3.3. MOLECULAR MODELLING AND GRAPHICS

3.3.2. Molecular modelling, problems and approaches

This section is concerned with software techniques which permit a set of atomic coordinates for a molecule to be generated *ab initio*, or to be modified, by reference to some chosen criterion, usually the electron density. Software that can change the shape of a molecule must be cognizant of the connectivity of the molecule and the bonding characteristics of atom types. It must also have means of regaining good stereochemistry if current coordinates are poor in this respect, or of performing its manipulations in ways which conserve essential stereochemical features. Approaches to some of these problems are outlined below. Many of the issues involved, including the topics outlined in Section 3.3.1.4 above, have been excellently reviewed by Hermans (1985), though with little reference to graphical aspects, and a comprehensive treatment of modelling methods based on energies is given by Burkert & Allinger (1982).

3.3.2.1. *Connectivity*

It is necessary to distinguish three different kinds of connectivity, namely structural, logical and drawing connectivities. Structural connectivity consists of the specification of the chemical bonding of the molecule and, as such, is an absolute property of the molecule. Logical connectivity consists of the specification of what part or parts of a molecule are moved, and in what way, if some stereochemical feature is altered. Logical and structural connectivity are closely related and in simple cases coincide, but the distinction is apparent, for example, if the puckering of a five-membered ring is being modelled by permitting folding of the pentagon about a line connecting non-adjacent corners. This line is then a logical connection between two moving parts, but it is not a feature of the structural connectivity.

Drawing connectivity consists of a specification of the lines to be drawn to represent the molecule and often coincides with the structural connectivity. However, stylized drawings, such as those showing the α-carbon atoms of a protein, require to be drawn with lines which are features neither of the logical nor of the structural connectivity.

3.3.2.1.1. *Connectivity tables*

The simplest means of storing connectivity information is by means of tables in which, for each atom, a list of indices of other atoms to which it is connected is stored. This approach is quite general; it may serve any type of molecular structure and permits structures to be traversed in a variety of ways. In this form, however, it is extravagant on storage because every connection is stored twice, once at each of the nodes it connects. It may, however, provide the starting material for the algorithm of Section 3.3.1.5.2 and its generality may justify its expense.

From such a list, lists of bonds, bond angles and dihedral angles may readily be derived in which each entry points to two, three or four atoms in the atom list. Lists of these three types form the basis of procedures which adjust the shape of a molecule to reduce its estimated potential energy (Levitt & Lifson, 1969; Levitt, 1974), and of search-and-retrieval techniques (Allen *et al.*, 1979).

Katz & Levinthal (1972) discuss the explicit specification of structural connectivity in terms of a tree structure in which, for each atom, is stored a single pointer to the connected atom nearer to the root, virtual atoms being used to allow ring structures to be treated as trees. An algorithm is also presented which allows such a tree specification to be redetermined if an atom in the tree is newly chosen as the root atom or if the tree itself is modified.

Cohen *et al.* (1981) have developed methods of handling connectivity in complicated fused- and bridged-ring systems.

3.3.2.1.2. *Implied connectivity*

In cases where software is required to deal only with a certain class of molecule, it may be possible to exploit the characteristics of that class to define an ordering for lists of atoms such that connectivity is implied by the ordering of items in the list. Such an ordering may successfully define one of the three types of connectivity defined in Section 3.3.2.1 but it is unlikely to be able to meet the needs of all three simultaneously. It may also be at a disadvantage when required to deal with structures not part of the class for which it is designed. Within these limitations, however, it may be exceedingly efficient. Both proteins and nucleic acids are of a class which permits their logical connectivity to be specified entirely by list ordering, and the software described in Section 3.3.3.2.6 uses no connectivity tables for this purpose. The ordering rules concerned are given by Diamond (1976*b*).

