Molecular
Simulations

BIOGRAF

Tutorial

Version 3.0

June 15, 1992

Molecular Simulations, Inc.
200 Fifth Ave.
Waltham, MA 02154

Contents

1 Introduction 1-1

1.1 Starting BIOGRAF 1-2
  1.1.1 Executing a Macro on Start Up 1-2
1.2 The Graphical User Interface (GUI) 1-4
  1.2.1 The Graphics Window 1-5
  1.2.2 The Logo Window 1-5
  1.2.3 The Modules/Messages Window 1-5
  1.2.4 The Menu Pad 1-5
  1.2.5 The GRAPHICS, TEXT, and PLOTS Buttons 1-5
  1.2.6 Interfacing with the Program 1-6
1.3 Manipulating the Graphics Display Image 1-7
  1.3.1 The Dials Unit 1-7
  1.3.2 The Mouse 1-9
  1.3.3 The Function Keys 1-12
1.4 Viewing the Image in Three Dimensions 1-14
  1.4.1 Using Titan's Stereo Monitor 1-14
  1.4.2 Using the Personal Stereo Viewer on the IRIS 1-15
1.5 Selecting Menu Boxes and Buttons 1-16
1.6 Using the Menus 1-17
  1.6.1 The Functions Menu Tree 1-17
  1.6.2 The TOP MENU Button 1-18
  1.6.3 The Items on the Functions Menus 1-18
  1.6.4 Basic Conventions for Using the Menus 1-20
1.7 Inputting Information via Dialog and Alert Boxes 1-22
1.8 Picking Atoms in the Graphics Window 1-23
  1.8.1 Picking a Range of Atoms 1-23
1.9 Displaying Groups - The DISPLAY Button 1-24
1.10 Entering File Names 1-25
  1.10.1 The Browser Box 1-25
  1.10.2 The New Dir and New File Buttons 1-28
1.11 Resetting the Program 1-28
1.12 Using Command Mode - The COMMAND Button 1-29
1.13 Running the Program in the Background 1-30
1.14 Basic Functions 1-31
1.15 Conclusions 1-32

2 Displaying Molecular Structures 2-1

2.1 Getting Started 2-1
2.2 Reading the Data 2-2
2.3 Looking at the Groups 2-3
2.4 Creating Groups 2-5
2.5 Creating the Trace 2-7
2.6 Choosing the Primitives 2-9
2.7 The Half-Bond Representation 2-10

2.8 Creating the Main Chain 2-10
2.9 Changing the Group Color 2-11
2.10 Creating the Side Chain, Carbonyl, Water, and Labels Groups 2-12
2.11 Non-standard Groups 2-13
2.12 Specifying Residues 2-13
  2.12.1 Keying the Residues 2-14
  2.12.2 Looking at the Ligands Group 2-14
  2.12.3 Picking the Residues 2-15
2.13 Using Structural Descriptors (Selectors) 2-16
2.14 Saving Information 2-17
  2.14.1 Group Files 2-17
  2.14.2 Coordinate Files 2-18
  2.14.3 Testing Our Work 2-19
2.15 Conclusion 2-20
2.16 Advanced Notes 2-21

3 Small Molecule Building and Drawing 3-1

3.1 Getting Started 3-1
3.2 Setting Up the Builder 3-2
3.3 The Half-Bond Representation 3-3
3.4 Setting up the Base Fragment 3-4
3.5 Adding the Functional Groups 3-5
3.6 Deleting Hydrogen Atoms 3-6
3.7 Changing Element Types 3-7
3.8 Drawing Structures 3-9
3.9 Adding Double Bonds with the Bond Operation 3-10
3.10 Adding the Missing Hydrogen Atoms 3-10
3.11 Getting the Correct Stereochemistry 3-12
3.12 Minimizing the Energy 3-14
3.13 Saving Information 3-16
3.14 Conclusions 3-17
3.15 Advanced Notes 3-18

4 Creating Surfaces 4-1

4.1 How Molecular Surfaces are Created 4-1
4.2 The Surface Menu 4-2
4.3 Surface Operations 4-3
  4.3.1 Calculate 4-3
4.4 Creating a van der Waals Surface for CLT 4-4
  4.4.1 Reading the CLT Data 4-4
  4.4.2 Calculating the van der Waals Surface 4-6
4.5 Display Operations 4-6
  4.5.1 Setting Up Groups for Display 4-7
  4.5.2 Color-Coding the Surface to Show Properties 4-7
4.6 Saving the Dot Surface 4-10
4.7 Using Dot Surfaces to Study An Enzyme/Inhibitor Complex 4-10
  4.7.1 Reading in the Thermolysin and CLT Data 4-10
  4.7.2 Looking at the Molecular Structures of Thermolysin and CLT 4-11
  4.7.3 Reading and Displaying the CLT Dot File 4-12
  4.7.4 Reading and Displaying the Active Site Dot File 4-13

4.8 Conclusion 4-14

5 Docking 5-1

5.1 Starting the Program and Reading the Data 5-1
5.2 Changing the Group Color 5-4
5.3 Creating Groups 5-5
  5.3.1 Making the AT Base Pairs Group 5-6
  5.3.2 Making a Group to Hold the Original Conformation 5-8
5.4 Docking 5-9
  5.4.1 Selecting the Groups to be Docked 5-9
  5.4.2 Rotating the Image 5-10
  5.4.3 Docking the Netropsin 5-11
  5.4.4 Storing the New Coordinates 5-11
5.5 Calculating Energies 5-13
  5.5.1 Setting up an Energy Expression 5-13
  5.5.2 Minimizing the Energy 5-15
5.6 Saving Information 5-17
  5.6.1 Group Files 5-17
  5.6.2 Coordinate Files 5-18
  5.6.3 Testing Our Work 5-19
5.7 Advanced Notes 5-20

6 Molecular Dynamics of Small Peptides 6-1

6.1 Building the Peptide 6-2
6.2 Minimizing the Energy 6-4
6.3 Running Dynamics 6-7
  6.3.1 Setting the Defaults 6-8
  6.3.2 Doing the Calculation 6-11
6.4 Minimizing the Structure Again 6-12
6.5 Saving Information 6-13
6.6 Rendering a Group in Monochrome 6-14
6.7 Animating a Dynamics Trajectory 6-15
  6.7.1 Defining the Trajectory File and Listing the Conformers 6-15
  6.7.2 Reading in the Positions of the Conformers 6-16
  6.7.3 Animating the Conformations 6-17
  6.7.4 Using Auto Scan 6-17
  6.7.5 Looking at the Low Energy Conformers 6-18
  6.7.6 Showing All the Conformations at Once 6-18
  6.7.7 Extracting the Best Conformer 6-20
6.8 Creating a Macro and Running It in the Background 6-21
6.9 Conclusion 6-23

7 Site-Directed Mutagenesis 7-1

7.1 Executing a Macro on Start Up 7-1
7.2 Looking at the Groups 7-2
7.3 Creating Groups 7-4
  7.3.1 Defining the Tyrosine 82 Group 7-5
  7.3.2 Choosing the Primitives 7-7
7.4 Replacing the Side Chain 7-8

7.5 Changing the Center of Rotation 7-10
7.6 Rotating the Side Chain 7-11
7.7 Adding Hydrogens 7-12
  7.7.1 Hydrogen Atom Types 7-12
  7.7.2 Heteroatoms 7-13
7.8 Creating Groups Based on Distance Criteria 7-14
7.9 Saving Information 7-15
  7.9.1 Group Files 7-15
  7.9.2 Coordinate Files 7-16
  7.9.3 Testing Our Work 7-17
7.10 Calculating Energies 7-19
  7.10.1 Setting Up an Energy Expression 7-19
  7.10.2 Fixing the Cytochrome C Backbone 7-20
  7.10.3 Partitioning the Energy 7-22
  7.10.4 Minimizing the Energy 7-23
7.11 Doing a Conformational Energy Search 7-24
  7.11.1 Creating a Search Group and Setting up the Energy Expression 7-25
  7.11.2 Defining the Search Space 7-26
7.12 Displaying the Search Results 7-27
  7.12.1 The Best Conformations Method 7-27
  7.12.2 Rendering a Group in Monochrome 7-28
  7.12.3 Animating a Conformational Energy Surface 7-29
7.13 Conclusion 7-35
7.14 Advanced Notes 7-35

A Details of Molecular Dynamics A-1

Chapter 1

Introduction

BIOGRAF is a menu-driven program for biological and chemical simulations. The program builds molecular structures, optimizes conformational energies, calculates the dynamics of motion for molecular systems, and analyzes molecular structure and dynamics. BIOGRAF can be run interactively or as a batch job (in the background). It employs an X-Windows based graphical user interface (GUI), making it highly portable and capable of running on a wide range of platforms. The interface is designed to be both easy to learn and easy to use. The program has the power to handle structures with more than ten thousand atoms, yet is simple enough to use on diatomic molecules.

BIOGRAF is driven by selecting operations from the Functions menus at the left of the Menu Pad. Other menus are also used to display information or choices which are needed to execute a particular function. The program may also be driven by entering commands from the keyboard. This command mode feature is particularly useful in that it allows the user to create macros so that a series of commands can be automatically executed.

This chapter describes the human interface. It includes information on how to use the menus, dialog boxes, and buttons. It also provides a general introduction to the program, including information on basic operations such as starting and resetting the program, manipulating the graphics display, viewing the image in three dimensions, using the menus, inputting data, and selecting atoms. A brief description of the top level functions is also included at the end of the chapter. Subsequent chapters illustrate the use of selected functions. All tutorials assume that you have read Chapter 1. However, the remaining chapters are independent, *i.e.,* each may be done without knowledge of the others.

It is assumed that you are familiar with basic windowing operations including how to open, close, push down, pop up, reposition, and resize windows, as well as how to use scroll bars and slider bars.

1.1 Starting BIOGRAF

BIOGRAF is X-Windows-based and runs on systems using the UNIX operating system. It is therefore highly portable and presents a common interface on a wide range of platforms. If you have not yet installed BIOGRAF, go through the procedures on how to use the COMSTRUCT program described in the *Molecular Simulations Installation Guide*. COMSTRUCT is used to create the command procedures and defaults files that assign all the directories, input and output files and devices that BIOGRAF needs to operate and then run the program.

To start up the program, you would generally log on and type in the start up command (***biograf*** is commonly used, but this command may be different at your site). Note that UNIX is case sensitive. For example, the name "biograf" is different than "BIOGRAF" or "Biograf". All standard BIOGRAF commands and file names are in lower case with UNIX. You may also read in and display a molecule on start up. If the molecule is in a BioDesign General Format (*bgf*) file it can be read in by including the file name at the end of the start up command (*e.g., biograf filename.bgf*).

1.1.1 Executing a Macro on Start Up

Macros may also be specified in the command line used to start up and run the program (*e.g., biograf filename.macro*). If the macro is not located in the current working directory, then directory specifications must also be included. We have a macro file which reads in the files needed for this tutorial. It is called *dna.macro* and is stored in your distribution macro directory. We'll specify this macro as part of the start up command. First, we'll find out where your distribution macro directory is located so that we can include the proper path specifications.

Typically the full specifications for the distribution macro directory are */usr/msi/biogv300/macro* The distribution macro files are always located in the */biogv300/macro* directory; however, the first part of the path specifications may differ on your system. You can determine the location of the */biogv300/macro* directory on your system by typing *echo $MSIUSR* from a systems window.

1. Log on and type ***echo $MSIUSR***. The directory listed in response to this command (*e.g., /usr/msi*) should precede */biogv300/macro* when giving specifications for your distribution macro directory.

2. Type in the start up command followed by the specifications for your distribution macro directory and the macro file name ***dna.macro***.

   Example: ***biograf /usr/msi/biogv300/macro/dna.macro***

A start up window appears displaying the logo and copyright notice. The defaults, parameter conversion files, and other pertinent data are read in and the five windows of the final start up display appear (see Figure 1.1).

The program then quickly executes the series of commands in the *dna.macro* file. Messages appear while the appropriate coordinate and group files are read in. The two chains of the DNA double helix appear in the graphics window along with the anti-viral drug netropsin. We are looking down the DNA spiral from the top. The last command in the macro file ("menu") switches us to interactive mode. The Top menu is displayed in the Menu Pad and the program is ready to accept input.

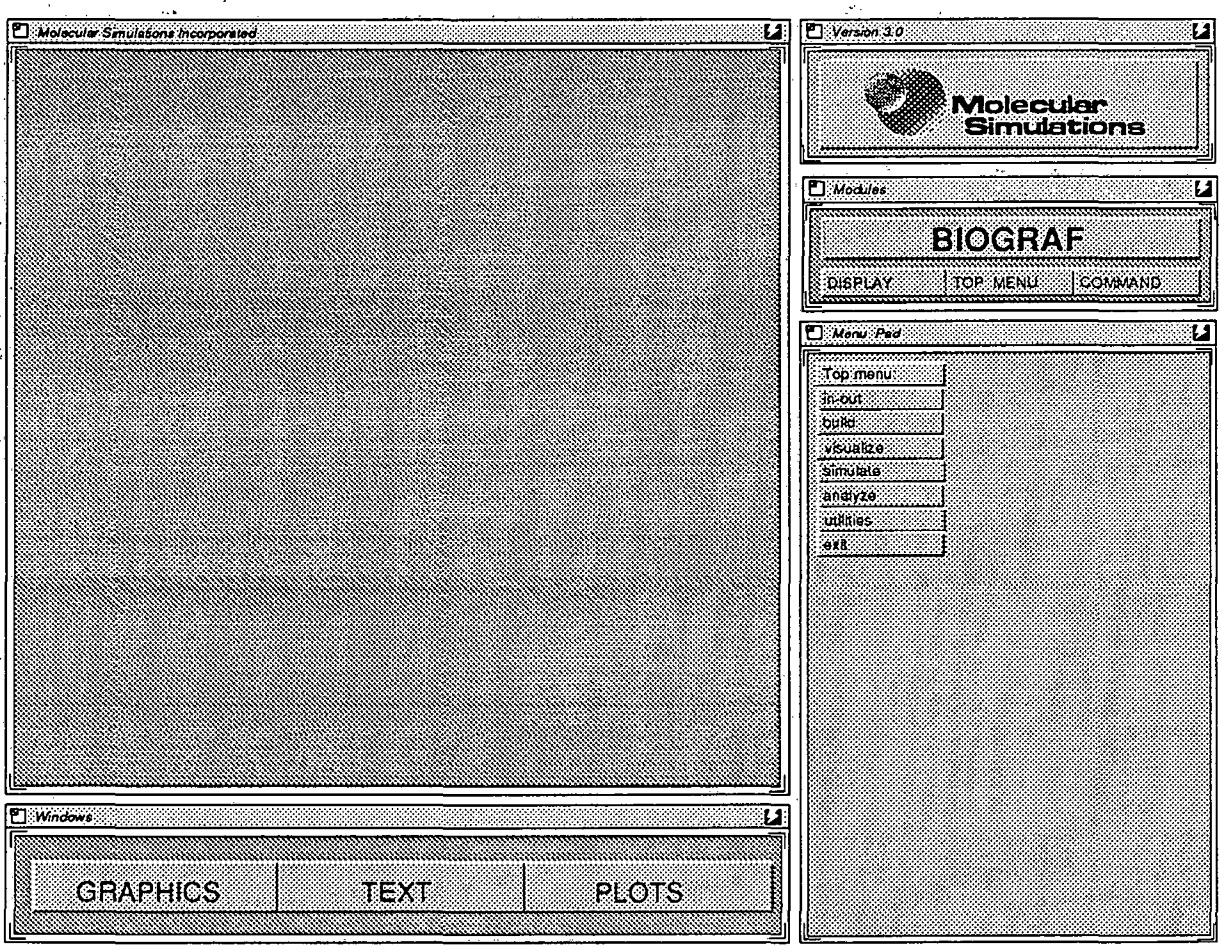


Figure 1.1: BIOGRAF start-up display

1.2 The Graphical User Interface (GUI)

The program employs an X-Windows-based graphical user interface (GUI) which uses standard widgets including file browsers, dialog boxes, alert boxes, slider bars, and buttons. The program is basically menu-driven, easy to learn, and easy to use. Commonly used functions are readily accessible via constantly displayed buttons. Buttons have also been provided which make it easy for the user to switch between displaying the text window and the graphics window. A file browser feature speeds up file input and makes it easy to change your default directory.

The screen is divided into five sections as shown in Figure 1.2:

1) A graphics window for displaying molecules (upper left)
2) A logo window (upper right)
3) A Modules/Messages window directly beneath the logo which contains a Modules button and the DISPLAY, TOP MENU, and COMMAND buttons (middle right)
4) A Menu Pad window which contains the menus (lower right)
5) A window containing the GRAPHICS, TEXT, and PLOTS buttons (lower left)

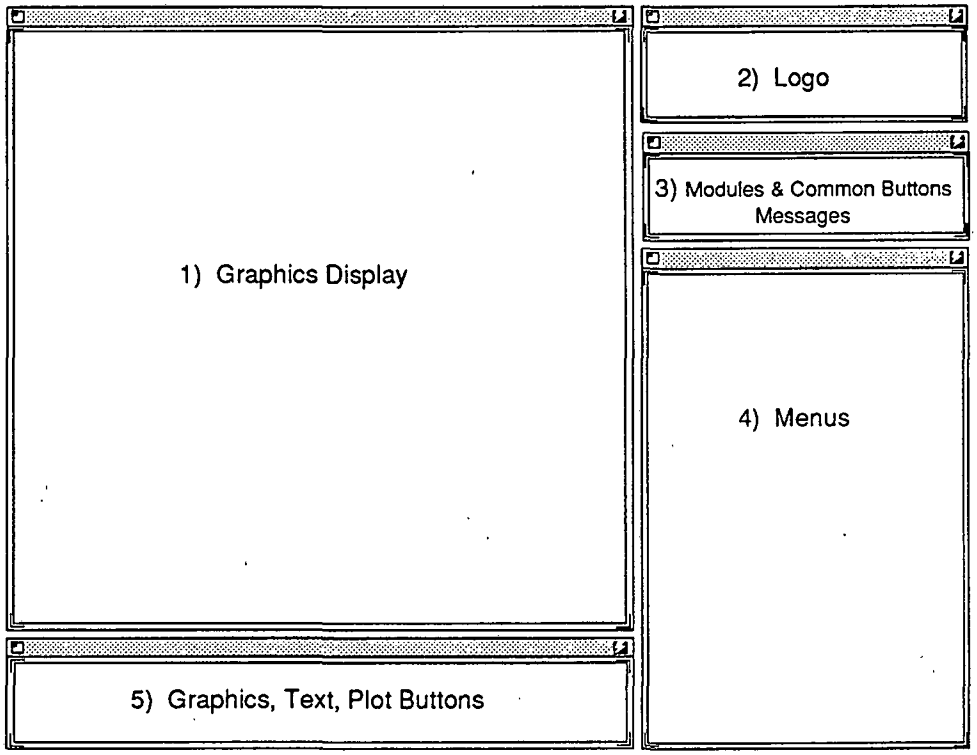


Figure 1.2: BIOGRAF screen layout

Each of these windows can be relocated, resized, or closed if desired. The system window (used to issue the start up command) in addition to displaying system messages also displays results and information for a number of program functions. It is referred to here as the text window and can be readily viewed by clicking the TEXT button.

