Journal of Computer-Aided Molecular Design, 2 (1988) 165–178
ESCOM

J-CAMD 039

Symposium Overview
The Shell Conference on Computer-Aided Molecular Modelling

G.R. Hays and D.P. de Bruijn

Koninklijke/Shell-Laboratorium Amsterdam (Shell Research B.V.), P.O. Box 3003, 1003AA Amsterdam, The Netherlands

The 'Shell Conference on ...' series began in 1985 and meetings are held approximately twice a year. The idea behind the conferences is to bring together invited scientists from both universities and industry, and representatives from different Shell Research laboratories, to create a forum to discuss the future directions of the chosen research area. These meetings have embraced a wide range of topics of interest to Shell Research as a whole.

This particular conference, organised by the Analytical Department of the Koninklijke/Shell-Laboratorium, Amsterdam (KSLA), was held on 4-6 October, 1987 at Hoenderloo in the Netherlands. The aim was to review the state-of-the-art and to discuss the future of molecular modelling and design. The programme itself consisted of a series of presentations on prescribed topics, panel discussions, and software and hardware demonstrations. Many of the presentations given consisted of overviews, experiences, advice and predictions for the future. The panel sessions, which involved the speakers within that session and a discussion leader who summarised some of the points made in an introduction, encouraged even further discussion and speculation. This overview attempts to catch the flavour of the meeting and convey some personal views that were expressed and conclusions drawn.

1. SETTING THE SCENE

The two opening presentations, by Dearing (Shell Research, Rijswijk) and Kollman (Univ. California, San Francisco, UCSF), laid the groundwork. The 'tools of the trade' were introduced (3-D graphics, empirical and ab initio calculations, databases, statistical analysis) and many recurring themes of the conference were touched upon. The two different perspectives of the academic and the industrial modeller were clearly presented. At times these views can be diametrically opposed: the academic who can choose the most interesting problem for fundamental research versus the industrial researcher who needs complete and reliable tools to tackle real-world problems. Although both speakers presented general overviews, they illustrated their points with biochemical applications of Computer-Aided Molecular Modelling (CAMM), the native ground and still the major area of application.

0920-654X/88/$3.50

Dearing, having withdrawn from CAMM research for the challenge of oil exploration and production a number of years ago, was able to review the history and status of CAMM from some distance [1]. He offered his views on the general stage of development by comparing CAMM with the more mature field of seismic computing and modelling. His conclusion is that CAMM is now semi-mature, which implies on the one hand that others in addition to specialists apply the tools, but on the other hand that these tools are not yet robust enough to allow this to be carried out on a wide scale. A new type of scientist – the modeller or computational chemist – has however already been born. A problem with which one is now faced is large scale software development and standardisation and the future will see further commercialisation of this effort. Dearing believes that a plateau has been reached in the development of empirical methodologies to calculate molecular properties (molecular mechanics, molecular dynamics, quantitative structure activity relationships), while the problems of developing optimal parameters for these techniques, and their application to areas other than biochemistry, remain.

In his presentation, Kollman stressed the enormous progress that has been made in recent years. Whilst in 1980 only small molecules in the gas phase could be treated completely, these molecules have considerably 'grown' in size and much larger systems can be tackled in solution, via a combination of distance geometry, molecular dynamics (MD) and experimental constraints. The strength of CAMM is, in Kollman's eyes, the powerful combination of tools [2]. The value of molecular graphics was demonstrated by a film on MD calculations on DNA. Besides being able to produce pretty pictures (an aspect of CAMM often heard during the conference), the UCSF MIDAS system was used to show qualitative features of steric and electronic effects [3]. For example, electrostatic potentials on Connolly [4] surfaces were visualised by colour coding and arrows showing the gradients.

As part of 'rational drug design', a newly developed method by Kuntz (UCSF) was presented to illustrate the approach to the design of inhibitor molecules for a known receptor site. The method is based on distance geometry and starts with filling the available volume with atoms, gradually refining this towards a candidate molecule.

Free Energy Perturbation as applied in MD was perceived to be the major theoretical tool for drug designers [5,6]. Since the free energy can be calculated directly, rather than being obtained as the difference of two large numbers, this method offers better accuracy to, for example, calculate differences in solvation energies or site-specific mutagenesis. MD on macromolecules remains difficult however as for example, one does not know the time span over which the calculation has to be prolonged. The role of theory, in prediction as well as in interpretation, depends on the problem and data available and at the present stage no general recipe can be prescribed.

2. THE FUTURE OF ORGANIC MODELLING

Despite the variety in the speakers' backgrounds, viewpoints and styles, a number of general conclusions can be drawn on which a wide consensus was apparent.

* CAMM can be defined, as was done explicitly by Vinter (Smith, Kline & French, Welwyn), as the summation of Computational Chemistry and Molecular Graphics using the fundamental measure of energy, however yielding more than the sum of the parts.

