3.3.3.2.10. *O*

Jones *et al.* (1991) have developed a modelling system for proteins with a radically different approach to any of the foregoing, in that they begin by reducing the available electron-density map to a skeletal representation (Greer, 1974; Williams, 1982) which consists of a line running through the density close to its maximal values, this being the basis of a chain trace. Provisional α-carbon positions are also estimated at this stage. A database of known structures is then scanned for pentapeptides which may be superimposed on five successive positions in the chain trace, the best fit so found being taken to provide coordinates for the three central residues of the developing model. The process advances by three residues at each step, the first and last residues of the pentapeptide being used only to ensure that the central residues are built in a manner compatible with what precedes and follows.

The process ensures that conformations so built are free from improbable conformations, and the whole forms an adequate starting structure for molecular-dynamics procedures, even though some imperfect geometry is to be expected where each tripeptide joins the next.

3.3.3.3. *Molecular-modelling systems based on other criteria*

Systems described within this section mostly have some form of energy minimization as their objective but some are purely geometrical. The optimization of molecules through empirical force fields has been reviewed by Allinger (1976), Burkert & Allinger (1982) and Boyd & Lipkowitz (1982). Some of these systems are in the academic domain, others are commercial. Most have capabilities exceeding the features referred to here and, of necessity, the list cannot be complete. No attempt at comparative evaluations is attempted or implied.

3.3.3.3.1. *Molbuild, Rings, PRXBLD and MM2/MMP2*

Liljefors (1983) has described a system for constructing representations of organic molecules. The system develops the molecule with plausible geometry and satisfied valencies at all stages of the development with explicit recognition of lone pairs and the various possible hybridization states. Growth is generally by substitution in which a substituent and the atom it is to replace are both nominated from the screen. The bond which is reconstructed in a substitution is generally a single bond. Double and triple bonds are introduced by the substitution of moieties containing them. Atom types may be changed after incorporation in the growing molecule, so that although the menu of substituents includes —CH_3 but not —NH_2 the latter may be obtained by incorporating —CH_3, then changing C to N and one of the hydrogens to a lone pair. Facilities are also provided for cyclization and acyclization.

van der Lieth *et al.* (1984) have described an extension to this that is specialized to the construction of fused-ring systems. It permits the joining of rings by fusion of a bond, in which two adjacent atoms in one ring are superposed on two in another. It also permits the construction of spiro links in which one atom is common to two rings, or the construction of bridges, or the polymerization of ring systems to form, for example, oligosaccharides. Again the satisfaction of valencies is maintained during building and the geometry of the result is governed by superposition of relevant atoms in the moieties involved.

PRXBLD is a molecular model-building program which accepts two-dimensional molecular drawings in a manner similar to *Script* (Section 3.3.3.3.2) and constructs approximate three-dimensional coordinates from these. It is the model-building component of *SECS* (Simulation and Evaluation of Chemical Synthesis) (Wipke *et al.*, 1977; Wipke & Dyott, 1974; Wipke, 1974). See also Anderson (1984).

All three of these programs produce output which is acceptable as input to *MM*2(82)/*MMP*2 which are developments of Allinger's geometrical optimization based on molecular mechanics (Allinger, 1976).

3.3.3.3.2. *Script*

This system, described by Cohen *et al.* (1981), is specialized for fused-ring systems, especially steroids, but is not limited to these classes. The system allows the user to draw on the screen (with a light pen or equivalent) a two-dimensional representation of a molecule using single lines for single bonds, double lines for double bonds, and wedges to indicate out-of-plane substituents. The software can then enumerate the possible distinct conformers, each of which is expected to be near an energy minimum on the conformational potential surface. Each conformer may then be annealed to reach an energy minimum using an energy estimate based on bond lengths, bond angles, torsion angles and van der Waals, electrostatic and hydrogen-bonding terms. An example is given of the identification of an unusual conformer as the most stable one from twelve possibilities for a four-ring system.

The program is a development of similar work by Cohen (1971) in which the molecule was defined in terms of a tree structure and an optimizer based on search techniques rather than gradient vectors was used. The method included van der Waals terms and hence estimated energy differences between stereoisomers in condensed ring systems arising from steric hindrance.

3.3.3.3.3. *CHARMM*

This system, due to Brooks *et al.* (1983), is primarily concerned with molecular dynamics but it includes the capability of model-building proteins and nucleic acids from sequence information and values of internal coordinates (bond lengths, bond angles and dihedral angles). The resulting structure (or a given structure) may then be optimized by minimizing an empirical energy function which may include electrostatic and hydrogen-bonding terms as well as the usual van der Waals energy and a Hookean treatment of the covalent skeleton. Hydrogen atoms need not be handled explicitly, groups such as —CH_2— being treated as single pseudo atoms, and this may be advisable for large structures. For small or medium proteins hydrogens may be treated explicitly and their initial positions may be determined by *CHARMM* if they are not otherwise known.

3.3.3.3.4. *Commercial systems*

A number of very powerful molecular-modelling systems are now available commercially and we mention a few of these here. Typically, each consists of a suite of programs sharing a common data structure so that the components of a system may be acquired selectively.

The *Chem-X* system, from Chemical Design Ltd, enables models to be developed from sketch-pad input, provides for their geometrical optimization and interfaces the result to *Gaussian*80 for quantum-mechanical calculations.

MACCS, from Molecular Design Ltd, and related software (Allinger, 1976; Wipke *et al.*, 1977; Potenzone *et al.*, 1977) has similar features and also has extensive database-maintenance facilities including data on chemical reactions.

Sybyl, from Tripos Associates (van Opdenbosch *et al.*, 1985), also builds from sketches with a standard fragment library, and provides interfaces to quantum-mechanical routines, to various databases and to *MACCS*.

Insight II (Section 3.3.3.1.7) is available from Biosym and *GRAMPS* (Section 3.3.3.1.4) is available from T. J. O'Donnell Associates.

references