MANOSK: A graphics program for analyzing and modeling molecular structure and functions

J. Cherfils, M. C. Vaney,* I. Morize,† E. Surcouf,† N. Colloc'h,* and J. P. Mornon*

Laboratoire de Biologie Physicochimique, Bâtiment 433, Université de Paris-Sud, 91405 Orsay, France
Laboratoire de Minéralogie-Cristallographie, Universités de Paris VI et VII, CNRS UA09, Tour 16, 4 place Jussieu, 75252 Paris Cedex 05, France*
Rhône-Poulenc Santé, Centre de Recherche de Vitry, 13 quai Jules Guesde, 94407 Vitry/Seine Cedex, France†

A program written for the Evans and Sutherland PS300 graphic displays is described in this paper. Called MANOSK, it provides a flexible and powerful environment for displaying, manipulating and analyzing small molecules and macromolecules from atomic coordinates. Translations and rotations of up to four independent molecules are available from the dials, and screen-relative orientations of the molecules are used in all geometrical and energetical calculations. A gradual progression of functions complexity makes the program easy to use for occasional works and efficient for more sophisticated studies.

Keywords: interactive computer graphics, macromolecules, modeling

Received 15 March 1988
Accepted 12 April 1988

Molecular graphics has become a familiar tool in the study of biological molecules. The molecular biologist needs to be able to get started with the program with minimal training and needs to have various functions available for subsequent studies. Thus, a number of basic functions must be easily accessible using standard atomic files, and functionality should not be too dependent on very specific applications, but rather remain open to a wide range of subjects. This point of view applies to the molecular types as well as to the properties that can be studied or the operations to be performed on the structures, such as conformational analysis, docking between molecules, and study of a crystal packing or display of a molecular dynamics simulation.

We have been developing a graphics package called MANOSK[1] with a multidisciplinary approach including various aspects of graphic modelization. The software was devised to manipulate several independent molecular models while taking into account their respective spatial positions for each computed property. Although all types of molecules — proteins, nucleic acids, and small molecules — are treated the same way from the user's point of view, the program will treat them in a specific manner when necessary.

MANOSK's functions fall into several categories: display, spatial manipulations, comparisons, structural modifications, analysis of physical and chemical properties, coordinates and file maintenance (Table 1a). We describe below the main features of the program and list some applications that demonstrate its potential.

GENERAL FEATURES

MANOSK is written for the Evans and Sutherland PS300 graphic systems. The host program contains about 27 000 FORTRAN instructions. It can be adapted to any machine that supports the graphic routines distributed by the constructor. The user interacts with the program via a set of menus that contain graphic and functional items. Graphic items are used to turn on and off the predefined objects that constitute a so-called molecule (Figure 1). These items are defined in a rather directive way by the program, but in return they relieve the user of the task of naming and extensively defining what he or she is interested in. A "molecule" contains molecular objects, including backbones and surfaces, but also nonmolecular objects, such as labels, axes or physical properties.

Molecules can be translated and rotated either individually in a referential of their own, or together in the referential of the screen. In either case, the rotation axes remain parallel to the canonical axis of the screen. The switch between the different referentials is performed by function keys that assign dials to the manipulation of one or all molecules.

Several other functions are controlled by function keys — stereographic or orthogonal views, automatic movements, display of information, menus and axis on the screen, for instance. These functions are distributed over

Table 1. (a) List of main commands available in MANOSK. (b) Types of atomic coordinates files