* The future of CAMM is bright despite some speakers' assertions that CAMM was already a mature and indispensible technique in structural organic chemical research. Problems remain to be solved and many tools still have to be improved or created; the development of forcefields is especially important.

* Valuable use of CAMM in organic chemical work can already be made for: the search for interesting analogues of active compounds; gaining insight into crystal packing, inter- and intra-molecular interactions (e.g., hydrophobic interactions, hydrogen bonding, strains etc.); exploring possible reaction pathways; searching for unique conformations of flexible systems; and, most importantly, aiding the creation of ideas.

* CAMM has to be regarded, at least at the present stage, as a mainly qualitative tool. As Goodford (Univ. Oxford) stressed, perhaps its most important function is to provide insight into the three-dimensional structure of molecules. Therefore interactive access to CAMM facilities is important for chemists as well as specialised modellers. The use of CAMM to inspire the practical chemists' thinking was emphasised by all the speakers in this session.

* Crystallographic data bases, providing the link between theory and experiment, are one of the essential ingredients of CAMM, especially in organic modelling, and have to be used effectively. Murray-Rust (Glaxo, Middlesex) showed that the value of the Cambridge Structural Data Base [7] to both academic research and industry is beyond doubt. With appropriate search, statistical and display software, coupled if necessary to user-supplied programs, it is possible to combine the use of experimental data with model building techniques.

Many of these points were beautifully illustrated by case studies during the different presentations. Müller (F.Hoffman-La Roche & Co., Basel) described the integrated modelling of the (8Z)-photoisomer of fusarin C, a toxic and mutagenic fungal metabolite. The Roche Interactive Molecular Graphics system, the Roche relational data bases and their own united atom forcefields with a very restricted set of parameters, were employed [8]. The basic structure was built up from templates taken from the template data base, the data base was used to find analogues for the ring structure, and the model was checked for torsional and conformational strain (colour coded for hot spots such that partial relaxation could be carried out). Hydrophobic contacts and hydrogen bonding were then built into the system. This gave a good representation of the X-ray diffraction data, and the structure was then further modelled via the introduction of spacer groups to yield potentially new molecules. The modelling of peptides was also illustrated along similar lines.

Other pharmaceutical companies have also invested in their own software development. Vinter described the in-house Smith, Kline & French molecular graphics and computing software systems COSMIC and ASTRAL, which have been built up over the last 12 years [9]. Examples presented were of a docking simulation (18-crown-6/formamide complex), the modelling of rotational barriers (vancomycine aglycone and vancosamine on vancomycine with and without water) and ion-pair solvation. A powerful feature of both the COSMIC program and the Roche system is the ability to perform a full search through the conformational space with MM to find the set of minimum energy conformations. Müller calculated 50 conformations all within 8 kJ/mol of the lowest energy for a 15-ring molecule. This clearly demonstrates that 'the' conformation of a molecule might not exist at room temperature, a fact one has to be aware of when modelling.

Odell (Shell Research, Sittingbourne) described CAMM approaches to pesticide design [10], also using (initially) in-house written software. The basic problem here is that the receptor site itself is usually unknown. Illustrations of methods for modelling active site components (azole EBI fungicides which inhibit P450 enzymes); the investigation of electronic properties (acetylcholine analogues); the analysis of factors regulating reactivity such as redox potentials (heteropentalenes); and methods for depicting intermolecular interactions such as solvation and binding site problems were shown. In many of these cases, Goodford's GRID and GRIN programs (which search for and display binding sites for specific functional groups) were applied.

Goodford himself described his GRID program [11] as a tool designed to help chemists think constructively as regards to non-covalent interactions in condensed systems, which may contain both large and small molecules in one or more aqueous or hydrocarbon phases. Examples presented included the protein collagen where different probe groups could be distinguished according to chemical property; the disaccharide maltose, and cholesterol and lysozyme in water (assessments of the interactions of small molecules); and specific ligand binding sites in haemagglutinin, a large enzyme of therapeutic interest from the influenza virus with ca. 20 000 atoms. Although GRID was originally devised with macromolecules in mind, the energy factors are currently being defined to deal with different target types including agrochemicals, polymers and zeolites and is already being used to investigate a much wider range of research problems than was originally envisaged. Goodford also demonstrated a further point of his, that although modelling is still qualitative, the insight gained can be enormous. This he illustrated with a past project of his, where a potent anti-sickle cell compound was designed based on modelling, even though they were looking at a very poor structure of haemoglobin that was not even the correct structure; in addition entropy was ignored and nothing was quantified. The moral of the story is that if one uses the tools sensibly they can help one think about what is going on.

A number of other topics raised more debate. Many of these items were related to the question of trustworthiness of CAMM results. Most participants agreed that CAMM outcomes have to be confronted with experiment. Here nuclear magnetic resonance spectroscopy and X-ray diffraction were often mentioned as important techniques in organic modelling.