Drawing connectivity needs explicit specification in such a case; this may be done using only one 16-bit integer per atom, which may be stored as part of the atom list without the need of a separate table. This integer consists of two signed bytes which act as relative pointers in the list, positive pointers implying draw-to, negative pointers implying move-to. As each atom is encountered during drawing the right byte is read and utilized, and the two bytes are swapped before proceeding. This allows up to two bonds drawn to an atom and two bonds drawn from it, four in all, with a minimum of storage (Diamond, 1984*a*).

Brandenburg *et al.* (1981) handle drawing connectivity by enlarging the molecular list with duplicate atoms such that each is connected to the next in the list, but moves and draws still need to be distinguished.

Levitt (1971) has developed a syntax for specifying structural connectivity implicitly from a list structure which is very general, though designed with biopolymers in mind, and the work of Katz & Levinthal (1972) includes something similar.

3.3.2.2. *Modelling methods*

Fundamental to the design of any software for molecular modelling are the choices of modelling criteria, and of parameterization. Criteria which may be adopted might include the fitting of electron density, the minimization of an energy estimate or the matching of complementary surfaces between a pair of molecules. Parameterizations which may be adopted include the use of Cartesian coordinates of atoms as independent variables, or of internal coordinates, such as dihedral angles, as independent variables with atomic positions being dependent on these. Systems designed to suit energetic criteria usually use Cartesian coordinates since all aspects of the structure, including bond lengths, must be treated as variables and be allowed to contribute to the energy estimate. Systems designed to fit a model to observed electron density, however, may adequately meet the stereochemical requirements of modelling on either parameterization, and examples of both types appear below.

Inputs to modelling systems vary widely. Systems intended for use mainly with proteins or other polymeric structures usually work with a library of monomers which the software may develop into a polymer. Systems intended for smaller molecules usually develop the molecular structure atom-by-atom rather than a residue at a time, and systems of this kind require a very general form of input. They may accept a list of atom types and coordinates if measurement and display of a known molecule is the objective, or they may accept 'sketch-pad' input in the form of a hand-drawn two-dimensional sketch of the type conventional in chemistry, if the objective is the design of a molecule. Sketch-pad input is a feature of some systems with quantum-mechanical capabilities.

377

3.3.2.2.1. *Methods based on conformational variables*

Suppose that **t** represents a vector from the current position of an atom in the model to a target position then (see Section 3.3.1.1.3), to first order, the observational equations are

$$t_{IA} = D_{IpA}\theta_p + v_{IA}$$

in which $\boldsymbol{\theta}$ represents changes to conformational variables which may include dihedral angles, bond angles, bond lengths, and parameters determining overall position and orientation of the molecule as a whole. If every such parameter is included the model acquires $3n$ degrees of freedom for n atoms, in which case the methods of the next section are more appropriate, but if bond lengths and some or all bond angles are being treated as constants then the above equation becomes the basis of the treatment.

$$D_{IPA} = \frac{\partial r_{IA}}{\partial \theta_P} = \varepsilon_{Ijk} n_{jP}(r_{kA} - r_{kP})$$

in which $\mathbf{n}_P$ is a unit vector defining the axis of rotation for an angular variable θ_P, $\mathbf{r}_A$ and $\mathbf{r}_P$ are position vectors of the atom A and the site of the parameter P, and $\mathbf{v}_A$ represents a residual vector.

$\sigma = v_{ia}v_{ia}$ is minimized by

$$\boldsymbol{\theta} = \boldsymbol{M}^{-1}\mathbf{V}$$

in which

$$M_{PQ} = D_{iPa}D_{iQa}$$
$$V_P = D_{iPa}t_{ia}.$$

More generally, if σ represents any scalar quantity which is to be minimized, *e.g.* an energy, then

$$V_P = -\frac{1}{2}\frac{\partial \sigma}{\partial \theta_P}$$

$$M_{PQ} = \frac{1}{2}\frac{\partial^2 \sigma}{\partial \theta_P \partial \theta_Q}.$$

It is beyond the scope of this chapter to review the methods available for evaluating $\boldsymbol{\theta}$ from these equations. Difficulties may arise from two sources:

(i) Inversion of M may be difficult if M is large or ill conditioned and impossible if M is singular.

