

3. DUAL BASES IN CRYSTALLOGRAPHIC COMPUTING

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 (Purísima & 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).

3.3. MOLECULAR MODELLING AND GRAPHICS

The algorithm developed by Porter (1978) for shading spheres to be darkened near their peripheries also computes the proper appearance of the line of intersection of two spheres wherever interpenetration occurs, in contrast to some simpler systems which draw a complete disc for whichever sphere is forward of the other. Provided that all opaque spheres are drawn first, the system is also capable of representing transparent spheres by darkening the colour of the existing background inside, and especially near the edge of, discs representing transparent foreground spheres.

Other systems that produce space-filling pictures of a similar general character have been produced by Motherwell (1978), by Sundaram & Radhakrishnan (1979) and by Lesk (next section).

3.3.3.1.3. *Lesk & Hardman software*

The complexity of macromolecules is a formidable obstacle to perceiving the basic features of their construction and the stylized drawings produced by this software following the artistry of Richardson (1977, 1981, 1985) enables the internal organization of such molecules to be appreciated readily. The software is capable of mixing several styles of representation, among them the Richardson style of cylinders for α -helices, arrows for β -strands and ribbons for less-organized regions, or the creased-ribbon technique for the whole chain, or a ball-and-stick representation of atoms and bonds, or space-filling spheres. All these styles are available simultaneously in a single picture with depth cueing, colour and shading, and hidden-feature suppression as appropriate. It is also able to show a stylized drawing of a complete molecule together with a magnified part of it in a more detailed style. See Lesk & Hardman (1982, 1985).

3.3.3.1.4. *GRAMPS*

This system, due to O'Donnell & Olson (O'Donnell & Olson, 1981; Olson, 1982) provides a high-level graphics language and its associated interpretive software. It provides a general means of defining objects, drawable by line drawings, in such a way that these may be logically connected in groups or trees using a simple command language. Such a system may, for example, define a subunit protein of an icosahedral virus and define icosahedral symmetry, in such a way that modification of one subunit is expressed simultaneously in all subunits whilst preserving the symmetry, and simultaneously allowing the entire virus particle to be rotated. Such logical and functional relationships are established by the user through the medium of the *GRAMPS* language at run time, and a great diversity of such relationships may be created. The system is thus not limited to any particular type of structure, such as linear polymers, and has proved extremely effective as a means of providing animation for the production of cine film depicting viral and other structures. *GRAMPS* runs on all Silicon Graphics workstations under IRIX 4.0 or above.

3.3.3.1.5. *Takenaka & Sasada's system*

Takenaka & Sasada (1980) have described a system for the manipulation and display of molecular structures, including packing environments in the crystal, using a minicomputer loosely coupled to a mainframe. Their system is also capable of model building by the addition of groups of one or more atoms with a facility for monitoring interaction distances while doing so.

3.3.3.1.6. *MIDAS*

This system, due to Langridge and co-workers (Gallo *et al.*, 1983; Ferrin *et al.*, 1984) is primarily concerned with the display of existing structures rather than with the establishment of new ones, but it may modify such structures by bond rotations under manual control. It is of particular value in the study of molecular

interactions since two or more molecules may be manipulated simultaneously and independently. Visual docking of molecules is greatly facilitated by the display of van der Waals surfaces, which may be computed in real time so that the turning of a bond in the underlying structure does not tear the surface (Bash *et al.*, 1983).

3.3.3.1.7. *Insight*

This system, originally due to Dayringer *et al.* (1986), has a functionality similar to *MIDAS*. It has been replaced by *Insight II* (current version 2.3.5). It appears to be well suited to the study of intermolecular relationships in docking and in structural comparisons, and it is able to make modifications to structures. Objects for display may be molecular or non-molecular, the former having an atomic substructure and the latter consisting of a vector list which may not be subdivided into referable components. Map fitting with the current version has been reported.

3.3.3.1.8. *PLUTO*

PLUTO was developed by Motherwell (1978) at the Cambridge Crystallographic Data Centre (CCDC) for the display of molecular structures and crystal-packing diagrams, including an option for space-filling model style with shadowing. The emphasis was on a free format command and data structure, and the ability to produce ball-and-spoke drawings with line shadowing suitable for reproduction in journal publication. Many variant versions have been produced, with essentially the 1978 functionality, its popularity deriving from its ease of use and the provision of default options for establishing connectivity using standard bonding radii. It was distributed as part of the CCDC software associated with the Cambridge Structural Database, with an interface for reading entries from the database.

In 1993 Motherwell and others at the CCDC added an interactive menu and introduced colour and PostScript output. New features were introduced to allow interactive examination of intermolecular contacts, particularly hydrogen-bonded networks, and sections through packing diagrams (Cambridge Structural Database, 1994).

3.3.3.1.9. *MDKINO*

This system, due to Swanson *et al.* (1989), provides for the extraction and visualization of selected regions from molecular-dynamics simulations. It permits stereo viewing, interactive geometric interrogation and both forwards and backwards display of motion.

3.3.3.2. *Molecular-modelling systems based on electron density*

Systems described in this section require real-time rotation of complicated transparent scenes and all used vector-graphics technology in their original implementations for that reason, though many are now available for raster machines. In every case the graphics are the means of communication between the user and software possessing high functionality, capable of building a representation of a molecule *ab initio* and to alter it, change its shape and position it optimally in relation to an electron-density map, with due attention being paid to stereochemical considerations, by one or more of several approaches.

3.3.3.2.1. *CHEMGRAF*

Katz & Levinthal (1972) have developed a powerful modelling and display system for macromolecules known as *CHEMGRAF*. This system permits the definition of many atom types which includes bonding specifications, so that, for example, four types of carbon atom are included in the basic list and others may be added.