Although more rigorous methods i.e. ab initio and semi-empirical calculations, are becoming more accessible due to faster computers and better algorithms, most people agreed (with the notable exception of von Ragué Schleyer, Univ. Erlangen), that MM will continue to play an important role in CAMM. For cases where it can be applied with confidence, such as large bio-organic molecules, MM often yields the best conformations and it can be used interactively. If MM is applied in conjunction with data bases and experiments one can obtain a 'self-consistent system', even if parameterisation and subsequent application of MM is really only a 'successful curve-fitting technique'.

One of the important drawbacks is that only energies, rather than wave functions and charges, can be calculated. Von Ragué Schleyer pointed out that MM produces results on conformations but no explanation. A better approach might be to optimise a structure with MM and then to calculate the wave functions with a molecular orbital program such as MOPAC/AMPAC [12]. Furthermore, as Odell added, there is an urgent need for more accurate forcefields for heteroatoms (e.g., O, S, N, P), in conjugated systems, organometallic compounds and synthetic polymeric systems, and to describe the factors involving solvation, desolvation and polarisability.

A lively and lengthy discussion took place during the panel session chaired by Mulheirn (Shell

Research, Sittingbourne), on the usefulness (or otherwise) of QSAR (quantitative structure activity relationships) and its relationship to modelling. Here opinion was firmly divided between those who believed QSAR to be of little practical use (the difficulty of defining criteria sufficiently; the lack of negative data for a thorough QSAR evaluation as uninteresting compounds tend to be dropped immediately) and those who wholeheartedly believed it has a role to play (especially at the beginning of a design procedure where it can help in the evaluation of experimental design and indicate limitations at an early stage). A compromise that could be agreed upon was that QSAR input to a CAMM study could be just as much value as input from a data base, if the field is narrowed in the search for a new active analogue and the number of compounds that actually have to be synthesised is thereby reduced. Even here though the practical chemists still remain to be convinced that such a route is faster in the long term than direct synthesis, although the same people are more than happy with any technique that delivers new options even if it is not the best possible or if the technique behaves as a 'black box'.

3. IS OUR THEORETICAL KNOWLEDGE SUFFICIENT FOR MOLECULAR MODELLING?

It was clear from the contributions on the theoretical background underlying modelling that developments in this area are particularly important in two respects. Theoretical progress is required first to improve the trustworthiness of structural properties derived in the microscopic world, thereby increasing the predictive power, and second to translate these microscopic results to macroscopic properties (e.g., modulus, strength) and behaviour (e.g., transport phenomena), thereby improving the explanative understanding.

Berendsen (Univ. Groningen) described MD simulations and Monte Carlo calculations, both empirical techniques employed to find a representative ensemble of 3-D structures [13]. The environment of the system, the energy and temperature can be included to enable the calculation of meaningful physical properties. The full determination of the entropy (and hence free energy) of a more complex ensemble, however, is still seldomly possible. Furthermore, there is no generalised scheme to find all local minima and the global minimum of the energy in an arbitrary system, and it is doubtful whether such a scheme could ever exist. For these reasons, systems known in detail have to be used to probe the significance of the strategies used for ill- or un-defined ones. Nevertheless, those involved in the field appear optimistic about future developments and judge lack of computer power as the main bottleneck. Counterchecks of MD-derived conformations of crystalline biomolecules generally deviate less than one Ångstrom from those determined by X-ray diffraction. Water molecules in the structure present a complication, but have to be taken into account; otherwise errors in the Coulombic terms of up to an order of magnitude might result.

Von Ragué Schleyer introduced the quantum chemical toolbox by stating that computational chemistry yields program packages rather than new concepts. He defined the task of quantum chemistry as predicting better than experiments can measure, i.e., within 4 kJ/mol. He illustrated his statements by predictions for the existence of putative small organic molecules, which might be found experimentally at a later stage, like the dilithium methyl cation with penta-coordinated carbon [14]. In the field of large molecules, however, approximations have to be made. Hückel, extended Hückel, CNDO and INDO have already been largely superceded: MM, semi-empirical methods (MINDO/3, MND, AM1) and ab initio have taken over. Allinger's complete set of force-

fields as contained in MM2 [15], allows very rapid calculations. Semi-empirical methods are at least an order of magnitude slower and are on average about 25-50 kJ/mol in error. They cannot be much improved, have a lack of rigour and it is hard to predict which program will work for which molecule. Ab initio methods, on the other hand, produce excellent results (for example on heats of formation), but the calculations take 3 to 4 orders of magnitude more CPU time. The quality of the various types of ab initio methods can be estimated by looking at the basis sets used, at how geometry is handled, and at the approximation chosen for electron correlation. The predictions for vibrational frequencies offer a suitable means of judging the calculation performance. Von Ragué Schleyer stressed the value of the Handbook of Computational Chemistry [16] and the Carnegie-Mellon Quantum Chemical Archive [17], which with its 20 000 parameters contains an enormous amount of information on small molecules.

