RasMol V2.2

A Molecular Visualisation Program

Roger Sayle
Biocomputing Research Unit
University of Edinburgh

last updated 14th October 1993

Contents

1 Introduction

RasMol is an menu- and command-driven interactive utility for the visualisation of molecular structures. The program will run on any BSD-like UNIX system, optionally displaying frames to either an 8 bit or 24 bit colour frame buffer available over the X Window System. Support is also provided for a dials box (via the XInput extensions) if one is connected to the system.

RasMol v2.1 accepts atomic co-ordinates in the standard Brookhaven Protein Data Bank (PDB) format, optionally supplemented David Bacon's Raster3D colour scheme specifications. The program interactively displays the molecule on the screen in a variety of representations, including wireframe, backbone, union-of-spheres, sticks and ball-and-stick. Different portions of the molecule may be rendered in any representation and colour independent of the rest of the model. The molecule may be rotated, translated, zoomed and z-clipped interactively, either from the keyboard, scroll bars or dials attached to the workstation. Finally, the rendered image may be written to a file in a variety of file formats.

2 General Operation

2.1 Running RasMol

To invoke the RasMol molecular visualisation tool, the user should type the command '`rasmol`' from the unix prompt. Immediately upon starting, the program displays the following message to identify the version number of the running program:

```
RasMol Molecular Renderer
Roger Sayle, June 1991
Version 2.1B
```

Immediately underneath this banner message, appears the program's command line prompt '`RasMol> `'. If the program is being executed under the X Window System, the program determines the type of the display being used. If the screen has either an 8 bit or 24 bit colour frame buffer, RasMol creates another window, which is used to display menu options and the rendered images. If a suitable screen is not available, RasMol may only be used from the command line. Commands may be typed to manipulate the model, and to output the generated image to a raster file.

If an optional argument is given on the unix command line, this is taken to be a Brookhaven Protein Data Bank (PDB) file which is automatically loaded by the program (see the `load` command). If the file is found, the program displays the usual statistics after reading in the file, otherwise the error message '`Error: File not found!`' is displayed, before the user is presented the RasMol prompt.

In order to leave RasMol, the user can type the command `quit` or `exit` at the RasMol prompt, and the program will return the user to the familiar unix prompt. Alternatively, if a prompt other than the main RasMol prompt is being displayed, the user may hit control-C (`^C`) to leave the program. The message '`*** Quit ***`' will be output to the terminal, before the usual unix prompt is redisplayed. The program may also be terminated by selecting the **Quit** menu option, on the bottom of the main menu.

2.2 Display Window

If the program is run under the X Window System environment with a suitable colour screen, RasMol creates an additional window to display the rendered molecule interactively, as it is manipulated. This window is subdivided into three main regions. The left hand side of the window, called the 'canvas', is used to draw the images of the molecule. By default this area initially displays a black background. Both below and to the right of the canvas are two scroll bars used to rotate the molecule interactively. Finally, to the far right of the window are the buttons that form the menu.

While the mouse pointer is located within the canvas area of the display window, the mouse pointer is drawn as a cross-hair cursor, to enable the 'picking' of objects being displayed (see later); otherwise the mouse pointer is drawn as an arrowhead. Any characters that are typed at the keyboard while the display window is in 'focus' are redirected to the command line in the terminal window.

The display window may be resized at any point during the session. This has the effect of simply rescaling the image displayed on the canvas. RasMol imposes limits on the size of the display window such that the window must be large enough to display the menu and scroll bars and yet small enough to fit on a single screen. Attempts to enlarge the screen may fail owing to insufficient memory on the host machine, in which case RasMol reports the error message '`Renderer Error: Unable to allocate frame buffer!`' or some similar error.

On eight bit displays, when the number of colours required by the program exceeds the number of free colours on the screen, the program uses its own colourmap. This has the effect of temporarily displaying all windows other than the display window in false colours while the mouse pointer is within the display windows. If the mouse pointer is moved outside the display windows, the original colours of the other windows return, and the image on the canvas is shown in 'false colour'. Once the number of colours required by the program drops again, the presentation of colours returns to normal.

2.3 Scroll Bars

The scroll bar across the bottom of the canvas area is used to rotate the molecule about the y-axis, i.e. to spin the nearest point on the molecule left or right; and the scroll bar to the right of the canvas rotates the molecule about the x-axis, i.e. the nearest point up or down. Each scroll bar has a 'indicator' to denote the relative orientation of the molecule, which is initially positioned in the centre of the scroll bar. These scroll bars may be operated in either of two ways. The first is by clicking any mouse button on the dotted scroll bar background to indicate a direct rotation relative to the current indictor position; the second is by clicking one of the arrows at either end of the scroll bar to rotate the molecule in fixed sized increments. Rotating the molecule by the second method may cause the indicators on the scroll bars to wrap around from one end of the bar to the other. A complete revolution is indicated by the indicator travelling the length of the scroll bar. The angle rotated by using the arrows depends upon the current size of the display window. No facility is provided for depressing the mouse button and 'dragging' the indicator along the scroll bar.

2.4 Menu Buttons

The menu consists of a vertical column of large buttons that are selected by depressing and releasing any of the mouse buttons while the mouse pointer is over a menu button. Each menu option either directly changes the system options, or displays a submenu from which the user can select a further option. The number of menu buttons displayed on the screen varies from menu to menu. Most menus have the option **Cancel** to allow the user to return to the menu without affecting any of the current options. When selection of a menu option causes the program to prompt the user for additional information in the terminal window, the current command line being edited is erased (see later). When RasMol first starts executing, the menu area of the display window presents the main menu. The bottom option from this menu, **Quit**, allows the user to terminate the program directly, and return to the UNIX prompt.

2.5 Picking

In order to identify a particular atom or bond being displayed, RasMol allows the users to 'pick' objects on the screen. The mouse is used to position the cross-hair cursor over the appropriate item, and then any of of the mouse button is depressed. Provided that the pointer is located close enough to a visible object, the program determines the identity of the nearest atom to the point identified.

The program will display, in the terminal window, the atom's type, serial number, residue name and residue number. If the atom is a member of a named chain, the chain identifier is also displayed. Two examples of the output generated by selecting an atom are displayed below:

```
Atom: CA 349  Group: SER 70
Atom: O  526  Hetero: HOH 205  Chain: P
```

The first line describes the alpha carbon of the serine-70 amino acid in a protein. The unique Brookhaven serial number for this atom is 349. The following line describes the oxygen atom in a water molecule attached to the P chain of the main molecule. The word 'Hetero' distinguishes heterogenous molecules (such as cofactors) from the residues in the main molecule, noted by 'Group'. [These two atoms are referred to by the two atom expressions "`SER70.CA`" and "`HOH205P.O`", respectively, when using the RasMol commands `select` and `restrict`.]

2.6 Dials Box

If RasMol detects a 'dials box' attached to the user's workstation, it also allows the molecule to be manipulated interactively by the dials. Once RasMol starts up, it labels the LED displays above each dial, "ROTATE X", "ROTATE Y", "ROTATE Z" and "ZOOM" across the top row from left to right, and "TRANS X", "TRANS Y", "TRANS Z" and "SLAB" from left to right across the bottom row. Rotating any of the knobs will automatically transform and redisplay the molecule interactively. The dials only have effect while the mouse pointer is within the display window. If more than one application is using the dials box at a time, care must be taken to remember the dial labels assigned by each program, as each application may overwrite the dial-label LEDs.

The rotation about the X and Y axes automatically updates the indicators on the appropriate scroll bars. All the rotation dials rotate the molecule 180° for a complete revolution of the dial. All the remaining dials clamp their values to permissible ranges; turning these dials past their limits has no effect.

The “ZOOM” dial allows the interactive zooming of the molecule between 10% and 200% of the original default magnification. Rotating the dial clockwise magnifies the molecule and anticlockwise shrinks it. A complete revolution of the dial corresponds to a 100% change in scale.

The “SLAB” dial, which is only effective when slabbing is enabled, allows the user to move the front z-clipping plane from the nearest point on the molecule to the furthest. A complete rotation of the SLAB dial corresponds to moving the clipping plane half the distance between the front and back of the molecule. Turning the SLAB knob clockwise moves the clipping plane closer to the viewer (increasing the number of objects displayed), and turning it anticlockwise moves it further away (preventing more objects from being displayed). Slabbing mode is enabled by selecting the **Slab** menu button on the **Options** menu, or by using the `slab` command on the command line.

The translation along the X and Y axis allows the centre of the molecule to be moved within the canvas area of the screen. Rotation and zooming are still performed relative to the centre of the molecule, which may often not be at the centre of the canvas. The TRANS Z dial currently has no effect.

