PCDRA: PC interactive molecular representation and modeling system

M. Afshar, M. Nanard and J. Sallantin

Centre de Recherche Informatique de Montpellier, Montpellier, France

J. Haiech

LCB, CNRS, Marseilles, France

PCDRA was designed to provide the average biologist with a user-friendly molecular display on a low-cost personal computer. The package is menu driven and is built so that a biologist, with little or no computing knowledge, finds it easy to use. The system gives a color representation with depth cueing of a protein whose atomic coordinates are stored as a PDB file. Moreover, the system presents several features similar to HYDRA[1] and therefore is a good introduction to molecular graphics, especially for beginners in protein modeling.

Keywords: personal computer, computer graphics, molecular modeling, protein engineering

INTRODUCTION

Since the advent of computerized molecular graphics, new techniques, algorithms and machines have flourished in biological laboratories. In the last two years, the machines have made a quantum leap[2] in calculation and display speed. Although the price of "super workstations" is going down, there still is a big gap between this price and that of personal computers. At the same time, PCs are becoming more and more powerful and inexpensive, and so very common in most biological laboratories.

It is clear that in most biological fields, the study (or at least the awareness) of the structure/activity relationship of proteins has become essential.[3] So now most researchers, as opposed to the few specialized researchers in the recent past, need to have rapid access to an easy-to-use molecular graphics program. Unfortunately it is not yet possible, even for the best equipped laboratories, to consider having a super workstation for each researcher.

Address reprint requests to Mr. Afshar at the Centre de Recherche Informatique de Montpellier, 860 route de St. Priest, 34090 Montpellier, France.
Received 6 June 1989; accepted 10 October 1989

The new advances in microcomputer technology and man–machine interface allow us to build an easy to use 3-D environment with more functionalities on the most widespread PC: IBM compatible PCs, which can be used for routine visualization of proteins.

In the past few years man–machine interface has also made a quantum leap. Command inputs with long and difficult syntaxes are things of the past. Our program, called PCDRA, tries to take advantage of these advances. The user does not need to remember any particular syntax. All operations are activated with the mouse by clicking on screen, a menu item, an icon, a button etc. in a Macintosh-style pop-up window. The cursor is designed to further help the user. Even the file access environment is fully graphical and menu driven. Thus the keyboard is rarely used, except for typing atom names or residue types.

The software has been designed to allow developers to add their own modules to the program easily. Another important feature of PCDRA is that is can be easily interfaced to any other program, locally on the PC, or on a remote VAX, using Digital's DecNet DOS.

In this paper, we describe the architecture of the PCDRA system, the implementation on PC/AT with an EGA/VGA card or PS/2 and the functions of this 3-D editor. Doing this we will compare the results with existing softwares implemented on PCs.[4–6]

HARDWARE CHOICE

The computer used is the new IBM PS/2 model 50. It has 1 megabyte (Mb) RAM, a 20 Mb Hard Disk, a 1.4 Mb floppy disk, a mouse, a serial and a parallel port and a VGA color graphics adapter. Two screen resolutions are available:

(1) 600 × 350 with 16 colors out of a palette of 256 K and double buffering;
(2) 600 × 480 with 16 colors out of 256 K, but no double

buffering and available only on a VGA card. Its advantage is that spheres do not appear elliptical.

The serial port is used for the connection to the VAX via DECNET. The parallel port is connected to a laser or color jet printer.

The program source code is written in C (using the MicroSoft C Compiler). It relies on a very portable, MS Pascal written, object oriented, graphics toolbox;[7](the same toolbox is available on the Vax-Station). This architecture assures that the program is highly portable on most machines with any operating system that supports memory allocation and basic graphic primitives such as WRITE_DOT and DRAW_LINE; a version of PCDRA has been implemented on our Vax-Stations.

The program can run on any IBM PC compatible microcomputer with at least 512 K RAM, an EGA (with 256 K of graphics memory) or VGA graphics adapter, the appropriate color monitor and a serial mouse. A floating-point co-processor and 1 Megabyte on hard disk are strongly recommended, but not essential.