1.2.1 The Graphics Window

The graphics window is used for the display and visualization of molecules. The graphics window must be active in order to manipulate molecules with the mouse or to use the function keys. One can usually activate the graphics window by simply placing the cursor in it. However, some systems also require you to click in the graphics window in order to activate it.

1.2.2 The Logo Window

The logo window is used to display the Molecular Simulations logo.

1.2.3 The Modules/Messages Window

The Modules/Messages window includes a modules button (top) which is used to switch between the different Molecular Simulations products or modules that have been installed on the system (*e.g.*, BIOGRAF, POLYGRAF, NMRgraf, and Polaris). The name of the module currently being used is displayed. You must be at the top level menu in order to switch. Beneath the modules button are three buttons used to implement commonly used program functions. These are the DISPLAY, TOP MENU, and COMMAND buttons (these are described later). This window is also used to display prompts and messages. When this is required, the other buttons are overwritten with the message or with a dialog box requesting input from the user. Prompts and messages are used extensively by the program and are the primary means of communication with the user. The user should therefore check this window regularly.

1.2.4 The Menu Pad

The menus are the means by which BIOGRAF accepts commands. The Functions menus are used to drive the program. They are always located at the left of the Menu Pad. Additional menus appear adjacent to the Functions menus which display additional operations or choices which are needed to execute a particular function. These additional menus appear only when input is needed from the user. They include the Files menu, which is used to display the names of the files which are currently in the system, the Groups menu, which displays the groups currently in the system, and Selections menus which display the choices available for the currently active function. The Energy results menu is also displayed here during minimization and dynamics calculations; it lists several key energy-related values. The use of the menus is described in more detail in Section 1.6.

1.2.5 The GRAPHICS, TEXT, and PLOTS Buttons

The GRAPHICS, TEXT, and PLOTS buttons allow the user to easily switch between viewing the graphics window, text window, and plot window. For example, when running minimizations or molecular dynamics calculations, information related to the energy calculations appears in the text window. This simulation information consists of several lines of text and is often more fruitfully viewed in a *large* text window. However, a large-sized text window might obscure visualization of the molecules in the graphics window. One would probably like to display a large graphics window when building and visualizing molecules, but bring up a large text window when running energy calculations. The GRAPHICS and TEXT buttons allow the user to quickly switch between these two alternatives. These buttons are not operative, however, when in COMMAND mode or while running energy calculations. The TEXT button should therefore be used *before* going into command mode or starting any energy calculations.

1. Use the mouse to position the cursor over the TEXT button. Click the TEXT button to bring the text window forward (if there are any other system windows currently open, they will also be brought forward).

2. Reposition the text window so that it is located at the lower left of the screen but is above the GRAPHICS, TEXT, and PLOTS buttons window.

3. Click the GRAPHICS button to bring the graphics window forward.

   Note: On Silicon Graphics IRIS 4-D workstations, the F11 key may also be used to bring the text window forward.

Similarly, the PLOTS button will be used to bring up a window displaying plots of various functions. (The PLOTS button feature is not available in this release.)

1.2.6 Interfacing with the Program

The program communicates with the user by printing messages in the Modules/Messages window and by bringing up dialog and alert boxes. Also for certain operations, results and other informative data are displayed in the text window.

The user communicates with the program in three ways: by clicking in menu boxes and buttons, by typing information or making selections in dialog boxes, and by picking atoms in the graphics window. When in command mode, the user communicates by entering commands in the text window. The images in the graphics window are manipulated using the dials or mouse. Special function keys may also be used. These methods of interfacing with the program are described in more detail in the following sections.

Throughout this tutorial, all menu titles, subtitles, and menu items in the "on" state appear on the menu in **bold** format. Also, **bold** is used when referring to a menu item (*e.g.*, To abort the function, pick **return**). Buttons that appear in uppercase are referred to in UPPERCASE (*e.g.*, click the DISPLAY button); otherwise lowercase is used (*e.g.*, Click the Cancel button). Information that is typed in by the user is referred to in **bold** and *italics* (*e.g.*, Enter ***test.bgf*** for the file name) and prompts and messages are referred to in *italics* (*e.g.*, The message *Working...* appears while the calculations are being performed).

1.3 Manipulating the Graphics Display Image

Most of the graphics systems that run BIOGRAF provide a mouse, special function keys, and dials which allow you to manipulate the image shown in the graphics window. These differ somewhat depending on the graphics system.

1.3.1 The Dials Unit

A dials unit is optional on Silicon Graphics IRIS 4-D workstations and Stardent's Titan Graphics Supercomputer. A dials unit is also available for the IBM RISC System/6000, but is currently not supported. (Henceforth, these systems will be referred to respectively as the IRIS, the Titan, and the RS6000. The Evans & Sutherland ESV, Digital Equipment's DS5000, and Hewlett Packard's HP9000 will be referred to respectively as the ESV, the DS5000, and the HP9000). The dials unit consists of eight dials which allow you to move an image within the molecular display area. These include six dials which are used to either rotate or translate the image along the x, y, or z axis, a zoom dial which expands or shrinks the image, and a depth cueing dial.

The orientation and position of the dials differs depending on the graphics system; diagrams of the different types of dials units are shown in Figure 1.3 below.

x rotate 1	x translate 5
y rotate 2	y translate 6
z rotate 3	z translate 7
zoom 4	depth 8

A) Dials unit for Silicon Graphics IRIS 4-D Workstations

x translate 5	y translate 6	z translate 7	depth 8
1 x rotate	2 y rotate	3 z rotate	4 zoom

B) Dials unit for Stardent's Titan Graphics Supercomputer

Figure 1.3: Dials units for different graphics systems

The dials are labeled and numbered. Note the position of dial 8 for your system. This dial has multiple functions. It is normally used for depth cueing. However, when animating a trajectory, dial 8 is used to cycle through the different conformers being displayed and on some systems it controls the scan speed during auto scan. It is also used to rotate a portion of a molecule about a designated bond. Dials 1 through 7 may also take on this function when multiple bonds are being rotated simultaneously.

The x, y, and z axes are defined as shown in Table 1.1.

x	⇨	horizontal	positive to right
y	⇨	vertical	positive up
z	⇨	out of the screen	positive toward you

Table 1.1: Definition of x, y, and z axes

The edges of the graphics display define the limits of x and y. The x, y, and z axes are fixed in their own frame of reference instead of on the image: *e.g.*, turning the y rotate dial always makes the image pirouette around the vertical axis of the graphics display regardless of its previous orientation.

When an image is depth cued, its intensity decreases with its apparent distance from the viewer, creating a three-dimensional effect. The parts of the molecule "closest" to you look the brightest and the parts farthest away look the faintest. On depth cued systems there are two *clipping planes* parallel to the display that define the allowed range of z. The front clipping plane is usually fixed, while the rear clipping plane can move. The intensity of the distant atoms is proportional to the separation between the rear and front clipping planes. The depth cueing dial controls the position of the rear clipping plane. Moving the rear clipping plane away from you causes distant parts of the molecule to become brighter; moving it toward you progressively dims the back of the molecule until only a thin slab can be seen.

Warning (Titan users only): We have observed that rotating the dials on the Titan during an operation that updates the display (*e.g.*, energy minimization, dynamics calculations, auto scan) sometimes causes the program to crash. It would be prudent to minimize the use of the dials while these operations are being performed.

1.3.2 The Mouse

The mouse may also be used to manipulate the image in the graphics window. The graphics window must be active, however. Table 1.2 summarizes the button and key presses and mouse movements required for each function.

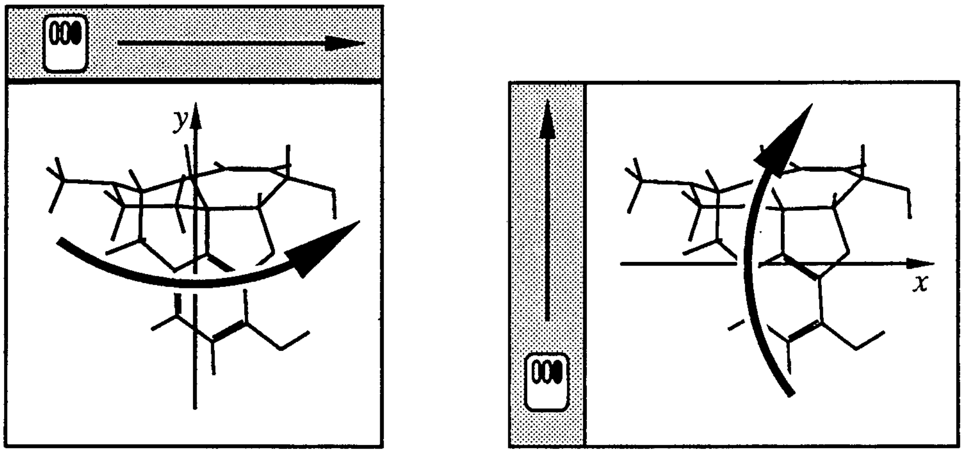
Depressing the middle mouse button and moving the mouse allows you to drag the image around the screen (this corresponds to the x and y translate functions). To translate the image along the z axis, the shift key and the middle mouse button must both be depressed while moving the mouse; movement up or to the right translates the image in the +z direction (toward the viewer), while moving the mouse down or to the left translates it in the -z direction (away from the viewer).

Similarly, depth cueing may be altered by depressing the control key and the middle mouse button; moving the mouse up or to the right dims the back portions of the image, while moving it down or to the left intensifies them. (When animating a dynamics trajectory, depressing the middle mouse button and control key and moving the mouse allows you to display earlier or later conformations, while when auto scanning, it allows you to change the scan speed.)

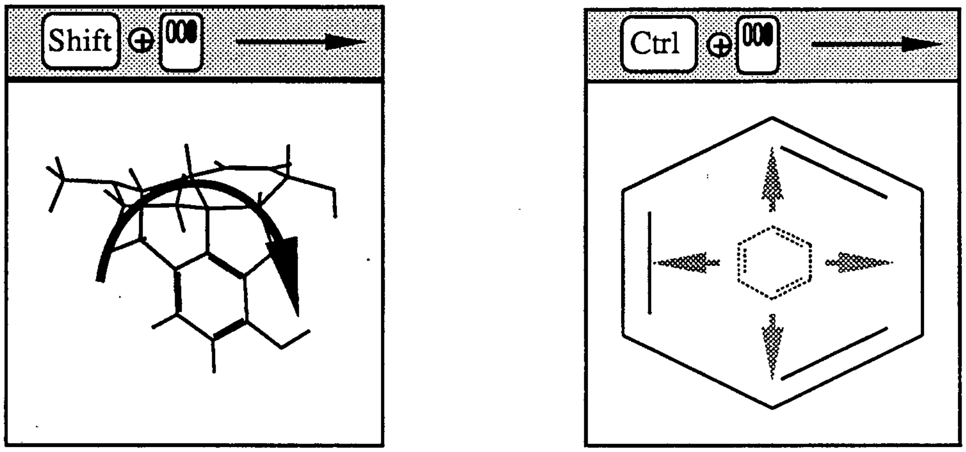
Function	Mouse Button	Key	Mouse Motion
x translate	middle	-	right /left
y translate	middle	-	up/down
z translate	middle	shift	up or right (toward viewer), down or left (away from viewer)
depth cueing	middle	control	up or right (dims distant objects), down or left (intensifies distant objects)
x rotate	right	-	up/down
y rotate	right	-	right /left
z rotate	right	shift	up or right (clockwise), down or left (counter-clockwise)
zoom	right	control	up or right (enlarges), down or left (reduces)

Table 1.2: Using the mouse to manipulate the image in the graphics window

The right mouse button is used for the rotate and zoom functions (see Figure 1.4). Depressing the right mouse button and moving the mouse to the right or left rotates the image around the y axis, while moving the mouse up or down rotates the image around the x axis. For z rotations, the shift key must also be depressed. In this case, moving the mouse up or to the right rotates the image clockwise around the z axis, while moving the mouse down or to the left rotates it in the counter-clockwise direction. To resize the image, the control key must be depressed in conjunction with the right mouse button. Moving the mouse up or to the right enlarges the image, while movement down or to the left makes it smaller.



Y Rotate X Rotate



Z Rotate Zoom

Figure 1.4: Use of mouse for rotate and zoom functions

1. If using the mouse, activate the graphics window by placing the cursor in it and if necessary clicking with the mouse .
2. Rotate the DNA/netropsin around the y axis until it is lengthwise and horizontal on the display.
3. Rotate the DNA around the z axis until it is vertical on the display.
4. Translate the molecules along the different axes, then return them to the center of the display.
5. Zoom the molecules "in" and "out".

6. If zooming the molecules moves them off the display, use the translate functions to recenter them.

   The remaining steps (7 - 10) illustrate the depth cueing function.

7. Translate the image towards you using either the z translate or zoom functions. Note how the brightest sections of the molecules disappear as they pass forward of the front clipping plane.

8. Use the depth cueing function to bring the rear clipping plane forward. (If using the mouse, depress the middle mouse button and the control key and move the mouse up or to the right; if using the dials, turn the depth cueing dial counterclockwise.) The most distant parts of the molecule grow fainter and fainter as the rear plane approaches; they finally disappear altogether when it passes them.

9. Use the depth cueing function to push back the rear clipping plane. (If using the mouse, depress the middle button and the control key and move the mouse down or to the left; if using the dials, turn the depth cueing dial clockwise.) As the plane recedes, the parts of the image that appear farthest away brighten until the entire display is uniform in intensity. Note how the DNA molecule no longer looks three-dimensional.

10. Set the depth cueing to your satisfaction. By adjusting the position of the image and the clipping planes, you can selectively remove front and back portions of the molecule to examine cross-sections of its structure.

1.3.3 The Function Keys

Several function keys are provided which can be used to modify the display. These are function keys F1 - F5, and F11. They are described in Table 1.3 below. In addition, the program makes use of an abort key. This is used to stop any interactive molecular mechanics calculation after completing the current step or to abort charge equilibration calculations. The abort key is F8 for the HP9000, F10 for the Titan, and F12 for all the other machines. The graphics window must be active in order to use the function keys.

Key	Description	Iris	Titan	RS6000	HP9000	ESV	DS5000
F1	Removes menus, enlarges graphics window	Yes	Yes	Yes	Yes	Yes	Yes
F2	Displays two images for split screen stereo viewing	Yes	Yes	Yes	Yes	Yes	Yes
F3	Rotates right image displayed by F2 key by 90°	Yes	Yes	Yes	Yes	Yes	Yes
F4 or F5	Activates hardware for stereo viewing	Yes	Yes	No	No	No	No
F11	Brings text window forward	Yes	---	Yes	---	No	No
F8, F10, F12	Aborts molecular mechanics and charge equilibration calculations	Yes	Yes	Yes	Yes	Yes	Yes

Table 1.3: Function keys

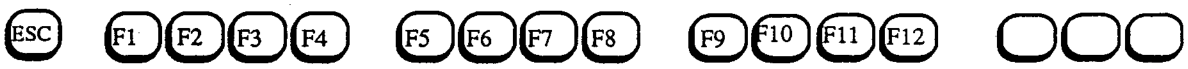
The position of the function keys on the keyboard differs depending on the graphics system. On the IRIS, RS6000, and ESV there are 12 function keys along the top of the keyboard, while the DS5000 has 14 function keys at the top of the keyboard. The Titan has 10 function keys in a keypad at the left, and the HP9000 has 8 function keys at the top of the keyboard (see diagrams A through D in Figure 1.5).

1. Make sure the graphics window is active (place the cursor in it, and if necessary, click with the mouse).
2. Press the F1 key to remove the menus and enlarge the graphics window.
3. Press the F2 key to display two images for split-screen stereo viewing. Cross your eyes to see the image in three dimensions.
4. Press the F3 key to rotate the image on the right by 90° about the x axis.

5. Translate the molecule along the various axes and note how the orthogonal projections give different views of the image. This feature is especially useful when you want to move different molecules relative to each other (*e.g.*, when docking molecules).

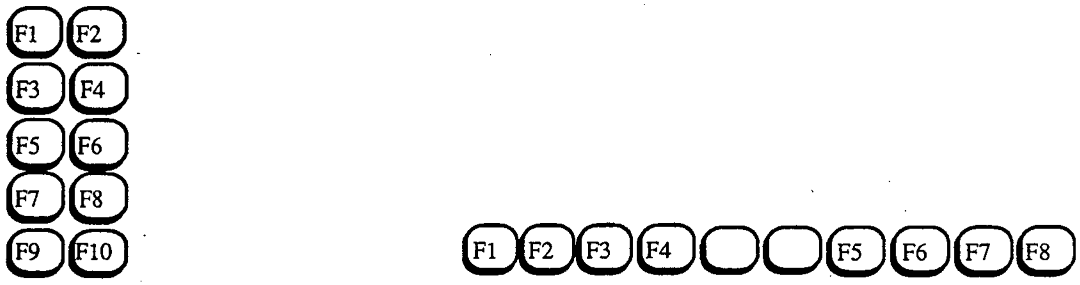
6. Press the F3, F2 and F1 keys to restore the display to normal.