2.7 Command Line Interface

RasMol allows the execution of interactive commands typed at the RasMol prompt in the terminal window. Characters typed into either the terminal or the display window are processed on the command line. Each command must be given on a separate line terminated by a newline or carriage return character. Keywords are case insensitive and may be entered in both lower and upper case letters. All whitespace (space, tab and formfeed) characters are ignored, except to separate the keyword and the arguments of a command. Blank lines (those containing only whitespace) are ignored. There is an internal restriction that command lines are limited to a maximum of 256 characters. Strings may be delimited by matching single or double quotation marks. Placing a hash ‘`#`’ character anywhere outside quotes terminates the line. RasMol will ignore the rest of the line, which may be used to comment on the command.

If a syntax error is detected on entering an interactive command, RasMol indicates the location of the error on the command line by placing the ‘`^`’ character under the offending word or character, and writing an error message on the following line. If a command is not recognised by RasMol, the program will generate an ‘`Unrecognised command!`’ error and redisplay the main prompt. If surplus information is given at the end of a command line, RasMol will execute the recognised command, but issue the warning message ‘`Warning: Ignoring rest of command!`’. Some commands may prompt the user for more information. These commands display a different prompt and are discussed in the command reference.

Whenever RasMol outputs diagnostic or error messages to the screen owing to selecting options from the menu or picking objects on the screen, the current command line is cleared. And the prompt redisplayed after any text has been displayed.

2.8 Command Line Editing

RasMol allows basic editing of the command line. Pressing either backspace, delete or `^H` will delete the previous character, and the key `^D` may be used to delete the character under the cursor. Several characters may be used to move the cursor along the command line. The characters `^B`, `^F`, `^A` and `^E` move the cursor back a single character, forward a single character, to the beginning of the line and to the end of the line, respectively. When the cursor is not at the end of the command line, typed characters are inserted into the line and do not overwrite existing characters. After a command line has been edited, a newline or carriage return will enter the entire line, regardless of where the cursor is positioned. Because RasMol is unable to move the cursor up to the previous line, care must be taken when editing commands that wrap over several lines. In the event that another process overwrites or corrupts the command line, the character `^L` may be used to redisplay the line on the screen.

RasMol maintains a history of recently used commands, so the that the user never needs to type the same commands repeatedly. Typing `^P` on the command line will display the previous command in the history and `^N` will display the following command. These commands may be edited using the features described above. Moving forward or backward through the command history undoes the modifications made to the current line. The number of commands retained in the history depends upon their length. RasMol can retain more short command lines and fewer long ones.

For users of "vt100" and compatible terminals (such as an "xterm"), RasMol also understands the cursor control characters on the keyboard. The right and left cursor keys have the same affect as `^F` and `^B`, moving the cursor forward and back a single character respectively. Similarly, the up and down cursor keys have the same function as `^P` and `^N`, producing the previous and next entries in the command history respectively.

2.9 Start-up File

When RasMol is first run, it searches for an initialisation file of commands to run before the command prompt is presented to the user. The file is called **.rasmolrc** on UNIX systems, and **RASMOL.INI** on MS-DOS and Microsoft Windows Systems. The format and execution of the this file is identical to that of the `script` command described in section 5.9 of this manual.

RasMol first looks for the initialisation file in the current directory, and if it is not found will look for it in the user's home directory. On MSDOS systems the environment variable '`HOME`' may be used name the appropriate directory. Unlike the command '`script ".rasmolrc"`', the program will not generate an error message if the file is not found.

3 Menu Options

3.1 Main Menu

When RasMol is started up at the beginning of a session, the menu area on the right hand side of the display window presents the main menu. This is the root menu for the RasMol menu system, and this menu is redisplayed after each operation selected from a menu is performed.

The main menu of RasMol Version 2.1 contains six options.

3.1.1 Load

This option allows the user to specify the Brookhaven Protein Data Bank (PDB) file to be read in. The function of this button is identical to the interactive command '`load`' typed from the RasMol command line. The PDB file contains the Cartesian co-ordinates of each atom in the molecule to be displayed. A complete description of the PDB file format and records understood by the program (including the Raster3D colour scheme records) are given in section 6.2 of this document. If a molecule has already been loaded from a PDB file, clicking the **Load** option has no effect.

On pressing **Load** the user is prompted by the message '`PDB file name:`' in the terminal window. The filename must be a valid UNIX filename without wildcard characters, and must be entered without string delimiters. Entering an empty filename will abort the operation and return the user to the main menu and RasMol command line prompt. If the file exists it is loaded by the program, otherwise RasMol outputs the error message '`Error:  File not found!`', before returning the user to the main menu and the RasMol command line prompt.

Once the PDB file is loaded, the RasMol program determines the connectivity of the molecule by proximity of atoms: Two atoms within an appropriate interval of one another are considered to be bonded. Once these calculations have been performed, the program outputs statistics on the molecule together with any appropriate information found in the PDB file. Once all the required preprocessing has been performed, RasMol 'selects' all the atoms in the molecule and displays a monochrome depth-cued wireframe representation on the display window canvas. The exact details of the molecule preprocessing steps are discussed under the '`load`' command, section 5.4 in the command reference section of this document.

3.1.2 Display

Selecting the **Display** menu button from the RasMol main menu presents the user with the RasMol display options submenu. This submenu allows the current representation of the molecule to be changed. The operation of the options on this submenu is described in detail in section 3.2, "Display Menu". These options allow the molecule to be represented as wireframe, alpha carbon backbone, stick bonds, space filling union-of-spheres or ball-and-stick models.

3.1.3 Colours

Clicking the **Colours** option on the RasMol main menu displays the colour submenu in the display window. These submenu options allow the user to modify the colour scheme of the atoms and bonds in the currently displayed molecule. The function of the buttons on the colour submenu are described in detail in section 3.3, "Colours Menu". The options allow the molecule to be shown in monochrome or coloured according to atom type, amino acid type, residue number, polypeptide chain, temperature factor or a user-defined colouring scheme supplied in the molecule's PDB file.

3.1.4 Options

The **Options** menu is used to access the RasMol miscellaneous options submenu. This submenu allows the user to modify the state of the regularly used RasMol variables. Details of each of the options available via the RasMol options menu are described in depth in section 3.4, "Options Menu". These options are used to turn slabbing mode, shadowing and specular highlights on and off. In addition, the inclusion/exclusion of hydrogren atoms and heterogenous groups from the display may be altered by buttons on this submenu.

3.1.5 Save

The **Save** menu option on the RasMol main menu is for writing the currently displayed image to an output raster file. The functionality of this menu option is also available through the commands '`write gif`', '`write epsf`', '`write ppm`' and '`write sunrle`'. Selecting the **Save** button displays a submenu of raster file formats in which the current image may be stored. This menu has the five options **GIF**, **EPSF**, **PPM**, **Rast** and **Cancel**. The last option allows the user to abort writing the current image to a file, and return to the main menu. The first four options specify the file format to be used in writing the current frame, which correspond to compuserve GIF format, Encapsulated PostScript, raw portable pixmap format and run length encoded SUN rasterfile format.

Selecting any one of these three format options causes RasMol to prompt the user '`Output file name:`'. The filename must be a valid UNIX filename without wildcard characters and must be entered without string delimiters. Entering a empty filename causes the program to abort the save operation and redisplay the main menu and RasMol command prompt. Conventionally, GIF images have the filename extension '.gif', PostScript files have the extension '.ps', raw PPM files have the extension '.ppm' and RLE SUN rasterfile images the extension '.ras'. If the program is unable to create the given filename for writing, RasMol aborts the command with the error message '`Output Error: Unable to create file '···'!`'. Once the file has been written (which may take a short time) the main menu and command line prompt are automatically redisplayed.

If the 8-bit version of RasMol version 2.1 is running, all four file formats may be output; however, the 24-bit versions of RasMol 2.1 may only output EPSF, PPM and SUN rasterfile formats. If the user attempts to write out a 24-bit GIF image, the program warns the user with the message '`Output Error: 24 bit GIF files unsupported!`'. Although the portable pixmap format represents raster images with three bytes per pixel, the eight bit version of RasMol produces a maximum of 250 unique RGB triples.

3.1.6 Quit

The bottom button on the RasMol main menu is **Quit**. The function of this button is identical to that of the commands '`quit`' and '`exit`' from the command line. **Quit** terminates the RasMol program and returns the user to the UNIX prompt. When exiting, RasMol closes the display window and the image currently being displayed is lost. If a dials box is attached to the workstation, all the LED labels above each knob are turned off.

3.2 Display Menu

The display submenu allows the current representation of the molecule to be changed. This submenu is displayed by selecting the option **Display** on the RasMol main menu. The display submenu contains seven menu options. The first six change the representation of the current active zone (see section 5.10 on the `select` command) and the final option, **Cancel**, is used to return the user to the main menu without modifying the current display.