Van der Avoird (Univ. Nijmegen) presented his experience with calculating macroscopic properties of materials starting from the Schrödinger equation. He distinguished three steps in the process. First the electronic structures have to be determined by quantum chemistry. These in turn have to be modelled by potential functions (forcefields), which is possible with an accuracy of at best 10%. Finally lattice dynamics yields the relevant macroscopic properties. For the molecular solids of nitrogen and oxygen, however, the standard approach fails. Therefore a new method was developed which has the advantage that no assumptions have to be made about the orientation of the molecules in the crystal and that large amplitudes of motion can be handled [18]. To this end the potentials were expanded in translational motions and the orientational dependences were retained. For the wave functions, the librations (orientational motions) were expanded in spherical harmonics, and the translational vibrations in three-dimensional oscillator functions. The orientational melting in solid nitrogen could thereby be calculated in addition to the relationship between the magnetic properties and the molecular motions in solid oxygen, the latter being even more difficult due to the triplet ground state of the oxygen molecule.

Results on ab initio calculations on zeolite stability were presented by Van Santen (Shell Research, Amsterdam) [19]. Beginning with calculations on the deformed dihedral angle in the oxygen-bridged silanol dimer, subsequent calculations on higher-membered rings revealed a major energetic jump and accompanying changes in atom distances and angles between three- and four-membered rings, after which stability became virtually independent of ring size. These ab initio results were judged to be very useful for understanding zeolite chemistry, in particular if expanded in the future by combining them with statistical mechanics. Catlow (Univ. Keele) added that also here the inclusion of water in the stability calculations (for which good forcefields are again lacking) is extremely important. Van der Avoird remarked that experimental verification is possible through infrared spectroscopy on occluded water.

Van Santen continued to introduce the discussion by stressing the 'triangular relationship' between computational chemistry, spectroscopy and chemistry. The main theme of the discussion centered on which theoretical approach can be applied in which case. Kollman thought that in industrial practice the costs for MM applications would already be roughly equal to those for small ab initio studies, which are to be preferred if quantitative results are required. Weber (Univ. Geneva), however, commented that ab initio methods still have difficulties in handling even simple transition metals. Again the problem of unavailability of many forcefields was raised: the theoreticians were questioned why this situation persists if the derivation from ab initio calculations is as easy as they had seemed to imply. Von Ragué Schleyer answered that the task is actually far from

trivial and complained that experimental support is often lacking. It may be concluded that computational chemistry has not yet fully bridged the gap between theoretical and practical chemistry. The tools are still hard to handle and the practical chemist may sometimes favour approaches that better match his intuitive feeling. The danger with this, however, as Müller believes he has frequently spotted, is that people are going to rely too much on pure graphics alone.

4. THE DESIGN OF CATALYSTS ON A COMPUTER SCREEN

Despite the somewhat later entrance of CAMM into the fields of inorganic and organometallic catalysis, it is clear that many parallels can be drawn with the biological and organic applications and there was general agreement that more can already be achieved than might have previously been anticipated.

In his introduction to the panel session, Catlow summarised the situation [20]. There are three main areas in the 'design' of catalysts – whether heterogeneous or homogeneous – where CAMM can play a role:

* substrate modelling (stability and structure, heterogeneous and homogeneous) where one can envisage:
  - testing the stability of hypothetical compounds and refining approximate structures
  - evaluating the role of critical parameters

* sorption or docking: examining the sites and energies of sorbed species in heterogeneous catalysis and the coming together of catalysts and reactant in homogeneous catalysis

* the reaction of sorbed or docked molecules, including reaction pathways and influence of critical parameters.

Modelling techniques that can be used for catalysis are the same as for life sciences and drug design:

* effective potentials (energy minimisation, MM, MD)
* quantum mechanical methods (especially in reaction pathways)
* interactive graphical displays.

In the case of substrate modelling it was generally agreed that the CAMM community is potentially in a good position. Modelling of inorganic structures is becoming possible and potentials are becoming available; critical parameters can also be studied in detail (e.g., Si/Al ratio in a zeolite). As Schenk (Univ. Amsterdam) showed, the range of sophistication varies from the relatively simple personal computer representation to real-time interactive graphics, where structures can be generated from the asymmetric units, atoms can be moved whilst preserving the overall symmetry and the graphics can be linked directly to powder X-ray diffraction simulations [21]. Weber, however, indicated that homogeneous organometallic structures are slightly more difficult: the problem lies with the need for a large and well-balanced electron basis set for metal atoms; as one is talking about in excess of 200 electrons for a transition metal complex, extended Hückel is the way

to approach interactive modelling. In this area in particular forcefields are generally not available and good quantum mechanical models are scarce. The development of forcefields for an MM approach, although difficult, was not considered impossible. Mixed feelings were expressed on the ability of current methodology to tackle the modelling of catalyst surfaces, Catlow's assertion that methods are already in place for surface modelling not being universally accepted.