SUMMARY

PCDRA is capable of representing, rotating and translating one or two molecules, in 3-D. It also has the following basic features:

- reading BROOKHAVEN standard files, locally, or from a remote DecNet node (2 files, up to 1400 atoms each);
- atom selection and color attribution;
- depth cueing;
- half-bond coloring;
- CPK representation;
- geometry calculations;
- structure modification;
- superposition of two structures using LSQKAB;[8-9]
- generation of output file in PDB format; or in any user definable format;
- hard copy of all or portion of screen to a laser printer or color printer;
- running another program or entering a DOS command from within PCDRA (DOS SHELL);
- all messages output by the system (warnings, menu labels, etc.) can be customized by the user by editing the resource file (PCDRA.RES).

FILE ACCESS ENVIRONMENT

The file access environment is fully graphical and menu driven. The program displays a list of files that are relevant to the application (by filtering the file extensions). For example, when the program prompts for an atom selection file name, only the files with *.SEL* extension (SELect) will be displayed as small folders (Figure 1). Up to 30 file names are displayed per page. If more than 30 relevant files are available in the listed directory, then file names are paged, and the user goes through them by pressing **F1**.

If the desired file does not have a standard extension (i.e., not the one expected by the program) the user can display all the current directory files by pressing the **BACKSPACE** button. In our previous example, if the user wanted to select a file with the *.ATM* extension, he first presses **BACKSPACE** and then clicks the appropriate folder.

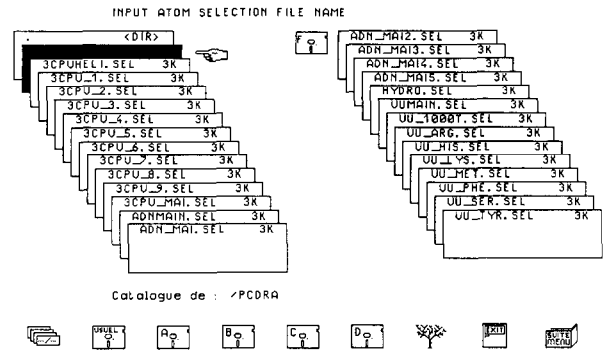

Figure 1. The fully mouse driven File Access Environment. The folders represent the available files. Clicking the ". .⟨DIR⟩" folder ascends the directory tree; and selecting a "xxx ⟨DIR⟩" folder displays the files in that sub-directory. All standard file operations are available through the displayed icons

It is possible to input a file from a directory, or a volume different than the current ones, by merely clicking the relevant icons displayed at the bottom of the screen. In the same way, the user can set the default volume and directory; rename or delete a file; delete or create a sub directory. The deletion and renaming operations must be confirmed by the user. Files can be also read from a remote DecNet node.

The user exits the file access environment by either:

- selecting a file by clicking it;
- typing the file name and pressing **CR**;
- clicking the EXIT pictogram;
- pressing **ESC**.

DATA INPUT FORMAT

At present, two file formats are available:

(1) standard Brookhaven PDB format;
(2) internal PCDRA binary format.

The program's modularity allows one to easily add new formats by adapting the existing ones, or by creating original procedures.

If the user chooses PDB format input, the system reads the PDB file, calculates the connectivity, assigns to each atom a reference to an internal radius table and optionally saves the coordinates and the calculated information in PCDRA binary format. In this case the file extension name is changed to *.PCD*.

The connectivity options are:

(1) CA joining;
(2) Distance search.

If the CA joining is selected, then each atom read is connected to the next. This is specially useful for backbone models.

If Distance search is chosen, the distances from each atom in the residue under consideration to all other atoms in the residue and in the next residue are compared against the

maximum allowed bonding distance (D). The bond is generated if the calculated distance is less than or equal to (D). The default for (D) is 1.8 Å. The system also recognizes disulphur bonds. Reading 1144 atoms from a PDB file takes 17 seconds, and 3 more seconds are needed to generate the connectivity.

If the user selects to input a PCDRA format file, all the data is loaded in less than 1 second into memory, no matter how big the structure is. It is then displayed. If the system has at least 640 K of processor memory a second data set could be loaded.

GENERAL SETUP

Once data is read, the last input structure is displayed as a simple stick diagram along with the main display menu on the right-hand side of the screen. The main chain is displayed in green, the main chain oxygens in red and the side chains in white.

Menus

There is one main menu: M_MENU. It is very much like those found in most other molecular graphics packages except that depending on which box is clicked, either an action is taken or a "pop-up" menu appears on the screen. This pop-up menu may be a simple, single column menu, or a menu with sub menus (Tree-Menu) (Figure 2).

