Journal of Computer-Aided Molecular Design, 2 (1988) 179–189
ESCOM

J-CAMD 040

Computer-aided molecular modelling: Research study or research tool?

A. Dearing*

Koninklijke Shell Exploratie en Produktie Laboratorium, 2280 AB Rijswijk, The Netherlands

Key words: Modelling techniques; Computational chemistry; Scientific software engineering; Molecular graphics

SUMMARY

Developments in computational methods and equipment have produced a new type of research chemist, who prefers to calculate properties as well as measure them, either to gain a better understanding of microscopic molecular behaviour per se, or to guide a broader scientific study using a so-called 'rational' approach. While there is good reason to believe that significant results can be obtained this way, it is clear that only some of the 'tools of the trade' are sufficiently robust to present to those who are not experts in the field.

This paper discusses the underlying basis for molecular modelling techniques, describing their history, state of maturity and prospects for future development. The intention is to outline the scope that these play in an industrial research environment, and to examine how they can successfully be incorporated as routine research tools.

Computer-based molecular modelling is not a unique discipline. Some comparisons are drawn between it and other computer-based techniques that have reached a greater degree of maturity, in order to highlight the points made.

INTRODUCTION

The term computer-aided molecular modelling (abbreviated to CAMM from now on) covers a range of techniques that can be used by the chemist, biochemist or physicist to simulate and then predict molecular properties. Each of these separate techniques became feasible and received attention as a research topic in its own right at different times from the 1960s onwards. There now seems to be a general acceptance that the discipline (whatever it may be) provides a viable approach, usable by those who are not necessarily experts, for studying certain problems that arise in pharmaceutical and agrochemical research. Interest in applying modelling techniques to other areas of chemistry, such as polymer or catalyst science, has also developed, and it is possible that these will become at least as important as the biochemical applications.

This paper examines how valid is the claim that CAMM is a mature methodology, suitable for widespread use in industrial research. The first part provides a basic overview of the subject, dis-

*Present address: Shell Research Limited, Sittingbourne Research Centre, Sittingbourne, Kent ME9 8AG, U.K.

cussing some of the tools available to a 'molecular modeller'. We consider what these tools have been used for, and how the applications have changed as scientific understanding and apparatus have improved.

In many respects, the use of computational chemistry has been only loosely integrated into an overall industrial R&D process for generating new products. It (along with other 'rational techniques') has been seen as a specialised tool useful for answering specific questions encountered during a research programme. This is partly because the expertise required to use the tools is not yet an integral part of mainstream scientific training, and partly because the tools of the trade are only emerging from the era in which individual groups hand-crafted separate techniques to their own requirements. The second part of the paper examines how the discipline must develop in order to provide integrated, reliable skills and apparatus that allow the non-expert to work productively.

THE TOOLS OF THE TRADE

CAMM became a subject worthy of discussion in the media around 1981, when computer graphics hardware became capable of displaying in colour and manipulating interactively (in 'real-time') pictures of large molecules such as proteins. This transformed people's attitudes to the use of the more theoretical aspects of chemistry and biochemistry. However, computer graphics only provides a window through which things are seen. Underlying the use of interactive graphics are a wide range of computational, information handling or conventional experimental techniques that provide the data being displayed, and it is important that the quality and reliability of these data are sound and well-understood.

The term 'modelling' conveys too narrow an impression. The following are generally treated as related techniques, although we could add any computer-based technique used to process, investigate or predict molecular properties:

- Presenting structural information and other molecular properties using computer graphics equipment, to manipulate and understand information available from different sources.
- Determining properties of small- or medium-sized molecules related to interatomic interactions, and modelling inter- and intra-molecular interactions, by use of empirical force-field calculations.
- Following molecular properties through time using molecular dynamics calculations, to gain information on molecular motions, free energies and other ensemble properties.
- Determining molecular properties related to electronic structure, using empirical or non-empirical quantum mechanical methods.
- Predicting structural properties of larger molecular systems, including the use of 'rule-based' tools, e.g. for predicting protein folding.
- Relating molecular properties to one another using statistical techniques, especially to obtain parametric formulae that express an important property in terms of more easily measurable (or calculable) quantities, or classify structures into distinct groups.
- Using computer-based databases to identify systems with common features or to predict properties.

Several symposia have covered the use of these techniques in the biochemical and biomedical fields, and the published proceedings [1–3] from these give a good introduction to the scope of CAMM.

