Computer-aided molecular design

W Graham Richards

Computer-aided molecular design has come of age. In the pharmaceutical and fine chemical industries computer graphics techniques allied to theoretical calculations are proving successful in suggesting to synthetic chemists molecules which will merit synthesis in terms of desired specific properties. Expert systems can also be used in suggesting the synthetic route by which the compounds may be synthesised. The techniques currently available can deal both with the relatively small molecules which tend to be the products of industry and the macromolecules with which they often interact to achieve desired results such as interference with biological control mechanisms.

molecular graphics, molecular structure, synthesis

CAMD (computer-aided molecular design) has not yet joined the universally recognised acronyms CAD and CAM, but activities which can be grouped under this heading are now widespread in industry and academia. There has been a proliferation of conferences on the subject, a number of general reviews[1–8] and a new journal largely devoted to the topic[9]. Despite this activity and the large overlap with computer-aided design of mechanical and architectural structures there has been very little cross-fertilization between the disciplines. This article aims to expose the nature of molecular design to the wider audience in the hope that some of the hardware and software can be adapted to the mutual advantage of both areas.

MOLECULES AND THEIR DESIGN

Chemists write the structure of a molecule as a chemical formula, such as $CH_3COOCH_2CH_2N(CH_3)_3^+$ for the neurotransmitter acetylcholine. This indicates that certain atoms are joined to each other giving a uniquely defined substance. More information is coded in the form of a representation of the formula which gives some impression of the relative positions of the atoms in space, found experimentally by techniques such as *X*-ray crystallography or nuclear magnetic resonance spectroscopy. Figure 1 shows an example of the sort of picture of a molecule which chemists have traditionally carried in their minds or built from kits of balls and sticks.

Some 10 million different compounds have been characterized. All are unique substances and, most importantly, if novel, they are very strong candidates for being patented. Several major industries exist on the basis of producing specific molecular products. All would like to design molecules with the properties they seek, rather than use trial and error or what has been termed 'molecular roulette'.

Physical Chemistry Laboratory, South Parks Road, Oxford, OX1 3QZ, UK

MOLECULAR INDUSTRIES

The pharmaceutical industry is most advanced in terms of designing molecules and the application of computer techniques. Here research generally starts with a lead compound with desired qualities – a known biological neurotransmitter, or perhaps a competitor's product. The molecules are thought to act by binding to very specific macromolecular receptors, proteins, rather as a hand fits a glove. The chemist synthesises variations on the lead compound which he or she hopes will have the appropriate shape and electron distribution, and the skill of medicinal chemists is in deciding which compounds to synthesize next, given that a range of compounds vary in activity. Traditional success rates have been very low, perhaps one valuable compound for every thousand or so synthesized. This is expensive. $100 million may be spent developing a drug, but potential rewards are similarly astronomic. One product may have annual sales in excess of $1 billion. The attractions of designing molecules are obvious.

The agrochemical world of herbicides, pesticides and fungicides very much follows the pattern of pharmaceuticals. Their products need to bind to the complex molecular architecture of proteins involved in the biochemical mechanisms of target organisms.

Small molecules with specific properties are the products of fine chemical industries such as those manufacturing flavours or perfumes which act at the molecular level by binding to specific molecular receptors. Dyes need particular spectroscopic properties and are again small molecules.

Work in large molecular species is most advanced in designing catalysts such as zeolites and metal oxide catalytic surfaces, but the promise lies in designing novel variants of enzymes which can be synthesized using genetic engineering techniques and in the design of molecular electronic devices.

DISPLAYS OF MOLECULAR STRUCTURE

The simplest, but in many ways most powerful use of computer graphics has been merely displaying computer-generated pictures of molecules rather like the wire, ball