For example, to save the selected atoms in PDB format, the user clicks **FILES** in M_MENU: a tree menu pops. Among the available functions, he selects **SAVE>** by rolling the cursor to the appropriate menu item. The message in the desired box is video inversed, and a ">" shows that this box leads to a sub-menu. To pop the sub-menu, the user rolls the cursor to the right, over the ">" sign and past the right-hand edge of the menu. A sub-menu appears, listing the many available output formats. In our example the user selects the **BROOKHAVEN>** box and here again the ">" sign shows that there is more to come. Rolling to the right pops the final menu, displaying two options for saving respectively all atoms read, or the selected atoms only. Clicking the second choice asks for a file name, through the file access environment, and writes on disk.

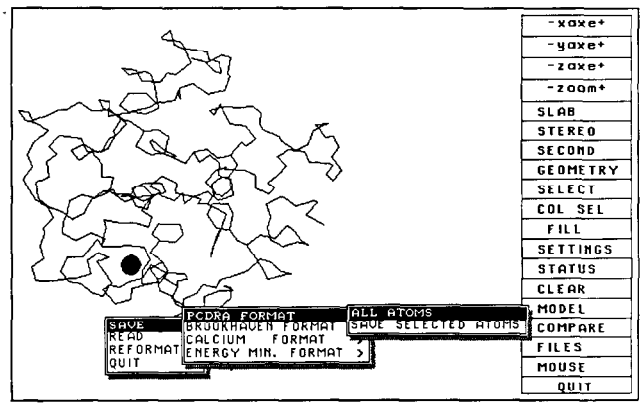

Figure 2. An example of a tree-menu used to access the file Input-Output module. This tree-menu is popped by clicking the right-hand side main menu. The system uses tree-menus as a navigation tool which leads to PCDRA commands. All irrelevant choices are hidden from the user, providing a simple command input

This example illustrates our user-friendly command input strategy. It is important to note that while exploring the menu, the user did not need to know any particular syntax, and that nothing was typed.

The menu items are self explanatory, allowing the novice user to easily perform what he intends; they also allow the experienced user to access the desired function quickly.

At any stage, the user can cancel a command by clicking outside the menus or by pressing **ESC**.

The cursor

When in the main menu (M_MENU) zone, the cursor is represented by a pencil. The rotations and zoom boxes have a special cursor: when in these special boxes, the cursor is represented by a vertical bar (rectangle) which is at its smallest length at the middle of the box. As the mouse is rolled either to the right or to the left, the cursor gradually grows taller. In the rotation boxes, the vertical bar represents the rotation angle: minimal at the center (2°) and maximal at the edges (90°). When the cursor is in the right half of the box, the angle is considered positive; negative if in the left half. In the zoom box, the cursor represents the scaling factor. The right half of the box corresponds to "zoom-in," and the left half to "zoom-out." If the user opens a pop-up menu, then the cursor becomes the common arrow if it is in a valid box, or a bird if it is not (i.e., outside the pop-up menu). In the display window, the cursor is an enlarged plus sign ("+").

The display window

When in the display window, clicking and holding the right mouse button displays at the bottom of the screen the identification of the nearest atom to the cursor position. This method was preferred to associating identification stickers to atoms and rotating them with the structure. This would have meant slower rotations and the standard sticker method does not seem to be necessary for most applications.

If the second data set is loaded, the user can swap between main and second by clicking **MAIN** which becomes **SECOND**, in M_MENU. The two data sets can only be viewed simultaneously in the comparison head (LSQ).

Clicking **STEREO** in M_MENU switches to stereo view display mode.

BASIC FUNCTIONS

There are 3 basic functions available:

(1) translation;
(2) rotation;
(3) zooming.

Translation

Clicking the left mouse button in the display window translates the structure so that the cursor position is moved to

the center. The method is simple and any point on the screen can become the display center. It has the disadvantage that the Z coordinates are not directly affected. Each translation is performed only in the screen plane. A remedy to this is to make a translation, followed by a 90° rotation around the X or Y axes and another translation.

Rotation

Rotations are possible about the three orthogonal axes. The rotation center is always the display window center. Its position regarding the structure can be changed using the translation procedures described above.