Use of computational chemical techniques by companies with interests in chemical or biological fields is not new, and some of the above methods were used well before CAMM was recognised as a specific technology. For example, the literature contains a substantial number of relevant publications from staff working for the same company as the author during the period 1965–1975. However, a great expansion of interest took place around 1980–1982, and many companies acquired specialist hardware intended for use by a dedicated group responsible for CAMM.

In general these groups have survived and thrived, supported by societies such as the Molecular Graphics Society, close contacts with academic research groups, the availability of academics' software via groups such as QCPE [4], and a growing acceptance by company management that the service being provided is of value.

Interactive computer graphics

A broad discussion of the development of interactive computer graphics for molecular studies has been presented elsewhere [5–11]. Much of the process has been associated with the names of Robert Langridge and Richard Feldmann, the first pursuing interactive vector drawing techniques and the second raster-based space-filling techniques. These two approaches, necessarily distinct in the early 1980s, are now converging. Developments in raster techniques have made vector-based hardware more or less obsolete, and one type of display terminal can now perform both roles.

Costs of hardware are falling, although not yet to a level where the hardware can be placed freely on everyone's desk. Reasonable quality software is available at an acceptable price, and in the near future, it seems unlikely that cost and capability of graphics workstations will restrict the availability of adequate interactive molecular graphics equipment, which, coupled with appropriate three-dimensional structural databases, will provide useful and uncontentious tools for both specialists and others.

However, there are still problems with the use of computer graphics: for example, it has been difficult to develop adequate techniques to handle more complex data structure manipulations, especially to investigate interactions with flexible or irregularly shaped cavities or to determine possibilities for conformational mobility at the nanosecond timescale or longer. Effective direct interaction between user and display becomes much more difficult when the necessary manipulations go beyond bulk movements about a few torsional angles.

In order to handle this type of problem, a very close tie-up is needed between graphics display software and powerful underlying computational facilities that can take fuller account of molecular properties. Alternatively, an extensive rule-base must be available based on detailed practical study. Development of both approaches is taking place at present, and it is likely that, during the next few years, these techniques will be closely integrated into the presentational tools.

Empirical energy calculations: Molecular mechanics and dynamics

Empirical expressions for representing non-covalent interactions have existed for a long time [12]. They form the basis for molecular mechanics, which combines numerical refinement algorithms with a parametric forcefield to calculate interaction energies and refine structures to some minimum energy conformation, and molecular dynamics, which allows molecular motions to be followed through time. Allinger [13] and others developed molecular mechanics into a precise tool for handling a restricted class of small hydrophobic structures, and Levitt, Hagler and others [14–

20] probed the application of forcefield calculations (both molecular mechanics and dynamics) to larger molecular systems and to a wider variety of atom types.

The required parameters are intended to satisfy two conditions simultaneously: to treat molecular forces in terms of conceptually-straightforward quantities such as bond-stretching, angle-bending, etc., and to model the molecular potential energy surface to the required accuracy with the minimum number of terms.

By the early 1980s, adequate forcefields were available for specific systems with restricted chemical types. For example, Kollman et al. [21–23] were modelling peptide and nucleotide systems of 200–500 atoms, and were able to reproduce experimental binding affinities for related substrates to these systems by making broad assumptions about energetic terms such as solvation.

Since then, there has been less progress in forcefield development, but much more widespread access to, and use of, programs implementing molecular mechanics, especially alongside an interactive graphics facility. The scope of molecular mechanics calculations has not changed greatly, although it is significantly easier and quicker to carry out calculations on a few hundred atoms, provided an adequate model structure can be developed. In addition, refinement algorithms have not developed sufficiently to optimise complex structures except to nearby local energy minima, and as a result the technique has not proved useful for handling extensive refinement of large structures, e.g. for studying protein folding.

On smaller molecular systems, programs such as MM2 [13] provide high quality results on defined structural types, and can perform conformational energy scans semi-automatically, although the time taken to do these becomes excessive if the molecule is too flexible. Commercially-available packages [24, 25] are able to display results so as to highlight similarities in the conformational properties of analogues. Work has been done to extend the types of system for which parameters are available, for example to inorganic compounds.

The rapid growth in use of molecular mechanics has led to a corresponding growth in the range of chemical types of interest. Virtually everyone who has used a molecular mechanics program finds that some required parameters are not available, and probably more unguided effort is spent resolving this than on any other problem. Although this shortcoming may not matter in some cases, it is important to remember that the value of a molecular mechanics or dynamics calculation rests on the quality of the model parameters used.