(a) Graphics and analysis		
	MACRO	Display of macromolecules
	MICRO	Display of small molecules
	ENVEL	Display of van der Waals surfaces
	CALPHA	Alpha carbon chains
	RIBBON	Ribbon like alpha carbon chains
	SYM	Molecular symmetries
	COLOR	Coloring options
	PART	Selects and displays subsets of the molecule
	NAME	Labels
	SNDOBJ	Displays a user-defined object
	ORIG	Changes rotations origins
	CENTER	Centers the picture
	ANIM	Animation of a user-defined set of pictures
	MOVE	Animation of molecular vibrations
	VALENCE	Valence angles
	DIHED	Dihedral angles
	DIST	Intra-Inter molecular distances
	UPDATE	Interactive distance monitoring
	PHIPSI	PHI and PSI angles
	HBOND	Intra-Inter molecular hydrogen bonds
	ENER	Nonbonded energies
	SUBST	Amino-acids substitutions
	TORS	Modification of torsional angles
	CONNECT	Creates a bond
	BREAK	Breaks a bond
	HGEN	Generates theoretical hydrogen positions
	DNA	Generates DNA coordinates
	RMS	Root mean square differences
Interfaces with external programs		
	SURF	Solvent-accessible surface (3)
	SIW	Van der Waals signature (4)
	ELEC	Electrostatic potentials on surfaces (5)
	SPLINE	Smoothed surfaces (6)
	FIT	Superimposition of molecules (7)
	SKELETON	BONES density maps (11)
	WBACK	Creates PLOT files from the screen
Miscellaneous		
	LOAD	Reads coordinates files
	SAVE	Writes coordinates files
	COPY	Internal duplication of fragments
	CREATE	Stores a fragment under a user-defined name
	PARAM	Lists and modifies molecular parameters
	INFO	Information about current coordinates sets
	HELP	On-line extended help
	TUTORIAL	Getting started
	@file	Executes a command file
	HISTORY	Copies output to a file
	TEXT	Puts text on the screen
	VMS	Creates a VMS sub-process (VAX only)
(b) File type		Origin
ASCII		Protein Data Bank (2)
		PROTIN (18)
		Diamond (19)
		Charmm (8)
		CHEM-X (20)
Binary		MANOSK
		BRUGEL(10)

menus that can be interchanged by using the numeric keypad.

There are two complementary ways to interact with the program. The first one uses the graphic tablet, via a mouse or a graphic stylus, for picking atoms or function items in menus. Alternatively, the user can type in a function name or any nonambiguous abbreviation of it, or a list of atoms or residue names (see the examples in Table 2). The switch between these two modes is totally transparent to the user and offers a convenient way to choose the best way to interact with the program. A journal file, which contains every instruction and parameter, is created after each run of the program. It can be used to generate command files containing a repetitive sequence of instructions. Both an extended HELP utility and the VAX/VMS command mode are accessible at any time.

COORDINATES

The program can manipulate up to four data files at the same time. Several types of formatted and unformatted files are allowed (Table 1b). Standard coordinates for MANOSK are Cartesian coordinates in Ångströms, but fractional coordinates can be converted by giving the crystal cell parameters. The program treats internal items in a hierarchical way starting from atoms, residues (amino acids, nucleic acids, ligands, molecules, and so on), chains and molecules. The same hierarchy is used for the graphic representation and allows very quick and natural treatment and display of any subset of interest.

The connectivity of residues that have standard Brookhaven Protein Data Bank[2] residue and atom names is computed from patterns described in data libraries. Because misspelled names and nonstandard structures are frequent (unknown atomic or residual names, duplication of a residue, and so on), a set of

Table 2. Examples of atomic selections. (a) Selections of atoms and residues using a coded list. Fields separated by —, stand for the data set number, the atom names and/or residues numbers, and the chain number, which is optional. Residue numbers can also be selected according to residue names by using the TYP keyword. (b) Use of the CREATE function to select, name and store a list of residues or atoms. The item, which is called here Helix1, can be retreived by placing = before its name

(a)	1__1,15 20, 25__1	Data set #1, residues 1 to 15 and 20 to 25 in chain 1
	1__Typ ARG	Data set #1, all arginines
	1__CA__235,245	Data set #1, CA α atoms for residues 235 to 245
	2__ZN	Data set #2, all ZN atoms
(b)	CREATE	Call to the function
	Helix1	Name of the item
	1__120,130	Data set #1, residues 120 to 130
	=Helix1	Retrieval

