ELSEVIER

Journal of Molecular Structure (Theochem) 401 (1997) 219–225

THEO CHEM

Evidence that there is a future for semiempirical molecular orbital calculations

Donald B. Boyd

Department of Chemistry, Indiana University–Purdue University at Indianapolis (IUPUI), Indianapolis, IN 46202-3274, USA

Received 15 October 1996; revised 5 November 1996

Abstract

The frequency of use and discussion of semiempirical and ab initio software is traced with bibliometric data from the Current Journals of the American Chemistry Society (JCACS) database, which has complete papers published in 19 journals in various fields of chemistry. Not only is the use of all types of computational chemistry software increasing apace, but also the number of papers mentioning semiempirical programs as a percentage of all papers mentioning quantum chemistry programs is growing. This percentage referring to semiempirical molecular orbital software grew from 29% in 1989 to 34% in 1993. © 1997 Elsevier Science B.V.

Keywords: Semiempirical molecular orbital calculations; CJACS database; MOPAC; GAUSSIAN program; GAMESS; AMPAC

1. Introduction

We address the question the organizer of this symposium, Professor Andrew J. Holder, posed somewhat rhetorically: "Semiempirical Methods: Is There a Future?" Other papers presented at the symposium describe recent advancements that ensure the continued vitality of semiempirical molecular orbital methods. We take a different tack: we retrospectively examine the frequency of use of semiempirical software in the recent chemical literature. From these data, we can try to prognosticate the immediate future.

One way to gather information on the use of semiempirical calculations is to manually scan the dozens of journals that may have some prospect of having relevant papers. However, such a search would not only be tedious, but prone to human error. Consequently, we undertook a computerized search of a database with pertinent information. The Current Journals of the American Chemical Society (CJACS) database is produced by the Chemical Abstracts Service (CAS) and is available through the online service STN. CJACS covers 19 journals (Table 1) representing a wide cross-section of the chemical literature.

The CJACS database is large with more than 145 000 papers. Corresponding to the burgeoning scientific literature, more than 16 000 new papers are added each year. Of these, approximately 2000 (in 1993) mentioned computational chemistry software, including not just quantum chemistry programs but also molecular modeling and other software used by computational and other chemists interested in these tools [1]. About 1 in 8 papers in the online database relates to computational chemistry.

Of the 19 journals covered in CJACS, the most computational chemistry is found on the pages of

PII S0166-1280(97)00023-7

Table 1
Journals in the Current Journals of the American Chemical Society (CJACS) Online Database of the Chemical Abstracts Service (CAS)

Accounts of Chemical Research
Analytical Chemistry
Biochemistry
Chemical Reviews
Environmental Science and Technology
Industrial and Engineering Chemistry: Fundamentals
Industrial and Engineering Chemistry: Process Design and Development
Industrial and Engineering Chemistry: Product Design and Development
Inorganic Chemistry
Journal of Agricultural and Food Chemistry
Journal of Chemical and Engineering Data
Journal of Chemical Information and Computer Sciences
Journal of Medicinal Chemistry
Journal of Organic Chemistry
Journal of Physical Chemistry
Journal of the American Chemical Society
Langmuir
Macromolecules
Organometallics

Journal of the American Chemical Society, Journal of Medicinal Chemistry, Journal of Chemical Information and Computer Sciences, Journal of Physical Chemistry, Inorganic Chemistry, and Journal of Organic Chemistry. This is consonant with a recent manual search showing that approximately 1 in 4 papers in both *Journal of the American Chemical Society and Journal of Medicinal Chemistry* involve at least some computational chemistry [2]. In contrast, some of the other journals listed in Table 1 have little or none.

CJACS has the entire contents of each paper including text, tables, and references. Thus by searching for keywords tailored to identify specific computer programs, it is possible to retrieve those papers that use, discuss, cite, or otherwise mention those programs. No attempt was made to differentiate between full articles, notes, and communications. Details of the database searches have been reported elsewhere [1]. Because there are literally hundreds of semiempirical and ab initio programs, we made our task manageable by focusing on some of the most well-known programs (Fig. 1). Most of this software has been commercialized in the last 10 years. However, versions of two semiempirical programs that are important to our survey, MOPAC [3] and AMSOL [4], are still available inexpensively from QCPE [5,6]. The versatile ab initio program GAMESS remains free [7].