(ii) Successful evaluation of $M^{-1}\mathbf{V}$ will not minimize σ in one step if $\mathbf{t}$ is not linearly dependent on $\boldsymbol{\theta}$ or, equivalently, $\partial^2\sigma/\partial\theta_P\partial\theta_Q$ is not constant, and substantial changes $\boldsymbol{\theta}$ are involved. Iteration is then necessary.

Difficulties of the first kind may be overcome by gradient methods, for example the conjugate gradient method without searches if M is available or with searches if it is not available, or they may be overcome by methods based on eigenvalue decompositions. If non-linearity is serious less dependence should be placed on M and gradient methods using searches are more valuable. In this connection Diamond (1966) introduced a sliding filter technique which produced rapid convergence in extreme conditions of non-linearity. These topics have been reviewed elsewhere (Diamond, 1981, 1984*b*) and are the subject of many textbooks (Walsh, 1975; Gill *et al.*, 1981; Luenberger, 1984).

Warme *et al.* (1972) have developed a similar system using dihedral angles as variables and a variety of alternative optimization algorithms.

The modelling of flexible rings or lengths of chain with two or more fixed parts is sometimes held to be a difficulty in methods using conformational variables, although a simple two-stage solution does exist. The principle involved is the sectioning of the space of the variables into two orthogonal subspaces of which the first is used to satisfy the constraints and the second is used to perform the optimization subject to those constraints.

The algebra of the method may be outlined as follows, and is given in more detail by Diamond (1971, 1980*a*,*b*). Parametric shifts $\boldsymbol{\theta}_1$ which satisfy the constraints are solutions of

$$\mathbf{V}_1 = \boldsymbol{M}_1\boldsymbol{\theta}_1$$

in which $\mathbf{V}_1$ and $\boldsymbol{M}_1$ depend only on the target vectors, $\mathbf{t}_1$, of the atoms with constrained positions and on the corresponding derivatives. We then find a partitioned orthogonal matrix $(\boldsymbol{A}_1\boldsymbol{B}_1)$ satisfying

$$(\boldsymbol{A}_1\boldsymbol{B}_1) = (\boldsymbol{E}_\lambda\boldsymbol{E}_0)\begin{pmatrix} \boldsymbol{R}_\lambda & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{R}_0 \end{pmatrix}$$

in which $\boldsymbol{E}_\lambda$ are the eigenvectors of $\boldsymbol{M}_1$ having positive eigenvalues, $\boldsymbol{E}_0$ are those having zero eigenvalues, and $\boldsymbol{R}_\lambda$ and $\boldsymbol{R}_0$ are arbitrary orthogonal matrices. Then

$$\begin{pmatrix} \boldsymbol{A}_1^T \\ \boldsymbol{B}_1^T \end{pmatrix}\mathbf{V}_1 = \begin{pmatrix} \boldsymbol{A}_1^T \\ \boldsymbol{B}_1^T \end{pmatrix}\boldsymbol{M}_1(\boldsymbol{A}_1\boldsymbol{B}_1)\begin{pmatrix} \boldsymbol{A}_1^T \\ \boldsymbol{B}_1^T \end{pmatrix}\boldsymbol{\theta}_1$$

$$= \begin{pmatrix} \boldsymbol{A}_1^T\boldsymbol{M}_1\boldsymbol{A}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}\begin{pmatrix} \boldsymbol{A}_1^T \\ \boldsymbol{B}_1^T \end{pmatrix}\boldsymbol{\theta}_1$$

$$\boldsymbol{A}_1^T\boldsymbol{\theta}_1 = (\boldsymbol{A}_1^T\boldsymbol{M}_1\boldsymbol{A}_1)^{-1}\boldsymbol{A}_1^T\mathbf{V}_1$$

