Vol. 39, pp. 835-45.

Rusinko, A., Farmen, M.W., Lambert, C.G., Brown, P.L. and Young, S.S. (1999), "Analysis of a large structure/biological activity data set using recursive partitioning", *Journal of Chemical Information and Computer Sciences*, Vol. 39, pp. 1017-26.

Schaftenaar, G. and Noordik, J.H. (2000), "Molden: a pre- and post-processing program for molecular and electronic structures", *Journal of Computer-Aided Molecular Design*, Vol. 14, pp. 123-34.

Suhr, C. and Warr, W.A. (1992), *Chemical Information Management*, Wiley-VCH, Weinheim.

Taylor, R.D., Jewsbury, P.J. and Essex, J.W. (2002), "A review of protein-small molecule docking methods", *Journal of Computer-Aided Molecular Design*, Vol. 16, pp. 151-66.

Wang, R.X., Lai, L.H. and Wang, S.M. (2002), "Further development and validation of empirical scoring functions for structure-based binding affinity prediction", *Journal of Computer-Aided Molecular Design*, Vol. 16, pp. 11-26.

Warr, W.A. (1999), "Balancing the needs of the recruiters and the aims of the educators", paper presented at the 218th American Chemical Society National Meeting, New Orleans, LA, available at: www.com/warrzone2000.html (accessed 5 October 2007).

Warr, W.A. (2005), "Twenty five years of progress in chemoinformatics", paper presented at the 229th American Chemical Society National Meeting, San Diego, CA, available at: www.warr.com/25years.html (accessed 5 October 2007).

Wessel, M.D., Jurs, P.C., Tolan, J.W. and Muskal, S.M. (1998), "Prediction of human intestinal absorption of drug compounds from molecular structure", *Journal of Chemical Information and Computer Sciences*, Vol. 38, pp. 726-35.

Wild, D.J. and Wiggins, G.D. (2006), "Challenges for chemoinformatics education in drug discovery", *Drug Discovery Today*, Vol. 11, pp. 436-9.

Willett, P. (2003), "A history of chemoinformatics", in Gasteiger, J. (Ed.), *Handbook of Chemoinformatics*, Wiley-VCH, Weinheim, pp. 6-20.

Willett, P., Barnard, J.M. and Downs, G.M. (1998), "Chemical similarity searching", *Journal of Chemical Information and Computer Sciences*, Vol. 38, pp. 983-96.

Corresponding author
Peter Willett can be contacted at: p.willett@sheffield.ac.uk

To purchase reprints of this article please e-mail: **reprints@emeraldinsight.com**
Or visit our web site for further details: **www.emeraldinsight.com/reprints**

Downloaded by Indian Institute of Technology Kharagpur At 02:09 10 May 2018 (PT)

Downloaded by Indian Institute of Technology Kharagpur At 02:09 10 May 2018 (PT)

This article has been cited by:

1. ChenQiuhong, Qiuhong Chen, GengNing, Ning Geng, ZhuKan, Kan Zhu. 2018. Review and bibliometric analysis of Chinese agricultural economics research: 2006-2015. *China Agricultural Economic Review* **10**:1, 152-172. [Abstract] [Full Text] [PDF]
2. Peter Willett. Molecular Similarity Approaches in Chemoinformatics: Early History and Literature Status 67-89. [Crossref]
3. Valerie J. Gillet, John D. Holliday, Peter Willett. 2015. Chemoinformatics at the University of Sheffield 2002-2014. *Molecular Informatics* **34**:9, 598-607. [Crossref]
4. Junping Qiu, Hong Lv. 2014. An overview of knowledge management research viewed through the web of science (1993-2012). *Aslib Journal of Information Management* **66**:4, 424-442. [Abstract] [Full Text] [PDF]
5. Kamal Lochan Jena, Dillip K. Swain, K.C. Sahoo. 2012. Scholarly communication in Journal of Financial Crime, 2006-2010: a bibliometric study. *Journal of Financial Crime* **19**:4, 371-383. [Abstract] [Full Text] [PDF]
6. Haijun Wang, Qingqing He, Xingjian Liu, Yanhua Zhuang, Song Hong. 2012. Global urbanization research from 1991 to 2009: A systematic research review. *Landscape and Urban Planning* **104**:3-4, 299-309. [Crossref]
7. Kamal Lochan Jena, Dillip K. Swain, Sada Bihari Sahu. 2012. Scholarly communication of The Electronic Library from 2003-2009: a bibliometric study. *The Electronic Library* **30**:1, 103-119. [Abstract] [Full Text] [PDF]
8. Peter Willett. 2011. Chemoinformatics: A history. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **1**:1, 46-56. [Crossref]
9. Peter Willett. Chemoinformatics 920-929. [Crossref]
10. Peter Willett. 2009. A Bibliometric Study of Quantitative Structure-Activity Relationships and QSAR & Combinatorial Science. *QSAR & Combinatorial Science* **28**:11â12, 1231-1236. [Crossref]
11. Peter Willett. 2009. Similarity methods in chemoinformatics. *Annual Review of Information Science and Technology* **43**:1, 1-117. [Crossref]