2. Results of the literature survey

Fig. 1 tracks the number of papers mentioning quantum chemistry programs in the 5-year period, 1989–1993. Several important conclusions are apparent from these findings. The use of the GAUSSIAN ab initio programs [8,9] far and away exceeds that of any other program compared. In 1993 almost 450 papers mentioned one of the GAUSSIAN series of programs (through GAUSSIAN 92 because the survey ended in 1993 before GAUSSIAN 94 appeared). The second most mentioned program was the semiempirical program MOPAC. About 150 papers in the CJACS journals mentioned MOPAC in 1993. Applications and discussions of GAUSSIAN and MOPAC have grown monotonically. In third place is another ab initio program GAMESS. It is noteworthy that programs available free or at nominal cost, such as MOPAC and GAMESS, are highly used. Below the top three curves, the other programs tracked in our survey bunch together and were mentioned in about 50 or fewer papers per year over the entire 5-year period.

With the data in Fig. 1, one can ask the additional question of whether the use of semiempirical methods is decreasing or increasing compared with the overall number of papers pertaining to quantum chemistry. It might be tempting to speculate that the proportion devoted to semiempirical calculations is decreasing because the speed of computers is growing and the cost of computers is declining. These two trends should make ab initio calculations more accessible to a wider number of users and applicable to a broader range of research problems. In fact, however, what we find is that the percentage of papers mentioning semiempirical programs is increasing (Fig. 2). In 1989 the percentage stood at 29%; by 1993 it had steadily grown to 34%. The important point of this graph (and of the other figures in this paper) is the qualitative trend, rather than the quantitative details.

Whereas quantum chemistry is important, it is only one aspect of computational chemistry. Today in academic and industrial research laboratories alike, chemists are using molecular modeling, molecular simulations, molecular databases, and other software