in which the matrix to be inverted is positive definite. $\boldsymbol{A}_1$, however, is rectangular so that multiplying on the left by $\boldsymbol{A}_1$ does not necessarily serve to determine $\boldsymbol{\theta}_1$, but we may write

$$\boldsymbol{\theta}_1 = (\boldsymbol{A}_1\boldsymbol{B}_1)\begin{pmatrix} \boldsymbol{\varphi}_1 \\ \boldsymbol{\psi}_1 \end{pmatrix}$$

giving

$$\begin{pmatrix} \boldsymbol{A}_1^T \\ \boldsymbol{B}_1^T \end{pmatrix}\mathbf{V}_1 = \begin{pmatrix} \boldsymbol{A}_1^T\boldsymbol{M}_1\boldsymbol{A}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}\begin{pmatrix} \boldsymbol{\varphi}_1 \\ \boldsymbol{\psi}_1 \end{pmatrix}$$

$$\boldsymbol{\varphi}_1 = (\boldsymbol{A}_1^T\boldsymbol{M}_1\boldsymbol{A}_1)^{-1}\boldsymbol{A}_1^T\mathbf{V}_1$$

and $\boldsymbol{\psi}_1$ is indeterminate and free to adopt any value. We therefore adopt

$$\boldsymbol{\theta}_1 = \boldsymbol{A}_1\boldsymbol{\varphi}_1 = \boldsymbol{A}_1(\boldsymbol{A}_1^T\boldsymbol{M}_1\boldsymbol{A}_1)^{-1}\boldsymbol{A}_1^T\mathbf{V},$$

which is the smallest vector of parametric shifts which will satisfy the constraints, and allow $\boldsymbol{\psi}_1$ to be determined by the remaining observational equations in which the target vectors, $\mathbf{t}$, are now modified to $\mathbf{t}_2$ according to

$$\mathbf{t}_2 = \mathbf{t} - \boldsymbol{D}_2\boldsymbol{\theta}_1,$$

$\boldsymbol{D}_2$ and $\mathbf{t}_2$ being the derivatives and effective target vectors for the unconstrained atoms. We then solve

$$\mathbf{V}_2 = \boldsymbol{M}_2\boldsymbol{\theta}_2$$

in which $\boldsymbol{\theta}_2$ is required to be of the form

$$\boldsymbol{\theta}_2 = \boldsymbol{B}_1\boldsymbol{\psi}_1$$

giving

$$\boldsymbol{\theta}_2 = \boldsymbol{B}_1(\boldsymbol{B}_1^T\boldsymbol{M}_2\boldsymbol{B}_1)^{-1}\boldsymbol{B}_1^T\mathbf{V}_2$$

and apply the total shifts

$$\boldsymbol{\Theta} = \boldsymbol{\theta}_1 + \boldsymbol{\theta}_2$$

to obtain a structure which is optimized within the restrictions imposed by the constraints.

It may happen that $B_1^T M_2 B_1$ is itself singular because there are insufficient data in the vector $\mathbf{t}_2$ to control the structure and the parametric shifts contained in $\boldsymbol{\theta}_2$ fully. In this event the same process may be applied again, basing the solution for $\boldsymbol{\theta}_2$ on

$$\begin{pmatrix} A_2^T \\ B_2^T \end{pmatrix} B_1^T M_2 B_1 (A_2 B_2)$$

so that the vectors in B_2 represent the degrees of freedom which remain uncommitted. This method of application of constraints by subspace sectioning may be nested to any depth and is completely general.

A valid matrix A_1 may be found from M_1 by using the fact that the columns of M_1 are all linear combinations of the columns of E_λ and are void of any contribution from E_0. It follows that A_1 may be found by using the columns of M_1 as priming vectors in the Gram–Schmidt process [Section 3.3.1.2.3 (i)] until the normalizing step involves division by zero. A_1 is then complete if all the columns of M_1 have been tried. $(A_1 B_1)$ may then be completed by using arbitrary vectors as primers.