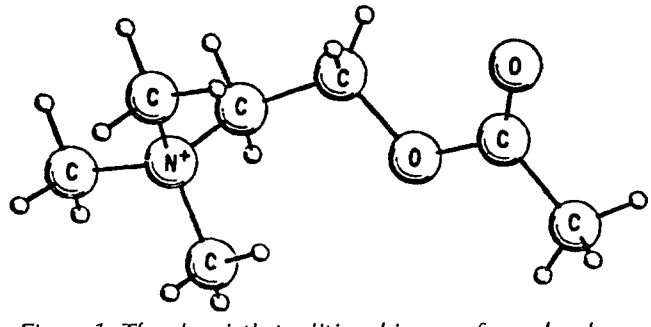

Figure 1. The chemist's traditional image of a molecule: acetylcholine

 0010–4485/85/050215–04 $03.00

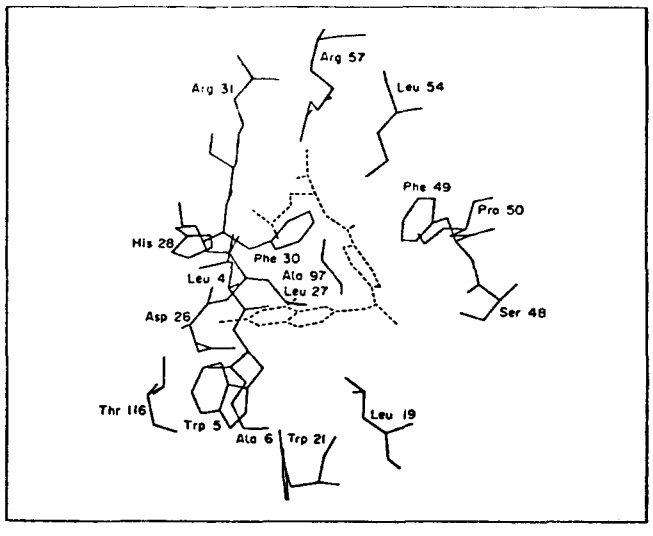

Figure 2. The drug methotrex (dotted) located in the binding site of the enzyme dihydrofolate reductase, drawn using an ICL PERQ compiler using software developed by R Hubbard of York University, UK and C B Naylor of Oxford University, UK

and stick or space-filling models used by generations of chemists. Colour plates 1 and 2 (see p C1) give some typical examples.

The computer offers several important extra features which cannot be realized with sets of physical models. The bond lengths and angles may be taken from crystallographic databases and are thus accurate and not mere crude averages. Pictures of two molecules may be superimposed and distinguished using colour. Gravity does not cause problems with large structures with tens of thousands of atoms such as enzymes. Parts of a molecule may be rotated with respect to other parts and indications of atoms clashing clearly highlighted. If enzyme crystal structures are known then small molecules may be fitted into active or binding sites, as in Figure 2.

For these displays there are a number of software packages available[10,11], all of which allow superimpositions to be performed interactively, permitting the chemist to incorporate his or her own knowledge and experience in the modelling.

The extension of molecular display to the large periodic structures of the solid state permit visualization not only of perfectly crystalline materials but also the less conventional but important structural phenomena involving ingrowths, epitaxy or surface reorganization and absorption. Of particular current interest are the zeolite structures which contain channels of molecular dimension with important implications for catalysis. Colour plate 3 gives an example of the display of a zeolite.

THEORETICAL COMPUTATIONS

Beyond simple displays of molecular structure comes the incorporation of theoretical calculations into the graphical representation. Calculations are broadly of two types; energetics and electronic properties.

Energy

The energy of a molecule will vary as bond lengths and angles are changed. If one is seeking either the most stable shape of a molecule, or an indication of the shapes into which a molecule could be distorted by a given binding energy, then it is important to be able to compute the energy as a function of molecular shape in real time. For this purpose the methods of so-called molecular mechanics have been developed. This type of calculation, like the simple displays, treats a molecule as if it were a set of balls and bonds (or springs) with a library of potential energy functions for each type of atom–atom distance and for bending bond angles or torsion angles. The methods are purely empirical, but have reached a high level of sophistication and reliability for some restricted ranges of molecular types.

The computationally more demanding approach which is as yet restricted to small molecules is the use of quantum mechanics. Packages are available to solve the Schrödinger equation in the Hartree-Fock approximation for molecules of up to perhaps fifty atoms[12]. These methods may be of the *ab initio* variety, making no approximations with regard to the integrals necessary or semi-empirical, incorporating some experimental data to reduce the computational demands of vast numbers of integrals.

Electronic properties

The quantum mechanical calculations come into their own when the electronic properties of molecules are the point of interest. Since at the microscopic level a molecule is a blob of electron density and its chemistry is a feature of the electronic behaviour, this is an increasingly important area and one where computer display is of paramount importance. Solving the Schrödinger equation, written cryptically as $H\psi = E\psi$, yields not only the energy, E, of a molecule for a given set of positions of the atomic nuclei, but also the associated wave function, ψ. From this wave function all electronic properties may be calculated. The only problem is how to display this information in a form which is readily assimilated by a synthetic chemist.

The simplest electronic property is electron density. A particular contour level may be displayed as in Colour plate 4.

The virtue of such an image is that it may give an idea of molecular size and shape which is more realistic than merely taking standard-sized spheres for each atom type. This kind of display may also permit the superposition of molecules which all fit the same unknown receptor and allow a tentative mapping of the all-important target site. One problem in such a procedure is just how to superimpose the molecules if they are structurally diverse. Current work is attempting to automate this by relying on a criterion of maximum electronic similarity in the form of a similarity index.

In fact molecules do not have a defined edge. The distance at which a molecule is strongly repelled by another will depend on the relative natures of the two species. A water molecule will experience different interaction energies with another molecule depending on whether the hydrogen atoms or the oxygen is pointing towards the other molecule. To go some way towards recognizing this fact it has become clear that a more revealing expression of electronic properties is a display of electrostatic molecular potential – the interaction energy between a proton and the molecule. Because there are regions of attraction and repulsion this property is a better discriminator between molecules. A particularly appealing way of displaying this information has been developed by Quarendon[13] using constructive solid geometry to produce a representation of a molecule which is instantly recognizable by a chemist and

then to colour-code the electrostatic potential on this surface as illustrated in Colour plate 5 (see p C2).