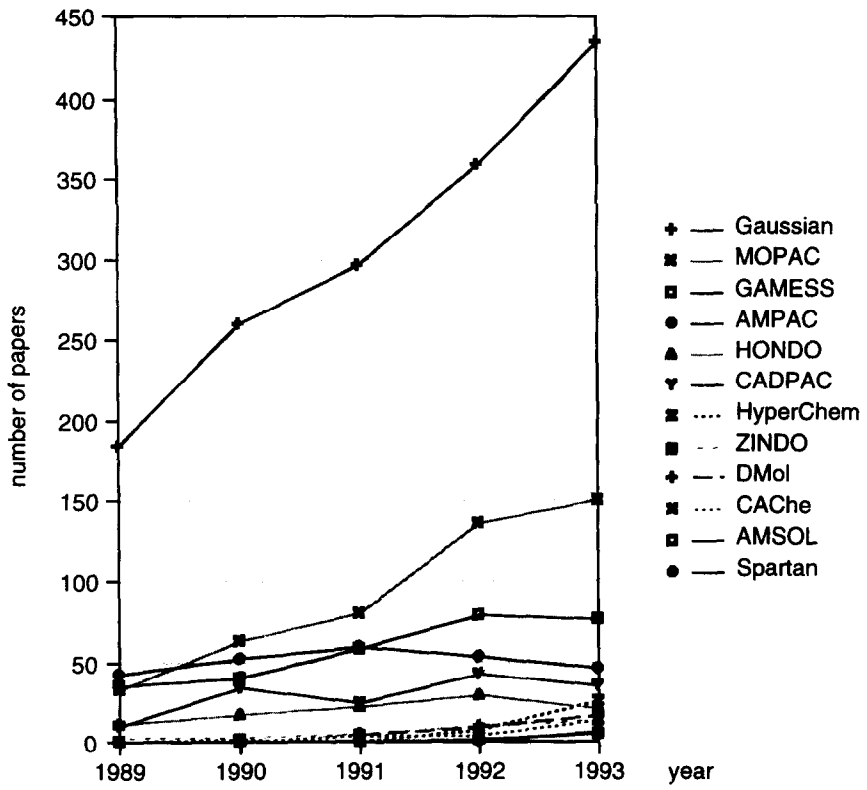

Fig. 1. Number of papers appearing in the CJACS journals from 1986–1993 that mention programs for molecular orbital calculations. The programs that were surveyed include all versions of the following: GAUSSIAN, MOPAC, GAMESS, AMPAC, HONDO, CADPAC, HYPERCHEM, ZINDO, DMOL, CACHE, AMSOL, and SPARTAN. Some of these programs are used for only quantum mechanical calculations, whereas others, namely HYPERCHEM, CACHE, and SPARTAN, have more versatile molecular modeling capabilities. GAUSSIAN, SPARTAN, and DMOL can do density functional theory calculations. The functionality of some of the programs overlap, e.g. GAUSSIAN, MOPAC, AMPAC, HYPERCHEM, ZINDO, CACHE, AMSOL, and SPARTAN, all have implementations of the various popular semiempirical Hamiltonians.

tools in the quest to understand and make predictions about the behavior of compounds. A wider analysis of the chemical literature showing the increased use of all sorts of programs is given elsewhere [1]. One interesting highlight of those results is summarized in Fig. 3, where it can be seen that the number of papers mentioning quantum chemistry programs is growing. Papers mentioning other categories of computational chemistry programs are also becoming more frequent. In fact, the growth rate for using force field-based software is greater than that for quantum chemistry. In 1989 and 1990, papers mentioning the quantum chemistry programs were more numerous than those mentioning force field approaches, but after 1991, the force field based programs have been more highly mentioned. A misconception that is unfortunately too pervasive among theoreticians is that all quantum mechanical methods, including semiempirical ones, are perforce superior to force field methods for all research questions. As this misconception is nullified, use of force method methods will continue to increase.

3. Discussion of the role of semiempirical methods

It is probably safe to project from our analysis that there is a future for semiempirical molecular orbital methods. This projection will hold at least for the foreseeable future until some other method or set of methods comes along that exceeds semiempirical methods in terms of speed, accuracy, and versatility. Two trends are apparent: all facets of computational chemistry are becoming more prevalent in the chemical literature, and the proportion of semiempirical-based