The parameters are an attempt to express real behaviour in terms of a small number of determining factors, and their meaning must not be over-interpreted. A parameter set that produces good results on the set of compounds used for calibration may not be reliable for other compounds, and it is not sensible to combine parameters determined from separate 'fitting' exercises and expect these to give reliable results. In order to ensure that parameter sets are stable with respect to a wide range of problems, or to know the extent to which the functional equations need to be extended to determine specific properties (e.g. to produce good vibrational frequencies) requires skill and experience, plus substantial detailed study.

The methodology underlying molecular mechanics has probably developed close to its limits. Advances that should be expected include the provision of good and reliable force-field parameters for a much wider range of systems than are available at present, defined measures of quality and reliability, and more standardised techniques for defining and developing new force-field parameters so that these can be incorporated alongside existing values. Activities such as the Biosym consortium (whereby a number of interested companies jointly finance this research) are notable,

and it will be interesting to see whether this approach produces adequate acceptance and standardisation of method among both commercially and academically-oriented groups.

Electronic structure calculations

Quantum mechanical programs were among the first tools to be used which are now considered part of the CAMM toolkit. There is a sizable literature [26–29] exploring the use both of semi-empirical and ab initio methods, including many papers from within industry.

Ab initio calculations provide very powerful facilities for studying the electronic distribution around a molecule and its properties in the presence of disturbing perturbations. However, they are not suitable for routine application by novices; the computing resources required even for simple computations remain very high, and considerable experience is needed to know what type of calculation will give meaningful results.

Much of the development effort applied in the 1980s on standard ab initio programs has been to make post Hartree–Fock techniques more available. The applicability of these techniques to problems of relevance to industry at present is limited.

By contrast, usable semi-empirical methods have developed very significantly in recent years, with Dewar's group [30, 31] providing considerable progress with the MINDO, MNDO and AMO sequence of programs. Similar remarks apply to a semi-empirical molecular orbital calculation as to molecular mechanics; one is using a parameterised technique that can give very reliable results within some boundary of problems but may be quite unreliable outside that region.

Problems arise in the use of quantum mechanical calculations because of misunderstanding of the meaning of results, or of the extent of calculation needed to determine quantities interest. For example, it is popular to use properties such as a Mulliken population analysis [32] or orbital coefficients to describe the electron distribution without questioning whether differences in values mean anything. On the other hand, sometimes a properly used extended Huckel or Frontier Orbital [33] calculation gives quick insight into a property that would be obscured by a MNDO or ab initio calculation. Problems disappear if the user has adequate insight and knowledge of the relevant literature, but it is not clear how to impart these to the non-specialist who is making casual use of a computer program.

An interesting and important technique is to calculate the electrostatic isopotential field that the molecule would exhibit in the presence of another charged entity. Displayed graphically, the calculated interaction energy between a proton and the molecule throughout the space surrounding the molecule illustrates powerfully those regions to which charged and polarised species can or cannot easily approach.

Use of statistical methods

Traditional QSAR studies were based on a limited set of parameters, representing partition, electronic and steric properties expressed in an empirical manner. The major goal was generally to express a target property in terms of these basic parameters by the use of a regression equation. Hansch analysis is the best known example of QSAR, and this and related techniques have proved popular for highlighting the properties (usually partition coefficient) apparently responsible for activity, and for suggesting optimal values for these properties.

QSAR has frequently been criticised as a wholly interpolative, non-explanatory technique, which is only useful when sufficient data has been collected to make its use unnecessary. However,

the methodology used for QSAR has developed and combined with other techniques used in statistics and computational chemistry to provide a more generally-useful set of tools. It is becoming common to use parameters derived from the more formal CAMM techniques (especially simple parameters from molecular orbital calculations), and to use, e.g., factor analysis rather than multiple regression to express results [34].

Especially when the underlying mechanisms governing a particular activity are poorly understood, statistical techniques are essential to highlight significant structural properties or to determine the extent to which a series of experimental tests (e.g. compounds produced in a synthesis series) do span an acceptable spread of key parameters. It is also appropriate to use statistical methods to characterise the large quantities of data available from databases of molecular data. As usual, more problems are caused by improper expectations than by fundamental inadequacies in the technique.