Manipulation of a ring of n atoms may be achieved by treating it as a chain of $(n + 2)$ atoms [having $(n + 1)$ bond lengths, n bond angles and $(n - 1)$ dihedral angles] in which atom 1 is required to coincide with atom $(n + 1)$ and atom 2 with $(n + 2)$. $\mathbf{t}_1$ then contains two vectors, namely the lack-of-closure vectors at these points, and is typically zero. A_1 is then found to have five columns corresponding to the five degrees of freedom of two points of fixed separation; $\boldsymbol{\theta}_1$ contains only zeros if the ring is initially closed, and contains ring-closure corrections if, through non-linearity or otherwise, the ring has opened. B_1 contains $(p - 5)$ columns if the chain of $(n + 2)$ points has p variable parameters. It follows, if bond lengths and bond angles are treated as constants, that the seven-membered ring is the smallest ring which is flexible, that the six-membered ring (if it can be closed with the given bond angles) has no flexibility (though it may have discrete alternatives) and that it may be impossible to close a five-membered ring. Therefore some variation of bond angles and/or bond lengths is essential for the modelling of flexible five- and six-membered rings. Treating the ring as a chain of $(n + 1)$ atoms is less satisfactory as there is then no control over the bond angle at the point of ring closure.

A useful concept for the modelling of flexible five-membered rings with near-constant bond angles is the concept of the pseudorotation angle P, and amplitude θ_m, for which the jth dihedral angle is given by

$$\theta_j = \theta_m \cos\left(P + \frac{4\pi j}{5}\right).$$

This formulation has the property $\sum_{j=0}^{4} \theta_j = 0$, which is not exactly true; nevertheless, θ_j values measured from observed conformations comply with this formulation within a degree or so (Altona & Sundaralingam, 1972).

Software specialized to the handling of condensed ring systems has been developed by van der Lieth *et al.* (1984) (Section 3.3.3.3.1) and by Cohen *et al.* (1981) (Section 3.3.3.3.2).

3.3.2.2.2. *Methods based on positional coordinates*

Modelling methods in which atomic coordinates are the independent variables are mathematically simpler than those using angular variables especially if the function to be minimized is a quadratic function of interatomic distances or of distances between atoms and fixed points. The method of Dodson *et al.* (1976) is representative of this class and it may be outlined as follows. If $\mathbf{d}$ is a column vector containing ideal values of the scalar distances from atoms to fixed target points or to other atoms, and if $\mathbf{l}$ is a column vector containing the prevailing values of these quantities obtained from the model, then

$$\mathbf{d} = \mathbf{l} + D\delta\mathbf{x} + \boldsymbol{\varepsilon}$$

in which the column matrix $\delta\mathbf{x}$ contains alterations to the atomic coordinates, $\boldsymbol{\varepsilon}$ contains residual discrepancies and D is a large sparse rectangular matrix containing values of $\partial l/\partial x$, of which there are not more than six non-zero values on any row, consisting of direction cosines of the line of which $\mathbf{l}$ is the length. $\boldsymbol{\varepsilon}^T W \boldsymbol{\varepsilon}$ is then minimized by setting

$$D^T W(\mathbf{d} - \mathbf{l}) = D^T W D \delta\mathbf{x},$$

which they solve by the method of conjugate gradients without searches. This places reliance on the linearity of the observational equations (Diamond, 1984*b*). It also works entirely with the sparse matrix $W^{1/2}D$, the dense matrix $D^T W D$, and its inverse, being never calculated.