3.2.1 Wireframe

This menu option displays the current active zone of the molecule as a depth-cued wireframe model. Each bond is represented as a narrow line between bonded atoms. Bonds nearer the viewer are brighter than those towards the back of the molecule, providing an impression of depth. The colour of a bond, or `none`, may be set with the `colour bond` command (section 5.2). If the bond has no colour of its own, it is represented as two half bond segments, each of which is given the colour of the nearest bonded atom. This menu option is equivalent to the RasMol commands `spacefill off`, `backbone off` and `wireframe on`.

3.2.2 Backbone

The **Backbone** menu option displays the current active zone of the molecule as a set of cylinders connecting the alpha carbons along the polypeptide backbone of the protein. Each 'bond' is drawn as a 0.48Å diameter cylinder. The colour of the backbone may be set using the RasMol `colour backbone` command (section 5.2). If the colour of the backbone segment is set to `none`, the two halves of the cylinder are drawn the colours of the alpha carbon atoms that they join. This representation is especially useful for revealing the secondary structure and folding of a protein. This menu option is equivalent to the RasMol command sequence `spacefill off`, `wireframe off` and `backbone 80`.

3.2.3 Sticks

The **Sticks** menu option displays the current active zone of the molecule as a set of sticks model. Each bond is represented as a cylinder of 0.48Å diameter.

The colour of a bond, or `none`, may be altered with the `colour bond` command (section 5.2). If the bond has no colour of its own, it is represented as two half bond cylinders, each of which is given the colour of the nearest bonded atom. This menu option is equivalent to the two RasMol commands `spacefill off` and `backbone off`, followed by the command `wireframe 80`.

3.2.4 Spacefill

This command represents the current active zone of the molecule as a union-of-spheres surface. Each atom is represented as a sphere of the appropriate van der Waals radius centred at the atom's position. The colours of the atoms may be changed by either the **Colours** submenu (section 3.3) or the RasMol `colour atom` command (section 5.2). This menu option is equivalent to the RasMol commands `spacefill` and `wireframe off`, `backbone off`.

3.2.5 Ball & Stick

The RasMol **Ball & Stick** menu option draws the current active zone of the molecule as a ball-and-stick model. Each atom is represented as a sphere of 0.48Å radius, and each bond is represented as a narrow cylinder of 0.16Å radius. This representation is combination of both a spacefilling and stick bond models. The colours of the atoms and bonds may be altered with the `colour atom` and `colour bond` commands respectively. This menu option is equivalent to the four RasMol commands `spacefill 120`, `wireframe 40`, `backbone off` and `ribbons off`.

3.2.6 Ribbons

The RasMol **Ribbons** menu option draws the current active zone of the molecule as a ribbon model. This menu option is equivalent to the four RasMol commands ribbons 380, `spacefill off`, `wireframe off` and backbone off.

3.3 Colours Menu

The colours submenu options allow the user to modify the colour scheme of the atoms and bonds in the currently displayed molecule. This submenu is displayed by selecting the option **Colours** on the RasMol main menu. The colours submenu contains eight menu options. The first seven select a colour scheme for the currently active zone (see section 5.10 on the `select` command) and the final option, **Cancel**, is used to return the user to the main menu without modifying the current colouring.

3.3.1 Mono

The top menu button on the RasMol colours menu, **Mono**, is used to colour the current active zone of the molecule monochrome white. This is the default colour scheme on loading a PDB file using the `load` command. This menu option is equivalent to the command '`colour white`'.

3.3.2 CPK

The **CPK** option on the colours submenu is used to shade the current active zone of the molecule by atom type, using the Corey, Pauling and Kultun (CPK) colouring scheme. These are the colours commonly used by chemists to denote atom type. In this scheme, carbon appears light grey, hydrogen appears white, oxygen appears red, sulphur appears yellow, nitrogen appears light blue and phosphorous appears pink. Any unrecognised element is displayed in magenta. This menu option is equivalent to the command '`colour cpk`'.

3.3.3 Shapely

The **Shapely** menu button on the colours submenu, changes the colour of each atom in the current active zone to a shade determined by the atom's residue type. For nucleic acids, each base is given a unique colour; for proteins, each amino acid sidechain is given a unique colour, and the the atoms of the polypeptide backbone are drawn in light grey. This is the colour scheme used by "Shapely models". This menu option is equivalent to the command '`colour shapely`'.

3.3.4 Group

The menu button **Group** on the RasMol colours submenu colours each atom of the current active zone in a colour based on its position along the macromolecule chain it belongs to. The actual colours use a 'ramped' colour map, beginning with dark blue at the start of each chain then slowly changing through cyan, green and yellow to red at the opposite end of the chain.

The matching of atom to colour is dependent upon the current status of the RasMol `hetero` variable (see the `set hetero` command). When the `hetero` variable is reset, and heterogenous groups are excluded from the default active zone, the colours are based directly on their position along the chain. Hence the N-terminal residue of proteins and the $5'$-terminus of nucleic acids are coloured blue, and the C-terminal of proteins and $3'$-terminus of nucleic acids are coloured red. If the `hetero` variable is set, and HETATM atoms are included in the default active zone, each heterogenous molecule associated with a chain is coloured as if it were appended to the end of the chain. Hence the last heterogenous molecule in the PDB file for each chain is drawn in red, and all atoms are coloured based on residue number.

Currently the functionality of this menu option is not replicated in the RasMol command line interface.

3.3.5 Chain

The **Chain** menu option on the colours submenu colours each atom in the current active zone using a colour based on the macromolecular chain to which it belongs, or with which it is assocaited. The chains are coloured arbitrarily by choosing the appropriate number of colours from a blue-to-red colour map as described in section 3.3.4 above.

Currently the functionality of this menu option is not replicated in the RasMol command line interface.

3.3.6 Temperature

The colours menu option **Temperature** is used to colour the currently active atoms of a molecule in a colour based on their anisotropic temperature factors, as defined in the PDB file from which it was read. The actual colours use a 'ramped' colour map, beginning with the 'coldest' atoms drawn in dark blue, then slowly changing through the colours cyan, green and yellow to red, as the atoms get warmer, with red representing the 'hottest' atoms in the file.

The matching of atom to colour depends upon the current value of the RasMol `hetero` variable (see section 3.4.3 and section 5.11 on the `set hetero` command). If the heterogenous atoms are currently excluded from the default active zone, i.e. the `hetero` variable has the value `false`, the temperature of the red and blue atoms are determined by the maximum and minimum temperature factors of the atoms in the main molecule. Otherwise, if heterogenous groups are included in the default active zone because the `hetero` variable has the value `true`, then the colours are taken from the temperature range of all the atoms in the PDB file.

Currently the functionality of this menu option is not replicated in the RasMol command line interface.

3.3.7 User

The last colouring option on the RasMol colours submenu, **User**, allows the current active zone of the molecule to be coloured according to a user-defined colour scheme. This colour scheme is provided by including additional 'COLO' records in the PDB file containing the molecule. These colour records, which are not supported by Brookhaven National Laboratories, are based upon those used by David Bacon's Raster3D program [1] and specify the colours of individual atoms by using a pattern matching algorithm. The exact format of these supplemental records is described in detail in section 6.2 of this document.

This menu option is equivalent to the command '`colour user`'.

3.4 Options Menu

The options submenu allows user to alter commonly used display options and parameters used in rendering molecules. This menu option is presented by clicking the **Options** menu button on the RasMol main menu. The options submenu displays six menu options. The first five are used to toggle the values of parameters and the final option, **Cancel**, may be used to exit the options submenu without modifying the current state of the program.

3.4.1 Slab

The **Slab** menu option on the RasMol options submenu is used to toggle the front z-clipping function, slabbing. When slabbing is enabled, RasMol only renders those objects (atoms and bonds) that are greater than a given distance from the viewer. This allows the user to 'cut away' the front part of the molecule to reveal internal or occluded detail. All wireframe bonds that intersect the clipping plane are subdivided, such that only the portion behind the z-clipping plane is drawn. The representation of spheres that intersect the clipping plane is governed by the `slabmode` variable (see section 5.11 on the `set slabmode` command). By default, `slabmode` has the value `solid` which depicts clipped atoms as solid filled spheres.

When RasMol is first started, the position of the z-clipping mode is 50% of the way through the molecule (drawing only the furthest half of the structure), with slabbing switched off. The current position of the z-clipping plane may be altered by executing a `slab` command, or by using the "SLAB" dial on the dials box (if attached). The position of the slabbing plane is moved by turning the "SLAB" dial even when slabbing is disabled.

This menu button works as a toggle button, alternatively turning on and off slabbing depending on the current setting. The user should avoid setting the position of the slab plane to the value 100, as the operation of the **Slab** will appear ineffective, since the whole molecule remains displayed. Similarly, because there is a performance degradation associated with rendering z-clipped images, it is recommended that slabbing is switched off, rather than moving the slabbing plane in front of the molecule. Shadowing may not be performed while slabbing is enabled, so shadowing is automatically disabled when slabbing is switched on. This menu option performs the same task as the two RasMol commands `slab on` and `slab off`.