Each time the user clicks **−XROT+**, **−YROT+** or **−ZROT+** in M_MENU, the structure is rotated around the desired axis by the selected angle. Continuous rotation is performed by holding down the mouse button or activating the **SCROLL** button. The action is stopped by disactivating **SCROLL** or depressing the mouse button. Rotating 1000 atoms with half bond coloring and using depth cueing with 7 shades is performed in 1 second.

Zooming

Zooming in and out of the structure is performed by selecting the desired scaling that appear as the cursor is moved horizontally in the **−ZOOM+** box.

ATOM SELECTION AND COLOR ATTRIBUTION

Care has been taken to develop this section as much as possible for we think it is one of the most important features in any molecular graphics system, especially when working with as complex a structure as proteins. Our system is quite powerful and very flexible at the same time. No syntax needs to be learned, since all selection is done through menus. Both functions use the same pop-up menu (in fact a tree menu).

The functionalities offered here are very similar to those found on HYDRA.[1]

Atom selection

Atoms can be selected (or removed from a previous selection) by structural component, calculated attribute or structural feature.

Clicking **SELECT** in M_MENU pops the selection menu, and a window on the top left-hand corner informing whether the program is selecting (**INCLUDE**) or removing (**EXCLUDE**) atoms.

Going through the menus, the user can:

- select *residues in a range,* by typing the boundary residue numbers, or clicking them directly on screen;
- define a *sphere* on screen, and select the atoms or residues that fall in it;
- select by *atom name* or *residue type;*
- select by coordinate value. This is specially interesting when using the values in the "fourth" data field, namely the B-FACTOR field in the Brookhaven's PDB format. It is possible to replace the set of values in this field by another set using **APPEND FOUR** in M_MENU. The new set can be of any origin, and only the user needs to know its significance. This is an extremely flexible way to interface PCDRA with any other program that produces numeric values associated with the structure. Once the data set (X, Y, Z or FOUR) is selected, the user must choose one of the 6 standard logical operators (EQ,NE,LT,LE,GT,GE) and a value. The system finally ends up with the following expression: (X/Y/Z/FOUR) **op** value. All atoms satisfying this expression will be selected;
- combine any of these functions together (e.g., select all the Tyrosines between residue 2 and 53);
- save, restore and edit the selection file.

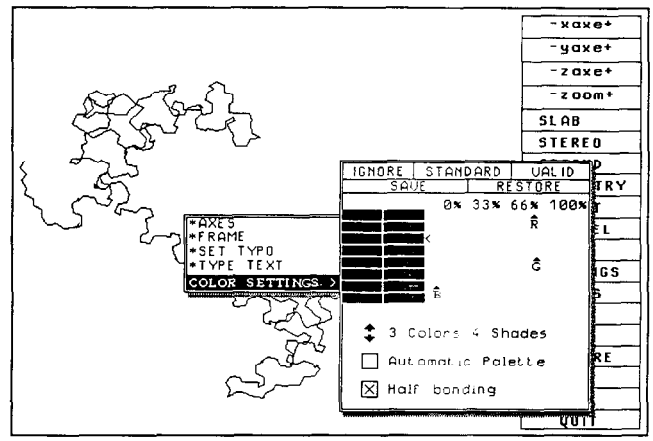

Figure 3. The color palette setting window. This is an example of a PCDRA dialog box, where the user sets the color registers either automatically, or individually, by using scroll-bars, push-buttons, etc.

Color selection

The color selection is performed using the same input procedure as for the atom selection. However, each command is ended by a supplementary color menu. The picked color is associated to the **COL SEL**ected atoms.

Owing to the limited number of available colors, we had to design a system to let the user choose a compromise between the number of different colors used (1 to 7), and the number of different shades for each color (1 to 7). If the shading level is greater than 1, then depth cueing will be performed for both stick and fill representation. If the shading is increased, less colors will be available.

Each color can be set individually from a set of 256,000 available colors, or by setting the number of desired colors and pressing the **AUTO PAL** button (Figure 3). The user then chooses the basic colors, and the shading is automatically generated. The color palette can be saved to and retrieved from disk. Using this same window, it is possible to set ON or OFF the **half-bond coloring.**

CPK REPRESENTATION

It is possible to combine space filling and simple stick representations. The CPK head has the following options:

(1) **MOD SEL**>: leads to the selection menu. Each selection ends with a menu of available representations

(**FILL** and **STICK**). Choosing **FILL** will select the atoms for the van der Waals figure drawing procedures.