7. IRIS users only: press the F11 key to bring the text window forward. Press it again to push the text window back.



A) Function keys on Silicon Graphics IRIS 4-D, IBM RISC System/6000, and Evans & Sutherland ESV Workstations (top of keyboard).

B) Function keys on DS5000 Workstation (top of keyboard)



C) Function keys on Stardent's Titan Supercomputer (left side of keyboard)

D) Function keys on Hewlett Packard's HP9000 Workstation (top of keyboard)

Figure 1.5: Function keys for the different graphics systems

1.4 Viewing the Image in Three Dimensions

The program provides a **stereo** function which renders the image in the molecular display area as a stereo pair so that it can be seen in three-dimensions. A trained user can easily see the three-dimensional image (you must either cross your eyes (proximal mode) or rotate them outward (distal mode). The **stereo** function is described under Visualize Operations, in Section 4.6 of the *BIOGRAF Reference Manual.* Alternatively, special equipment may be purchased for some systems which allows you to view images in three dimensions (without crossing your eyes). On the Titan, a special stereo monitor may be purchased, while on the IRIS, special glasses (Personal Stereo Viewer) may be used. This stereo hardware is activated by pressing the F5 key on the Titan and the F4 key on the IRIS. On the Titan, however, a special alias command must be set up first. This is described below.

If you don't have a Titan with a stereo monitor, or an IRIS with the special glasses (Personal Stereo Viewer), skip to Section 1.5.

1.4.1 Using Titan's Stereo Monitor

In order to enable the stereo monitor on the Titan, an alias command which will execute the stereo version of the program must be set up. This is done by editing a start up macro to include the stereo option, running COMSTRUCT with the desired alias command, and editing the defaults and .cshrc files. This procedure is described in the *Molecular Simulations Installation Guide.* This procedure needs to be performed only once. If you have a Titan with the stereo monitor and this alias has not yet been set up, do so now (or have your system supervisor do it for you).

A few more steps must also be done *after every log in* to get into stereo mode.

1. Enter the X-Windows environment by typing ***xstart −stereo***

2. Type ***xterm −display unix:0.1&***

3. Hold down ALT and press the F2 key to go into the stereo screen environment.

4. Type ***setenv DISPLAY unix:0.1***
   awm&

5. Run the program in stereo mode by typing the alias command set up previously, followed by the macro file name for reading in DNA (*e.g.*, ***biostereo dna.macro***).

   When the DNA molecule comes up, you should be able to see it in three dimensions.

6. Use the F5 key to activate the stereo mode (F5 toggles you back and forth between stereo and normal modes). It may be necessary to reposition the window.

7. Rotate and translate the molecule along the different axes to get a good three-dimensional view.

8. If desired, use the depth-cueing function.

1.4.2 Using the Personal Stereo Viewer on the IRIS

1. Press the F4 key to enable the stereo glasses (Personal Stereo Viewer).

2. Flip the appropriate switches on the glasses to turn them on, get into stereo mode, set the phase, *etc.* (consult the appropriate SGI manuals for further details).

   When set properly, you should be able to see the DNA molecule in three dimensions.

3. Rotate and translate the molecule along the different axes to get a good three-dimensional view.

4. If desired, use the depth-cueing function.

5. When finished viewing the molecule, press the F4 key to return to normal mode. Turn the glasses off when finished as they are battery-powered.

1.5 Selecting Menu Boxes and Buttons

The user interacts with the program primarily by selecting menu boxes and buttons and by picking atoms on the graphics display with the mouse. All three of the mouse buttons may be used when selecting menu items and buttons, but only the left button is operative when picking atoms. It is important to note that the position of the mouse on the mousepad is irrelevant; it is the x and y movement of the mouse that determines the position of the cursor. Each menu label appears bright when the cursor is positioned over it. The menu item is not selected however until you click in the box with the mouse, causing it to be highlighted (*i.e.*, the background becomes bright).

1. Place the cursor over the menu items in the Top menu. Note how each of the menu labels appears bright when the cursor is positioned over it.

2. Click in the **analyze** box. You have just selected the **analyze** function. The analyze box is momentarily highlighted and the Analyze submenu immediately appears, displaying additional functions.

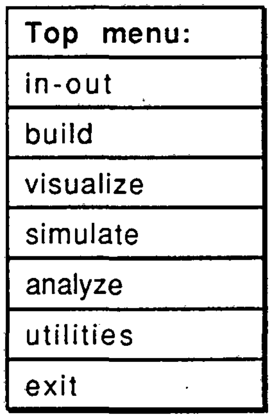


Figure 1.6:
Top menu

1.6 Using the Menus

The menus are the means by which BIOGRAF accepts commands. The Functions menus are used to drive the program. They are always located at the left of the Menu Pad. Additional menus appear adjacent to the Functions menus which display additional operations or choices which are needed to execute a particular function. These additional menus appear only when input is needed from the user. They include the Files menu, which is used to display the names of the files which are currently in the system, the Groups menu, which displays the groups currently in the system, and the Selections menus which display the choices available for the currently active function. The Energy results menu is also displayed here during minimization and dynamics calculations; it lists several key energy-related values.

1.6.1 The Functions Menu Tree

Operations are initiated by picking one of the functions from the top level Functions menu which appears at start up (the Top menu). When one of these functions is selected, the top level menu is replaced by a submenu displaying additional functions. These functions may also have their own submenus, *etc*. Thus, we have a Functions menu tree which consists of several levels. We started the session at the Top menu, then picked **analyze** to bring up the Analyze submenu, so we are currently down one level in the Functions menu tree.

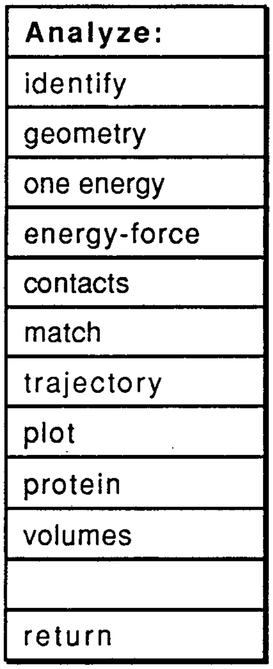


Figure 1.7: Analyze menu

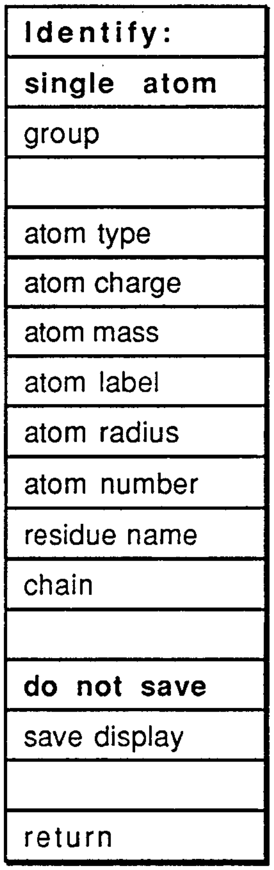


Figure 1.8: Identify menu

1. Pick **identify** from the Analyze menu. The Identify menu shown in Figure 1.8 appears. We are now down two levels in the Functions menu tree.

   The top box of every Functions menu is always partially highlighted (the background is light gray) and contains the current menu's name which always ends in a colon. Picking the menu title has no effect. Every Functions menu also has a **return** function at the bottom. Picking the **return** box restores the previous Functions menu, taking you up one level in the menu tree.

2. Pick **return**. We are brought up one level to the Analyze menu.

3. Pick **return** again. We are brought up another level to the Top menu.

1.6.2 The TOP MENU Button

As you have seen, picking the **return** box in any Functions menu always restores the previous Functions menu and takes you up one level in the menu tree. However, if you are several submenus down in the tree, this must be done repeatedly in order to get back to the top level menu. Picking the TOP MENU button in the Modules/Messages window speeds this process up and immediately returns you to the top level no matter where you are in the Functions menu tree.

1. Pick **analyze,** then pick **identify** to bring you down two levels as before.
2. Click the TOP MENU button located in the Modules/Messages window. You are immediately returned to the Top menu.

1.6.3 The Items on the Functions Menus

The items on each of the Functions menus do one of four things:

1) bring up a submenu displaying additional functions.

2) execute the function indicated by the menu item's name. This may be done immediately or after additional information is obtained from the user.

3) indicate a parameter value (usually in the box directly beneath the item). The parameter values are non-pickable menu items and are displayed in a different color so they may be easily recognized. Generally, picking the menu box containing the parameter name brings up prompts which allow you to change the associated value.

4) indicate the current state or mode which will be applied when another operation is performed. This includes items which are either in the "on" or "off" state where "on" is indicated by highlighting the menu item, and items which cycle through different possibilities where the one currently displayed is the one which is applied.

The In-Out menu provides some examples of the different types of menu items.

1. Pick **in-out** to bring up the In-Out menu (see Figure 1.9).

   As mentioned previously, menu titles always appear at the top, end in a colon, and are partially highlighted (a light gray background is used). Picking them has no effect. In addition to the menu title at the top, we have two subtitles on this menu (Read defaults: and Write defaults:). Subtitles are distinguishable from regular menu items since they are displayed in a different color (the default label color is blue) and always end with a colon. Picking subtitles also has no effect.

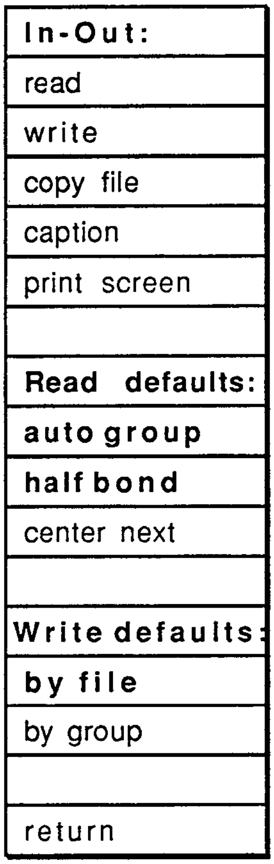


Figure 1.9:
In-Out menu

2. Pick **In-Out:**, **Read defaults:**, and **Write defaults:**. Notice that nothing happens.

   The **caption** and **print screen** menu items are examples of options which bring up their own submenus. These submenus display additional options used in implementing the caption and print screen functions.

3. Pick **print screen**. Notice that the Print screen submenu appears displaying additional options which are used in implementing the **print screen** function. Pick **return** to get back to the In-Out menu.

   The In-Out menu also has examples of on/off menu items. Two of the read defaults (**auto group** and **half bond**) are on/off menu items that are currently in the "on" state, indicating that these options will be applied when the operation they refer to (the **read** operation) is performed. This means that a group will automatically be created for the file to be read in and that the new molecule will be displayed with the half-bond colors (rather than in monochrome). Toggling these items turns them off (they no longer appear highlighted).

4. Toggle **auto group** and **half bond** off and on. Notice how the highlighting changes. Leave them both at their default settings (on).

   The third read default (**center next**) is an example of a cyclic menu item. It controls the centering of molecules on the display. Toggling this box cycles through three different centering methods: **center next, center all**, and **no center**. The one which is currently shown in the menu box is the one which is applied when the read operation is executed. The default (**center next**) centers the next molecule (the one being read in) in the graphics window.

5. Pick the **center next** box and cycle through the three centering possibilities. Leave it with **center next** displayed.

This menu also has an example of two mutually exclusive menu items. These are the write defaults, **by file** and **by group**. Only one of these may be on (highlighted) at a time. Toggling the item which is off turns it on and at the same time turns the other item off.

6. Toggle **by group** on. Notice that as **by group** goes on, **by file** goes off (it is no longer highlighted). Toggle **by file** on again.

The rest of the items on the In-Out menu are all functional, that is, picking them causes an operation to be performed. Most of them, however, require further input from the user before the operation may be completed. The program employs certain conventions in using the menus which make this process easier. These are described in the next section.

1.6.4 Basic Conventions for Using the Menus

The program employs the following conventions for using the menus:

1) Once a function has been selected from the Functions menu, additional information may be needed in order to complete the operation. The program requests input in one of three ways:
   (a) It displays another menu with a list of selections. A prompt may appear in the Modules/Messages window. If still visible, the original menu may become dimmed indicating that it is currently inactive (picking menu items has no effect).
   (b) It surrounds the graphics display window with a red border and prints a message in the Modules/Messages window asking you to perform an operation within the graphics window.
   (c) It displays a dialog box requesting you to make a selection or enter information. The cursor is automatically positioned in the dialog box.
2) If the program requires you to make only one selection from a menu, then the operation continues as soon as you pick the item. When it is possible to make more than one selection, however, you can select and deselect items until you have picked the set of items you want. The operation is then completed by picking **return**.
3) Most item selection operations can be aborted by picking **return** or **abort**.
4) The program can indicate completion of an operation in several ways: by restoring the Modules and Common Buttons (if there was a message in the Modules/Messages Window), by removing the red border surrounding the graphics window (if picking atoms), and by restoring the menus and/or dialog boxes to their previous condition.

Some of these basic conventions for using the menus can be illustrated by using the **render groups** function to render the netropsin molecule as balls and sticks.

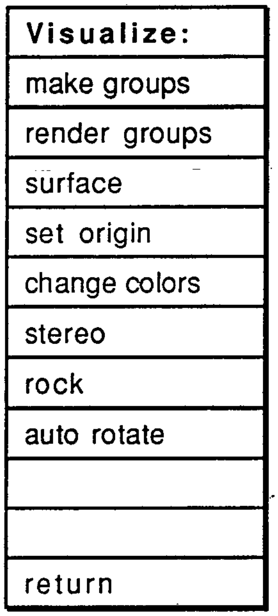


Figure 1.10:
Visualize menu

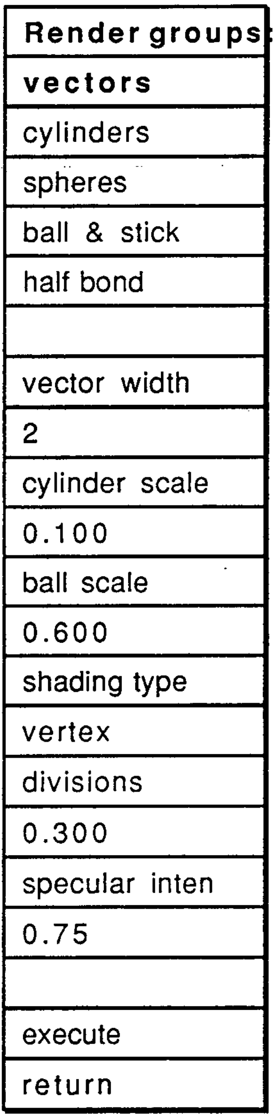


Figure 1.11:
Render groups
menu

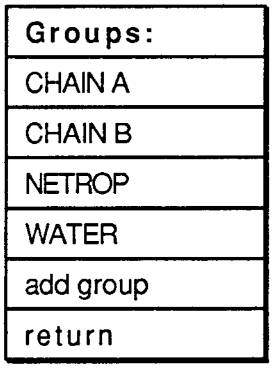


Figure 1.12:
Groups menu

1. Pick **return** to go back to the Top menu, then pick **visualize**. The Visualize menu shown in Figure 1.10 appears.

2. Pick **render groups** to bring up the Render groups menu (see Figure 1.11). The Groups menu is displayed with a list of selections (the groups currently available in the system) as shown in Figure 1.12. The program requests input by displaying the following prompt in the Modules/Messages window *Pick the group(s) to be rendered.*

3. Pick **NETROP** from the Groups menu. The NETROP menu box becomes highlighted indicating that it has been selected. It is possible to pick more than one group, so the Groups menu remains displayed.

4. Pick **CHAIN A**. It also becomes highlighted. We only want to render the NETROP group, so pick **CHAIN A** again to deselect it.

5. Pick **return** to complete the operation. The Groups menu disappears indicating that the group selection has been completed. The prompt also disappears and the Modules/Messages window is restored to its normal appearance.

6. On the Render groups menu, toggle **ball & stick** on (highlight it). The default rendering mode (**vectors**) is automatically turned off (not highlighted).

7. Pick **execute**. The message *Working...* appears in the Modules/Messages window while the program performs the required calculations. When complete, the NETROP group appears in the graphics window in ball and stick mode and the Modules/Messages window is restored to normal.

1.7 Inputting Information via Dialog and Alert Boxes

Several program functions require the user to enter information via dialog boxes. We'll use an option on the Geometry menu to illustrate this.

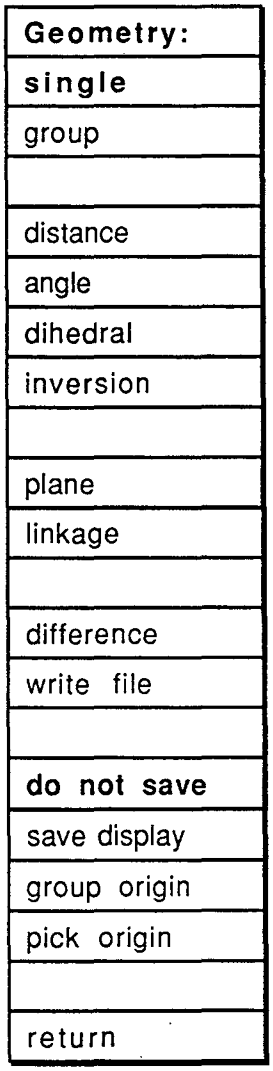


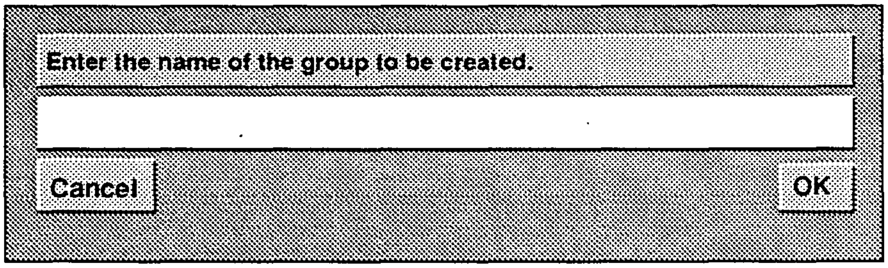
Figure 1.13: Geometry menu

1. Click the TOP MENU button to return to the Top menu, pick **analyze** to bring up the Analyze menu, then pick **geometry**. The Geometry menu shown in Figure 1.13 is displayed.