3.4.2 Hydrogen

The **Hydrogen** option on the options menu allows the user to enable and disable the display of hydrogen atoms from the menu interface. Normally hydrogen atoms are not resolved by X-ray crystallography and only appear in data files generated by energy minimization programs. In order to enable the underlying structure of the molecule to be visualised, it is often necessary to disable the large number of hydrogen atoms that obscure the molecule. This is done by changing the current value of the `hydrogen` variable and selecting or restricting the current active zone as required. By default, the `hydrogen` variable has the value `true` and all hydrogen atoms and bonds are displayed along with the rest of the model. The default active zone, as defined by the `select` and `restrict` commands without parameters, relies on the current setting of `hydrogen`.

If **Hydrogen** is clicked when RasMol is currently permitting the display of hydrogen atoms, then the `hydrogen` variable is assigned the value `false`, the `restrict` command is used to force only the default active zone to be rendered and the message '`Hydrogens removed!`' is output to the terminal window. The `restrict` command causes all objects (atoms and bonds) not in the default active zone to be removed from the display. In normal operation, this means that all the hydrogen atoms/bonds disappear from the screen. If the `hetero` flag is currently `false`, any displayed heterogenous groups will also disappear from the display. This has the same effect as the command sequence `set hydrogen off` followed by the command `restrict`.

If the RasMol `hydrogen` variable is `false` when the **Hydrogen** menu option is clicked, then the `hydrogen` flag is set to `true`, the `select` command is used to set the current active zone to the new default active zone and the message '`Hydrogens selected!`' is written to the terminal window. The actual image drawn in the display window canvas does not change, but all subsequent representation and colouring commands will affect the molecule's hydrogen atoms. If the `hetero` variable is currently true, all heterogenous atoms will also be 'selected'. This has the same effect as the command sequence `set hydrogen true` followed by the command `select`.

3.4.3 Het Atoms

The option menu **Het Atoms** button allows the user to enable and disable the display of heterogenous groups from the menu interface. Typically water atoms and cofactors are also resolved by X-ray crystallography, along with the atoms of the protein or nucleic acid. Sometimes these heterogenous groups are important in the functioning of the macromolecule, but more generally they are solvent moieties that co-crystallize with the main molecule. The **Het Atoms** option allows the user to define whether these atoms, contained as HETATM records in the PDB file, should currently be displayed. This is done by changing the current value of the `hetero` variable and selecting or restricting the current active zone as required. By default, the `hetero` variable has the value `true`, and all heterogenous atoms and bonds are displayed along with the rest of the model. The default active zone, as defined by the `select` and `restrict` commands without parameters, relies on the current setting of `hetero`.

If **Het Atoms** is clicked when RasMol is currently permitting the display of heterogenous atoms, then the `hetero` variable is reset to the value `false`, the `restrict` command is used to limit the display of objects to the default active zone and the message '`HETA atoms selected!`' is printed below the RasMol prompt. The `restrict` command causes all objects (atoms and bonds) not in the default active zone to be removed from the display. In normal operation, this means all the heterogenous atoms/bonds disappear from the screen. If the `hydrogen` variable is currently `false`, any displayed hydrogen atoms will also disappear from the display. This has the same effect as the command sequence `set hetero off` followed by the command `restrict`.

If the RasMol `hetero` variable is `false` when the **Het Atoms** menu option is clicked, then the `hetero` flag is assigned `true`, the `select` command is used to set the current active zone to the new default active zone and the '`HETA atoms selected!`' diagnostic is displayed in the terminal window. The actual image drawn on the display window canvas does not change, but all subsequent representation and colouring commands will affect the molecule and the associated hetero groups. If the RasMol `hydrogen` variable is set, all hydrogen atoms will also be 'selected'. This menu option is identical to the command sequence `set hetero true` followed by the command `select`.

3.4.4 Specular

This menu option performs the same task as the two RasMol commands `set specular on` and `set specular off`.

3.4.5 Shadow

The **Shadow** button on the RasMol options menu toggles the value of the `shadow` variable. When this parameter has the value `true`, the macromolecular structure is rendered with cast shadows from a single light source, otherwise the shadowing is not calculated. Although by comparison RasMol is extremely fast at rendering shadowed images, testing which parts of a molecule are occluded from the light source is computationally expensive. It is recommended that, on slower machines, the correct orientation of the molecule is performed using the normal lighting model and then, once the desired position is determined, shadowing is enabled.

Slabbing may not be performed while shadowing is enabled, and so slabbing is automatically disabled when shadowing is switched on.

This menu option performs the same task as the two RasMol commands `set shadow on` and `set shadow off`.

4 Atom Expressions

4.1 Primitive Expressions

The main constituent of an atom expression is the primitive expression. A primitive expression uniquely identifies a related group of atoms within a molecule. A primitive expression consists of four major fields, only the first of which is compulsory. The first field specifies the residue type, the second specifies the residue number, the third identifies the chain (if applicable) and the final field identifies the atoms within the residue. A primitive expression may not contain any whitespace. Each of the fields is described in detail below.

Syntax: `<residue-type>{<residue-number>}{<chain>}{.<atom-type>}`

The first field consists of one to three alphabetic or question mark ('?') characters that identify the type of residue to be selected. The characters are matched against the residue name of each atom in the molecule, with the question mark acting as a wildcard character that matches any letter. For proteins, this means that the three letter codes for amino acids must be used. This field is case insensitive with all characters being converted to uppercase before the comparison is performed. A single asterisk ('`*`') character may be used as a abbreviation for the field "`???`". It is recommended that the asterisk is used whenever possible, for efficiency reasons. Unfortunately, residue names containing digits, such as "`SO4`", cannot be specified.

The second field contains a number that identifies the residue number within the molecule. This field may be any number of numeric characters, or a single asterisk to match all residues. Normally, the first field contains an asterisk if this field is given. A numeric value in this field uniquely identifies a single residue (base or amino acid) in the macromolecule, and hence a residue type given in addition would be redundant.

The third field is a single alphabetic character used to identify a chain of the molecule (if appropriate), or either an asterisk or a question mark to indicate that all chains should be used in the match. If a character is specified, this is converted to uppercase before any comparison. If the residue number field is an asterisk, a digit may be used in this field if the chains are denoted by numerals.

The final field is a period or full-stop ('`.`') character followed by one to four alphabetic or question mark characters that identify the atom types within a residue. All alphabetic characters are converted to capitals, and an asterisk character may be used in place of four question mark characters. Because spaces are not permitted in atom types, care must be taken in distinguishing ambiguities, such as `*.CA` meaning all alpha carbons or calcium atoms.

All fields that are specified in a primitive expression have to match a given atom in order for that atom to be selected. The flexabilty permitted by this kind of pattern-based template, allows the user to specify areas of interest in a molecule very quickly.

Examples:

`*`	All atoms
`CYS`	Atoms in cysteines
`AS?`	Atoms in asparagine and aspartic acid
`*120`	Atoms in residue 120
`*P`	Atoms in chain P
`*.N?`	Nitrogen atoms
`CYS.S?`	Sulphur atoms in cysteines
`SER70.C?`	Carbon atoms in serine-70

4.2 Predefined Sets

In addition to the primitive expression, RasMol atom expressions may also contain predefined sets. These sets are single keywords that represent portions of the molecule of interest. Using predefined sets is a way of abbreviating primitive expressions, and in some cases of selecting areas of a molecule that would otherwise be difficult to distinguish. Each of the currently implemented predefined sets is described below.

acidic The set of acidic amino acids. These are the residues Asp, Glu and Tyr.

alpha This set contains all the α-carbons in the protein molecule.

amino This set contains all the atoms contained in amino acid residues. This is useful for distinguishing the protein and DNA subunits in the current molecule database.

backbone This set contains the four atoms of each amino acid that form the polypeptide N-C-C-O backbone for proteins, or the sugar-phosphate backbone of nucleic acids.

basic The set of basic amino acids. These are the residues Lys, Arg and His.

cystine Once the disulphide bridges of a protein have been determined (using the RasMol `ssbond` command) this set contains all the atoms of cysteine groups that are disulphide bonded, i.e. half cystines. The set off free cysteines may be found using the expression '`cys and not cystine`'.

hetero This set contains all the heterogenous atoms in the molecule. These are all the atoms specified by HETATM entries in the PDB file. These typically contain cofactors and other ligands.

hydrogen This predefined set contains all the hydrogen and deuterium atoms in the molecule.

hydrophobic This set contains all the hydrophobic amino acids. These are the amino acids Ala, Leu, Val, Ile, Pro, Phe, Met and Trp.

neutral The set of neutral amino acids. These are all the residues that are `amino` but are neither `acidic` nor `basic`.

polar The set of polar amino acids. These are all the residues that are in the set `amino` but not `hydrophobic`.

purine The set of purine ribonucleotides. These are the bases adenosine and guanosine (A and G respectively).

pyrimidine The set of pyrimidine ribonucleotides. These are the bases cytidine and thymidine (C and T respectively).

selected This set contains the set of atoms that are members of the current active zone. These are defined by the previous `select` or `restrict` commands, and not by any part of the given expression.

sidechain The functional side chain of the amino acid. This set contains the atoms not in the N-C-C-O polypeptide backbone of proteins or the sugar-phosphate backbone of nucleic acids. This predefined set is equivalent to the RasMol expression `(not backbone)`.