Sorption ('docking') in heterogeneous catalysis can also be tackled. As Cheetham (Univ. Oxford) indicated, a zeolite for example can be thought of as an inorganic enzyme. Parameterisation should not even be a problem as all zeolites whatever their specific structure, should look the same to organic molecules, allowing transfer of parameters. The docking of a hydrocarbon molecule within an internal zeolite cavity has parallels with acceptor/receptor bio-organic problems and can be approached in the same way by energy minimisation. This, Cheetham demonstrated with predictions of heat of adsorption and locations of adsorbed molecules [23]. One should beware, however, about 'global minima' at 4K: for catalysts which only act as such at elevated temperatures, such global minima will be meaningless and one needs to carry out the calculations under reaction conditions. This then becomes a more time-consuming process for which Monte Carlo approaches can be applied, as Cheetham showed for methane in zeolite Y and the position of pyridine adsorbed in zeolite L [24]. A further complication is that the cavities in such three-dimensional inorganic systems are normally filled with water molecules and cations, the cations playing an important role in the catalysis and not just there to make up the charge. This too has to be taken into account, as was shown for benzene in zeolite L.

In connection with these problems, concern was expressed as to how important (catalyst) framework relaxation might be. The question as to whether the catalyst may be thought of as rigid or not or whether the use of zeolites at high temperatures made them flexible enough to worry about, occupied many people. It was suggested that one might be in a position to test this problem, but first it will be necessary to verify the potentials being used. Nevertheless a difference of opinion was noted as to how important relaxation might be: Cheetham pointed out that some structures (zeolites ZSM-5 and θ-1) display phase transitions upon molecule adsorbance and hence framework structures can be responsive to sorption; others suggested that the change upon relaxation would not be so extreme and could be calculated. In any case universal concern was expressed as to the host-sorbate potentials and how general they really are.

In the field of homogeneous catalysis the same problem exists as with substrate characterisation: the paucity of forcefields. Nevertheless such processes as nucleophilic addition to unsaturated hydrocarbons and electrophilic addition to metallic complexes can be tackled. Weber's results from application of the extended Hückel method were in good agreement with experiment for ferrocene and similar systems [22].

Reactions of adsorbed molecules is the least researched area and views differed both as to how much can now be achieved and can be expected in the future. Quantum mechanical calculation and embedded cluster techniques were considered to be the approaches to pursue, although Weber suggested that the extended Hückel methods used for the organometallics could also be applied in aluminosilicates.

The possibilities of modelling diffusion and transportation phenomena also gave rise to some debate, although again the consensus was that the problem can indeed be tackled with the tools now available. Catlow summarised the situation thus: if diffusion is fast, MD could be applied; if slow then an activated state (e.g., saddle point) could be defined and energies and entropies of ac-

tivation calculated. Potential energy plots might then be converted into diffusivities, although again the problem of framework relaxation becomes important if molecules are 'squeezing through' the lattice. Cheetham illustrated the representation of hydrocarbon diffusion pathways by means of energy contoured diagrams, comparing the behaviour of benzene in θ-1 and silicalite.

General conclusions from this session were therefore that modelling is in good shape for substrate work; an effective methodology remains to be developed for sorption; but that much more work has to be done before the modelling of reaction pathways becomes anything like routine. As regards the theme of the discussion – is 'design' really possible? – it is clear that in the case of homogeneous catalysis we still have a long way to go, but such a thought is not impossible; the problem lies with the improvement in the potentials. As far as heterogeneous catalysis is concerned much can be calculated for a 'theoretical' structure, even including computation of the activity with the aid of embedded clusters (which is in effect borrowing from enzyme modelling where one uses a local approach looking at just one part of the molecule rather than by the application of a holistic view). Again the key is good potentials, which are lacking.

As in other sessions then, the problem of potentials was again cited as the stumbling block to better modelling. It was agreed that the modelling community has to let it be known that working in this area is actually rewarding and not uninteresting. One approach to this problem is manifested in the various industrial consortia being set up. Much concern was expressed that if rigourous forcefields were forthcoming from such consortia they would be proprietary and witheld from the scientific community. This was strongly denied by the consortia representatives present. Due to the apparent interest in forcefield development at the conference, Thacher (Biosym Technologies, San Diego), although initially intending to speak about polymer modelling, explained the approach taken in Biosym's consortium research. At the time of the symposium, Biosym's consortium consisted of ten American companies (both chemical and computer) and had been in existence for eighteen months. Starting with biopolymers, the existing forcefields were tested and relevant experimental and theoretical data incorporated to derive improvements. Ab initio calculations are carried out to provide 'missing' data and special programs are being developed to parameterise the potentials. Twelve functional groups had been treated in six months. Although criticism was strong as to the amount of forcefield parameters and the ab initio techniques chosen, it seemed too early to judge the practical value of the exercise. Industrial forcefield consortia, however, seemed destined to become popular: to date (Spring 1988) two more have been founded (a second from Biosym and one from Biodesign), both specialising in synthetic polymer modelling.