   We'll be measuring some distances between atoms shortly. The distances will be illustrated on the display. We have the option of saving these displayed values by creating a group for them. We'll tell the program to save them by toggling the **save display** option on.

2. Pick **save display** to highlight it. The **do not save** menu item is automatically turned off.

   A dialog box comes up as shown in Figure 1.14 prompting us to *Enter the name of the group to be created.*

   Enter the name of the group to be created.

   Figure 1.14: Dialog box example

   The dialog box always includes a prompt box at the top and a text entry box (highlighted) directly beneath it. The cursor is automatically positioned appropriately so you don't need to move it (text can be input as long as the cursor is located anywhere within the dialog box). You just type in the required information, then select the OK button. The OK button can be selected by clicking in it with the mouse or by simply pressing the <Enter> key. The Cancel button is used to abort the function.

3. Enter ***distance*** as the name of the group, then press the <Enter> key to select OK and verify the entry. The dialog box disappears.

The Alert box is similar to the dialog box in that it also contains a prompt or message and OK and Cancel buttons. It appears in the graphics window area, however, and is used for critical functions such as resetting or exiting the program.

1.8 Picking Atoms in the Graphics Window

We'll illustrate how to pick atoms with the **distance** option.

1. Pick **distance** on the Geometry menu. The program indicates that you are required to pick an atom by surrounding the graphics window with a red border. You are also prompted *Pick the first atom or anywhere else to continue.*

   Only the left mouse button is operative when picking atoms. If two atoms are displayed in the same area (within the tolerance of the small box of the cursor), the atom appearing closer to you may not be the one that is picked.

2. Zoom in or change the orientation of the molecules so that only the atom you want to pick lies within the area of the cursor.

3. Pick the atom. The picked atom is marked with a red cross. The prompt changes to *Pick the second atom or anywhere else to continue.*

4. If you were not able to pick the desired atom, it may be due to another atom which is displayed in the same area, but which is not visible because depth cueing causes it to be displayed with little or no intensity. Temporarily reduce the degree of depth cueing to see if additional atoms lie within your picking area.

5. Pick another atom. The second atom picked is marked with a yellow cross. A dotted blue line is drawn between the two atoms and the distance is displayed. We can continue measuring distances if desired. The prompt reads *Picked distances are shown. Pick the first atom or anywhere else to continue.*

4. Pick a non-atom to complete the operation. The graphics window is no longer surrounded with a red border, the prompt disappears, and the Modules/Messages window returns to normal.

1.8.1 Picking a Range of Atoms

There are instances when you will need to pick a "range" of atoms, that is, a subset of the total atoms displayed. They must all be connected to one another through a series of covalent bonds. Basically, a range of atoms is defined by picking a base atom and then delimiting the range by picking stopping atoms. Section 1.5.1 of the *BIOGRAF Reference Manual* describes this in more detail.

1.9 Displaying Groups - The DISPLAY Button

The DISPLAY button brings up the Display menu which lists the groups that are currently in the system. Groups are collections of atoms and bonds chosen in such a manner as to reduce a molecular structure into functional parts. Program functions can then be applied to selected groups of atoms. For example, groups can be rendered differently (*e.g,* displayed in different colors) enhancing visualization. The DNA molecule shown here has been partitioned into two separate groups, CHAIN A and CHAIN B. These have been rendered in different colors, making it easier to distinguish them. Groups are created using the **make groups** function under **visualize**. This ability to create groups is one of the most powerful features of the program. You can turn a group on or off (have it displayed or not) by toggling on the group name in the Display menu.

1. Click the DISPLAY button in the Modules/Messages window. The Display menu appears and lists the groups currently in the system as shown in Figure 1.15. Note that all the groups are turned on (appear highlighted in the menu) except the **WATER** group. The **distance** group we created in the last section is listed at the bottom of the menu.
2. Toggle each of the groups on and off. Note that the groups are no longer displayed in the graphics window when toggled off. Leave all but the **WATER** group toggled on.
3. Pick **return** to complete the function. The Display menu is replaced with the currently active Functions menu.

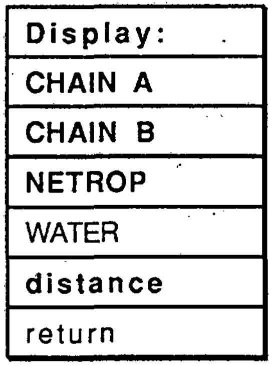


Figure 1.15: Display menu

1.10 Entering File Names

File names must be specified when reading in and writing out files. When writing files, the file name is entered directly in the dialog box which appears. If the file is not being saved in the current working directory, the full pathname including disk and directory specifications must be included ("../" may be used to indicate the previous directory). Logical names cannot be used. The program automatically adds on the appropriate three- or four-letter extension if not already included.

When reading in files, a special file name entry dialog box appears (see Figure 1.17). The file name may be entered directly (using the New Dir and New File buttons) or it may be selected from the listings provided in the browser box. These features are described below.

1.10.1 The Browser Box

A special file name entry dialog box is used for file input (see Figure 1.17). This dialog box is displayed when reading in files (*e.g.*, when using the **read** option on the In-Out menu) and when using the **browser** function on the Utilities menu to search through files or change the default directory.

1 Click the TOP MENU button to return to the Top menu, then pick **in-out** to bring up the In-Out menu.

2. Make sure the **auto group** and **half bond** options are toggled on, and that **center next** is displayed.

3. Pick **read** from the In-Out menu. The prompt *Pick the file type to be input* appears in the Modules/Messages window and the File types menu shown in Figure 1.16 is displayed.

   Ten different coordinate file formats are available. These are listed at the top of the menu. They provide information about the positions and connectivities of the atoms in the structure. The other file types provide information about force field parameters, groups, coordinate transformations, solvent-accessible surfaces (dot files), and vectors (*e.g.*, for electron density contours). Our data is in a BioDesign General format file.

2. Pick **BioDesign** to indicate that we want to read a BioDesign General format file.

   The file name entry dialog box is displayed (see Figure 1.17) and you are prompted to *Choose the file to be input.*

   The full pathname for the current working directory is given at the top of the dialog box next to "Dir:", and the full pathname for the cur-

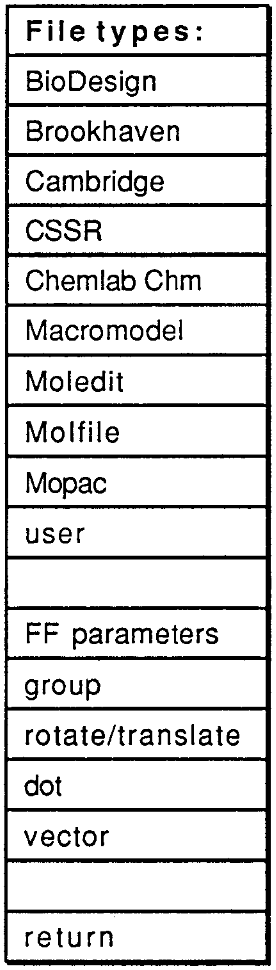
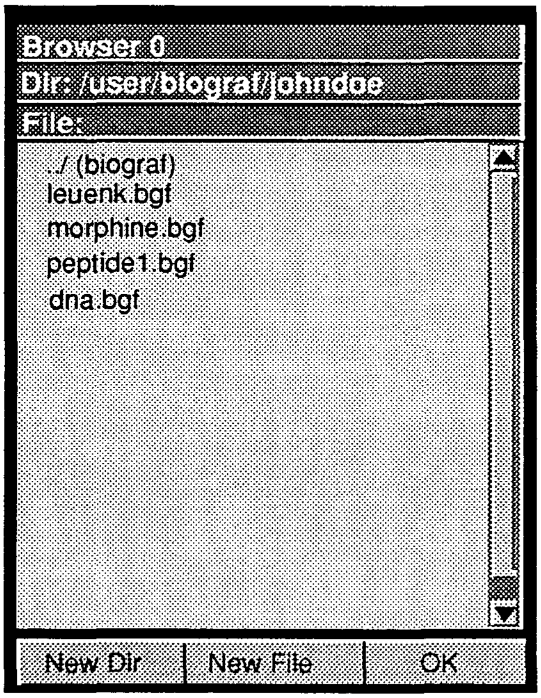


Figure 1.16:
File types menu



Browser 0

Dir: /user/biograf/johndoe

File:

../ (biograf)
leuenk.bgf
morphine.bgf
peptide1.bgf
dna.bgf

Figure 1.17: Example of Filename Entry and Browser Box

rently selected file is shown next to "File:" (initially this is blank since no file has been selected yet). The subdirectories and files within the current working directory are listed in the browser box. The directories are always shown in red, while the files are shown in black and have a 3- or 4- letter extension. When reading in a file, the listings only include files of the filetype specified (*e.g.*, if Mopac was chosen as the filetype, only *.mpc* files are listed). If using the browser function on the Utilities menu, however, all filetypes are shown.

The previous directory (up one level) is always shown at the top of the list in parentheses ("../" indicates the previous level). Selecting this allows you to move up in the directory path, while selecting one of the subdirectories moves you down. Selection is done by clicking on the desired directory or file name with the mouse (file names become highlighted when selected).

3. Select the previous directory from the top of the list (shown in red and in parentheses).

4. Note that the browser box now displays listings for the directory chosen. The directory name shown at the top next to "Dir:" also changes appropriately.

The file we want to read in is located in your distribution data directory. Typically the full specifications for the distribution data directory are */usr/msi/biogv300/data*. The distribution data files are always located in the */biogv300/data* directory. However, as noted at the beginning of the tutorial, the first part of the path specifications may differ on your system.

5. If you do not recall the location of the */biogv300* directory on your system, then type *echo $MSIUSR* from a systems window to obtain it.

6. Use the browser box to move up and down in the directories and locate the */biogv300/data* directory. Several data files are listed.

Using the Scroll Bar

The scroll bar at the right of the browser box can be used to scroll through the listings. The three mouse buttons each scroll the list in a different way.

1. Position the cursor within the scroll bar box, then click the right mouse button. This scrolls the list down.

2. Click the left mouse button and note that this scrolls the list up.

3. Position the cursor in the middle of the scroll box and click the middle mouse button or press the middle mouse button and drag the scroll bar to this spot. This moves you to a particular position in the list (1/2 way down in this case).

4. Use the three mouse buttons to scroll the list until you are comfortable with their use.

   These scrolling features are particularly useful when working with long lists. However, the use of exceedingly long lists may cause delays (it takes longer for the browser box to display them). This may be avoided by keeping no more than 50 files in any directory.

5. Locate the *chex.bgf* file in the listings, then select it by clicking on the file name. It becomes highlighted and the file name selected appears next to "File:" at the top.

6. Click the OK button to verify the file selection. The dialog box disappears and the message *Working...* appears while the program reads the data file. When finished, the cyclohexane molecule appears centered in the graphics window.

1.10.2 The New Dir and New File Buttons

You can also change the default directory or select a new file using the New Dir or New File buttons. Clicking these buttons with the mouse brings up a dialog box requesting you to enter the name of the new directory or file desired. The New Dir dialog box also lists the current directory at the top. If the new file or directory is not in your current working directory, the full pathname including disk and directory specifications must be included ("../" may be used to indicate the previous directory). We'll illustrate the use of the New File button by reading in the *toluene.bgf* file which is located in your distribution solvent directory (*/biogv300/solvent*).

1. Pick **read**, then select **BioDesign** for the file type as before.
2. Click the New File button at the bottom of the browser box. A dialog box comes up prompting you to enter the name of the file. Since you are currently in the */biogv300/data* directory, you can use "../" to indicate that you want to go up to the previous directory (*/biogv300).*
3. Enter ***../solvent/toluene.bgf*** in the dialog box, then press <Enter>. You are returned to the browser box. The file name you entered appears next to "File:" at the top.
4. Click the OK button at the bottom of the browser box to verify the selection. The data file is read in and the structure for the solvent toluene appears in the graphics window.

1.11 Resetting the Program

There may be times when you would like to work on a different molecule and would like to clear out all the atoms, groups, and files left over from the current session so that you can "start from a clean slate". This can be done using the program reset function. This allows you to start over without taking the extra time to quit BIOGRAF and restart it. The program reset option is found on the Utilities menu.

1. Click the TOP MENU button to return to the Top menu, then pick **utilities** to bring up the Utilities menu shown in Figure 1.18.
2. Pick **program reset**. An alert box appears asking you to verify the reset.
3. Click the OK button. BIOGRAF resets all the atom, group, and file arrays. When finished, the structures are all cleared from the graphics window and the Top menu is again displayed.

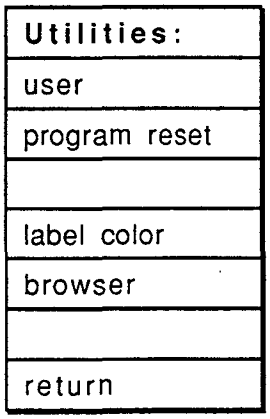


Figure 1.18: Utilities menu

1.12 Using Command Mode - The COMMAND Button

In addition to using the Functions menus to initiate operations, the program provides another mode of operation where you may issue commands by entering text in the text window. This is called command mode. Command mode allows the use of some very powerful special commands and macros. Command mode can be initiated using several different methods. The most direct method is to click the COMMAND button in the Modules/Messages window after the program has started.

1. Click the TEXT button in the lower left window to bring the text window forward. If necessary, resize or reposition the windows so that your views of the graphics window and the Menu Pad are not obscured.

2. Click the COMMAND button. The text window displays the current menu name (Top menu:) and a message appears in the Modules/Messages window reminding us that the cursor must be in the graphics window to use command mode.

3. Move the cursor into the graphics window to activate it (if necessary, also click with the mouse). Macros are executed by typing in the character "@" followed by the macro file name. If the macro is not located in your current working directory, then full directory specifications must also be included.

   We have a macro file which reads in the coordinate data for the morphine molecule. It also reads in a dot file containing surface information for morphine. It is called *intro.macro* and is located in your distribution macro directory. Refer to Section 1.1.1 if you do not know the location of your distribution macro directory.

4. Type @ followed by the macro file name including full directory specifications (*e.g.*, ***@/usr/msi/biogv300/macro/intro***). Do *not* include the *.macro* extension at the end of the file name.

   The program quickly executes the series of commands in the *intro.macro* file. The coordinate and dot file data for morphine are read in and the morphine molecule appears in the graphics window. It is surrounded by a dotted molecular surface.

   You can exit command mode and return to graphics (interactive) mode by entering one of the following commands after the command prompt: *menu*, *quit*, or *stop*. The last command in the macro file we just executed is *menu* so we are now back in interactive mode.

   Command mode can also be specified upon start up by either adding a "command mode" line to the defaults file (L_CMND YES) or by specifying a start up macro on program invocation as we did at the beginning of this tutorial. The use of command mode and macros is described in more detail in Appendix K of the *BIOGRAF Reference Manual.*

1.13 Running the Program in the Background

One of the easiest ways to run a job offline or in the background is to begin the job as an interactive session as we did in this tutorial. The program automatically creates a macro command file of the session called *logfile.macro* and places it in your current working directory. One can rename this file, edit it as desired, disable the graphics displays, and then run the macro in the background. It is usually a good idea to save the file information sent to the text window. This can be done by redirecting output to a file on the command line which is used to execute the program. The stored output file can be followed by running "tail -f" on the output file from another window.

See Section 1.8 of the *BIOGRAF Reference Manual* for more details. Chapter 6 of this tutorial also illustrates how to create a macro and run it in the background.

1.14 Basic Functions

The basic functions available via the items on the top level Functions menu are listed below with a very brief description of their use. The most commonly used functions are illustrated in detail in subsequent chapters of this tutorial.

In-Out	Transfer files to and from disk, copy files, add captions, arrows, borders *etc.* to the graphics display, generate PostScript files from the images on the screen for output to a printer.
Build	Sketch new structures and build polypeptides, lipids, DNA, RNA, and carbohydrates from monomeric units and organic molecules from organic fragments, build crystals from crystallographic data, and solvate molecules. Additional functions are used to rotate bonds, dock structures and interactively monitor interatomic distances, force vectors, and energy, add or delete hydrogens or modify their mass, calculate moments for a structure, convert coordinate files, and generate connectivity for input files which don't have bonding information.
Visualize	Partition molecules into functionally useful subsets of atoms (make groups), change rendering mode (vector, solid, cylinder, ball and stick), create ribbons, calculate and display dotted surfaces that show the structure's contact surface, electrical charge, or electrostatic potential, change the center of rotation of the display, change the color of structures, display images in stereo, cause objects to rock or rotate about the y axis.
Simulate	Perform energy minimizations, conformational searches and molecular dynamics, animate the sequences generated, perform statistical analysis on them, specify the parameters that govern the energy calculations, and modify the force field parameters.
Analyze	Identify atoms and linkages, calculate bond distances, angles and planes, analyze energies and forces, check for "bad" van der Waals contacts, measure molecular volumes, check proteins for irregular structural features, match structures by least-squares fits, analyze and plot trajectory file data, display, merge, and change the appearance of plot files.
Utilities	Reset the program, call up user-defined functions, browse through files in the system, change the default directory, and change the color of the labels.
Exit	Exit the program.