Predefined Set	ALA **A**	ARG **R**	ASN **N**	ASP **D**	CYS **C**	GLU **E**	GLN **Q**	GLY **G**	HIS **H**	ILE **I**	LEU **L**	LYS **K**	MET **M**	PHE **F**	PRO **P**	SER **S**	THR **T**	TRP **W**	TYR **Y**	VAL **V**
`acidic`				•		•														
`acyclic`	•	•	•	•	•	•	•	•		•	•	•	•			•	•			•
`aliphatic`	•							•		•	•									•
`aromatic`									•					•				•	•	
`basic`		•							•			•								
`buried`	•				•					•	•		•	•				•		•
`charged`		•		•		•			•			•								
`cyclic`									•					•	•			•	•	
`hydrophobic`	•							•		•	•		•	•	•			•	•	•
`large`		•				•	•		•	•	•	•	•	•				•	•	
`medium`			•	•	•										•		•			•
`negative`				•		•														
`neutral`	•		•		•		•	•	•	•	•		•	•	•	•	•	•	•	•
`polar`		•	•	•	•	•	•		•			•				•	•			
`positive`		•							•			•								
`small`	•							•								•				
`surface`		•	•	•		•	•	•	•			•			•	•	•		•	

Table 1: RasMol Amino Acid Classification

4.3 Comparison Operators

Different portions of the molecule may also be described by the use of equality, inequality and ordering operators on some of their properties. The format for these comparison expressions is a property name followed by a comparison operator, and then by an numeric value.

The atom properties that may be used in RasMol atom expressions are `atomno` for the PDB atom serial number, `resno` for the residue number, `radius` for the displayed radius of an atom (or zero if the atom is not represented as a sphere) and `temperature` for the anisotropic temperature factor in the PDB file (the value given in the comparison is actually 100 times the value specified in the file, to ensure the value is an integer).

The equality operator is denoted either '`=`' or '`==`' and the inequality operator as either '`<>`', '`!=`' or '`/=`'. The ordering operators are then '`<`' for less than, '`>`' for greater than, '`<=`' for less than or equal to, and '`>=`' for greater than or equal to.

Examples:

```
resno < 23
temperature >= 900
atomno == 487
```

4.4 Logical Connectives

All three of the previous forms of query may be combined using the standard boolean connectives **and**, **or** and **not**. Both **and** and **or** keywords are infix binary operators and the keyword **not** is a unary prefix operator. The order of precedence of these operators gives logical negation the highest priority, then conjunction and finally disjunction has the lowest priority. Brackets, ('`(`' and '`)`') may also be used to change the order of evaluation. Both the conjunction and disjunction are short circuit operators, hence the ordering of expressions may affect the performance of queries.

The characters '`&`' and '`|`' may be used to represent the keywords `and` and `or` respectively. For notational convenience the comma ('`,`') character may also be used for disjunction, thus allowing a comma separated list of queries to be treated as the union of the selected sets. The characters '`~`' and '`!`' may be used to represent the keyword `not`.

Examples:

```
backbone and not alpha
not (hydrogen or hetero)
not *.FE and hetero
arg, his, lys
```

5 Command Reference

5.1 Backbone

The RasMol `backbone` command permits the representation of a polypeptide backbone as a series of bonds connecting the adjacent alpha carbons of each amino acid in a chain. The display of these backbone 'bonds' is turned on and off by the command paramater the same as the wireframe command. The command `backbone off` turns off the selected 'bonds', and `backbone on` or with a number turns them on. The number can be used to determine the cylinder radius of the representation in 0.004Å units. Backbone objects may be coloured using the RasMol `colour backbone` command.

Syntax:

```
backbone {<boolean>}
backbone <value>
```

The maximum permitted backbone radius is 2Å. A parameter value of 500 or above results in an '`Error: Integer argument too large!`' error.

The reserved work backbone is also used as a predefined set (see section 4.2) and as a parameter to the 'set hbond' and 'set ssbond' commands.

5.2 Colour

Syntax:

```
colour {<object>} <colour>
color {<object>} <colour>
```

The current list of objects for the colour command includes `atom` for each space filling sphere, `bond` for each bond between two atoms, `backbone` the backbone 'bond' between alpha carbons, `hbond` for hydrogen bonds and `ssbond` for disulphide bridges. If no object specification is given on the command line the default object is assumed to be `atom`.

The predefined colour values of the RasMol program are `blue`, `black`, `cyan`, `green`, `greenblue`, `magenta`, `orange`, `purple`, `red`, `redorange`, `violet`, `white` and `yellow`. In additional to these predefined colours the user may specify the exact colour as an RGB (Red-Green-Blue) triple. The syntax of an RGB triple is a comma separated list of three numbers in square brackets. Each of these values must lie in the range 0 to 255 inclusive and denotes the relative amount of red, green and blue in the final colour respectively. As an example, the following table gives the RGB equivalent to each of the predefined colours.

Keyword	**RGB triple**
`blue`	`[0,0,255]`
`black`	`[0,0,0]`
`cyan`	`[0,255,255]`
`green`	`[0,255,0]`
`greenblue`	`[46,139,87]`
`magenta`	`[255,0,255]`
`orange`	`[255,165,0]`
`purple`	`[160,32,240]`
`red`	`[255,0,0]`
`redorange`	`[255,69,0]`
`violet`	`[238,130,238]`
`white`	`[255,255,255]`
`yellow`	`[255,255,0]`

If a colour is not recognised as being one of the RasMol predefined colours the program consults the X Window System server colour database. If the appropriate colour is not found by the server, RasMol displays the error message '`Error: Unknown or incorrect colour!`'.

The `cpk` colouring scheme is based upon the popular plastic space filling models, which were invented by Corey and Pauling and later improved by Kultun. These models consist of plastic truncated sphere parts, which can be fitted together to form union-of-spheres surfaces with rotatable joints. In this scheme, hydrogen appears white, oxygen appears red, sulphur appears yellow and nitrogen appears light blue. The standard molecular graphics alteration to the original colour scheme redefines carbon to appear as light grey instead of black. (Traditionally molecules are displayed on black backgrounds). The `cpk` colour scheme can not be used to colour `bond` or `backbone` objects.

The `shapely` scheme is based upon Bob Fletterick's 'Shapely Models', and colours atoms according to the type of the residue in which they appear. The backbone atoms of proteins and polypeptides are rendered in light grey. Each of the atoms in the side chains is coloured according to its parent amino acid. For nucleic acids, each base type is given a unique colour. The `shapely` colour scheme can not be used to colour `bond` or `backbone` objects.

The `none` option allows bonds, hbonds, ssbonds and backbone segments to be coloured according to the (alpha carbon) atoms which they connect. When a bond has no colour of its own, each half of its length is drawn in the colour of the atom to which it is connected, otherwise the 'bonds' are drawn in the specified colour. The `none` colour option can not be used to colour `atom` objects.

5.3 HBond

The RasMol '`hbond`' command is used to represent the hydrogen bonding of the protein molecule's backbone. This information is useful in assessing the protein's secondary structure. Hydrogen bonds are represented as dotted lines between the donor and acceptor residues. The first time that the '`hbond`' command is used, the program searches the structure of the molecule to find hydrogen bonded residues and reports the number of bonds to the user. The command '`hbond on`' displays the selected '`bonds`' as dotted lines, and the '`hbond off`' turns off their display. The colour of hbond objects may be changed by the '`colour hbond`' command. Initially, the dotted line has the colours of its connected atoms.

By default the dotted lines are drawn between the accepting oxygen and the donating nitrogen. By using the '`set hbond`' command the alpha carbon positions of the appropriate residues may be used instead. This is especially usefull when examining backbone structures.

Syntax:

```
hbond {<boolean>}
```

RasMol uses Kabsch and Sander's algorithm for determining the main chain hydrogen bonds of proteins, as implemented by their DSSP computer program. There is no generally corrected H bond definition, as there is no sharp border between the quantum-mechanical and electrostatic regimes (wave function overlap dominates at short distances and electrostatic interaction at larger distnaces) and no discontinuity of the interaction energy as a function of distance or alignment.