Data retrieval from on-line databases

Structural information on several classes of molecule have been provided in machine-readable form for some time. The Cambridge Crystallographic Database [35] provides the most complete example; virtually all small organic and organometallic crystal structures published in the open literature have been incorporated into this database, and the information is openly available at a very reasonable price. Other structural databases provide information on proteins (the Brookhaven database [36], carbohydrates and inorganic structures.

One of the most significant advances made possible by the availability of adequate on-line structural information is to identify the variability in particular structural elements among a large class of representatives [37, 38]. Searching a large database provides conformational information similar to that available from a calculation, with increased confidence in the validity of the information. Analogies can be made here to the highly developed computer-aided chemical synthesis programs [39–41] which rely on extensive internal databases of reactions.

Large on-line databases are also available of spectroscopic information, and companies are developing increasingly large internal databases containing measured properties (e.g. field results from pesticide trials). Less integrated use is being made of such databases within CAMM. A major difficulty lies in the provision of standard access techniques to the full range of information.

Summary

In some respects, the methodology underlying CAMM has reached a plateau, with reasonable tools available to handle a wide range of studies in different fields. In other respects, the situation is far from adequate. Although the individual tools can give reliable insight, the boundary conditions that limit the applicability of each tool can be obscure by comparison with the boundaries governing experimental methods. The more sophisticated and highly-developed CAMM methods (e.g. dynamics) are most applicable to events that occur at timescales shorter than 10^{-12} second. Whilst we have tools that probe different timescales, we cannot yet integrate these together into an all-purpose molecular modelling workstation that can be placed on the chemist's desk to provide reliable and relevant information.

CAMM is at its best at the two extremes of expertise. The specialist can obtain highly pertinent information to rationalise existing experimental work, whilst the novice is given insight by the combined use of simple display techniques, data retrieval and elementary statistical classification

methods. At both ends of the scale, the tool both encourages and requires thought about the type of experimental data needed to resolve hypotheses.

However, virtually all the available calculational tools rely explicitly or otherwise on a well-developed set of parameters. There is a great danger that enthusiasm for using the tool exceeds enthusiasm for providing a reliable data set. A common question asked by concerned molecular modellers is 'Who is providing the underlying science?'

Moderate progress is likely in developing the more traditional modelling tools such as QSAR, molecular mechanics and electronic structure calculations. More powerful developments can be expected in the application of database technology to provide on-line access to growing archives of general or proprietary data. Experience with using the Cambridge crystallographic database and the various statistical methods available from QSAR convinces the author that tools that allow access to, and analysis of, large numbers of detailed molecular 'fingerprints' (combining experimental and theoretically determined properties) will significantly improve our ability to extrapolate and predict molecular behaviour.

Traditionally, the same people have developed modelling techniques for use within their own research and supported the resulting tools for more general application by others. As use grows, this becomes less viable; it is inefficient and limits both the fundamental work of the specialist and the more widespread use by the general user. A different infrastructure is needed to support more extensive use of CAMM. This is the subject of the next part of the paper.

THE INFRASTRUCTURE

Most tools evolve through a development period, during which specialists handcraft some or all of the fabric themselves, to a mature phase in which better and cheaper tools are produced commercially, and users can be reasonably sure of the quality of what they buy. This mature phase produces much more widespread use of the tool, and this in turn allows users to place greater reliance on their suppliers for support.

There is an intermediate phase, during which individual users possess highly crafted tools that they trust and understand, but the commercial offerings (whilst having considerable gloss applied to attract customers) vary in quality and reliability, and may be badly constructed or supported. CAMM is currently at this phase.

This does not mean that our suppliers are a set of rogues! The process of creating adequate products, getting sufficient competition among suppliers and understanding among consumers of what is available, weeding out poorer quality material, and restoring an effective balance between use and support takes time.

CAMM is a discipline that touches the state of the art both in its scientific content and in the computational support it requires. It should be clear from the previous section that we expect continued progress in both aspects of the discipline, and need significant improvement in the quality and reliability of existing methodologies. Furthermore, CAMM is essentially a numerate rather than a qualitative discipline. It needs staff who understand the significance and nuances of computational techniques, but the results it produces must be integrated properly alongside more traditional techniques, and these latter may themselves need to develop as a result.

Although the changes needed to bring about greater maturity may be fairly obvious, it may be useful to examine another scientific discipline that makes extensive use of computational techniques to remind ourselves what these changes are.