The method is extremely efficient in annealing a model structure for which an initial position for every atom is available, especially if the required shifts are within the quasi-linear region, but is less effective when large dihedral-angle changes are involved or when many atoms are to be placed purely by interpolation between a few others for which target positions are available. Interbond angles are controlled by assigning d values to second-nearest-neighbour distances; this is effective except for bond angles near 180° so that, in particular, planar groups require an out-of-plane dummy atom to be included which has no target position of its own but does have target values of distances between itself and atoms in the planar group. The method requires a d value to be supplied for every type of nearest- and next-nearest-neighbour distance in the structure, of which there are many, together with W values which are the inverse variances of the distances concerned as assessed by surveys of the corresponding distances in small-molecule structures, or from estimates of their accuracy, or from estimates of accuracy of the target positions.

Hermans & McQueen (1974) published a similar method which differs in that it moves only one atom at a time, in the environment of its neighbours, these being considered fixed while the central atom is under consideration. This is inefficient in the sense that in any one cycle one atom moves only a small fraction ($\sim$3%) of the distance it will ultimately be required to move, but individual atom cycles are so cheap and simple that many cycles can be afforded. The method was selected for inclusion in *Frodo* by Jones (1978) (Section 3.3.3.2.7) and is an integral part of the *GRIP* system (Tsernoglou *et al.*, 1977; Girling *et al.*, 1980) (Section 3.3.3.2.2) for which it was designed.

3.3.2.2.3. *Approaches to the problem of multiple minima*

Modelling methods which operate by minimizing an objective function of the coordinates (whether conformational or positional) suffer from the fact that any realistic objective function representing the potential energy of the structure is likely to have many minima in the space of the variables for any but the simplest problems. No general system has yet been devised that can ensure that the global minimum is always found in such cases, but we indicate here two approaches to this problem.

The first approach uses dynamics to escape from potential-energy minima. Molecular-mechanics simulations allow each atom to possess momentum as well as position and integrate the equations of motion, conserving the total energy. By progressively removing energy from the simulation by scaling down the momentum vectors some potential-energy minimum may be found. Conversely, a minimization of potential energy which has led to a minimum thought not to be the global minimum may be continued by introducing atomic momenta sufficient to overcome potential-

energy barriers between minima, and subsequently attenuate the momenta again. In this way a number of minima may be found (Levitt & Warshel, 1975). It is equivalent to initializing a potential-energy minimization from a number of different conformations but it has the property that the minima so found are separated by energy barriers for which an upper limit is known so that the possibility exists of exploring transition pathways.

A second approach (Purisima & Scheraga, 1986) is relatively new. If the objective function to be minimized can be expressed in terms of interatomic distances, and if each atom is given coordinates in a space of $n-1$ dimensions for n atoms, then a starting structure may be postulated for which the interatomic distances *all* take their ideal values and the objective function is then necessarily at an absolute minimum. This multidimensional structure is then projected into a space of fewer dimensions, within which it is again optimized with respect to the objective function. The dimensionality of the model is thus progressively reduced until a three-dimensional model is attained at a low energy. This means that the minimum so attained in three dimensions is approached *from beneath*, having previously possessed a lower value in a higher-dimensional space. This, in itself, does not guarantee that the three-dimensional minimum-energy structure so found is at the global minimum, but it is not affected by energy barriers between minima in the same way, and it does appear to reach very low minima, and frequently the global one. Because it is formulated entirely in terms of interatomic distances it offers great promise for modelling molecules on the basis of data from nuclear magnetic resonance.

3.3.3. Implementations

In this section the salient characteristics of a number of systems are described. Regrettably, it cannot claim to be a complete guide to all existing systems, but it probably describes a fairly representative sample. Some of these systems have arisen in academia and these are freely described. Some have arisen in or been adopted by companies which now market them, and these are described by reference to the original publications. Other marketed systems for which originators' published descriptions have not been found are not described. Yet other systems have been developed, for example, by companies within the chemical and pharmaceutical industries for their own use, and these have generally not been described in what follows since it is assumed that they are not generally available, even where published descriptions exist.