Main chain hydrogen bonds, used in determining secondary structure, may be described by an electrostatic model. The electrostatic interaction energy between two

H-bonding groups is calculated by placing patial charges on the C, O $(+q_1, -q_1)$ and N, H $(+q_2, -q_2)$ atoms. This produces the energy equation

$$E = q_1 q_2 (1/r(\mathrm{ON}) + 1/r(\mathrm{CH}) - 1/r(\mathrm{OH}) - 1/r(\mathrm{CN})) * f$$

with $q_1 = 0.42e$ and $q_2 = 0.20e$, e being the unit electron charge and $r(\mathrm{AB})$ the interatomic distance from A to B. In chemical units, r is in angstroms, the dimension factor $f = 322$, and E is in kcal/mol. Typically a good hydrogen bond has a binding energy of -3 kcal/mol.

The default cutoff value is -0.5 kcal/mol, as described in Kabsch and Sander's paper [7]. This generous limit permits bifurcated H bonds and errors in coordinates.

5.4 Load

This command allows the user to specify the Brookhaven Protein Data Bank (PDB) file to be read in. The PDB file contains the Cartesian co-ordinates of each atom in the molecule to be displayed. A complete description of the PDB file format and records understood by the program (including the Raster3D colour scheme records) is given in section 6 of this document. If a molecule has already been loaded from a PDB file, this command displays the error message '`Error: Molecule database loaded!`'.

Syntax:

```
load {"<input-file>"}
```

The name of the file to load is given as a string parameter delimited by matching single or double quotes. If a filename string is not given, the user is prompted by the message '`PDB file name:`' in the terminal window. The required filename may be entered without string delimiters. If the user enters a blank line, the `load` command is aborted. The filename must be a valid UNIX filename without wildcard characters. If the file exists it is loaded by the program, otherwise RasMol outputs the error message '`Error:  File not found!`', before returning the user to the RasMol command line prompt.

Once the PDB file is loaded, the RasMol program determines the connectivity of the molecule by proximity of atoms: two atoms within an appropriate interval of one another are considered to be bonded. Currently two non-hydrogen atoms are considered to be bonded if they lie between 1.0Å and 1.9Å of one another, and an atom is considered bonded to a hydrogen if it is between 0.7Å and 1.2Å of the hydrogen's centre.

Once all the required preprocessing has been performed, RasMol 'selects' (see section 5.10) all the atoms in the molecule and displays a monochrome depth-cued wireframe representation on the display window canvas.

The RasMol `load` command displays status information on the contents of the file. An example of this output is presented below. The program determines the molecule name, classification and Brookhaven code from the appropriate records in the PDB file. A count of the number of chains, groups, atoms and bonds in the main molecule is displayed. If the data file contains heterogenous atoms, such as cofactors, the number of hetero atoms and groups are not included in these totals but presented in brackets after the main values.

```
Molecule name ..... CRAMBIN
Classification .... PLANT SEED PROTEIN
Brookhaven Code ... 1CRN
Number of Groups .. 46
Number of Atoms ... 327
Number of Bonds ... 379
```

If the `load` command is followed by anything other than a string, the program outputs a '`Error:  Filename string expected!`' error message.

5.5 Exit

The `exit` command is used to terminate a RasMol session. The behaviour of the `exit` command is identical to that of the `quit` command. When either an `exit` or a `quit` command is executed from the command line, the program closes the display window if the program is being used with a colour X Windows display and clears the LED labels on the dials box (if one is connected to the workstation), before terminating. Both the `exit` and `quit` commands finish executing the current script file when executed in either a script file or the '.rasmolrc' startup file. This command may also be executed by selecting the **Quit** option on the RasMol main menu.

5.6 Quit

The `quit` command is used to terminate a RasMol session. The behaviour of the `quit` command is identical to that of the `exit` command. See section 5.5.

5.7 Restrict

The `restrict` command is used to limit the display of the molecule to a specified portion. This command is similar to the `select` command (see section 5.10), except that any object not in the active zone after the command is executed is removed from the display.

Syntax:

```
restrict {<expression>}
```

5.8 Rotate

The `rotate` command is used to rotate the molecule on the screen. The molecule may be rotated about its centre around any of the three co-ordinate axes X, Y and Z. The first parameter of the `rotate` command specifies which axis to rotate. The final parameter to the `rotate` command is an integer specifying the angle to be transformed (in degrees). This value may be prefixed by a single unary minus sign to rotate the molecule in the opposite sense. For positive angles, rotating about X moves the closest point on the molecule down, rotating about Y moves the closest point to the right, and rotation about Z spins the molecule clockwise. Whitespace may appear between the unary minus and the first digit of the value. A unary minus is allowed to prefix a zero value.

If the RasMol program is being used interactively, the model may also be rotated by the two scroll bars to the right of and below the canvas area of the display window. The `rotate` command updates the position of the scroll bar indicators.

If there is a dials box attached to the workstation, the first three dials on the top row may also be used to rotate the molecule about the X, Y and Z axes respectively. Turning these dials rotates the molecule relative to its current position.

Syntax:

```
rotate <axis> {-} <value>
```

An axis parameter other than '`x`', '`y`' and '`z`' generates an '`Invalid command syntax!`' error. A missing or non-numeric integer value parameter causes an '`Integer value expected!`' error. Extremely large numeric values are internally truncated to 32 bit values.

5.9 Script

The RasMol 'script' command executes a set of commands that are contained in a text file as if they had been typed in interactively. Each command is placed on its own line in the text file terminated by a carriage return or newline character. The comment character, '#' may be used to place to annotate the file of commands, as described in section 2.7. The behaviour of the RasMol commands 'exit' and 'quit' within a script file cause, the program to cease execution of that file rather than terminate the RasMol program. RasMol script commands may be contained within script files, allowing command files to be nested upto a maximum of ten deep.

If any error occurs with a command within a script file, the error messages generated by RasMol are annotated with the name of the current script file and the line number in that file of the incorrect command.

Syntax:

```
script <filename>
```

Omitting the filename parameter from after the script command produces a 'Filename string expected' error message. If RasMol is unable to find and open the specified script file, they program will display the error message 'Cannot open script file '*name*'!', where *name* is the name of the file given on the command line.

If the command file contains any lines that contain more than 255 characters, the program issues the error message 'Script command line too long!'. Any commands on this line are ignored and execution of the file resumes on the following line of the file. If the nesting of script files exceeds the maximum depth, RasMol generates the error message 'Script command stack too deep!'.

5.10 Select

The select command is used to define the active zone of a molecule. The active zone of a molecule is that portion of the molecule model that is affected by commands that alter the representation and colour of atoms and bonds. The select command does not change the image displayed on the screen.

Syntax:

```
select {<expression>}
```

5.11 Set

The RasMol `set` command allows the user to alter various internal parameters of the rendering process. The syntax of an `set` statement specifies a 'variable' or 'parameter' to be modified followed by the value to be assigned to that parameter. The effects of each of these internal parameters is discussed separately in the following paragraphs.

Syntax:

```
set ambient [<value>]
set background <colour>
set bondzone <zonemode>
set hetero <boolean>
set hydrogen <boolean>
set shadow <boolean>
set slabmode <slabmode>
set specular <boolean>
set specpower <value>
```

Set Ambient

The RasMol `ambient` parameter is used to determine the amount of ambient light in the scene. Ambient light is the pervasive light not emitted from the primary light source. This is used to determine the intensity of surfaces that face away from the light, or areas hidden in shadow. The permisible integer values, betweem 0 and 100 inclusive, give the ambient light as a percentage of the primary light intensity. Omitting the argument returns the parameter to its default value.

Set Background

The RasMol `set background` command is used to change the colour of the canvas window background. By default, this colour is initially black. The background may be changed to any of the 'generic' colour values used by the `colour` command (see section 5.2). These colours include the predefined colours, RGB colour triples and entries in the X Window System server's colour database. Non-generic colours, such as `cpk`, `none` and `user` may not be specified. This command is usefull in preparing overhead slides from RasMol images that require light background colours.

Set BondZone

A RasMol bond zone mode is either '`and`' (`false`) or '`or`' (`true`). An '`and`" value selects a bond if if both bonded atoms are currently selected, whereas an '`or`' value selects a bond if either of the connected atoms is in the current active zone.

Set Slabmode

The current RasMol slab modes are `reject`, `half`, `hollow`, `solid` and `section`. In `reject` slab mode, any atom that is intersected by the z-clipping plane when slabbing is enable is not displayed at all. In `half` slab mode, each atom is represented as the front half of a hollow sphere. This allows atoms to be visible through the cut away section of a sliced atom. Slab mode `hollow` treats each atom as a full hollow sphere, and sliced atoms reveal the inside surfaces of their 'shell'. Finally, the `solid` slab mode represents the spheres as solid spheres: any sphere cut by the slabbing plane reveals a solid cross section of the interior. The colour of the cross-section is taken from the closest atom, hence two intersecting spheres have a colour boundry through their cross section. The `section` slab mode is identical to `solid` mode except that only the cross sections of atoms intersecting the slab plane are drawn. This gives a two dimensional slice through the molecule.