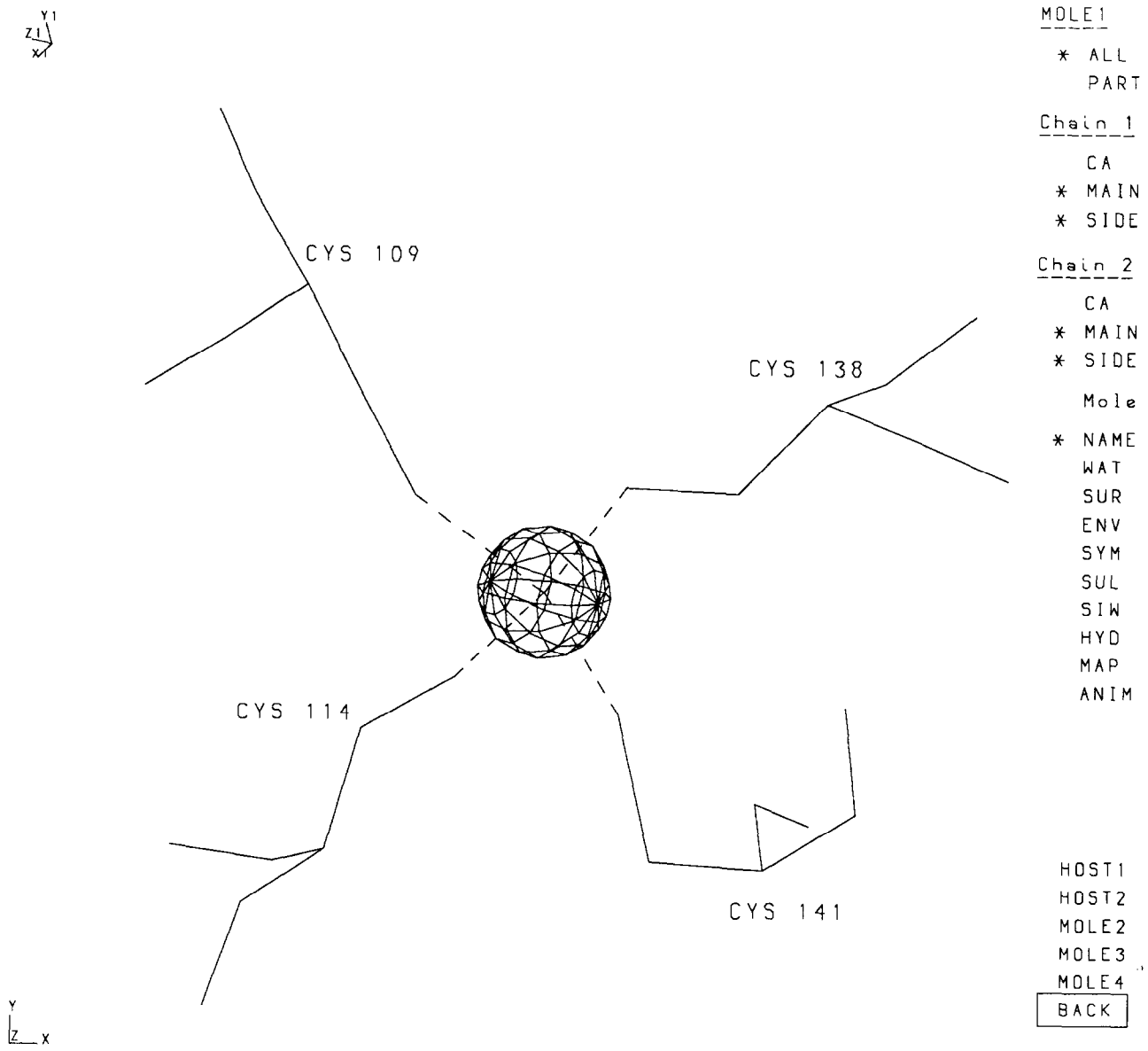

Figure 1. Display of a zinc atom connected to four cysteines in a regulatory chain of aspartate transcarbamylase.[21] *A graphic menu is displayed on the right of the picture. The stars correspond to the parts of the molecule that are currently displayed. The bottom items give access to other menus. The screen axes are on the left bottom, and the current orientation of the molecule relative to this referential is on the left top*

careful tests is applied to the data being read, and connectivity is computed from interatomic distances when necessary. Connectivity between consecutive residues is also checked. When nonstandard structures are found, the program lets the user define interactively new atomic types and parameters and compute connectivity entirely according to interatomic distances criteria. This treatment is efficient and offers easy recovery from errors.

Three different coordinates systems can be saved to disk files: coordinates in the same spatial orientation as the original file, coordinates displayed on the screen, and coordinates relative to the first data set, the latter being mainly used for docking a molecule to a reference molecule. Internal duplication of coordinates can be performed with a COPY command.

GRAPHIC OPTIONS

Various graphic styles are available, depending on whether molecules or properties are visualized and whether detailed or schematic display is required. Proteins and nucleic acids are displayed in standard wire models subdivided into residues or atoms. For other molecules, the wire representation is subdivided into atoms. This subdivision allows for individual treatment of either residues or atoms — for example, interactive coloration of a specific region of the molecule. Monoatomic ligands, such as metal ions, are displayed as balls (Figure 1), while water molecules are displayed as crosses. Alpha-carbons chains (or phosphate chains) are displayed as broken lines, smoothed lines or ribbons (Figure 2).