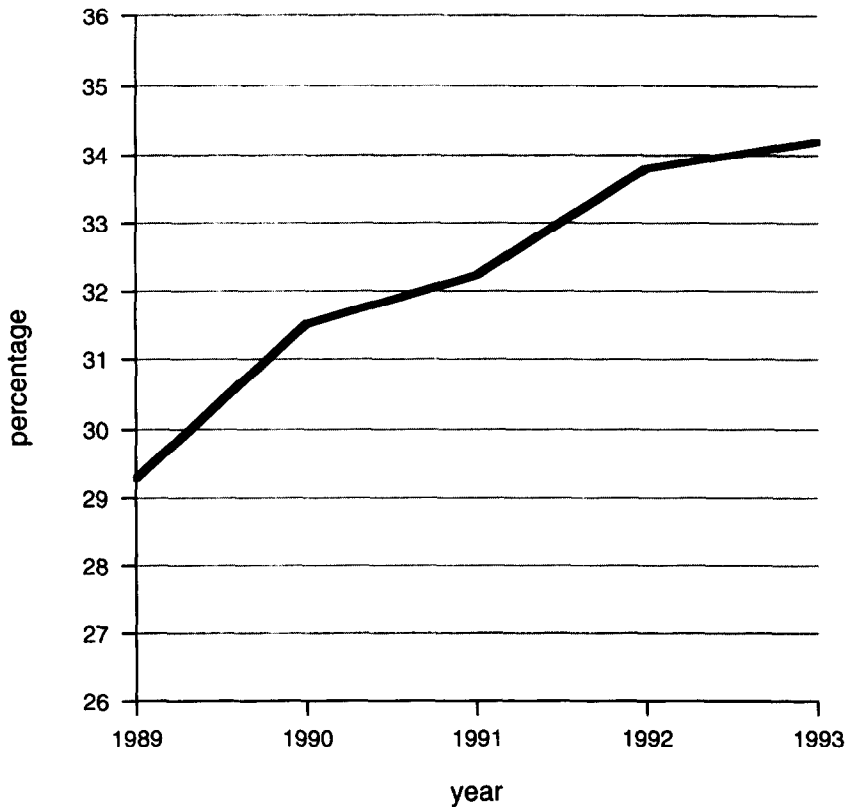

Fig. 2. Papers mentioning semiempirical programs as a percentage of all papers mentioning quantum chemistry software (Fig. 1) in the CJACS database. For purposes of classification, MOPAC, AMPAC, HYPERCHEM, ZINDO, CACHE, and AMSOL were taken to contribute to the semiempirical total, whereas GAUSSIAN, GAMESS, HONDO, CADPAC, DMOL, and SPARTAN were taken to be part of the quantum chemistry total. The search term QCPE was included in the semiempirical total because MOPAC is by far the program most distributed by QCPE. Even though SPARTAN can be used for semiempirical and ab initio calculations, our arbitrary classification does not distort the results significantly because, through 1993, SPARTAN was not highly cited in the literature in this period despite its popularity.

work is edging upward. We can put forward many reasons for the latter trend.

First is the spread of the use of computational chemistry software to other branches of chemistry. Many bench chemists have learned that these methods are practical research tools that can aid in understanding and predicting chemical phenomena. A proven track record has made nonexperts more likely to appreciate the potential value of output from the programs.

Also, nonexperts are less timid about using the computational chemistry tools because easy-to-use programs have become widely available. In the past, all the quantum chemistry codes were user unfriendly, but now it is possible, with a little training, for any chemist to set up and execute calculations and then analyze output all through convenient graphical user interfaces. The cost of high performance workstations has become within the reach of more chemistry departments.

Another reason for the increased use of semiempirical methods is improvements in the parameterization and further extension of the parameterization to more elements of the periodic table. Also, the wall between quantum mechanical and force field methods is gradually being removed as new hybrid methods are developed [10]. These important impetuses to the spread of semiempirical calculations were well covered in the other papers of this symposium.

In the following paragraphs, we give additional reasons for the increased use of semiempirical MO methods. These reasons all rest on the fact that these methods are significantly faster than either density functional theory or ab initio methods [11,12].

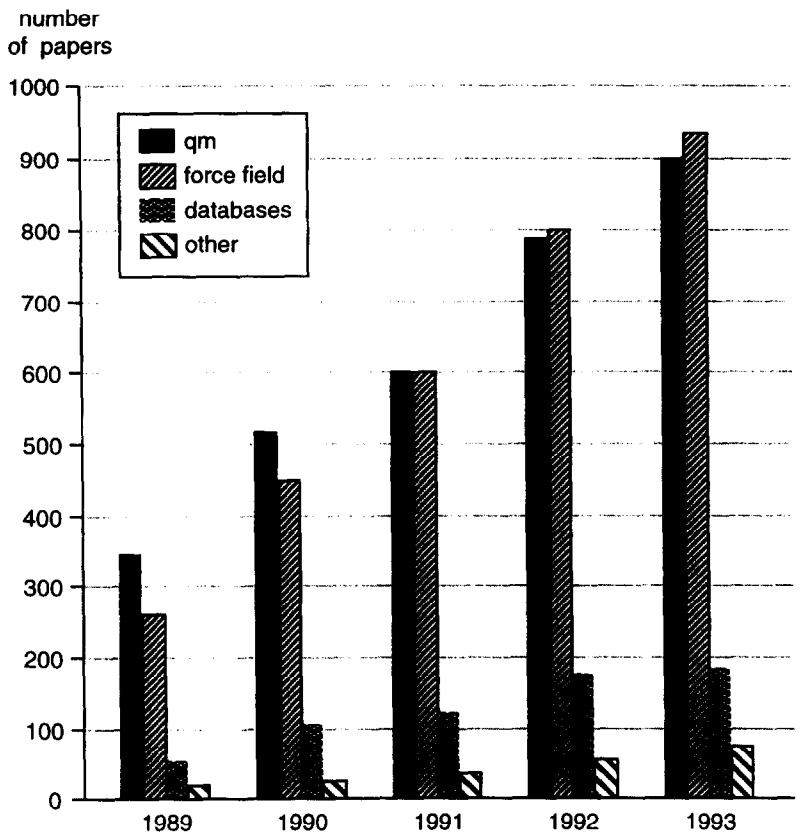

Fig. 3. Growing use of computational chemistry software in four categories: primarily quantum mechanically based (qm; i.e. those in Fig. 1), force field based (such as molecular mechanics and molecular dynamics), molecular databases (containing two- and/or three-dimensional structures), and other (molecular graphics, chemometrics, etc.).