Set SpecPower

The RasMol `specpower` variable describes the shininess of surfaces in the scene when the specular highlights are enabled by the `set specular` command. The larger the integer argument the shinier (or more reflective) the surfaces become. Low values produce large white dots on spheres (like those seen on plastic), while larger values produce smaller white dots (like those observed on metal or glass). Ommiting the argument from this command returns the `specpower` variable to its default setting.

Set Specular

The RasMol `specular` variable determines whether surfaces are to be rendered as shiny or diffuse surfaces. When the `specular` variable has the value `true` or `on`, all surfaces are considered shiny with typically a small white dot representing the reflection of the light source on the surface. If `specular` is `false` or `off` all surfaces are rendered with only diffuse lighting. The 'shininess' of the surface is determined by the value of the `specpower` variable.

5.12 Show

The RasMol '`show`' command displays usefull information about the currently loaded molecule. The '`show sequence`' command lists the sequence of residues that form the macromolecule; amino acids for protein and nucleotides for nucleic acids. The '`show info`' command displays the name of the molecule, the number of groups, atoms and bonds it contains together with some of the header information stored in the PDB file.

Syntax:

```
show sequence
show info
```

The `show sequence` command is particularly usefull for displaying the residue type and residue numbers of a macromoecule in order to prepare an atom expression for the required portion of the molecule. If the current macromolecule contains more than one chain, each the internal name of each chain is displayed on a separate line followed by the sequence that makes up that chain. The sequences are displayed ten residues to a line, giving the full residue name followed by its residue number. An example output for the protein *crambin* is given below.

```
THR1   THR2   CYS3   CYS4   PRO5   SER6   ILE7   VAL8   ALA9   ARG10
SER11  ASN12  PHE13  ASN14  VAL15  CYS16  ARG17  LEU18  PRO19  GLY20
THR21  PRO22  GLU23  ALA24  ILE25  CYS26  ALA27  THR28  TYR29  THR30
GLY31  CYS32  ILE33  ILE34  ILE35  PRO36  GLY37  ALA38  THR39  CYS40
PRO41  GLY42  ASP43  TYR44  ALA45  ASN46
```

The `show information` command displays the details of the size of the current molecule to the user. It presents data on the molecule's name, classification and brookhaven code (from the appropriate records in the PDB file), the number of chains, groups, atoms and bonds. If the data file contains any heterogenous groups, such as cofactors, these group and their atoms are not included in the above totals, but presented in brackets after the main values. An example output for the protein *crambin* is given below.

```
Molecule name ..... CRAMBIN
Classification .... PLANT SEED PROTEIN
Brookhaven Code ... 1CRN
Number of Groups .. 46
Number of Atoms ... 327
Number of Bonds ... 379
```

Giving an option other than `sequence`, `info` or `information` after the `show` command results in a RasMol "`Invalid command argument!`" error message.

5.13 Slab

The RasMol `slab` command may be used to turn z-clipping on and off, and positioning the z-clipping plane. Slabbing involves removing all the objects on the screen that are closer than a given distance from the viewer. This position is referred to as the location of the (front) z-clipping plane. When slabbing is enabled, all wireframe bonds that intersect the clipping plane are split, and only the section of the line furthest from the viewer is drawn. The representation of spheres that intersect the clipping plane is governed by the `slabmode` variable (see the `set slabmode` command). By default, `slabmode` is set to the value `solid` which depicts clipped atoms as solid filled spheres.

The position of the z-clipping (or slabbing) plane is represented as a value between 0 and 100. This represents a linear scale from the very back of the molecule to the front. This value may also be considered the percentage visibility, for example, a value of 25 typically only display one quarter of the atoms/bonds in the molecule. When RasMol is first started up, the initial value of the clipping plane is 50 (drawing only the back half of the molecule) and the slabbing mode is disabled.

There is a performance degradation associated with rendering images with slabbing enabled, hence it is recommended that slabbing is turned off rather than the position of the z-clipping plane being set to 100 (in front of the molecule). Similarly, the user should be aware that a blank screen may be caused by a low value of the slabbing plane while slabbing is enabled.

When the `slab` command followed by a boolean value, RasMol enables or disables slabbing without affecting the value of the clipping plane. Hence, if the `slab` command is followed by the keyword `on` or `true` slabbing is switched on, and if it is followed by either `off` or `false`, then slabbing is turned off. Executing the `slab` command with no arguments ensures that slabbing is currently enabled.

The `slab` command may also be given with a numeric parameter to specify the exact position of the slabbing plane. The value of this parameter must lie between 0 and 100. A slab value of 0 displays none of the molecule, a value of 50 displays the back half of it and a value of 100 shows all of it. Specifying a value to the slab command, automatically enables slabbing.

If there is a dials box connected to the workstation, the bottom right hand dial, labelled "SLAB", may be used to interactively move the clipping plane when slabbing mode is enabled. Turning the dial moves the location of the clipping plane relative to its current setting.

Syntax:

```
slab {<boolean>}
slab <value>
```

An unrecognised parameter to a `slab` command generates an '`Invalid command syntax!`' error message. A numeric value greater than 100 produces an '`Integer argument too large!`' error.

5.14 Spacefill

The RasMol `spacefill` command determines the radius of each atom in its spherical representation on the screen. The currently selected zone is represented space filling union-of-spheres model. An integer parameter may be used to specify the radius of each atom, given in 4nm units. The maximum permitted value is 500, which corresponds to a radius of 2Å. If no parameter is given, the radius is taken to be the Van der Waals radius of the appropriate atom.

The temperature option is used to set the radius of each selected sphere to the value in the temperature field of the molecule file. A zero or negative value causes no change in the selected atom. Temperature values greater than 2.00 are truncated to the maximum 2Å radius.

The user option allows the radius of the selected spheres to be determined by matching each atom against optional lines in the input data file. Details of the wildcard pattern matching used by Raster3D's COLOR records is given in section 6.2 of this document.

Syntax:

```
spacefill {<boolean>}
spacefill <value>
spacefill temperature
spacefill user
```

A numeric argument over 500 results in the RasMol error '`Integer argument too large`'.

5.15 SSBond

The RasMol '`ssbond`' command is used to represent the disulphide bridges of the protein molecule as dotted lines between the connected cysteines. The first time that the 'ssbond' command is used, the program searches the structure of the molecule to find half-cysteine pairs and reports the number of bridges to the user. The command 'ssbond on' displays the selected 'bonds' as dotted lines, and the 'ssbond off' turns off their display. The colour of ssbond objects may be changed by the 'colour ssbond' command. Initially, the dotted line has the colours of its connected atoms.

By default the dotted lines are drawn between the sulphur atoms within the cysteine groups. By using the 'set ssbond' command the position of the cysteine's alpha carbons may be used instead. This is especially usefull when examining backbone structures.

Syntax:

```
ssbond {<boolean>}
```

RasMol determines disulphide bonds, i.e. covalent links between the S^{γ} atoms of two Cys residues, to exist between two sulphur atoms separated by a distance of less than 3.0Å.

5.16 Translate

The RasMol `translate` command permits the molecule to be translated (shifted) along each of the three major axes. The integer argument specifies an absolute position along the given axis to move. Positive values move the molecule right and up for the 'x' and 'y' axes, and negative values left and down for the 'x' and 'y' axes respectively. An absolute value of 100 moves the molecule completely off of the canvas window, an absolute value of 50 places the center of the molecule at the very edge of the screen and a zero parameter places the center of the molecule in the middle of the screen with respect to the given axis.

Syntax:

```
translate <axis> {-} <value>
```

An axis parameter other than 'x', 'y' and 'z' generates an 'Invalid command syntax!' error. A missing or non-numeric integer value parameter causes an 'Integer value expected!' error. Integer values greater than 100 or less than -100 produce an 'Integer argument too large!' error.

5.17 Wireframe

The RasMol `wireframe` command determines the representation and radius of each bond on the screen.

Syntax:

```
wireframe {<boolean>}
wireframe <value>
```

5.18 Write

The `write` command is used to output the currently displayed image to a raster file. RasMol allows the format of the generated output file to be set to one of several standard formats. This file format may be optionally specified by following the `write` command by the keyword `gif`, `ppm`, `sun`, `sunrle` or `ps`, which denote compuserve GIF format, raw portable pixmap format, standard sun rasterfile, run length encoded sun rasterfile and PostScript, respectively. If a file format is not specified on the command line, the image is written in GIF file format.

The final parameter to the write command is a string containing the filename to which the image should be written. This string must be delimited by matching single or double quotation marks. If a filename is not given on the command line, the program prompts the user for the required filename with the prompt '`Output file name:`'. A filename typed into the terminal window at this prompt should not be delimited by quotation marks. Entering an empty filename at this prompt will cause RasMol to abort the `write` command. The filename must be a valid UNIX filename without any wildcard characters. Conventionally GIF images have the filename extension '.gif', raw PPM files have the extension '.ppm', SUN rasterfile images the extension '.ras' and PostScript files have the extension '.ps'.