5. HARDWARE FOR MOLECULAR MODELLING

The fast development of computer hardware, notably interactive graphics and supercomputers, has played a dominant role in the short history of CAMM. Computer science laboratories by their very nature can be years ahead of commercially available equipment (and software) and have indeed paved the way for the introduction and advancement of sophisticated techniques. Pique (Scripps Clinic, California) and Thorvaldsdottir (Univ. North Carolina, UNC) took the conference participants on some 'Scouting Expeditions in Molecular Graphics', from the pioneering work of the early seventies to the state-of-the-art (and hence the tomorrow for most CAMM users), in an often amusing but always thought provoking presentation. They showed astonishing graphics, created with off-the-shelf workstations (ray tracing images) or with newly-developed

hardware such as UNC's in-house Pixel-Planes multiprocessor, which is able to draw 13 000 spheres every second thereby allowing real-time manipulations of space-filled macromolecules [25,26]. Through experimental man-machine interfaces such as a head-mounted display and force feedback in molecular docking, and research into speech recognition and sound cues in molecular graphics, the foundations are being laid for the CAMM of the future.

A number of general conclusions were drawn on which agreement was apparent. The issue of vector- versus raster-graphics seems to be resolved in favour of the latter: the choice between the display of stick or surface models, and the use of stereo or depth cueing, is in the hands of the user and is no longer dependent on the hardware at one's disposal. Although the vast majority of modellers is still equipped with Evans & Sutherland PS300 graphics machines coupled to VAX computers, the once specialised field of CAMM is now merging with the mainstream of computing in the shape of new (mainly UNIX based) super-workstations. The direct coupling of graphics with a powerful CPU clearly offers new possibilities for truly interactive calculations, such as continuous energy monitoring. In addition mini-supercomputers are finding their way into the modelling community and are being utilised for the routine performance of extensive MD and ab initio calculations. A possibly cheaper alternative might be the transputer-based 'desk-bottom' supercomputer as presented by Davies (Chemical Design, Oxford). He also argued that the application has to be considered first, the software chosen accordingly and finally the hardware, to meet the requirements of the first two.

Hubbard (Univ. York) gave his personal views of the rapidly evolving commercial market. As recently as 1986, irrespective of how much money one had at one's disposal, the choice of hardware was straightforward. Since 1987, however, a large number of super-workstations with remarkable specifications have been introduced or announced, making the choice for the buyer harder than ever. Fortunately, Hubbard thinks that standards are emerging: PHIGS (3-D graphics), X-Windows and UNIX (although the battle with the VMS operating systems has not yet been conclusively decided). In conjunction with the trend towards coding in either standard Fortran or C, it can be expected that CAMM software will become portable between the various hardware systems. If the vendors of software packages can be further persuaded to create truly open architectures allowing the easy combination of programs from different sources, the life of a modeller would be much easier. This last point was generally considered as being crucial for the advancement of CAMM.

6. CONCLUDING SESSION

This concluding discussion was led by Vinter in the unfortunate absence of Richards (Univ. Oxford). A large number of topics were discussed, salient points made, worries expressed and advice given.

There was a certain amount of confusion amongst some of the participants as to the best (theoretical) path to follow when attempting to apply CAMM. When should one apply molecular mechanics? When should one use molecular dynamics? When should quantum mechanical approaches be invoked? Could polarisation be included? Is energy minimisation all that needs to be done? Given the lengths to which one has apparently to go to get useful parameters, does the whole approach in fact lack credibility? In short, what exactly is the role and status of theory in CAMM?

Perhaps the best answer, jointly provided by Professors Kollman, Goodford and von Ragué Schleyer, was that there are simply different ways of performing CAMM: all have their own concomitant advantages and disadvantages. Which approach one utilises depends on the context in which one is working and what exactly one is trying to do; in other words on the problem and the data available. There is also a hierarchy of approaches one might try. The problem will demand the level of theory to be applied: different people will be satisfied with different levels of answer. Useful insight will be gained for some problems without elaborate calculations, although it should not be forgotten that MM and graphics are essentially qualitative tools.

Two points to bear in mind are first, as emphasised by Goodford, that one should strive to express the parameters in energy equations in terms that are meaningful to the chemist and not as purely empirical coefficients without any obvious physical meaning. Second, that it is essential that predictions be related to real observations (either through experimentation or use of data bases).