1.15 Conclusions

BIOGRAF employs an X-Windows based graphical user interface that is designed to be both easy to learn and easy to use. The program is basically menu-driven and uses standard file browsers, dialog boxes, alert boxes, slider bars, and buttons. The program can be run interactively or in the background or commands may be entered from the keyboard. A powerful macro facility is also supported.

This concludes Chapter 1. To leave BIOGRAF, pick **exit** from the top level menu, then verify the exit by clicking the OK button.

You are now ready to go on to Chapters 2 through 7. The rest of the tutorials in this manual are independent of each other. This leads to some redundancy between the various chapters, but makes it possible for you to select the tutorials of most interest. Chapters 2 through 7 all assume that you have read Chapter 1.

Chapter 2

Displaying Molecular Structures

This chapter introduces the operations which allow you to display molecular structures in a useful manner. We will read in the data that describes the Cytochrome C molecule found in tuna and look at the ways in which it can be divided up into recognizable parts or *groups*. We will use the **make groups** operations to construct our own groups and then use the **in-out** operations to save the information.

It is assumed that you have already completed Chapter 1.

2.1 Getting Started

1. Enter the command ***biograf***. This starts up BIOGRAF without reading in information about any molecule. As explained in Chapter 1, this command may be different at your site.

   The Top menu will appear in the Menu Pad as shown in Figure 2.1.

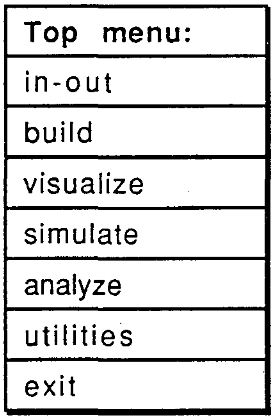


Figure 2.1:
Top level menu

2.2 Reading the Data

This section describes how to input data about the Cytochrome C molecule into BIOGRAF. The Cytochrome C data is located in two files, a coordinate file, *test.bgf*, which contains information about the positions and connectivities of the atoms, and a group file, *test.grp*, which contains group information. To read files, the program requires the user to input information which describes where these files are located.

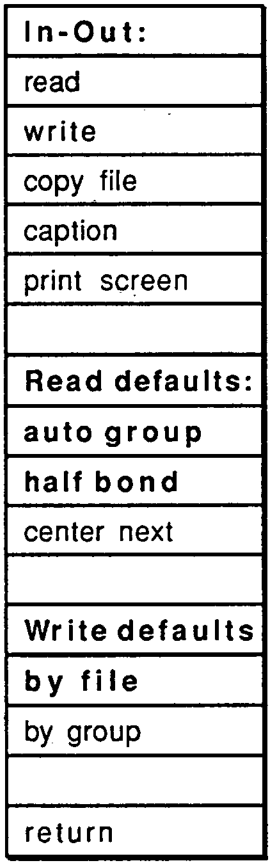


Figure 2.2:
In-Out menu

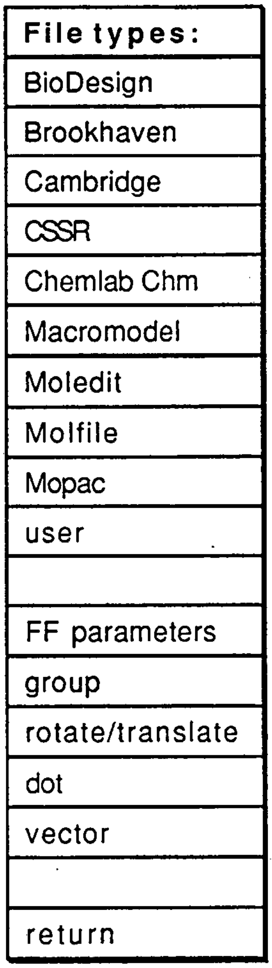


Figure 2.3:
File types menu

1. Pick **in-out** from the Top menu to display the In-Out menu (see Figure 2.2). This menu provides five basic functions: reading, saving, and copying files, adding captions, borders, *etc.* to the images in the graphics window, and creating PostScript files of the screen for output to a printer or to another application. These functions are performed using the first five options on the menu (**read**, **write**, **copy file**, **caption**, and **print screen**).

   Three read defaults also appear on the menu which determine how the structure will appear when it is read in. **Auto group** allows the user to automatically create a group for the file that is being read in, and **half bond** determines whether the structure will be drawn in a single color or with the half-bond colors. These options are either on or off. **Center next** is a three-way toggle which controls how the display is centered (on the file being read in, on all the files, or left as is). The read defaults apply only when reading in coordinate files. We will be reading in a BioDesign General Format coordinate file. We want to create a group for the file, use the half bond colors, and center it on the screen, so we'll keep the current settings for these options. The remaining options are write defaults which apply when saving files.

2. Pick **read**. The prompt *Pick the file type to be input:* appears in the Modules/Messages window and a new menu is displayed listing the possible file types as shown in Figure 2.3. The coordinate file types are listed first. These provide information about the positions and connectivity of the atoms in the structure. Other file types are also available which provide information about force field parameters, groups, coordinate transformations (rotation/translation files), solvent-accessible surfaces (dot files), and vectors (*e.g.*, for electron density contours). Our data is in a BioDesign General Format file.

3. Pick **BioDesign** from the File types menu to indicate that we want to read a BioDesign General Format file. The file name entry dialog box appears and we are prompted to *Choose the file to be input*. The Cytochrome C data is in the *test.bgf* file which is located in the distribution directory (*test* is the *filename*, and the *bgf* extension indicates that the file *type* is a BioDesign General Format file).

   Typically, the full specifications for the distribution data directory are */usr/msi/biogv300/data.* The distribution data files are always located in the */biogv300/data.* directory; however, the first part of the path specifications may differ on your system. To determine the location of the */biogv300/data* directory on your system, type *echo $MSIUSR* from a systems window. The directory listed in response to this command should be followed by */biogv300/data* when giving directory specifications for distribution data files.

4. Use the browser box to locate the distribution data directory (*e.g.*, ***/usr/msi/biogv300/data***, then select the ***test.bgf*** file from the listings. Click the OK button to complete the entry.

   The message *Working* appears while the program reads the data file. Since we had **auto group** toggled on, a new group is also created. It is defined as **Tuna** (the file descriptor name for the Cytochrome C data). The entire Cytochrome C system is then displayed as one group. From the display, it is clear why there is an advantage to subdividing the molecule into smaller groups. We'll be able to see these groups after we read in the group file.

5. Pick **read**, then pick **group** from the File types menu. You are prompted to *Choose the name of the groupings file.*

6. Select the ***test.grp*** file from the listings in the browser box, then pick OK to complete the entry. Messages appear while the groups are being read in. When BIOGRAF finishes reading the data, the **Tuna** group disappears and four of the groups that we read in appear in the graphics window.

2.3 Looking at the Groups

BIOGRAF handles molecular information as *groups*. A group is a collection of atoms and bonds chosen in such a manner as to reduce a molecular structure into recognizable parts. There are various ways to partition the Cytochrome C molecule into meaningful groups. We chose the ten groups shown in Table 2.1.

The four groups currently displayed are: Trace, Heme, Ligands, and Thio E. These groups are turned on, while the remaining groups (All, Main, Side, Carbonyl, Water, and Labels) are turned off and thus do not appear on the display. Groups may be turned on or off (displayed or not) by clicking the DISPLAY button and toggling on the group name in the Display menu.

Group	Color	Description
All	half-bonded	The entire molecule
Trace	red	A trace through all the α-carbons (C_α)
Main	red	The peptide backbone or main chain (-N-C_α-C-)
Side	cyan	The peptide residues or side chains
Carbonyl	yellow	The peptide C=O groups
Heme	purple	The heme group
Ligands	green	The histidine (HIS 18) and methionine (MET 80) residues which constitute the fifth and sixth ligands to the heme iron
Thio E	cyan	The two thioether bonds which attach the heme to the protein
Water	blue	The crystallographic water molecules
Labels	yellow	The residue labels and numbers

Table 2.1: Groups in Cytochrome C molecule

1. Click the DISPLAY button located in the Modules/Messages window (see Figure 2.4). The Display menu shown in Figure 2.5 comes up and lists all the groups currently in the system. Note that the four groups which are currently displayed in the graphics window appear highlighted in the menu (**Trace**, **Heme**, **Ligands**, and **Thio E**).

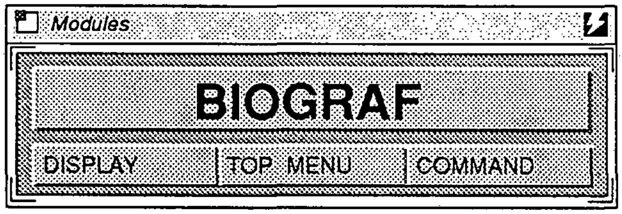


Figure 2.4 Modules/Messages Window with DISPLAY, TOP MENU and COMMAND Buttons

2. Toggle the various groups on and off to look at the molecule. Pick **return** when finished. The Display menu disappears and the In-Out menu comes up in its place.

The same atoms and bonds may appear in more than one group. This *does not* mean BIOGRAF has made more than one molecular structure that contains those atoms and bonds. For example, the **Ligands** group is a subset of the **Side** group, but BIOGRAF knows that the atoms and bonds contained in **Ligands** appear only once in the Cytochrome C molecule.

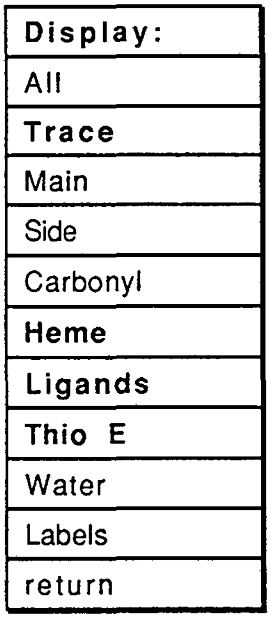


Figure 2.5: Display menu

2.4 Creating Groups

BIOGRAF expects all data to be specified in terms of groups. The following sections describe how to partition the Cytochrome C molecule into the ten groups as shown in Table 2.1.

Groups may be defined in either of two ways:

- with standard structural elements or *primitives* (*e.g.*, an amino acid residue), or
- with standard structural descriptors or *selectors* (*e.g.*, hydrophilic side chains).

Appendix A of the *Biograf Reference Manual* "Standard Group Elements" lists the primitives and selectors that may be used to describe biological macromolecules.

Groups are created using the **make groups** operations found under **visualize.** Before we create any groups for Cytochrome C, we'll first reset the program so that we can get rid of all our current groups and start from a clean slate.

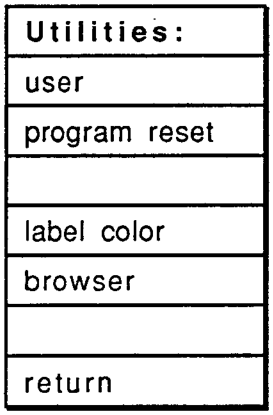


Figure 2.6: Utilities menu

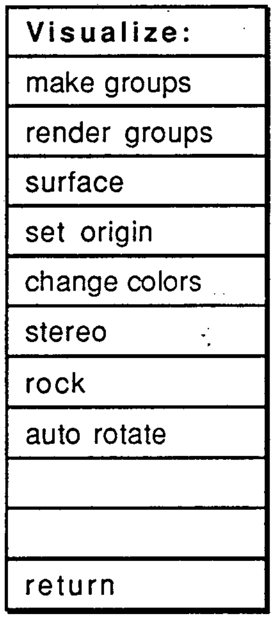


Figure 2.7: Visualize menu

1. Return to the top level menu, then pick **utilities**. The Utilities menu shown in Figure 2.6 appears.
2. Pick **program reset**. An alert box appears asking you to verify the reset.
3. Click the OK button. When BIOGRAF finishes the program reset it no longer remembers anything about Cytochrome C including the coordinate information, so we'll have to read the coordinate file back in again before we can create any groups. The Top menu is again displayed.
4. Pick **in-out** to bring up the In-Out menu, then toggle **auto group** off.
5. Read in the ***test.bgf*** file as we did before. (Pick **read**, pick **BioDesign** for the file type, then use the browser box to choose the ***test.bgf*** file from your distribution data directory (*e.g.*, ***/usr/msi/biogv300/data***).

   The coordinate data for Cytochrome C is read in. The molecule is not displayed, however, since no group was created (**auto group** was off).
6. **Return** to the Top menu, then pick **visualize**. The Visualize menu shown in Figure 2.7 appears.

7. Pick **make groups** on the Visualize menu to bring up the Make Groups menu shown in Figure 2.8.

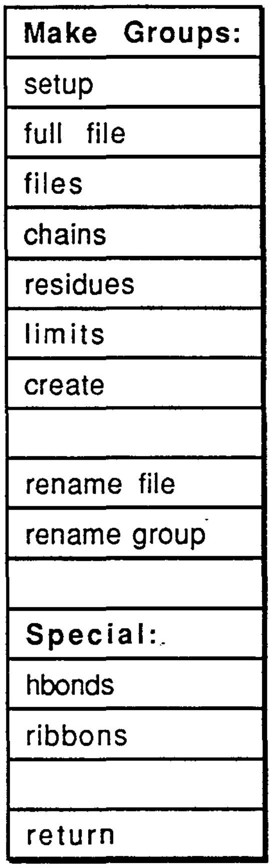


Figure 2.8: Make Groups menu

The first seven options on the Make Groups menu are used to define the portion of a structure which is to comprise the new group. It is essential to perform them in the proper sequence, starting with **setup** and moving down sequentially to **create.** **Setup** is used to specify the group which is to be created or redefined. **Files** is used to select the coordinate file(s) from which the group is to be built. **Full file** is an on/off switch which indicates whether the group is to include all the atoms in the file (on) or only part of the file (off).

The **chains, residues,** and **limits** operations allow the user to delimit the group in terms of standard structural elements or primitives for biological macromolecules, such as chains and residues or standard structural selectors such as hydrophilic side chains. One may also limit the group by specifying a range of residues to be included or by using a distance constraint. Once the group has been defined, the user then picks **create** to generate it. The sequence of steps used to construct a group is given below.

setup ➪ **files** ➪ **chains** ➪ **residues** ➪ **limits** ➪ **create**

If the whole file is to be used (**full file** is on), then the **chains, residues** and **limits** options are not applicable and are skipped.

setup ➪ **files** ➪ **create** (**full file** on)

The correct sequence is essential. However, it may not be necessary to repeat every step each time you define a group because **files, chains**, and **residues** always correspond to their last setting. You can make changes in the middle of the sequence by redoing the appropriate step and then repeating the subsequent steps in the proper order. The defaults for **files, chains**, and **residues** are often appropriate, so that constructing groups involves only the steps

setup ➪ **limits** ➪ **create**

Primitives and **selectors** also retain their previous settings, but the residues picked or keyed in under **limits** are lost each time you toggle **setup,** so that the limits include the entire molecule.

2.5 Creating the Trace

The first group we create is the trace of the Cytochrome C backbone.

1. Pick **setup** on the Make Groups menu. The Groups menu appears as shown in Figure 2.9 and we are prompted *Pick the group to be defined:*

2. Pick **add group**. A dialog box comes up and we are asked to *Enter a name for the group to be defined.*

3. Type ***Trace*** as the group name in the dialog box. To verify the entry click the OK button or simply press <Enter>.

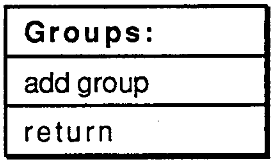


Figure 2.9: Groups menu

4. Pick **files** from the Make Groups menu.

   The Files menu appears with the file descriptor **Tuna** highlighted as shown in Figure 2.10 and we are prompted to *Pick the data file designators.*

   This is the file we want, so we'll leave it toggled on. Verify the file selection by picking **return.**

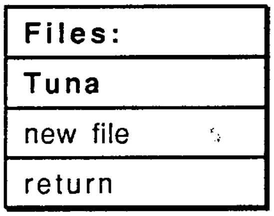


Figure 2.10: Files menu

5. Pick **chains**. A new menu appears listing the chain designators currently in the system as shown in Figure 2.11.

   Two chain designators are listed: ?, which indicates that the program will accept any chain designator, and **R** (for reduced Cytochrome C). The first time the Chains menu is entered, **?** is toggled on. If it is not on, toggle it on. Pick **return** to verify the selection.

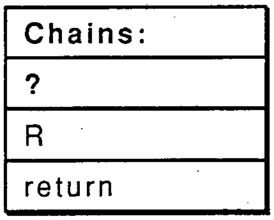
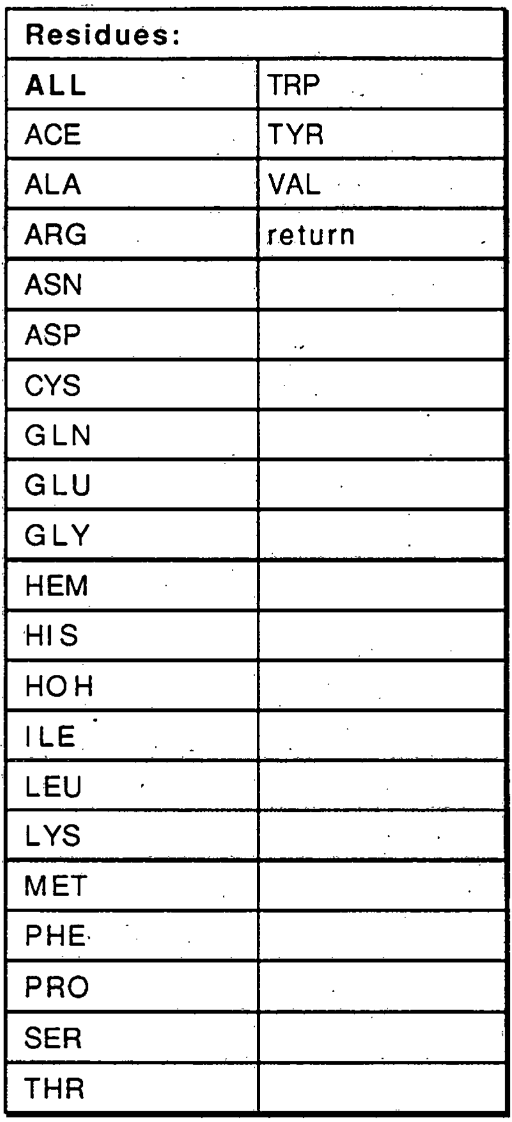


Figure 2.11: Chains menu

6. Pick **residues** from the Make Groups menu. The Residues menu appears as shown in Figure 2.12. **ALL** should be toggled on, indicating that any residue will be accepted. If **ALL** is not on, toggle it on. Complete the selection by picking **return**.