The resolution of the created image file is taken from the current size of the display window canvas area. Hence different sized images may be created by resizing the window appropriately before executing the `write` command. If the image is being prepared non-interactively (i.e. not from a suitable colour frame buffer), the image is generated at 576×576 resolution, which is the default size of the canvas area of the display window.

The current 24 bit version of RasMol, version 2.1, is unable output into the compuserve GIF file format for technical reasons. If the user attempts to write out a 24 bit GIF image, the program warns the user with the message '`Output Error: 24 bit GIF files unsupported!`'. Although the portable pixmap format represents raster images with three bytes per pixel, the eight bit version of RasMol only produces a maximum of 250 unique RGB triples.

Syntax:

```
write {<format>} {"<output-file>"}
```

If the `write` command is followed by an unrecognised file format, or if the output filename is not a valid string, RasMol issues a '`Filename string expected!`' error message. If the program is unable to write to the given file, RasMol aborts the command with the error message '`Output Error:  Unable to create file '...'!`'.

5.19 Zoom

The `zoom` command is used to modify the magnification of the molecule being displayed. The molecule is always magnified with respect to its centre. The zoom parameter may be used to scale the molecule between 10% and 200% of its default magnification. The default scale of the molecule is chosen to allow the molecule to be rotated about any angle and still fit within the display canvas (i.e. not require clipping). There is a performance degradation with displaying molecules that clip the edges of the screen.

The `zoom` command without any parameters or followed by the keywords `off` or `false`, turns off the magnification of the molecule and sets the scale to its initial value. The `zoom` command followed by the keyword `on` or `true` magnifies the molecule to 1.5 times its initial scale.

The `zoom` command may also be given with a numeric parameter to specify the exact magnification required. The value of this parameter must lie between 0 and 200. This value represents the percentage magnification required, with a value of 100 representing the default scale of the molecule and a value of 200 rendering the molecule twice its normal size. Numeric values less than or equal to 10 generate minimum scale images at one tenth the size of the original. The command '`zoom on`' has the same effect as the command '`zoom 150`'.

If there is a dials box connected to the workstation, the top right hand dial, labelled "ZOOM", may be used to magnify the molecule interactively. Turning the dial alters the magnification of the molecule relative to its current setting.

Syntax:

```
zoom {<boolean>}
```

```
zoom <value>
```

An unrecognised parameter to a `zoom` command generates an '`Invalid command syntax!`' error message. A numeric value greater than 200 produces an '`Integer argument too large!`' error.

6 File Formats

6.1 Brookhaven Data Bank Files

If you do not have the Brookhaven documentation, you may find the following summary of the PDB file format useful. The Protein Data Bank is a computer-based archival database for macromolecular structures [3]. The database was established in 1971 by the Brookhaven National Laboratory, New York, as a public domain repository for resolved crystallographic structures. The Bank uses a uniform format to store atomic co-ordinates and partial bond connectivities as derived from crystallographic studies.

PDB file entries consist of records of 80 characters each. Using the punched card analogy, columns 1 to 6 contain a record-type identifier, the columns 7 to 70 contain data. Columns 71 to 80 are normally blank, but may contain sequence information added by library management programs. The first four characters of the record identifier are sufficient to identify the type of record uniquely, and the syntax of each record is independent of the order of records within any entry for a particular macromolecule.

The only record types that are of major interest to the RasMol program are the ATOM and HETATM records (apart from COLO record type not sanctioned by Brookhaven), which describe the position of each atom. ATOM/HETATM records contain standard atom names and residue abbreviations, along with sequence identifiers, co-ordinates in Angstrom units, occupancies and thermal motion factors. The exact details are given below as a FORTRAN format statement.

```
FORMAT(6A1,I5,1X,A4,A1,A3,1X,A1,I4,A1,3X,3F8.3,2F6.2,1X,I3)
```

Column	Content
1-6	'ATOM' or 'HETATM'
7-11	Atom serial number (may have gaps)
13-16	Atom name, in IUPAC standard format
17	Alternate location indicator indicated by A,B or C
18-20	Residue name, in IUPAC standard format
23-26	Residue sequence number (order as below)
27	Code for insertions of residues (i.e. 66A & 66B)
31-38	X co-ordinate
39-46	Y co-ordinate
47-54	Z co-ordinate
55-60	Occupancy
61-66	Temperature factor
68-70	Footnote number

Residues occur in order of their sequence numbers, which always increase starting from the N-terminal residue for proteins and $5'$-terminus for nucleic acids. If the residue sequence is known, certain atom serial numbers may be omited to allow for future insertion of any missing atoms. Within each residue, atoms are ordered in a standard manner, starting with the backbone (N-C-C-O for proteins) and proceeding in increasing remoteness from the alpha carbon, along the side chain.

HETATM records are used to define post-translational modifications and cofactors associated with the main molecule. Optional TER records are interpreted as breaks in the main molecule's backbone.

If present, RasMol also inspects HEADER, COMPND and END records. Information such as the name, Brookhaven code, revision date and classification of the molecule are extracted from HEADER and COMPND records, and the end of the file may be indicated by an END record.

6.2 PDB Colour Scheme Specification

The RasMol v2.1 program also accepts the supplementary record-type COLO in the PDB file format. The record was introduced by David Bacon's Raster3D program for specifying the colour scheme to be used when rendering the molecule [1, 2]. This extension to the standard file format is not currently supported by Brookhaven.

The COLO record has the same basic record type as the ATOM and HETATM records described above, but with COLO in the first four columns:

Column	Content
1-6	‘COLO’
7-30	Mask (described below)
31-38	Red component
39-46	Green component
47-54	Blue component
55-60	Sphere radius in Angstroms
61-80	Comments

Note that the Red, Green and Blue components are in the same positions as the X, Y, and Z components of an ATOM or HETA record, and the van der Waals radius goes in the place of the Occupancy. The Red, Green and Blue components must all be in the range 0 to 1.

Colours are assigned to atoms using a matching process. The Mask field is used in the matching process as follows. First RasMol reads in and remembers all the ATOM, HETA and COLO records in input order. When the user-defined (‘User’) colour scheme is selected, RasMol goes through each remembered ATOM/HETA record in turn, and searches for a COLO record that matches in all of columns 7 through 30. The first such COLO record to be found determines the colour and radius of the atom.

In order that one COLO record can provide colour and radius specifications for more than one atom (e.g. based on residue, atom type, or any other criterion for which labels can be given somewhere in columns 7 through 30), a ‘don’t-care’ character, the hash mark “#” (pound sign, number sign, sharp sign) is used. This character, when found in a COLO record, matches any character in the corresponding column in a ATOM/HETA record. All other characters must match identically to count as a match. As an extension to the specification, any atom that fails to match a COLO record is displayed in white.

7 Contact

If you have any questions, problems, comments, complaints or suggestions please do not hesitate to call or write to the author:

Roger Sayle
Biocomputing Research Unit Phone: (+44) (0)31 650 5163
Room 707, Darwin Building EMail: `ros@biocomp.ed.ac.uk`
University of Edinburgh
Mayfield Road
Edinburgh EH9 3JR
UK

References

[1] David Bacon. *Raster3D programmer/user guide*, July 1988.

[2] David Bacon and Wayne F. Anderson. A fast algorithm for rendering space-filling molecule pictures. *Journal of Molecular Graphics*, 6(4):219–220, December 1988.

[3] Frances C. Bernstein et al. The Protein Data Bank: A computer-based archival file for macromolecular structures. *Journal of Molecular Biology*, 112:535–542, 1977.

[4] H. Iijima, J. B. Dunbar Jr., and G. Marshall. Calibration of effective van der waals atomic contact radii for proteins and peptides. *Proteins: Structure, Functions and Genetics*, 2:330–339, 1987.

[5] Adobe Systems Inc., editor. *PostScript Language Reference Manual*. Addison-Wesley, Reading, Mass., 1985.

[6] Adobe Systems Inc., editor. *PostScript Language Tutorial and Cookbook*. Addison-Wesley, Reading, Mass., 1985.

[7] Wolfgang Kabsch and Christian Sander. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22:2577–2637, 1983.

[8] Arthur M. Lesk, editor. *Protein Architecture: A Practical Approach*. IRL Press, 1991.

[9] Nelson Max. Computer representation of molecular surfaces. *IEEE Computer Graphics and Applications*, pages 21–29, August 1983.

[10] Roger Sayle. Parallel algorithms for molecular graphics. BSc(Eng) Project Report, Department of Computing, Imperial College, University of London, June 1990.

[11] Roger Sayle and Andrew Bissell. RasMol: A program for fast realistic rendering of molecular structures with shadows. In *Proceedings of the 10th Eurographics UK '92 Conference*, University of Edinburgh, Scotland, April 1992.

Index