The intention of this approach is to produce coloured displays which may lead the chemist to a rationalization of variable activity in a series of similar compounds in a qualitative way using a lot of data coded in colour. Quantitative correlations may then be sought using the numerical information which the quantum mechanical calculations provide. The electrostatic field surrounding a molecule may also be calculated and used in a parallel fashion although since this quantity has direction as well as magnitude the actual display is more complex[14].

COMPUTER-AIDED SYNTHESIS

The design aids discussed so far have largely been aimed at helping the synthetic chemist decide which small molecule it would be best to synthesize. Increasingly the computer is also being used to help decide how that compound should be synthesized. There are being developed a number of chemistry-based expert systems which are concerned with synthesis planning. The most well-known of these is the program with the acronymic name LHASA (logic and heuristics applied to synthetic analysis) developed by Corey and his collaborators at Harvard[15], USA. Systems known as SECS[16], CASP[17] and SYNCHEM[18] are similar. These programs operate in a retrosynthetic manner. The target compound is entered as a structural drawing and the program works backwards suggesting alternative precursors right back to possible starting materials. At each level the user decides which precursors merit further processing. The user may also impose constraints to limit the combinatorial problem.

The alternative approach which generates reactions by a formal procedure is the aim of a different class of synthesis design program represented by EROS[19]. Organic reactions are generated by rules about the formation and breaking of bonds between atoms. The use of formal generators of reactions will yield an impossibly large number of reactions including the totally implausible. The programs thus need some method of evaluation, such as estimates of the thermodynamic reasonableness of proposed schemes.

Both approaches have proved themselves and are moving from being considered rather esoteric to a situation where they are part of the synthetic chemist's armoury.

ENZYMES AS TARGETS

Often the chemistry design problem is that of producing a molecule which will bind to a largely hypothetical receptor. The increasingly important exception to this is the case where the target receptor is the complex binding site of an enzyme whose structure has been determined from *X*-ray crystallography. There are a number of cases where there is a crystal structure for an enzyme whose blockade would produce a desired biological effect such as lowering blood pressure, killing cancer cells or parasites, or preventing blood cells from sickling.

Design of inhibitors, or molecules which will bind into the enzymic active site so tightly as to prevent action on the natural substrate of this enzyme must start with a display of the enzyme site. Sometimes dotted surfaces are used to indicate available space or electronic properties like electrostatic potential. Colour plate 6 gives an example.

Any novel inhibitor must firstly fit into the site. This may be observed visually using docking routines which test for clashes between atoms. In addition to fitting, or avoiding strong repulsions, the putative inhibitor must bind as strongly as possible to the site. The interaction energy may be calculated using the methods of molecular mechanics or some modified quantum mechanical scheme[20,21].

The greatest weakness in this work is the failure to include the very important effects of solvation and desolvation of molecules when binding takes place between inhibitor and enzyme. These solvent effects dominate entropic effects. The problem is, however, to some extent circumvented by focusing attention on the differences in binding energies predicted for otherwise very similar compounds.

HARDWARE AND SOFTWARE

A wide range of hardware is currently being employed but in the most active area, that of pharmaceuticals, the chosen system appears to be the Evans and Sutherland PS-300 vector graphics system, frequently coupled to a Digital Equipment Corporation VAX 11/780. This type of vector graphics permits large numbers of vectors to be displayed, movement to be smooth in real time and sharp lines, nets and dotted surfaces to be drawn. The disadvantage of vector systems as opposed to raster technology is their inability to display solid colour, however. The dotted surfaces appear very confusing to the casual user while painted surfaces are much easier to grasp. For this reason as well as on grounds of cost, many applications use raster graphic terminals such as the Sigmex 6130 in preference to vector graphics terminals. As raster systems refresh at greater speeds the problems of including movement are diminishing[21]. The other major suppliers in molecular design are Chromatics, Megatek, Ramtech, Spectragraphics, Tektronix, Silicon Graphics and Vector General.

As well as these professional systems a huge variety of programs have been written for simple machines such as the Apple or BBC microcomputer. Basically the only requirement is a terminal capable of graphical output. These terminals need to communicate with the computer on a serial line at rates of between 30 and 960 characters/s. More efficient transfer speeds using parallel data transfer are of course possible but have to be reconciled with heavily used time-shared computers and the frequent need to transfer large amounts of data over long distances.

Surprisingly little use has as yet been made of individual workstations. Given that a typical pharmaceutical house may employ hundreds of synthetic chemists, it seems likely that future developments will include workstations in individual laboratories linked to a central system upon which theoretical computations can be performed and via which access to databases of structures may be obtained.

Two programs which can be used for molecular display are ORTEP[22] and PLUTO[23]. These have been modified in a number of academic environments for specialist purposes, but industry has widely gone to specialist software houses who market packages to perform many of the graphical designs features discussed here and in some cases systems for handling databases. The leading companies are Chemical Design Ltd[24], Molecular Design[25] and Tripos Associates[26].

The programs needed to run theoretical calculations on molecules are generally available at nominal cost from the admirable Quantum Chemistry Program Exchange[27].

THE STATE OF THE ART

New drugs or new products are unlikely to be discovered solely or even chiefly by the use of computer-aided