Although the latter two theoretical approaches generally give a better level of agreement with experimental data, semiempirical methods will continue to be a tool chemists turn to for at least initial exploration. Compared with molecular mechanics (force field) methods, semiempirical calculations enjoy the advantage of better versatility because they need fewer new parameters for difficult new molecules.

Users today want and expect interactivity with the computer. Despite the tremendous strides in speeding up ab initio calculations, semiempirical methods still hold the speed advantage. Many users are willing to sacrifice a little bit in accuracy perhaps, if results can be obtained while sitting at the workstation rather than having to come back days or weeks later to check on the progress of a job. Hence users will gravitate toward trying semiempirical methods.

Chemists are tackling tougher problems as time goes on. Users want to be able to model challenging problems such as a reaction mechanism taking place in an enzyme, the interaction of small molecules with metallic surfaces, the solution phase conformational energy of biomolecules, the electronic spectra of complex heterocycles, and so on. As researchers turn to new, more difficult problems, quick methods, if sufficiently reliable, will have a role. Many researchers are excited by exploring a topic that has never been investigated before. In contrast, some people can derive satisfaction being, say, the fifteenth person to publish an ab initio calculation on some small molecule, using 290 basis functions instead of 260 basis functions or using 12 000 configurations instead of 10 000 configurations, which a previous author had employed. Such incremental research is not without value but lacks the excitement of studying a new problem. Thus, the desire "to go where no one has gone

before'' will be an incentive for continued use of semiempirical methods.

Still another reason for the growth patterns in Figs. 1–3 is that the whole tempo of research in organizations with a mission to design commercially useful compounds has become even more intense. This is particularly true in the pharmaceutical industry. In traditional bench chemistry, medicinal chemists synthesized one new compound every 1–4 weeks on average [13,14]. Now in this era of combinatorial chemistry [14,15], when 2000 compounds can be synthesized each week, the computational chemist must use the fastest methods available to provide timely input in the design process.

A further reason for the growth patterns in Figs. 1–3 is that the users of computational chemistry software are gaining an appreciation of the importance of running some approximate model calculations before plunging into time-consuming higher level work. This indicates that educators [16] are doing a good job in teaching the value of scaling up a computational project. It makes sense to gain some familiarity with a research problem using quick, inexpensive methods before burning up a lot of computer time on higher level calculations.

Finally, users are also becoming accustomed to the fact that not all theoretical models give identical results. Different semiempirical calculations, just like different basis sets in ab initio work, can give different answers. In fact, even as one moves up the ladder from simple methods to the most sophisticated and costly methods [11], the computed results may not steadily converge toward a common answer. Depending on the size of the problem, application of the most sophisticated methods may be impractical. In such cases, the user may has no alternative other than to try several semiempirical methods in hopes that some consensus in the results is obtained. In any case, the following rule of thumb is appropriate for all individuals undertaking a computational chemistry project:[1]

Always
Model
Prior to
Accurate
Calculations

[1] Any resemblance of this acronym to a program name is purely coincidental.

In summary, we have presented bibliometric data showing the prevalence of the use of semiempirical molecular orbital software. The number of papers in the chemical literature mentioning these programs is increasing and constitutes roughly one-third of all papers mentioning software for quantum chemistry. Until ab initio, density functional, or some new quantum mechanical method can compete in terms of speed, semiempirical calculations will have a future.

Acknowledgements

The author thanks Professor Andrew J. Holder, a gentlemen with a fine sense of humor, for his invitation to participate in the symposium, ''Semiempirical Methods: Is There a Future?'', at the 211th American Chemical Society National Meeting, New Orleans, Louisiana, March 24–28, 1996.