Even so, the experts insisted that theory is at the cutting edge for classical organic molecules. Good results can be obtained and used intelligently; the theoretical tools available will solve problems in well defined areas. To some of the participants this was fine until the calculations broke down. Maybe one would only learn that the CAMM results were inaccurate after a lot of effort had been expended on synthesising the molecules. This problem is not unique to CAMM, however, and will always apply to every technique in synthetic chemistry research. Hence it was was suggested that the combination of QSAR and CAMM might be a sensible approach in these cases.

One aspect of this train of discussion to which everyone could agree, was that modelling does not negate the need for experimental results. Numerical experimental data are mandatory for validation and verification: everything should be tested against experiment. As Kollman emphasised, the quality of the model depends on the data set: errors in parameters will depend on the experimental data available. One should strive for a balanced and consistent model. Then the chance of being correct upon the extension of the parameterisation from one series of related molecules to another will be maximised.

In this respect experimentation is not the only possibility but data bases, and especially relational data bases are important. A number of crystallographic data bases are available (for both organic and inorganic structures) and these should be used effectively. Although as Cheetham reminded the symposium, in some areas where there is a dearth of reliable structural data (e.g., the extension of zeolite work to clays), calculation becomes paradoxically even more important where experiments are impossible or difficult to do.

The idea that molecular modelling is the golden panacea that will on its own invent or design new molecules was also discussed and dismissed. Experience has however shown that new ideas can be generated for further evaluation. CAMM is a tool like any other technique and should be treated as such. As an example the rhetorical question was posed as to which drug had ever been developed by nuclear magnetic resonance spectroscopy for instance. What is true claimed Müller, is that in all industrial environments there have been, since Day One, too many potential applications of CAMM to handle efficiently. Many firms can indeed point to examples where molecular modelling has played a large part in the synthesis of new molecules.

The conclusion therefore, as summarised by Vinter, is to be aware of misuse and the potential dangers. Bear in mind that the use of theory is not routine and that there are lots of gaps as far as the chemist is concerned. In many cases the results remain qualitative and rigour is lacking.

There will be no single or right answer when applying theory to molecular modelling and most especially to 3-D structure prediction. Understand the limitations; compare like with like; test everything against experiment; and exercise experience or go to someone with experience. This is the natural corollary of the fact that the theory is not always optimal for the problem being tackled. One has to develop one's own thoughts and place limited reliance on the numbers coming out of the computer (if one does not know what one is doing). The process should be considered to be a curve-fitting exercise which works extremely well for classical organic molecules but can work less well for more complex (and therefore often more interesting) structures.

This then begs the question as to who is going to do the basic theoretical work which underpins the future of CAMM. This was the aspect of CAMM which most worried both academia and industry. Energy parameterisation is necessary to predict unknown structural conformations and if the parameterisation is incomplete or obtained from a set of totally different molecules then the three-dimensional structure obtained for a compound will almost certainly be incorrect. Experimental data is needed to support parameterisation but very little work is being done on thermal chemistry (heats of combustion) or microwave measurements (which are the most useful types of experiments in this respect) because the academic world has apparently seen this type of work as unglamourous. Consortia are being founded by CAMM software vendors to do this work but the academic world is worried that this fundamental research – funded by industry – might remain proprietary and out of the public domain.

Given that experimental data are important to the future of CAMM in both academia and industry but that the work is not forthcoming, the question was raised as to whether industry is in a position to make a meaningful contribution in this direction. This can of course be done jointly and it was concluded that links between industry and universities are vital to the development of CAMM in both industry and academia and for industry to stay aware of new developments, as illustrated by Smith, Kline & French, for example. There are indeed areas where pre-competitive research can be carried out. From the university side, Goodford pleaded that industry should be prepared to add their structures to the generally available data bases.

A belief strongly expressed by both industry and academia was that commercial CAMM suppliers, and most especially software vendors, are (perhaps not surprisingly) competing and not collaborating. The wealth of new opportunities (hardware and software combined) was described by one participant as chaos. Furthermore the signs are that hardware developments will far outstrip software, which makes the question even more urgent as to how the software will cope. In this light it is clear that standards will be necessary and are thankfully beginning to emerge. The fact that molecular graphics hardware is merging with the mainstream of computing (e.g., workstations) may become advantageous, although at present this only causes uncertainty for buyers. Nevertheless, commercial suppliers of CAMM software were urged to create modular products with an open architecture. In this way packages will be able to be combined to suit everyone's personal needs, and software will no longer be hardware specific. The Molecular Graphics Society has considered trying to coordinate this type of activity in the past. It may be opportune to attempt a concerted effort to influence the vendors, perhaps through an industrial users' forum.