Residues:	
ALL	TRP
ACE	TYR
ALA	VAL
ARG	return
ASN	
ASP	
CYS	
GLN	
GLU	
GLY	
HEM	
HIS	
HOH	
ILE	
LEU	
LYS	
MET	
PHE	
PRO	
SER	
THR	

Figure 2.12: Residues menu

2.6 Choosing the Primitives

Next, we choose the primitives for the **Trace** group.

1. Pick **limits** on the Make Groups menu to display the Limits menu (see Figure 2.13).
2. Pick **primitives** on the Limits menu. The Primitives menu is then displayed listing the primitive structural elements as shown in Figure 2.14. **STD SET** should be toggled on, indicating that the standard set of primitive elements are included in the group. The standard set includes all the elements listed except the α carbons (**TRACE**) and the labels for the α carbons (**LABELS CA**).
3. Toggle **TRACE** on the Primitives menu, then pick **RETURN**. This tells BIOGRAF to include only the trace of the Cytochrome C backbone.
4. Pick **return** on the Limits menu to return to the Make Groups menu, then pick **create**.

   The message *Working*...is displayed in the Modules/Messages window. When the program finishes, a group corresponding to the trace of the Cytochrome C carbon backbone appears in the graphics window.

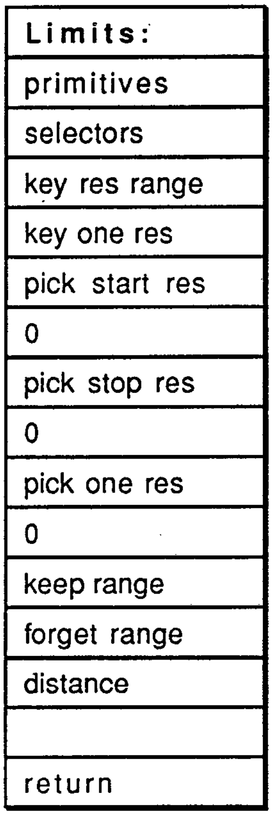


Figure 2.13:
Limits menu

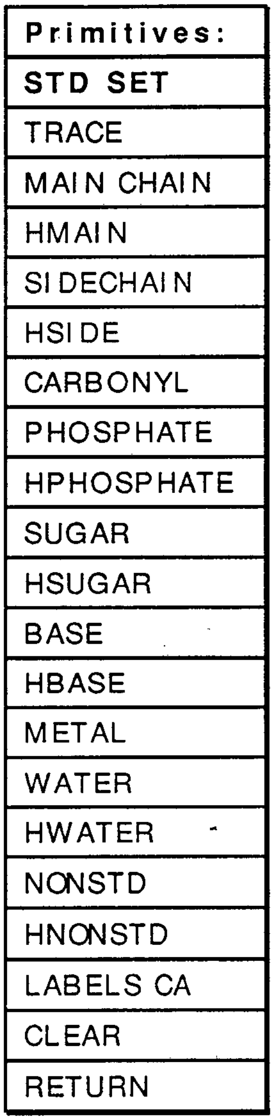


Figure 2.14:
Primitives menu

2.7 The Half-Bond Representation

A useful way to display a group is to color each of its atoms according to elemental type. The bonds between the atoms are displayed in *half-bond* mode such that the half of the bond closest to each atom is the same color as that atom. The colors used for some of the most common elements are shown in Table 2.2:

Element	Color	Element	Color
H	White	S	Yellow
C	Aqua	Cl	Green
N	Blue	Ca	Magenta
O	Red	Ti	Cyan (blue-green)
F	Green	Fe	Cyan (blue-green)
Na	Magenta	Zn	Magenta
Al	Magenta	Br	Green
Si	Aqua	Ru	Cyan (blue-green)
P	Yellow		

Table 2.2: BIOGRAF element colors

When BIOGRAF creates groups, it uses the half-bond colors. Since the trace of the Cytochrome C backbone consists only of carbon atoms, the **Trace** group we just created appears in a single color (aqua).

2.8 Creating the Main Chain

We next create a group that corresponds to the main chain of the Cytochrome C molecule. In addition to the alpha carbons of the trace, this group includes the nitrogens and carbonyl carbons in the peptide backbone.

1. Pick **setup** on the Make Groups menu, pick **add group** from the Groups menu, and define the new group to be ***Main***.

2. Verify that the settings under **files, chains** and **residues** remain unchanged.

3. Pick **limits** on the Make Groups menu, then pick **primitives** on the Limits menu.

4. Toggle **TRACE** off, toggle **MAIN CHAIN** on, then pick **RETURN**.

5. **Return** to the Make Groups menu, then pick **create**. The **Main** group appears on the display in the half-bond colors. Note the various colors for the different atoms: aqua for the carbons, blue for the nitrogens, and red for the oxygen (on the carbonyl carbon).

2.9 Changing the Group Color

Since we will be creating several more groups, it will be easier to differentiate them if each group is rendered in a different color. We can use the **change colors** option to change the color of any group, but the group must first be a single color (it cannot be represented with the half-bond colors). We can render a half-bonded group in monochrome with the **render groups** option. We can then use the **change colors** option to change the color.

Since the **Trace** group is only carbon atoms, it already appears in just one color (aqua). But the **Main** group we just created is displayed in three colors. We'll render the **Main** group in monochrome, then change the color to red.

1. **Return** to the Visualize menu, then pick **render groups**. The Groups menu appears and we are prompted to *Pick the group(s) to be rendered.*
2. Pick **Main**, then pick **return**. The Render groups menu appears (Figure 2.15). The first four options on the menu control the rendering mode. Groups can be rendered as vectors (lines), cylinders, spheres, or balls and sticks. Vectors is the default.
3. Make sure **half bond** is toggled off. When on, the group will be rendered in the half-bond representation; when off, the group will be rendered in a single color. The color used changes each time a new group is rendered.
4. Pick **execute**. The **Main** group is redisplayed in a single color.

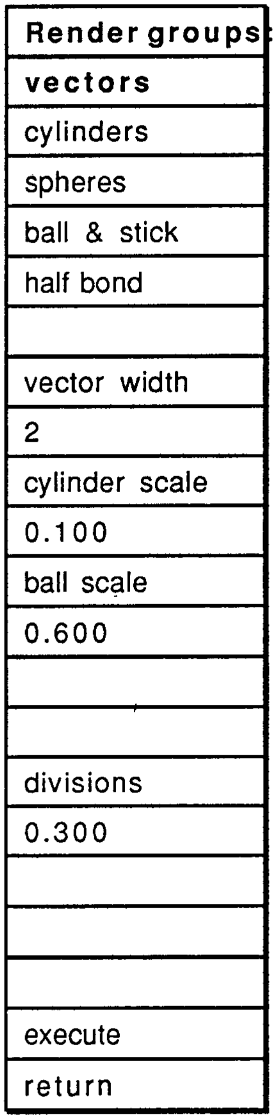


Figure 2.15: Render groups menu

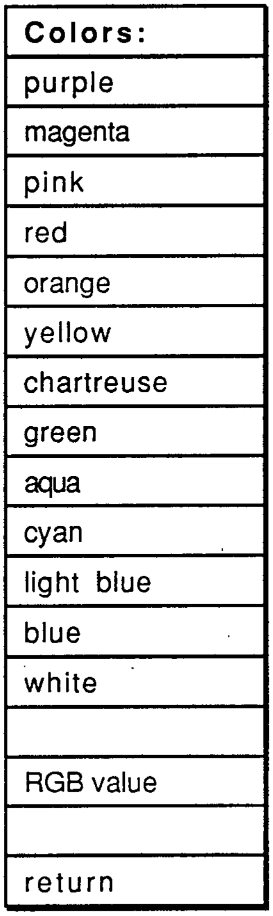


Figure 2.16: Colors menu

5. **Return** to the Visualize menu, then pick **change colors**. We are prompted to pick the group whose color is to be changed. Only one group can be modified at a time.

6. Pick **Main**. The Colors menu appears displaying 13 predefined colors (see Figure 2.16). An **RGB value** option is also available for specifying the exact color.

7. Pick **red**. The color of the **Main** group changes to red. Try different colors. Finish by picking **red**, followed by **return**.

2.10 Creating the Side Chain, Carbonyl, Water, and Labels Groups

We now create the side chain, carbonyl, water, and labels groups. We'll follow the same procedure as we did for **Trace** and **Main**. We'll also change the side chain and carbonyl groups to monochrome and color them cyan and yellow, respectively. When defining the primitives for each new group, remember to toggle off the primitives used to define the previous group.

1. Pick **make groups** on the Visualize menu.

2. Pick **setup**, choose **add group**, and define the new group as ***Side***.

3. Pick **limits**, then **primitives**, then toggle on the primitive **SIDE CHAIN**. Toggle all the other primitives off, then pick **RETURN**.

4. **Return** to the Make Groups menu, then pick **create**. The **Side** group is displayed in the half-bond colors.

5. **Return** to the Visualize menu, pick **render groups**, and select the **Side** group from the Groups menu. Make sure **half bond** is toggled off, then pick **execute**. The **Side** group is redisplayed in a single color.

6. **Return** to the Visualize menu, pick **change colors**, and select the **Side** group from the Groups menu. Pick **cyan** for the color, followed by **return**.

   A number of primitives have a name that starts with "H" (*e.g.*, HMAIN). These selections refer to the hydrogen atoms and bonds attached to the corresponding primitive (*e.g.*, MAIN). Very few X-ray crystallographic structures include data for H atom positions (H atoms are poor diffractors of X-rays), so it is often necessary to ask BIOGRAF to generate H atom positions.

7. Repeat steps 1-6 to create a new group defined as ***HSide*** using the primitives **SIDE CHAIN** and **HSIDE**. Note that the hydrogen atoms are now included. Color it **cyan**.

8. Repeat steps 1-6 to create a new group defined as ***Carbonyl*** using only the primitive **CARBONYL**. Color it **yellow**.

9. Repeat steps 1-4 to create a new group defined as ***Water*** using the primitive **HWATER**.

If you use the primitive **HWATER**, BIOGRAF produces water molecules in the line representation used for the other groups. However, if the primitive **WATER** is used, the water molecules are represented with pyramids.

10. Redefine the **Water** group using only the primitive **WATER** (repeat steps 2-4, but instead of choosing **add group** from the Groups menu, choose the **Water** group you just created). This time the water molecules appear as pyramids.
11. Repeat steps 2-4 to create a new group defined as ***Labels*** using the primitive **LABELS CA** ("CA" stands for α carbons). The **Labels** group contains the residue labels for the main chain of Cytochrome C; *e.g.*, the tyrosine amino acid that is the 82nd residue in the chain has the label TYR 82. Zoom in so that you can see the labels better.

2.11 Non-standard Groups

The set of "standard" selectors on the Primitives menu allows the normal components of proteins and DNA (such as amino acid residues and nucleic acid bases), as well as water molecules and metal ions, to be specifically included in groups. If we toggle **NONSTD** on the Primitives menu we can create groups that include atoms belonging to other kinds of molecular substructures. Cytochrome C contains a heme group, which is a non-standard residue. In order to create a group for it, the setting under **residues** must be changed from **ALL** to **HEM**, and the **NONSTD** primitive must be toggled on under **primitives**.

Our display is becoming rather cluttered, however. So, before creating any new groups, we'll turn the groups we've already created off.

1. Click the DISPLAY button, toggle off all the groups, then pick **return**.
2. Pick **setup** on the Make Groups menu, choose **add group**, and define the new group as ***Heme***.
3. Pick **residues**, toggle on **HEM** from the Residues menu, then pick **return**.
4. Pick **limits**, then **primitives**, then toggle **LABELS CA** off and **NONSTD** on. Pick **RETURN**.
5. **Return** to the Make Groups menu, then pick **create**. The **Heme** group appears on the display in the half-bond colors. You may have to zoom out to see it.
6. Use the **render groups** function to render the **Heme** group in monochrome, then use **change colors** to color it **purple**.

2.12 Specifying Residues

We now create a group corresponding to the histidine (HIS 18) and methionine (MET 80) residues which constitute the fifth and sixth ligands to the heme iron. This section demonstrates two methods for selecting HIS 18 and MET 80: *keying* the residues and *picking* the residues.

2.12.1 Keying the Residues

Keying a residue means telling BIOGRAF to which C_α it is attached.

1. Pick **make groups** to bring up the Make Groups menu, then pick **setup** and define a new group to be ***Ligands***.
2. Pick **residues**, toggle on **ALL** from the Residues menu, then pick **return**.
3. Pick **limits**, then **primitives**.
4. On the Primitives menu, toggle **NONSTD** off and **SIDE CHAIN** on. Complete by picking **RETURN**.
5. Pick **key one res** on the Limits menu. BIOGRAF prompts *Enter the residue number*.
6. Enter ***18*** in the dialog box, then select OK.
7. Pick **key one res** again and enter the number ***80***.
8. **Return** to the Make Groups menu and **create** the group. After a few seconds the **Ligands** group appears in the graphics window.
9. **Return** to the Visualize menu, use the **render groups** function to render the **Ligands** group in monochrome, then use **change colors** to color it **green**.

2.12.2 Looking at the Ligands Group

The **Ligands** group overlaps with the **Side** group. To see this more clearly, do the following:

1. Click the DISPLAY button, then toggle off all of the groups except **Ligands**.
2. Toggle on **Labels**. Rotate and zoom the molecule until it is clear which label goes with which residue on the **Ligands** group. You should be able to locate the HIS_18 and MET_80 labels on the **Ligands** group.
3. Toggle on **Side**.
4. Toggle off **Ligands** and note that the two residues remain visible as part of the side chain.

5. Toggle **Ligands, Main**, and **Heme** back on.

6. Toggle **Side** off. Pick **return** to complete the display operations.

2.12.3 Picking the Residues

Picking a residue means telling BIOGRAF which residue you want by toggling one of its atoms on the display. We will create the **Ligands** group again, this time by picking the HIS 18 and MET 80 residues from the display.

1. Pick **make groups** to bring up the Make Groups menu, pick **setup** and redefine the **Ligands** group using the same name.

2. Pick **limits**, then toggle **pick one res**. BIOGRAF outlines the graphics window in red and prompts *Pick a single residue:*

3. Pick any atom on the **HIS 18** residue. If necessary, reorient the molecule to find **HIS 18**. A red cross marks the picked atom. BIOGRAF displays the name of the selected residue directly beneath **pick one res** on the Limits menu. Information about the corresponding residue is also listed in the text window.

4. If you missed the HIS 18 residue, pick **pick one res** and try again.

5. Pick **keep range** to save it.

6. Toggle **pick one res** again and pick any atom on the **MET 80** residue. Toggle **keep range** to save it.

8. **Return** to the Make Groups menu and **create** the group. After a few seconds, **Ligands** blinks off and reappears on the display. It is shown in the half-bond colors.

9. Use either of the two techniques described in this section (picking or keying a residue) to define a new group as ***Thio E***. These are the two cysteine residues (**CYS 14** and **CYS 17**) that have thioether bonds connecting the main chain to the heme.

2.13 Using Structural Descriptors (Selectors)

The **selectors** operation on the Limits menu allows you to choose a group based on structural descriptors. For example, you can create a group consisting of the hydrophobic (water-hating) or hydrophilic (water-loving) residues.

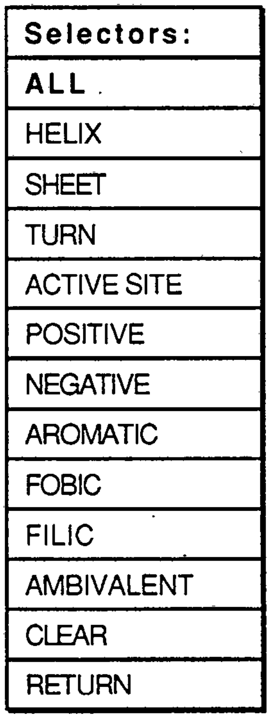


Figure 2.17: Selectors menu

1. Click the DISPLAY button, then toggle off all of the groups except **Main** and **Heme**.
2. Pick **setup** on the Make Groups menu and define a new group as ***Fobic***.
3. Pick **limits**, then **selectors**. The Selectors menu appears listing the structural descriptors as shown in Figure 2.17.
4. Toggle **FOBIC** on, then pick **RETURN**.
5. Confirm that **SIDE CHAIN** is the only entry selected on the Primitives menu.
6. **Return** to the Make Groups menu and **create** the group.
7. Click the DISPLAY button and toggle the **Water** group on and off. If necessary, zoom out so that the entire molecule is visible. Note how the hydrophobic residues cluster towards the center of Cytochrome C where water is excluded.
8. Define a new group as ***Filic*** using the **FILIC** selector. Make sure to toggle off the **FOBIC** selector.
9. **Return** to the Make Groups menu and **create** the group.
10. Click the DISPLAY button, then toggle the **Fobic** group off. Toggle the **Water** group on and off. Note how the hydrophilic residues generally occur on the surface of the protein near the water molecules. Pick **return** when finished.

For further description of the selectors, primitives and other aspects of the Make Groups Operations, see Section 4.1 of the *BIOGRAF Reference Manual.*

2.14 Saving Information

After you finish building a molecule, it is a good idea to write it to the disk. Saving information protects you from losing work and in subsequent sessions allows you to resume where you left off without repeating all of the steps of the previous session. This section describes how to save information in *group* files and *coordinate* files.