References

[1] D.B. Boyd, in K.B. Lipkowitz, D.B. Boyd (Eds.), Reviews in Computational Chemistry, vol. 6, VCH, New York, 1995, pp. 317–354. Information about this and other chapters of Reviews in Computational Chemistry can be found on the World Wide Web at http://chem.iupui.edu/.

[2] K.B. Lipkowitz, D.B. Boyd (Eds.), Reviews in Computational Chemistry, vol. 8, VCH, New York, 1996, pp. v–ix.

[3] J.J.P. Stewart, in K.B. Lipkowitz, D.B. Boyd (Eds.), Reviews in Computational Chemistry, vol. 1, VCH, New York, 1990, pp. 45–81; J.J.P. Stewart, QCPE Program No. 455.

[4] C.J. Cramer, D.G. Truhlar, in K.B. Lipkowitz, D.B. Boyd (Eds.), Reviews in Computational Chemistry, vol. 6, VCH, New York, 1995, pp. 1–72; C.J. Cramer, D.G. Truhlar, QCPE Program No. 606.

[5] QCPE (Quantum Chemistry Program Exchange), Indiana University, Bloomington, IN; e-mail qcpe@ucs.indiana.edu.

[6] For a compendium of chemistry software and its sources, see D.B. Boyd, in K.B. Lipkowitz, D.B. Boyd (Eds.), Reviews in Computational Chemistry, vol. 7, VCH, New York, 1995, pp. 303–380.

[7] M.W. Schmidt, K.K. Baldridge, J.A. Boatz, S.T. Elbert, M.S. Gordon, J.H. Jensen, S. Koseki, N. Matsunaga, K.A. Nguyen, S. Sun, T.L. Windus, M. Dupuis, J.A. Montgomery, Jr., J. Comput. Chem. 14 (1993) 1347.

[8] W.J. Hehre, L. Radom, P.v.R. Schleyer, J.A. Pople, Ab Initio Molecular Orbital Theory, Wiley, New York, 1986.

[9] M.J. Frisch, G.W. Trucks, M. Head-Gordon, P.M.W. Gill, M.W. Wong, J.B. Foresman, B.G. Johnson, H.B. Schlegel, M.A. Robb, E.S. Replogle, R. Gomperts, J.L. Andres, K. Raghavachari, J.S. Binkley, C. Gonzalez, R.L. Martin,

D.J. Fox, D.J. DeFrees, J. Baker, J.J.P. Stewart, J.A. Pople, GAUSSIAN 92, Gaussian, Pittsburgh, PA, 1992. Earlier versions, which were also included in the survery, generally came out in even numbered years.

[10] J. Gao, in K.B. Lipkowitz, D.B. Boyd (Eds.), Reviews in Computational Chemistry, vol. 7, VCH, New York, 1995, pp. 119–185.

[11] D. B. Boyd, in K.B. Lipkowitz, D.B. Boyd (Eds.), Reviews in Computational Chemistry, vol. 1, VCH, New York, 1990, pp. 321–354.

[12] L.J. Bartolotti, J. Flurchick, in K.B. Lipkowitz, D.B. Boyd (Eds.), Reviews in Computational Chemistry, vol. 7, VCH, New York, 1996, pp. 187–216.

[13] L.M. Balbes, S.W. Mascarella, D.B. Boyd, in K.B. Lipkowitz, D.B. Boyd (Eds.), Reviews in Computational Chemistry, vol. 5, VCH, New York, 1994, pp. 337–379.

[14] D.B. Boyd, in A. Kent, J.G. Williams (Eds.), Encyclopedia of Computer Science and Technology, vol. 33 (Suppl. 18), Marcel Dekker, New York, 1995, pp. 41–71.

[15] E.J. Martin, D.C. Spellmeyer, R.E. Critchlow Jr., J.M.Blaney, in K.B. Lipkowitz, D.B. Boyd (Eds.), Reviews in Computational Chemistry, vol. 10, VCH, New York, 1997, pp. 75–100.

[16] R.L. DeKock, J.D. Madura, F. Rioux, J. Casanova, in K.B. Lipkowitz, D.B. Boyd (Eds.), Reviews in Computational Chemistry, vol. 4, VCH, New York, 1993, pp. 149–228.