Software concerned especially with molecular dynamics has not been included unless it also provides static modelling capability, since this is a rapidly growing field and it has been considered to be beyond the intended scope of this chapter. Systems for which outline descriptions have already been given (Levitt & Lifson, 1969; Levitt, 1974; Diamond, 1966; Warme *et al.*, 1972; Dodson *et al.*, 1976; Hermans & McQueen, 1974) are not discussed further.

For some of the earliest work Levinthal (1966) still makes interesting reading and Feldmann (1976) is still an excellent review of the technical issues involved. The issues have not changed, the algorithms there described are still valuable, only the manner of their implementation has moved on as hardware has developed. A further review of the computer generation of illustrations has been given by Johnson (1980). Excellent bibliographies relating to these sections have been given by Morffew (1983, 1984), which together contain over 250 references including their titles.

The following material is divided into three sections. The first is concerned primarily with display rather than modelling though some of these systems can modify a model, the second is concerned with molecular modelling with reference to electron density and can develop a model *ab initio*, and the third is concerned with modelling with reference to other criteria.

Where software names are known to be acronyms constructed from initial letters, or where the original authors have used capitals, the names are capitalized here. Otherwise names are lower case with an initial capital.

While it is recognised that many of the systems here described are now of mainly historical interest, most have been retained for the second edition, some have been updated and some new paragraphs have been added.

3.3.3.1. *Systems for the display and modification of retrieved data*

One of the earliest systems designed for information retrieval and display was that described by Meyer (1970, 1971) which used television raster technology and enabled the contents of the Brookhaven Protein Data Bank (Meyer, 1974; Bernstein *et al.*, 1977) to be studied visually by remote users. It also enabled a rigid two-ring molecule to be solved from packing considerations alone (Hass *et al.*, 1975; Willoughby *et al.*, 1974). Frames for display were written digitally on a disk and the display rate was synchronized to the disk rotation. With the reduction in the cost of core storage, contemporary systems use large frame buffer memories thus avoiding synchronization problems and permitting much richer detail than was possible in 1970. A majority of the systems in this section use raster techniques which preclude real-time rotation except for relatively simple drawings, though *GRAMPS* is an exception (O'Donnell & Olson, 1981; Olson, 1982) (Section 3.3.3.1.4).

3.3.3.1.1. *ORTEP*

This program, the Oak Ridge Thermal Ellipsoid Program, due to Johnson (1970, 1976) was developed originally for the preparation of line drawings on paper though versions have since been developed to suit raster devices with interactive capability.

The program draws molecules in correct perspective with each atom represented by an ellipsoid which is the equi-probability surface for the atomic centre, as determined by anisotropic temperature factor refinement, the principal axes of which are displayed. Bonds are represented by cylindrical rods connecting the atoms which in the drawing are tapered by the perspective.

In line-drawing versions the problem of hidden-line suppression is solved analytically, whereas the later versions for raster devices draw the furthest elements of the picture first and either overwrite these with nearer features of the scene if area painting is being done or use the nearer features as erase templates if line drawings are being made.

3.3.3.1.2. *Feldmann's system*

R. J. Feldmann and co-workers (Feldmann, 1983) at the National Institutes of Health, Bethesda, Maryland, USA, were among the first to develop a suite of programs to display molecular structure using colour raster-graphics techniques. Their system draws with coloured shaded spheres, usually with one sphere to represent each atom, but alternatively the spheres may represent larger moieties like amino acids or whole proteins if lower-resolution representations are required. These workers have made very effective use of colour. Conventionally, oxygens have been modelled in red, but this system allows charged oxygens to be red and uncharged ones to be pink, with a similar treatment in blue for charged and uncharged nitrogens. By such means they have been able to give immediacy to the hydrophobic and electrostatic properties of molecular surfaces, and have used these characteristics effectively in studies of the binding possibilities of benzamidine derivatives to trypsin (Feldmann *et al.*, 1978).