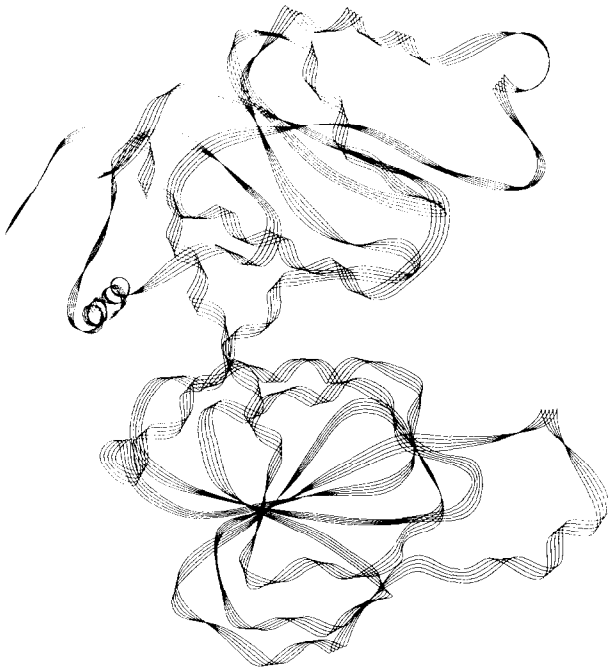

Figure 2. A ribbonlike representation of a catalytic chain of aspartate transcarbamylase[21]

Extensive computing and treatment of surfaces are available. Van der Waals surfaces are displayed interactively, with adjustable dot density. The Connolly's solvent-accessible surface program[3] is interfaced to MANOSK; MANOSK prepares data files for batch computation and displays the surface. Connolly's surface is used as an input to the computation of electrostatic potentials,[4] either on the surface or at some distance from it. A numerical value is associated with each surface dot. Display options let the user color the surface dots according to their associated numerical values and restrict the display to a given range of these values. This flexible display, together with an adequate coloration, contributes information to the analysis of macromolecular surfaces (Color Plate 1). Van der Waals signatures[5] could be computed and displayed by MANOSK. A smoothed overall shape of the molecule can also be computed and quantitatively analyzed using spline functions,[6] which emphasize significant features of the surface, especially for macromolecules (Color Plate 2).

COMPARISON OF ANIMATION

The MacLachlan's fitting algorithm,[7] which superimposes a sequence of points on a sequence of guide points by a root mean square procedure, is interfaced to MANOSK to provide flexible selection of the set of atoms to compare. Transformed coordinates can thus be used as a new internal data set for subsequent work. In a number of cases, the ability to alternate the display of structures to be compared enhances the perception of their differences and similarities. MANOSK provides various animation options that can be used for this purpose as well as to display a sequence of simulated conformations. The pictures can be described by the displacements of the coordinates referring to an original data set, together with an amplitude and a phase step for the animation. Another way to define an animation is to select a set of already displayed residues and/or Cα chains and copy them into different pictures, which will mainly be used together with the superimposition option. A third alternative is the linear interpolation of two different conformations, which can provide a rough insight into atomic trajectories. Films can be made easily from these animations and with automatic translations, rotations and scaling controlled by keys. Results of CHARMM's[8] molecular dynamics runs are also handled by these options.

STRUCTURAL MANIPULATIONS

Rotation of up to eight consecutive bonds is possible for any type of molecule, including models of proteins containing only α-carbons. Bonds are defined by four consecutive atoms that can either be picked on the screen or entered on the keyboard, or any combination of both. Torsional angles are modified by dials or by entering angle values on the keyboard. Van der Waals surfaces are attached to moving atoms.

Interactive substitutions of the 20 standard amino acids can be performed, either by picking any atom of the residue on the screen or by typing its name. The new side chain is fitted to the previous side chain. A graphic copy of the old amino acid structure can be kept for comparison. Insertions and deletions are performed manually by graphically adjusting peptide fragments that can derive from a homologous protein.

MANOSK lets users generate standard A and B forms of right-handed DNA.[9] When needed, hydrogen atoms can be added to a structure by computing their theoretical positions.