A digression: Seismic data processing

The example chosen concerns seismic data processing, in which information on the earth's subsurface obtained by recording the passage of acoustic waves is enhanced to provide a detailed picture of the subsurface, by combining signal processing and modelling techniques.

The oil exploration industry makes substantial use of the seismic data technique as an essential preliminary in determining where to drill. Seismic data is recorded over the surface of the area of interest, and a succession of complex data enhancement calculations refines these data to create a reasonably detailed picture of the various rock strata and faults. In ideal cases, it is possible to obtain information correct to a few metres at depths of several kilometres and to obtain accurate and realistic representations of the earth's surface millions of years ago.

Routine use of the seismic technique on a large scale dates back over twenty years. Some of its characteristics that may eventually apply to CAMM too include:

– *Manpower.* The necessary staff and expertise required for seismic data processing have generally been obtained by recruitment from the more numerical sciences such as physics, with specific training given by their employer, rather than by retraining geologists (the experimentally-inclined equivalent of chemists), who continue to have a separate and vital role that both precedes and follows the numerical work carried out by the geophysicists.

– *Organisation.* Most computational work is now carried out by contractor companies on behalf of a client, often within the client's own premises. Sophisticated software packages are also available under license for more general use. Sufficient competition exists between contractors who are all able to perform the same work using different packages to provide an incentive for adequate quality control and to achieve low costs.

Developments in the methods available to disciplines related to seismic data processing should make it possible to break down traditional boundaries between these disciplines. However, the use of different and incompatible computational tools causes problems that are difficult to overcome.

– *Development of the scientific methodology.* Commercial use of and support for the seismic technique greatly exceeds current academic involvement, and this imbalance often makes it hard to incorporate contributions from academic groups into industrial research programmes and into operational use. Incompatibilities between software packages exacerbate this problem.

Oil majors carry out substantial research of their own to provide additional techniques that they believe give a 'leading edge' in quality.

– *Technical.* Software packages are large, approaching 10^6 lines of FORTRAN code (approximately the same size as a fully-integrated molecular modelling package), sophisticated, but generally imcompatible with one another at the functional level, even though they provide similar facilities. Considerable use is made of specialist hardware such as array processors to improve performance at the cost of an increased support effort and lack of portability.

The cost of software development and support is high. A package intended to satisfy the (individual but not insignificant) needs of one oil major may require as much effort to develop and support as the effort actually expended in using it. There are few effective standards allowing simple transfer of software between packages [42]. However, the need for transfer of data between different groups has encouraged the development of reasonable standards for data transfer [43].

The cost of software development and support conflicts with a need for high computational performance. Maintenance costs associated with specialist hardware usually exceed the benefits

gained from using that hardware well within the life-time of the software. Unfortunately, use of the specialist hardware is often too intricately embedded into the software to remove.

A lot of similarities can be seen between the points outlined above and the development of CAMM, and problems that appear in an embryonic manner in the application of CAMM techniques have grown to serious proportions within seismic data processing.

Manpower

While it is reasonable to expect that new staff will have much greater awareness of the possibilities of using computer-based modelling techniques, it is not clear how to obtain people with high levels of skill in both modelling and more traditional methods. We may avoid the need to develop a new breed of industrial scientist to handle the computer-based tool, but we must obviously take care to balance a more widespread use among non-experts with adequate in-house application at a more advanced level.

Organisational and scientific

We currently encourage close ties between industrial and academic research groups, to ensure effective transfer of know-how and to direct experimental work such as macromolecular crystallography towards targets of importance to the industrial groups. In addition, much of the required software originates from academic sources. Liaison is only possible when the involvement and expectations of the two sides are well-balanced. The nature of this contact may change considerably if industrial involvement in CAMM increases substantially.

Provided industrial use of CAMM continues to grow, we can expect commercial support to take over in some of these areas. This is already apparent. Given that the market is sufficient to support adequate competition, we may be able to overcome problems of quality, be able to rely on the scientific content of the products, and expect that the scope of individual methodologies is properly defined and demonstrated. However, the situation seems rather precarious at present, and concerns about the underlying science will remain if there is insufficient market growth to make it worth someone's while to resolve this.