2.15.1 Group Files

A group is a collection of atoms and bonds chosen in such a manner as to reduce a molecular structure into recognizable parts. There are two types of files that save group information:

Type 1	saves all the sequences of menu picks that defined the various groups. When BIOGRAF reads a type 1 group file, it regenerates the groups based on the saved sequences of menu picks. This allows the same group file to work for different structures.
Type 2	saves the list of atoms and bonds for each group as well as the sequence of menu picks used to define them. A type 2 file regenerates groups faster than type 1 because it reconstructs them directly from the atom lists rather than from the menu picks.

Both types of group file contain the display status information that determines the color of the group and whether or not the group is toggled on or off.

Note: Type 2 files should only be used for coordinate files with the same set of atoms and connections. When in doubt, make a type 1 file.

To save group information about the molecule, proceed as follows:

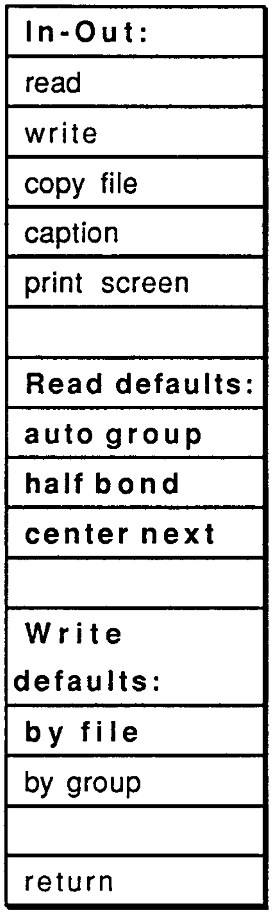


Figure 2.18: In-Out menu

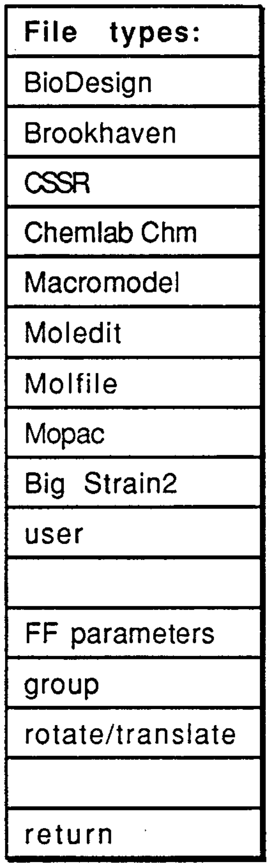


Figure 2.19: File types menu

1. Click the TOP MENU button to return to the Top menu, then pick **in-out** to bring up the In-Out menu (see Figure 2.18).
2. Pick **write**. The prompt *Pick the file type to be output:* is displayed and the File types menu appears in the Menu Pad (see Figure 2.19).
3. Pick **group** from the File types menu to let BIOGRAF know that the type of file you want to write contains group information. BIOGRAF brings up a dialog box and asks you to *Enter a name for the groupings file.*
4. Type ***tuna.grp*** (if you do not want to save the file in your current working directory, then include disk and directory information as well). Select OK to verify the entry. The *grp* extension indicates that the file is a group file. BIOGRAF prompts *Enter output file type [1 (normal) or 2].*
5. Enter the number ***1*** to indicate type 1 group information, then select OK. BIOGRAF then prompts you to enter remarks for the output file.
6. Type in remarks that will identify the file the next time you use it. BIOGRAF will print the remarks in the text window when it reads the file. Terminate the remarks with a blank line (Click OK or press <Enter>).

2.14.2 Coordinate Files

In order to reconstruct a group, BIOGRAF needs the information about the coordinates of the atoms in that group. The program saves coordinate information in coordinate files. Here we make a coordinate file of the BioDesign General Format type.

1. Make sure that the **by file** toggle is highlighted under the write defaults. The source of the coordinate information that we will be saving is in the DNA file that we read in.
2. Pick **write**. The prompt *Pick the file type to be output:* appears and the File types menu shown in Figure 2.19 again displays the file types. Ten different coordinate file formats are available.

3. Pick **BioDesign** to indicate the BioDesign General Format file type. This type is recommended for internal use with BIOGRAF since it contains all pertinent information.

   Since there is only one file in the system (DNA), it is automatically chosen as the source of coordinate information. BIOGRAF now asks you to *Enter a name for the output file.*

4. Type ***tuna.bgf*** in the dialog box (include disk and directory information if needed), then verify the entry by clicking OK. The *bgf* extension indicates that the file contains BioDesign General Format type data. BIOGRAF prompts you to enter remarks for the output file.

5. Type in remarks that will identify the file the next time you use it. Terminate the remarks with a blank line (Click OK or press <Enter>).

2.14.3 Testing Our Work

As an exercise, we test the new group and coordinate files by resetting the program and having BIOGRAF read them in again.

1. **Return** to the Top menu and pick **utilities**. The Utilities menu shown in Figure 2.20 is displayed.

2. Pick **program reset**. An alert box appears asking you to verify the reset.

3. Click the OK button. When BIOGRAF finishes the program reset it no longer remembers anything about the molecule. The Top menu is again displayed.

4. Pick **in-out** to bring up the In-Out menu.

5. Make sure the **auto group** and **half bond** options under the read defaults are toggled on. The **auto group** option treats the molecule as a single group and includes every atom and bond in the file. The **half bond** option uses the half-bond colors to display the molecule. Leave the centering option with **center next** displayed. This will cause the next file that we read in to be centered in the graphics window.

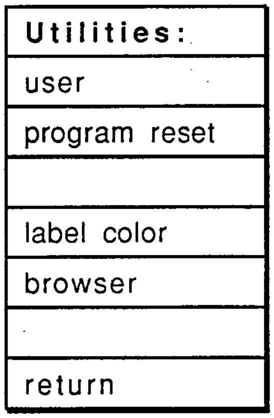


Figure 2.20: Utilities menu

6. Pick **read**. The File types menu appears and we are prompted *Pick the file type to be input:*

7. Pick **BioDesign**, since we saved our data in a *bgf* file. BIOGRAF brings up the file name entry dialog box and prompts *Choose the file to be input.*

8. Use the browser box to bring up the directory that you saved the files in, then select ***tuna.bgf*** from the file names listed. Click the OK button to verify the selection. The message *Working ...* is displayed as the coordinate file that we created in the previous section is read in. When finished, the Cytochrome C structures appear in the graphics window and a new group (**tuna**) is created.

Next we will read in the group file that we created to test whether it correctly saved the groups.

9. Pick **read** again from the In-Out menu and then pick **group** from the File types menu. BIOGRAF prompts *Enter the name of the groupings file.*

10. Select ***tuna.grp*** from the files listed in the browser box, then click the OK button. The **tuna** group disappears and the Cytochrome C groups that had been toggled on are displayed in the graphics window.

2.15 Conclusion

There are many ways in which a molecular structure can be broken up into recognizable parts, or *groups*. The following sequence of steps constructs a group from the Make Groups menu:

setup ➪ files ➪ chains ➪ residues ➪ limits (primitives or selectors) ➪ create

For many applications the defaults for **files**, **chains**, and **residues** are appropriate and constructing groups involves only the steps

setup ➪ limits ➪ create

In general, you want to choose groupings that are useful for your own application. The ten Cytochrome C groups are formed from a subset of the many options available under the Make Groups menu. You should study the menus brought up by **chains, residues, primitives**, and **selectors** to get a better idea of the different ways BIOGRAF can define groups. For example, groups may also be created using distance criteria to restrict the atoms included to a certain portion of the molecule. The distance criteria are set using the **distance** operation found on the Limits menu. This is discussed in Chapter 7. A special group containing hydrogen bonds can also be constructed using the **hbonds** option on the Make Groups menu, or **ribbons** can be constructed which are useful in illustrating the folding patterns of proteins and polypeptides. Section 4.1 of the *BIOGRAF Reference Manual* contains details of each menu.

To conclude this tutorial, complete the following exercise:

1. **Return** to the Top menu and pick **utilities**, then pick **program reset**. This erases all information about Cytochrome C in BIOGRAF.

2. **Read** in the coordinate file ***dna.bgf*** and the group file ***dna.grp*** from your distribution data directory. Four groups are read in: **Chain A, Chain B, Netrop**, and **Water**. All but the **Water** group are toggled on and appear in the graphics window,

3. **Create** four new groups with the same names, using the **make groups** options to divide the structure into chain A, chain B, netropsin, and water groups.

4. To check your work, use the DISPLAY button and toggle on both your groups and the original groups to see if they are the same.

   Note: The positions of the tetrahedra for the **Water** group you created should be identical to those on the **Water** group read in with the DNA file. However, the orientation of the tetrahedra may differ, since these are generated randomly.

2.16 Advanced Notes

1. BIOGRAF recognizes the standard files from the Brookhaven Protein Data Bank and the Cambridge crystallographic data base, as well as CHEMLAB CHM, CHEMLAB MOL, Macromodel, Moledit, and MOPAC files. The program also has its own internal file format which is similar to Brookhaven but contains additional force field information. (See File Formats in the *BIOGRAF Reference Manual,* Appendix D for further details.)
2. If you have groups created with the distance criteria and you've changed the original structure or coordinates, then you may want to recreate the distance groups to reflect these changes. This is done using the options under Make Groups.

This concludes Chapter 2. To leave BIOGRAF, pick **exit** from the Top menu, then verify the exit by clicking the OK button in the alert box.

Chapter 3

Small Molecule Building and Drawing

This chapter describes how to build and draw molecules with BIOGRAF. We will make the amino acid L-histidine as shown later in Figure 3.4.

BIOGRAF supports several builders. Currently, there are builders for organic molecules, peptides, DNA and RNA, lipids, carbohydrates, crystals, and solvents. Each of the builders automatically reads its own set of fragment files in a special directory. These fragment libraries provide short-cuts to building by taking advantage of the fact that specialized molecules such as peptides, DNA, lipids, and carbohydrates are made from well-defined fragments in a predefined way, *e.g.*, peptides are single chains of amino acids. The user can build up a large structure by combining peptide residues, fatty acids, small organic molecules, *etc.*

The Organic Builder may be thought of as the generic builder; it is the most flexible of the builders and can be used to make any molecule. Although L-histidine can be brought into BIOGRAF easily using the Peptide Builder, we use the Organic Builder to create the molecule as an exercise in using the functions available on the builder menus. Examples using the more specialized builders are given in subsequent chapters. It is assumed that you have read Chapter 1.

3.1 Getting Started

1. Enter the command ***biograf***. This starts up BIOGRAF without reading in information about any molecule. As explained in the introductory chapter, this command may be different at your site. The Top menu appears in the Menu Pad.

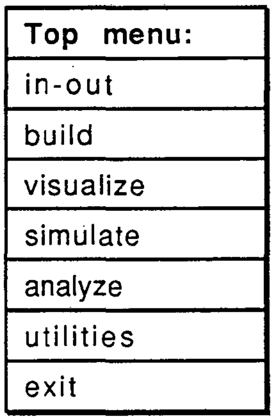


Figure 3.1:
Top menu

3.2 Setting Up the Builder

To build the L-histidine molecule, we use the Organic Builder found under **build**.

1. Toggle **build** on the Top menu to display the Build menu shown in Figure 3.2. The first eight options on the menu provide access to the Organic Builder and each of the specialized builders. The remaining options serve to complete the building process.

2. Pick **organic**.

   Before we begin building, we must set up an internal file which will hold information about the new molecule. The **setup** option is therefore automatically invoked and a dialog box comes up prompting us to *Enter a descriptor for the new file*.

2. Enter ***lhis*** as the descriptor for the file. Complete the entry by selecting the OK button. A file descriptor names the file that holds information about a molecule. A descriptor name can be anything up to 14 characters.

   The Organic menu appears as shown in Figure 3.3.

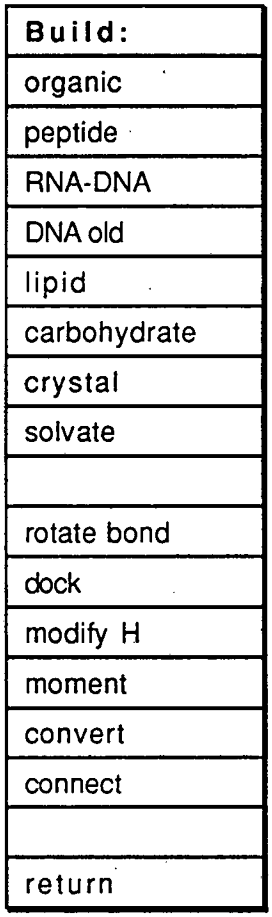


Figure 3.2: Build menu

The Organic Builder identifies every atom with a residue name, residue number, and chain designator regardless of whether or not it is part of a polypeptide. The program automatically uses the default values for these when a new file is created. The defaults are ORG for residue name, 1 for residue number, and A for chain designator. If the defaults are not acceptable, they can be changed using the **new atom ids** option. The defaults are fine for the molecule we are building, so no changes are needed.

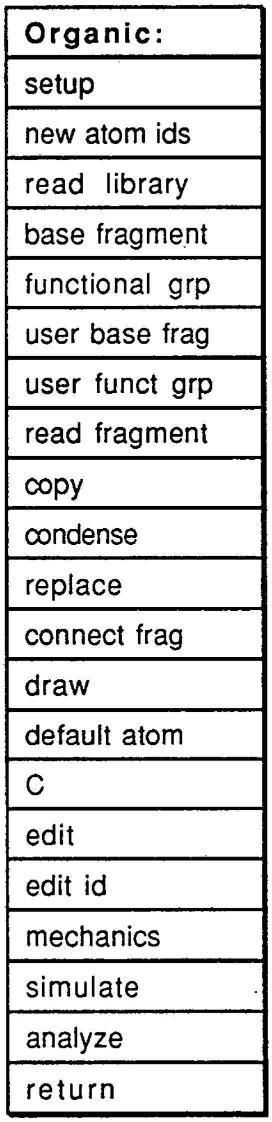


Figure 3.3: Organic menu

3.3 The Half-Bond Representation

The Organic Builder makes molecules in a half-bond representation by coloring each atom according to its element and then making the half of the bond closest to the atom the same color as the atom. The colors BIOGRAF uses for some of the most common elements are shown in Table 3.1.

Element	Color	Element	Color
H	White	S	Yellow
C	Aqua	Cl	Green
N	Blue	Ca	Magenta
O	Red	Ti	Cyan (blue-green)
F	Green	Fe	Cyan (blue-green)
Na	Magenta	Zn	Magenta
Al	Magenta	Br	Green
Si	Aqua	Ru	Cyan (blue-green)
P	Yellow		

Table 3.1: BIOGRAF Element Colors

3.4 Setting up the Base Fragment

Figure 3.4 shows a conformational formula of L-histidine. L-histidine consists of an imidazole ring attached to the β carbon of alanine. There are many ways to build L-histidine using the Organic Builder. We'll start building with a base fragment which serves as the primary building block to which we add other functional groups. The organic fragment library does not have an imidazole ring, but it does have the five-membered ring, cyclopentane, so we'll choose cyclopentane for our base fragment. We can then change two of the carbons in the ring to nitrogens and add two double bonds to make it an imidazole ring.

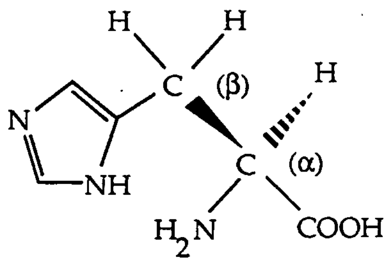


Figure 3.4: L-histidine

Selections:	
acetylene	furan
ammonia	h3po4
anthracene	hcl
benzene	hcn
cyclobutane	hono2
cyclohexane	indene
cyclopentane	methane
cyclopropane	naphthalene
decalin_cis	phenalene
decalin_trans	phenanthrene
ethane	steroid
ethylene	tetralin
fluorene	water
formaldehyde	
formic_acid	return

Figure 3.5: Base fragment library menu

1. Pick **base fragment** on the Organic menu. BIOGRAF prompts *Pick the base fragment to be read* and displays a new menu listing the library of organic fragments as shown in Figure 3.5.

2. Pick **cyclopentane** from the base fragment library menu. BIOGRAF retrieves the fragment, displays cyclopentane in half-bond colors in the graphics window, and repeats the prompt to pick a base fragment. We don't need any additional base fragments, so pick **return**.

   BIOGRAF always centers the base fragment in the graphics window and sets the origin of rotation for the display on it.

3. Enlarge the fragment if necessary.

We still need to replace two of the carbons with nitrogens and add two double bonds to make the imidazole ring. First, let's add the equivalent of alanine to the cyclopentane molecule.

3.5 Adding the Functional Groups

To get the equivalent of alanine on the cyclopentane ring, we'll use the functional group library to add an ethyl group, and then add a carboxy group to the ethyl. We could also use the functional group library to add the amino group, but we'll draw this on instead so that we can illustrate the Organic Builder's draw function.

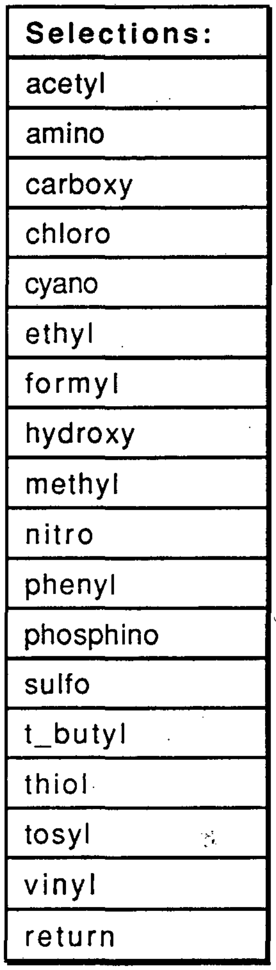


Figure 3.6: Functional group library menu

1. Pick **functional grp**. The menu shown in Figure 3.6 appears listing the functional groups available. BIOGRAF prompts *Pick the functional group to be read.*

2. Pick **ethyl**. BIOGRAF outlines the graphics window area with a red border and prompts *Pick hydrogen(s) to be replaced with functional groups. Pick anywhere else to terminate.*

3. Pick one of the hydrogen atoms. The ethyl group is attached to the cyclopentane ring at the picked location (replacing the picked atom) and the new structure is centered on the display. The covalent bond formed between the ethyl group and the cyclopentane is automatically given an appropriate bond length and angle.