Molecular and crystallographic symmetries can be defined either by director cosines of the axes or by entering explicit matrices, which can also be read from free-format data files. Internal coordinates or graphic copies of the original structure can be created (Figure 3).

CALCULATIONS

An important feature of MANOSK is that intermolecular geometrical and energetical calculations always refer to the relative screen orientation of the molecules. Interactive distance monitoring is available for a set of preselected pairs of atoms — for example, for all distances of an atom within a sphere. Hydrogen bonds are checked against criteria depending on whether hydrogen atoms are available or not.

Nonbonded energy, including electrostatics and Van der Waals interactions, can be computed. Van der Waals parameters and partial charges for standard amino acids are stored in a library that derives from the parameter library of the BRUGEL package.[10] Van der Waals parameters can be attributed to unknown residues according to atomic types, which are unambiguously defined by the nature of atoms neighborhood, while the user must

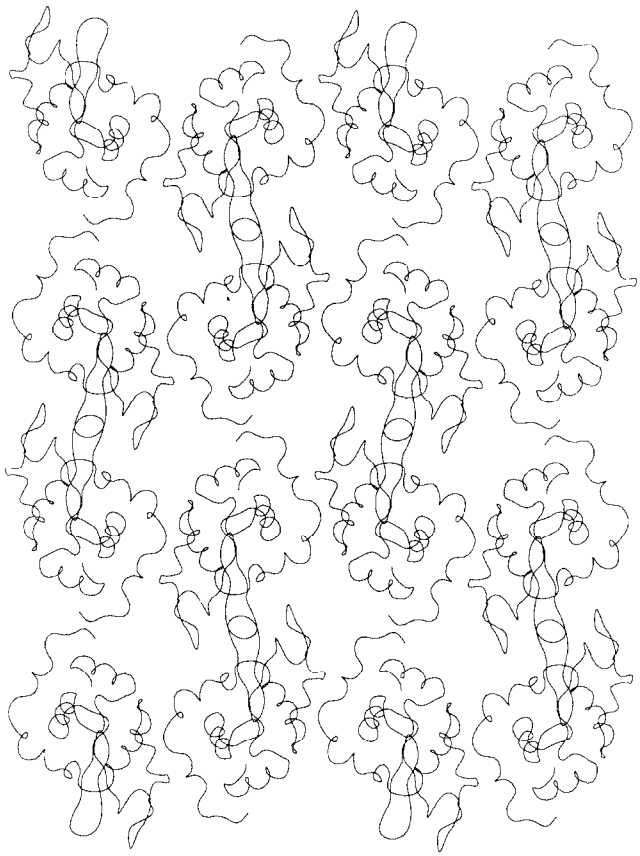

Figure 3. Orthorombic crystal packing of lysozyme[22] constructed from atomic coordinates of one molecule

define partial charges interactively. Undefined atomic types and parameters can be added interactively. Once the user has selected the site of interest and attributed the parameters, the energy computation can be called repeatedly. The lists of pairs are recomputed at every call, since movements between molecules can be quite important during docking manipulations. Bad contacts defined by a threshold value of the energy between two atoms are displayed. Nonbonded energy can also be computed when torsional angles are being modified.

INTERFACES

A number of functions proceed through interfaces to external programs, taking advantage of MANOSK's atomic selection facilities and multimolecular spatial manipulations. These interfaces are either included in the MANOSK task or submitted as subprocesses under VAX/VMS. Thus, data and command files for surface calculations can be entirely prepared and executed from MANOSK. Output files from the spline surface program[6] and from the electron density skeletonization BONES program[11] are only read and visualized.

A general interface has been developed for displaying results of user programs. Numeric values associated with coordinates must be stored in a file, which will be displayed as a graphic object whose name, coordinates range and value range to be visualized are chosen by the user. This feature requires little adaptation work and offers convenient graphic access to many user-specific applications.

An interface to the PS300 hard-copy facilities allows plots on Hewlett-Packard plotters and on Postscript Laserprinters.