Computational and technical

It is popular to suppose that the rate of advance of computer technology is adequate to resolve the problems that afflict computational science. It is true that computing power and packaging have improved at a rate that almost defies imagination, but these are still inadequate for many common computational chemical problems. More problematically, our productivity at software development and support has not improved in a corresponding manner, even with conventional computer hardware. New types of hardware (such as the transputer) offer beguiling solutions to complex computational needs, but considerable fundamental research will be required before these can or should be used too widely.

The scale of software effort required in CAMM is now too great for individual 'whizz-kids' to provide rapidly-developed tools for general use. We have moved into the (understandably unhappy) situation in which users wonder why three support staff are unable to keep up with error corrections and enhancements to a program that one person wrote in his spare time, or worse, in which new developments fail to appear within any realistic timescale. It is not uncommon to find

that the cost of correcting even a minor error in a large integrated package is of the same order of magnitude as the price a customer pays to use the entire package.

It is very important to understand, before making a large-scale commitment to develop computer-based science, the extent of the software support and development task that will result. There is a crucial need to use robust techniques to define in advance what software should do, instead of indefinitely adapting prototypes to satisfy supposed needs. Emphasis must be placed on the data handling framework that integrates software as well as on the computational facilities that software provides. These problems must be taken as seriously as the problem of providing good science, but experiences elsewhere show that this often does not happen. The consequences become apparent only after the new development has turned into an essential part of the infrastructure.

It may be appropriate for individual companies with large scope for applying the technology to develop and maintain their own CAMM toolkit. However, this is almost certainly not an approach that is generally desirable. Even for the situations in which it can be justified, much greater value will result if issues such as standard data exchange formats and information handling systems, and the availability of portable 'plug-compatible' software are pursued first.

CONCLUSIONS

We have tried in the preceding pages to introduce some of the possibilities available from using CAMM techniques, and also to show where significant difficulties will arise in turning a semi-mature discipline into a mature, routine methodology suitable for routine use in industrial research. We have not tried to cover all applications, and have probably erred on the side of caution rather than optimism.

This is not because we doubt the value of CAMM. The problems can be overcome and there is very considerable scope for scientific progress in many areas from the proper use of the various techniques that become available. The intention has been to raise questions about the best way in which this can be achieved, in order that appropriate answers may develop in your mind.

ACKNOWLEDGEMENTS

Several people have made useful suggestions that have improved the content of this paper. Thanks are due to them all, but especially to W.B.M. Roosdorp and J.G. Vinter.

REFERENCES

1 Weinstein, H. and Green, J.P. (Eds.) Quantum Chemistry in Biomedical Sciences, Ann. N.Y. Acad . Sci., Vol. 367, 1981.

2 Olsen, E.C. and Christofferson, R.E. (Eds.) Computer-Assisted Drug Design, ACS Symposium 112, American Chemical Society, Washington, D.C., 1979.

3 Thompson, J.S. and Robson, B. (Eds.) Computer Prediction of Molecular Conformation, 600th Meeting of the Biochemical Society, Oxford, Biochem. Soc.Trans., 10 (5) (1982).

4 Quantum Chemistry Program Exchange, Chemistry Department, Indiana University, Bloomington, IN 47405, U.S.A.

5 Langridge, R. and MacEwen, A.W., In IBM Scientific Computing Symposium on Computer-aided Experimentation, IBM, Yorktown Heights, NY, 1965, p. 305.