(2) **SET RAY**>: has two sub-menus.

- **% of VWR** lets the user set the percentage of van der Waals or "four" (see below) to be used when generating the disks or spheres.
- **f(four)** will make the program use the values in the fourth field (see atom selection) as van der Waals rays (Color Plate 1).

(3) **QUICK FILL**>: will draw atom sections on top of each other starting from the back. This well known method is fast and gives adequate results. The shading level is also used. A value of 7 produces a result which is as good as other much more complicated methods (Color Plate 2). A value of 3 or 4 will still give a very good depth impression. The sub menus let the user choose the number of superimposed disks used for each atom and whether or not the contour of each atom should be drawn in the background color. Not drawing the atom contours is a very simple way to have a rough idea of the contour of the structure (or part of it). Quick-filling 1000 atoms takes 10 seconds without depth-cueing and 12 seconds with a shading level of 7.

(4) **SPHERE**>: uses an algorithm due to R. Hubbard *et al.*[10] The drawing is slow (1 minute and 16 seconds for 1000 atoms), and memory consuming, but the effect is quite impressive with high shading levels (Color Plate 3). One problem here is that if the scale gets too big, the program runs out of memory and will produce truncated spheres; a warning message will appear. In the sub-menus, the user has the option of showing the grid used by the program. It is then possible to see the model being built up by square blocks.

Combining the **SET LINE WIDTH,** the **SET RAY**> **% of VWR** and **QUICK FILL** commands allows a "stick and ball" representation.

MODELING

PCDRA offers facilities for partial rotation of the structure around any axis, defined by selecting a pair of atoms on screen. Once an atom pair is picked, the axis joining the two points starts to flash, and the **MODEL ROT** box becomes active. Clicking the desired angle partially rotates the structure. Distances, angles and dihedral angles can be monitored during the modeling procedure. All basic operations (rotation, translation, etc.) are still available while modeling and here also the **SCROLL** button will generate a continuous partial rotation around the selected axis, or a global rotation about the main axis.

GEOMETRY CALCULATIONS

It is possible to calculate distances, angles and dihedral angles. One of the three operations is selected and the atoms are picked on the screen. The selected atom identifications, and the calculated value are displayed at the bottom of the screen until a key or a mouse button is hit.

LSQ

Eleanor Dodson's version of Kabsh algorithm[8,9] has been implemented. The main data set is treated as the working structure (the one that is going to be rotated and translated after the fitting), and the second data set as the reference structure. To perform the fitting, the user is first prompted for the working residue range (i.e., fit the residues from number $n1$ to residue $n2$), and chain name. Then the reference structure is drawn and the user is prompted for the reference residue range (i.e., fit the working residues to the reference residues starting with residue number n) and chain name. All this information can be typed in, or picked directly on screen. The following matching options are offered:

(1) match CA only;
(2) match main chain only;
(3) match side chain only;
(4) match all atoms.

Once the matching is performed, the fit statistics appear in a pop-up window. After the user has acknowledged the results by clicking the window, the screen starts rapidly flashing between the two optimized structures (Color Plate 4).

FILE OUTPUT

Two output formats are available at present: PCDRA binary and Brookhaven. Care has been taken to allow the user to define easily his own file format. Using **INPUT DATA**> it is possible to read data, using the same procedure as given in the section DATA INPUT. The **OUTPUT DATA**> submenu allows the user to save all read atoms, or the ones selected by the atom selection procedure, in the desired format. Each input and output format procedure is contained in an independent module. It is therefore very easy to make PCDRA read in a user defined format, by custom tailoring the provided, C written, input module sample, compiling it and linking it to the PCDRA object files. A new entry will then be added to the resource file.

SCREEN HARD COPY

At any moment during PCDRA run, it is possible to output a screen hard copy to a printer by pressing **ALT** and **F1**. By setting the buttons in the popped window, the user selects the printer driver. If laser printing is chosen, the selected screen portion is transformed to monochrome. The screen copy can either be saved into a file for later use or sent directly to the printer.