SUMMARY AND PERSPECTIVES

MANOSK contains a wide range of analysis and manipulation facilities. Basic functions are accessible in a short time with the help of a detailed notice containing various examples and an in-line tutorial. A second level contains more sophisticated options. The field of applications can be illustrated by some examples: comparison of molecules, homology modeling of proteins,[12] substrate docking, preparation and analysis of molecular mechanics or dynamics calculations,[13] analysis of normal modes of vibration,[14] and crystal packing analysis. Work in progress includes an interface with the molecules building program CONCORD[15] and with the quantum mechanics software SIBFA.[16] Also under way is a connection to the MANSEK software for analysis and comparison of proteins sequences and structures using the hydrophobic cluster-analysis method.[17] A crystallography/chemical-physics program is also under development. Other new specialized functions and current improvements are scheduled to be available with the forthcoming releases and will be described later.

The MANOSK package is available for academics and industries through an ANVAR license on request to Dr. J. P. Mornon.

ACKNOWLEDGEMENTS

We would like to thank C. Gaboriaud for her help in a part of this work and Professor J. Janin for his advice in the design of the program.

REFERENCES

1 Vaney, M. C. *et al.* MANOSK, a graphic program designed for macromolecules modelling. *J. Mol. Graph.*, 1985, **3**, 123–124
2 Bernstein, F. C. *et al.* The protein data bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.*, 1977, **112**, 535–542
3 Connolly, M. L. Analytical molecular surface calculations. *J. Appl. Crystal.*, 1983, **16**, 548–558
4 Weiner, P. K. *et al.* Electrostatic potential molecular surface. *Proc. Natl. Acad. Sci.*, 1982, **79**, 3754–3758
5 Surcouf, E., and Mornon, J. P. Signature Van der Waals: outil d'étude des associations moléculaires. *C. R. Hebd. Acad. Sci. B*, 1982, **295**, 923–926
6 Colloc'h N., and Mornon, J. P. Representation and analysis of protein surfaces using spline functions. *J. Mol. Graph.*, 1987, **5**, 170
7 McLachlan, A. D. Rapid comparison of protein structures. *Acta Cryst A.* 1982, **6**, 871–873
8 Brooks, B. R. *et al.* CHARMM, a program for macromolecular minimization and dynamics calculations. *J. Comp. Chem.*, 1983, **2**, 187–217

9 Arnott, S., and Hukins, D. W. L. Optimized parameters for A DNA and B DNA. *Biochem. Biophys. Res. Comm.*, 1972, **47**, 1504–1509
10 Delhaise, P., Bardiaux, M., and Wodak, S. Interactive computer animation of macromolecules *J. Mol. Graph.*, 1984, **2**, 103–106
11 Greer, J. Computer skeletonization and automatic density map analysis. *Meth. Enzymol.*, 1985, **115**, 206–224
12 Vaney, M. C. Doctoral thesis, University of Paris, VI, 1986
13 Cherfils, J. *et al.* The PAR5 mutation and the allosteric mechanism of E. coli Aspartate Transcarbamylase. *EMBO J.*, 1987, **9**, 2843–2847
14 Vergoten, G. et al. Assigning the 625^{-1} and other frequency Raman bands of the Z1 form of DNA using a normal modes analysis and computer graphics. *J. Mol. Graph.*, 1986, **4**, 187–189
15 Rusinko III, A., *et al.*, CONCORD, University of Texas, Austin. Distributed by Evans and Sutherland.
16 Chen, K. X., Gresh, N., and Pullman, B. A. Theoretical study of anthracene and phenanthrene derivatives acting as A–T specific intercalators. *Nucl. Acid. Res.*, 1986, **14**, 9103–9115
17 Gaboriaud, C. *et al.* Hydrophobic cluster analysis: an efficient new way to compare and analyse amino acids sequences. *FEBS Letters*, 1987, **244**, 149–155
18 Hendrickson, W. A. Stereochemically restrained refinement of macromolecular structures. *Meth. Enzymol.* 1985, **115**, 252–270
19 Diamond, R. A. mathematical model building procedure for proteins. *Acta Crys.*, 1966, **21**, 253–266
20 CHEM–X. Developed and distributed by Chemical Design Ltd., Oxford, UK
21 Krause, K. L., Volz, K. W., and Lipscomb, W. N. 2.5 Å structure of aspartate carbamoyltransferase complexed with the bisubstrate analog N-(phosphonacetyl)-L-aspartate. *J. Mol. Biol.*, 1987, **193**, 527–543
22 Berthou, J. *et al.* An X-ray study of the physiological temperature form of hen egg-white lysozyme at 2 Å resolution. *Proc. of Roy. Soc.*, 1983, *B217*, 471–489