6 Levinthal, C., Sci.Am., 214 (1966) 42–52.

7 Barry, C.D., Ellis, R.A., Graesser, S.M. and Marshall, G.R., In Faiman, M. and M. Nievergelt, J. (Eds.) Pertinent Concepts in Computer Graphics, University of Illinois Press, Champaign, IL, 1969, p.104.
8 North, A.C.T. and Barry, C.D., Biochem. J., 121 (1971) 121.
9 Diamond R., In Sayre, D. (Ed.) Computational Crystallography, Oxford University Press, Oxford, 1982, pp. 318–325.
10 Jones, T.A., In Sayre, D. (Ed.) Computational Crystallography, Oxford University Press, Oxford, 1982, pp. 303–317.
11 Pflugraph, J.W., Saper, M.A. and Quiocho, F.A., J.Mol.Graphics, 1 (1983) 53.
12 Westheimer, F.H., In Newman, M.S. (Ed.) Steric Effects in Organic Chemistry, Wiley, New York, NY, 1956, pp. 523–555.
13 Allinger, N.L., MM2 Program, QCPE 395, see Ref. 4.
14 Gelin, B.R. and Karplus, M., Biochemistry, 18 (1979) 1256.
15 Weiner, P.K. and Kollman, P.A., J.Comput.Chem., 2 (1981) 287–303.
16 Lifson, S., Hagler, A.T. and Dauber, P., J.Am.Chem.Soc., 101 (1979) 5111–5141.
17 McCammon, J.A., Gelin, B.R. and Karplus, M., Nature, 267 (1977) 585–590.
18 Levitt, M., J.Mol.Biol., 168 (1983) 595–621.
19 Van Gunsteren, W.F., Berendsen, H.J.C., Hermans, J., Hol, W.G.J. and Postma, J.P.M., Proc.Natl.Acad.Sci.U.S.A., 80 (1983) 4315–4319.
20 Dauber, P., Osgathorpe, D.J. and Hagler, A., Biochem.Soc.Trans., 10 (1982) 312–318.
21 Blaney, J.M., Weiner P.K., Dearing, A., Kollman, P.A., Jorgensen, E.C., Oatley S.J., Burridge, J.M. and Blake, C.C.F., J.Am.Chem.Soc., 104 (1982) 6424–6434.
22 Kollman, P.A., Weiner, P.K. and Dearing, A., Biopolymers, 20 (1981) 2583–2621.
23 Dearing, A., Weiner, P.K. and Kollman, P.A., Nucl. Acids Res., 9 (1981) 1483–1497.
24 Davies, E.K. and Prout, C.K., CHEMGRAF User Manual, Chemical Crystallography Laboratories, Oxford University, U.K., 1982.
25 Marshall, G.R., Barry, C.D., Bosshard, H.E., Dammkoehler, R.A. and Dunn, D.A., In Olsen, E.C. and Christofferson, R.E. (Eds.) Computer-Assisted Drug Design, ACS Symposium 112, American Chemical Society, Washington, D.C., 1979, pp. 205–226.
(See also many other generally available molecular modelling programs).
26 Schaefer, H.F., Applications of Electronic Structure Theory, Plenum Press, New York, NY, 1977.
27 Pople, J.A. and Beveridge, D.L., Approximate Molecular Orbital Theory, McGraw-Hill New York, NY, 1970.
28 Szabo, A. and Ostlund, N.S., Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory, Macmillan, New York, NY, 1982.
29 Naray-Szabo, G., Surjan, P.R. and Angyan, J.G., Applied Quantum Chemistry, D.Reidel, Dordrecht, 1987.
30 Dewar, M.J.S., Zoebisch, E.G., Healy, E.S. and Stewart, J.J.P., J.Am.Chem.Soc., 107 (1985) 3902–3909.
31 QCPE Program 506, see Ref. 4.
32 Mulliken, R.S., J.Chem.Phys., 23 (1955) 1833–1840.
33 Flemming, I., Frontier Orbitals and Organic Chemistry Reactions, Wiley, London, 1976.
34 Glen, R.C. and Rose, V.S., J.Mol.Graph, 5 (1987) 79–87.
35 Allen, F.H., Bellard, S.H., Brice, M.D., Cartwright, B.A., Doubleway, A., Higgs, H., Hummerlink, T., Hummerlink-Peters, B.G., Kennard, O., Motherwell, W.D.S., Rodgers, J.A. and Watson, D.G., Acta Crystallogr., Sect.B, 35 (1979) 2331.
36 Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer Jr., E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M., J.Mol.Biol., 112 (1977) 535–542.
37 Anderson, S., J.Mol.Graph., 2 (1984) 83–90.
38 Wipke, W.T., Dill, J.D., Peacock, S. and Hounshell, D., Search and Retrieval using an Automatic Molecular Access System, 182nd National Meeting of the American Chemical Society, New York, NY, 1981, American Chemical Society, Washington, D.C.
39 Chodosh, D.F. and Mendelson, W.L., Pharm.Technol., 7 (1983) 90–92.
40 Corey, E.J. and Wipke, W.T., Science, 166 (1969) 178.
41 Corey, E.J., Wipke, W.T., Cramer, R.D. and Howe, W.J., J.Am.Chem.Soc., 94 (1972) 421.
42 Hatton, L., Wright, A., Smith, S., Parkes, G., Bennett, P. and Laws, R., Software Practice and Experience, 18 (1988) 301–329.
43 Barry, K.M., Cavers, D.A. and Kneale, C.W., Recommended standards for digital tape formats, Geophysics, 40 (1975) 344–352.