CONCLUSION AND FUTURE DEVELOPMENTS

Table 1 summarizes our comparative study between 4 packages. Although it is difficult to assess improvements due to hardware and those due to software, this table gives a rough idea of the evolution of the molecular graphics software on PCs since 1983.

The PCDRA package has been tested in several laboratories and appears specifically useful in two domains:

Table 1. Comparative study of four packages. The study of these figures shows that better resolutions and more available RAM allows the study of much bigger structures on PCs; it is now possible to have a first step analysis of a protein on a low cost system. The increase of calculation speed, and a faster disk access help the interactivity and the user-friendliness

	Resolution	Maximum number of atoms	Available RAM	Rotation time per atom	Data preparation time/atom
PCDRA M. Afshar *et al.* (1989)	640 × 350	2 × 1400	640 K	1s/1000	18s/1000
Colour molecular graphics on a microcomputer R. E. Hubbard (1983)	300 × 150	830 (127 for fitting)	56 K	8s/830	10′/830
GRIMM K. Ogawa *et al.* (1984)	N.A.	300	N.A.	N.A.	N.A.
Computer graphic tool . . . for personal computers P. Cozzini *et al.* (1985)	640 × 225	N.A.	128 K	N.A.	30′/60

(1) for a skilled researcher who wants to use a simple display to communicate with a remote computer (VAX or other network) in order to set the computations and to visualize the preliminary results (later, when their access is possible, he will use the more powerful machines and graphics workstations);
(2) for educational purposes. Indeed, most biochemists are afraid of the complexity of molecular graphics packages, and of computers in general. Moreover, such packages and environments are not very common in average laboratories. Personal computers have invaded the biological laboratories, and have demystified computing. Therefore, PCDRA gives a simple introduction to the power of molecular graphics for the biochemist.

From our observations of the usage of PCDRA, we think that any such packages must develop in the following directions:

(1) inclusion of a communication module in order to download Brookhaven files, to comunicate with the mainframe in order to generate calculations (Connolly surface, energy minimizing . . .) and to exchange information with colleagues.
(2) implementation on the modeling head of a module to substitute an amino acid with another and to make a local energy minimization.

Finally, we have noted that our basic strategy to realize such a package (use of a multi-layer operating system, structured language programming: C and Pascal) could be generalized. We want to implement PCDRA on the SILICON GRAPHICS workstation to improve the portability of the software and to check the advantages to use an object-oriented approach in the realization of such softwarc.

The software and a short user's guide are available upon request.

ACKNOWLEDGEMENTS

We would particularly like to thank R. E. Hubbard who initiated this project. We should also like to thank E. J. Dodson and J. Nanard for their help and suggestions and T. Lukas for testing and suggestions on applications of PCDRA. The atomic coordinates for the SYNCAM–1 Calmoduline where obtained from Dr. F. R. Salemme and Dr. P. Weber, Dupont de Nemour Research Center, USA.

This work was carried out in the "Groupement Scientifique Developpement de l'Intelligence Artificielle en Languedoc Roussillon" (GS DIALR) and the project was financed by EEC BAP contract n° 0136–F.

REFERENCES

1 Hubbard, R. E. HYDRA: Current and Future Developments. In *Current communications in molecular biology: Computer Graphics and Molecular Modeling* (R. Fletterick and M. Zoller, Eds.) CSH, New York, 1986, pp 9–11
2 Truong Trong Thi, A., Vandeginste, P. *La recherche* 1988, **19**(204), 1412–1422
3 Haiech, J. Intelligence artificielle, vers des systèmes pour l'ingénierie des protéines. *Biofutur*, 1987, **62**, 44–55
4 Hubbard, R. E. *J. Mol. Graphics* 1983, **1**(1), 13–16
5 Ogawa, K., Yoshida, H. and Suzuki, H. *J. Mol. Graphics* 1984, **2**(4), 113–116
6 Cozzini, P., Pavesi, P. and Andreeti, G. D. *J. Mol. Graphics* 1985, **3**(3), 90–92
7 Nanard, M. and Nanard, J. "GTX TOOLBOX", Internal Technical Report, CRIM, 1987
8 Kabsch, W. *Acta Crystallogr.* A, 1976, **A32**, 922
9 Dodson, E. J., personal communication
10 Hubbard, R. E. and Fincham, D. *J. Mol. Graphics* 1985, **3**(1), 12–14