Another subject that created much discussion was the best method of approaching software: commercial packages, own packages or a combination of the two. Software development in-house is time consuming and expensive; it can, however, be an advantage to have in-house software. The programs can be better 'tuned' to the specific application and full advantage can be taken of hard-

ware developments. However the best approach at present seems to be to buy standard software and data bases and to adapt them to one's own problems and advantage. (Then one's software will be designed to address the problem rather than having to redesign the problem to suit the software). The drawback however is that this might not always be possible; most commercially available software coding is at present inaccessible. Whichever approach is taken, specialised manpower will be required and in-house experience is invaluable (which for newcomers to the field can be gained from a 'simple' graphics terminal and commercial 'off-the-shelf' software).

In conclusion, there was unanimous consensus that Computer-Aided Molecular Modelling is alive and kicking. State-of-the-art modelling is already at the stage where exciting things can be achieved in the areas of organic chemistry research and applications; furthermore the expectation is that such progress will be made in applications on organometallics, synthetic polymers and inorganic chemistry to render CAMM an indispensible key technology in such areas as catalysis. In short, the future of computer graphics in industry and academia is bright and assured.

ACKNOWLEDGEMENTS.

The authors wish to acknowledge the assistance of H.P.C.E. Kuipers and D.G. Streefkerk in preparing this article. Further we thank all the speakers at the conference for their valuable contributions and Sandy Abrahamsz for her indispensable support.

REFERENCES

1 Dearing, A., J. Comput.-Aided Mol. Design, 2 (1988) 179–189.
2 Singh, U.C. and Kollman, P.A., J. Comput. Chem., 7 (1986) 718.
3 Ferrin, T.E., Huang, C.C., Jarvis, L.E. and Langridge, R., J.Mol.Graph., 6 (1988) 2, 13.
4 Connolly, M.L., Science, 221 (1983) 709.
5 Bash, P.A., Singh, U.C., Langridge, R. and Kollman, P.A., Science, 236 (1987) 564.
6 Singh, U.C., Brown, F.K., Bash, P.A. and Kollman, P.A., J. Am. Chem. Soc., 109(1987) 1607.
7 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.
8 Müller, K., Ammann, H.J., Doran, D.M., Gerber, P. and Schrepfer, G., In Harms, A.F. (Ed.), Innovative Approaches in Drug Research, Elsevier, Amsterdam, 1986, p.125.
9 Vinter, J.G., Davis, A. and Saunders, M.R., J. Comput.-Aided Mol. Design, 1 (1987) 31–51.
10 Odell, B., J. Comput.-Aided Mol. Design, 2 (1988) 191–216.
11 Goodford, P., J. Med. Chem., 28 (1985) 849.
12 Dewar, M.J.S., Zoebisch, E.G., Healey, E.F. and Stewart, J.J.P., J. Am. Chem. Soc., 107 (1985) 3902.
13 Berendsen, H.J.C., J. Comput.-Aided Mol. Design, 2 (1988) 217–221.
14 a. Von Ragué Schleyer, P., Pure Appl. Chem., 56 (1984) 151.
b. Von Ragué Schleyer, P., J. Comput.-Aided Mol. Design, 2 (1988) 223–224.
15 Allinger, N.L. and Burkert, U., Molecular Mechanics, ACS Monograph No.177, American Chemical Society, Washington D.C., 1982.
16 Clark, T., Handbook of Computational Chemistry, Wiley, New York, 1985.
17 Carnegie Mellon Quantum Chemical Archive, Dept. Chemistry, Carnegie Mellon University, Pittsburgh, PA 15217.
18 Briels, W., Jansen, A.P.J. and Van der Avoird, A., Adv. Quantum Chem., 18 (1986) 131.
19 Ooms, G. and van Santen, R.A., Rec. Trav. Chim. Pays-Bas, 106 (1987) 69.
20 Catlow, C.R.A., J. Comput.-Aided Mol.Design, 2 (1988) 255–258.
21 Driessen, R.A.J., Loopstra, B.O., de Bruijn, D.P., Kuipers, H.P.C.E. and Schenk, H., J. Comput.-Aided Mol. Design 2 (1988) 225–233.

22 Weber, J., Fluekiger, P., Morgantini, P.-Y., Schaad, O., Goursot, A. and Daul, C., J. Comput.-Aided Mol. Design, 2 (1988) 235–253.
23 Cheetham, A.K., Nowak, A.K. and Betteridge, P.W., Proc. Ind. Acad. Sci. (Chem.Sci.), 96 (1986) 411.
24 Wright, P.A., Thomas, J.M., Nowak, A.K. and Cheetham, A.K., Nature, 318 (1985) 611.
25 Fuchs, H., Goldfeather, J., Hultquist, J.P., Spach, S., Austin, J.D., Brooks, F.P., Eyles, J.G. and Poulton, J., Comput. Graph., 19 (1985) 111–120.
26 Eyles, J., Austin, J., Fuchs, H., Greer, T. and Poulton, J., Pixel-planes 4: A Summary, Advances in Graphics Hardware 2: Proceedings of the Eurographics '87 Second Workshop on Graphics Hardware.