   The list of functional groups becomes dimmed in the menu (indicating they are currently non-pickable).

4. We don't want to add any more *ethyl* groups, so pick anywhere off the molecule. The red border disappears and we are again prompted *Pick the functional group to be read.*

5. Pick **carboxy** from the functional group library, then join it to the ethyl side-chain by picking one of the hydrogen atoms on the end carbon of the ethyl. Pick off the molecule, then pick **return** on the Functional group library menu to get back to the Organic menu.

3.6 Deleting Hydrogen Atoms

This section illustrates some of the editing operations available under the Organic Builder which are useful in the construction of L-histidine. Since we intend to add bonds to the cyclopentane ring and change the element type of some of the atoms, it is convenient to remove the hydrogen atoms for the time being and add them back later. We'll use the **select** function on the Edit menu to specify the hydrogen atoms, then use **delete** to delete them. Then we'll use **auto type** to change two carbons on the ring into nitrogens.

1. Pick **edit** on the Organic menu to display the Edit menu shown in Figure 3.7.

2. Pick **select**. The **select** option is used to specify which atoms are to be included when using the **drag**, **drag-clean**, **move**, **invert**, **add H**, and **delete** options. You may select a substructure or range of atoms, select atoms individually, or specify that only hydrogens, heteroatoms, or lone pairs be included. This is done by picking the appropriate option on the Selection methods menu which appears as shown in Figure 3.8. The default selection method is **continuous**, which allows the user to continue to perform the desired operation on individual atoms selected one at a time.

3. Pick **hydrogens** in response to the prompt.

   We can use the **show** option to verify that all the hydrogens have been selected.

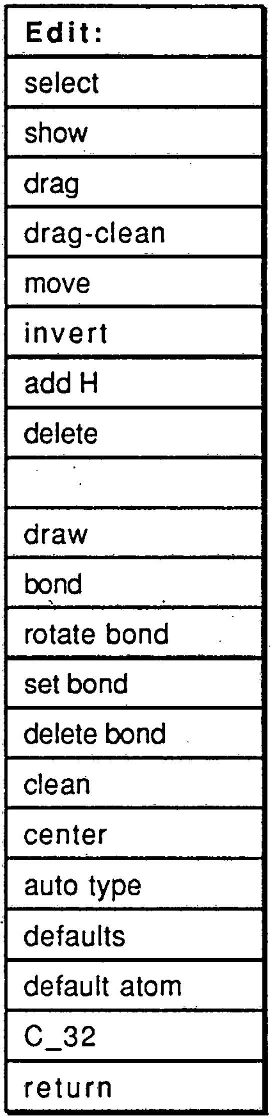


Figure 3.7: Edit menu

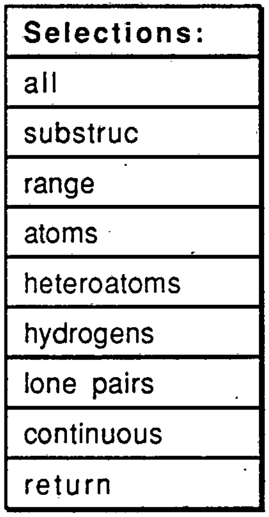


Figure 3.8: Selection methods menu

4. Pick **show**. Red crosses mark the selected atoms (all the hydrogens). Pick anywhere in the graphics window to continue. The red crosses disappear.

5. Pick **delete**. The message *Working...* appears. When finished, the hydrogen atoms disappear from the structure.

3.7 Changing Element Types

Next we change two carbon atoms on the ring into nitrogen atoms using the **auto type** function. When we are finished, the structure should be similar to the one shown below.

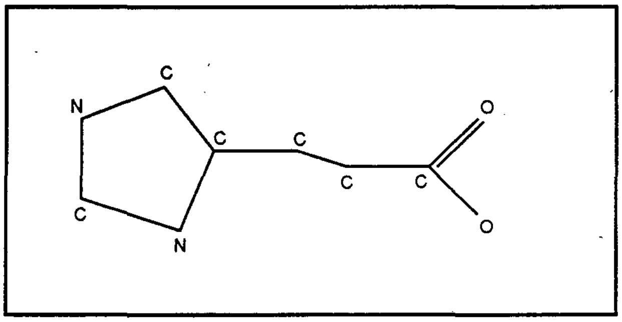


Figure 3.9: Edited structure (hydrogens deleted, two carbons replaced with nitrogens)

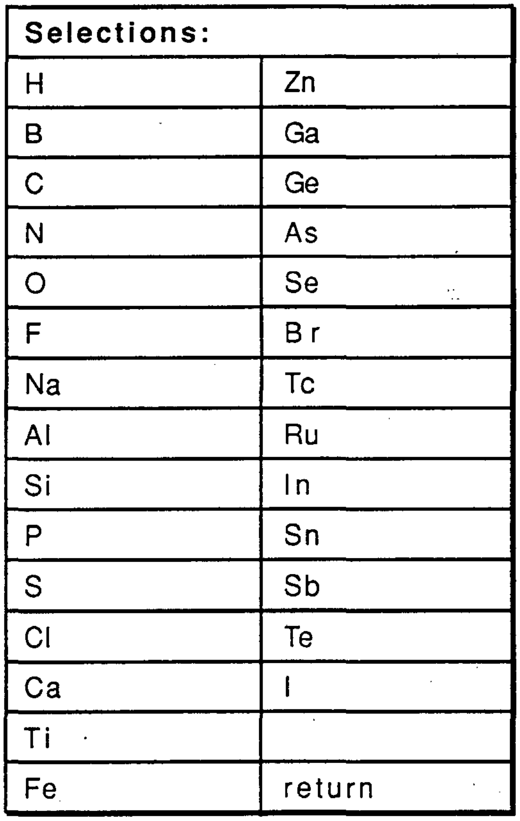
1. Pick **auto type**. Labels appear on each atom specifying its element type and a new menu appears listing the element symbols as shown in Figure 3.10. We are prompted:

   Pick element symbol and corresponding atoms.
   Pick off the molecule to return to the atom type menu.
   Pick RETURN to execute auto typing.

2. Pick **N** from the list of elements. This changes the element type to nitrogen. BIOGRAF surrounds the graphics window with a red border.

3. Going clockwise, pick the carbon atom in the cyclopentane ring which is adjacent to the ring carbon where we attached the ethyl group (refer to Figure 3.9 above). BIOGRAF changes the element type of the picked atom to N and gives it the appropriate half-bond color. A red cross marks the picked atom.

4. If you picked the wrong atom, change it back to carbon (pick off the molecule to return to the Atomic elements menu, pick C, then pick the atom again), then repeat steps 2 and 3 above, this time picking the correct atom.



Selections:	
H	Zn
B	Ga
C	Ge
N	As
O	Se
F	Br
Na	Tc
Al	Ru
Si	In
P	Sn
S	Sb
Cl	Te
Ca	I
Ti	
Fe	return

Figure 3.10: Atomic elements menu (Dreiding-II force field)

5. Continuing in the same direction, pick the second carbon over from the one you just picked (skipping one ring carbon in between). BIOGRAF changes it into a nitrogen atom and updates the half-bond colors. Terminate the atom selection by picking off the molecule.

6. Pick **return** on the Atomic elements menu to execute auto typing. BIOGRAF automatically generates the force-field atom type for each atom in the molecule and changes the atom labels from an element symbol to a type symbol. For atom types, a five character label is used with the nomenclature shown in Table 3.2. For example, the atom type C_32 corresponds to an sp^3 hybridized carbon atom that has covalent bonds to two implied (implicit) hydrogen atoms.

7. Verify that the atoms in the molecule have the correct type for the current structure.

8. Click anywhere to remove the atom type labels from the display.

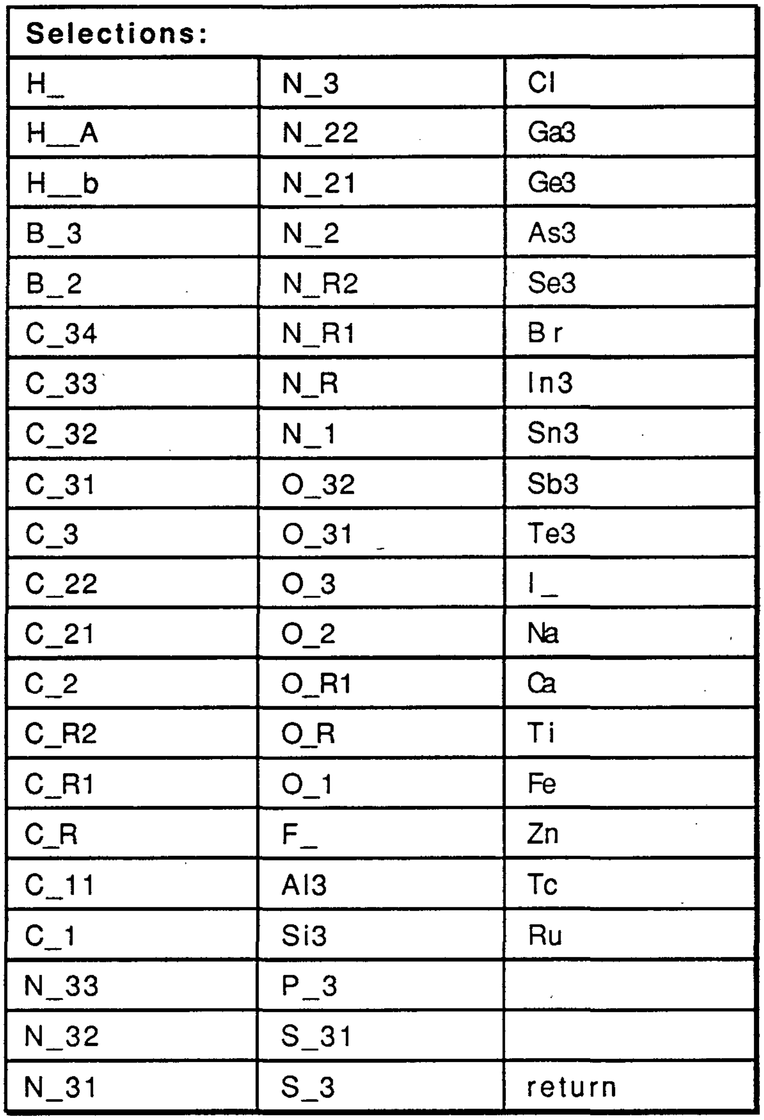
Char	**Meaning**
1-2	Element symbol: An underscore appears in the second column if the symbol has only one letter
3	Hybridization: 3 - tetrahedral 2 - trigonal 1 - linear R - resonant
4	Number of bound implicit hydrogens
5	Alternate parameters indicator (alphabetic)

Table 3.2: BIOGRAF atom types

3.8 Drawing Structures

In addition to making bonds by joining fragments, BIOGRAF also lets you draw structures in 3-D by using the mouse to sketch the bonds. Here we describe how to draw an amino group on what will be the α carbon of histidine. The draw operation draws in the atom type listed under **default atom** on the Edit menu. **C_32** is the default. We want to draw in a nitrogen atom, so we'll change this to **N_3**.

1. Pick **default atom**. The Atom types menu shown in Figure 3.11 appears listing the atom types for the current force field.
2. Pick **N_3**. This tells the program to make nitrogen atoms with sp^3 hybridization.
3. Pick **draw** on the Edit menu. BIOGRAF prompts *Draw the structure. Pick an atom twice to stop a connected sequence. Pick outside the red border to stop drawing.*
4. Pick what will be the α carbon of histidine (refer to Figure 3.13). A red cross marks the picked atom.



Selections:		
H_	N_3	Cl
H__A	N_22	Ga3
H__b	N_21	Ge3
B_3	N_2	As3
B_2	N_R2	Se3
C_34	N_R1	Br
C_33	N_R	In3
C_32	N_1	Sn3
C_31	O_32	Sb3
C_3	O_31	Te3
C_22	O_3	I_
C_21	O_2	Na
C_2	O_R1	Ca
C_R2	O_R	Ti
C_R1	O_1	Fe
C_R	F_	Zn
C_11	Al3	Tc
C_1	Si3	Ru
N_33	P_3	
N_32	S_31	
N_31	S_3	return

Figure 3.11:
Atom types menu (Dreiding-II force field)

To make bonds, **draw** connects the point that you pick on the display to the last point that you picked. It places the bond in a plane that contains the previously picked point and is parallel to the display, that is, the bond lies in an x-y plane at the z coordinate of the previously picked point.

5. Pick a point about one bond length away from the α carbon. A bond is drawn from the α carbon to a nitrogen atom at that point.

6. Pick the nitrogen atom you just added again. Picking an atom twice completes the bonding circuit.

7. Pick outside the red border surrounding the graphics window. This stops the drawing process.

   Note: The **draw** operation does not automatically add hydrogen atoms to a sketch. You must draw them in yourself or use the **add H** option. We'll do this later, but first we'll add the double bonds.

3.9 Adding Double Bonds with the Bond Operation

The Edit menu has a **bond** option which allows one to add bonds to a structure. The **bond** function differs from the **draw** function in that **bond** only works between existing atoms while **draw** may also be used to enter new atoms. We'll use the **bond** operation to add two double bonds to the ring and thus make it an imidazole ring.

1. Pick **bond** on the Edit menu. BIOGRAF outlines the graphics window in red and prompts *Pick the atoms of the bond(s) to be created. Pick off the molecule to quit.*

2. Pick the ring carbon that is joined to the side-chain. A red cross appears on the picked atom.

3. Pick the ring carbon adjacent to the one you just picked. BIOGRAF displays the message *Working...*, and draws in the extra bond between the two carbon atoms on the ring.

4. Now add the other double bond: pick the nitrogen atom adjacent to the last picked carbon, then continuing in the same direction, pick the carbon atom adjacent to it. We should now have an imidazole ring with two double bonds. Pick off the molecule to quit.

3.10 Adding the Missing Hydrogen Atoms

We now have a rather strange looking histidine molecule with its hydrogen atoms missing. To regularize the structure, we need to add the hydrogen atoms and then minimize the energy of the molecule. Before adding the hydrogens, we need to make sure all the atom types are correct.

1. Pick **auto type**. BIOGRAF again displays the Atomic elements menu and prompts *Pick element symbol and corresponding atoms. Pick off the molecule to return to the atom type menu. Pick RETURN to execute auto typing.*

2. Pick **return** (we don't need to change any element types this time). BIOGRAF generates the atom types and puts a label on each atom.

3. Check the atom types to verify that they are correct. Note that the nitrogen atom we drew in is labeled N_33, indicating that it will be bonded to three hydrogen atoms. This would give us a positively charged amino group. We'll change the atom type to N_32 so that only two hydrogens will be added to the amino group, making it neutral. We'll also change the atom

type of the carboxy oxygen (the single bonded one) from O_2 to O_R so that it will accept a hydrogen. This is done using the **atom type** option under **edit id**. Click anywhere to remove the atom type labels from the display.

4. **Return** to the Organic menu, then pick **edit id**. The menu shown in Figure 3.12 is displayed.

5. Pick **atom type**. Labels appear on the atoms, the Atom types menu is displayed and we are prompted *Pick the atom type and the corresponding atoms. Pick off the molecule to select a new atom type. Pick RETURN to continue.*

6. Pick **N_32** from the Atom types menu. The graphics window becomes surrounded with a red border.

7. Pick the nitrogen atom we drew in (labeled N_33). It is marked with a red cross and its atom type changes to N_32. Pick off the molecule.

8. Pick **O_R** from the Atom types menu, then pick the carboxy oxygen (the single bonded one). Its atom type changes to O_R. Pick off the molecule, then pick **return**.

9. Click anywhere to remove the atom type labels from the display.

   We are now ready to add the hydrogen atoms.

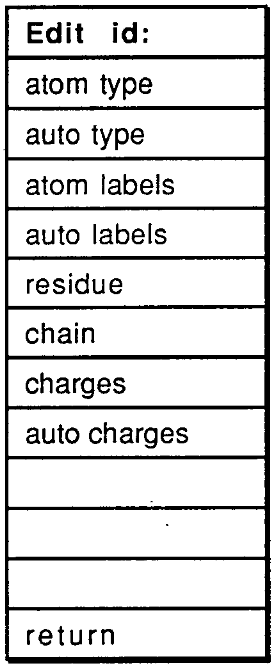


Figure 3.12:
Edit id menu

10. **Return** to the Organic menu, pick **edit** to bring up the Edit menu, pick **select**, then pick **all**. We want to add hydrogens to the entire structure.

11. Pick **add H**. Information about the molecule appears in the text window. After the program generates the positions, hydrogen atoms appear on the structure in the appropriate half-bond colors.

    BIOGRAF selects for optimal hydrogen bonding by pointing donor hydrogen bonds towards the nearest acceptor. If the L-histidine molecule were going to interact with other molecules, we could tell BIOGRAF to take the additional molecules into account when it generates hydrogen atom positions for the amino acid. The program would then add H atoms to the L-histidine such that they point towards neighboring hydrogen bond acceptors on the other molecules, if possible. However, at the moment we just have the L-histidine molecule, so we need not specify any other molecules.

3.11 Getting the Correct Stereochemistry

With the exception of glycine, all of the amino acids are chiral molecules. Histidine has one center of chirality about its asymmetric carbon (the α carbon), and thus comes in two forms, D and L. The D form rotates polarized light to the right, while the L form rotates it to the left. Figure 3.13 compares the D and L forms of histidine. The L form must have the side chain containing the imidazole ring projecting forward when the amino and carboxy groups are both in the plane of the display; in the D form, this side-chain points backwards. The L form is the naturally occurring one, so we'll build it this way. If the stereochemistry of our current structure is incorrect, we can use the **invert** function to correct it.

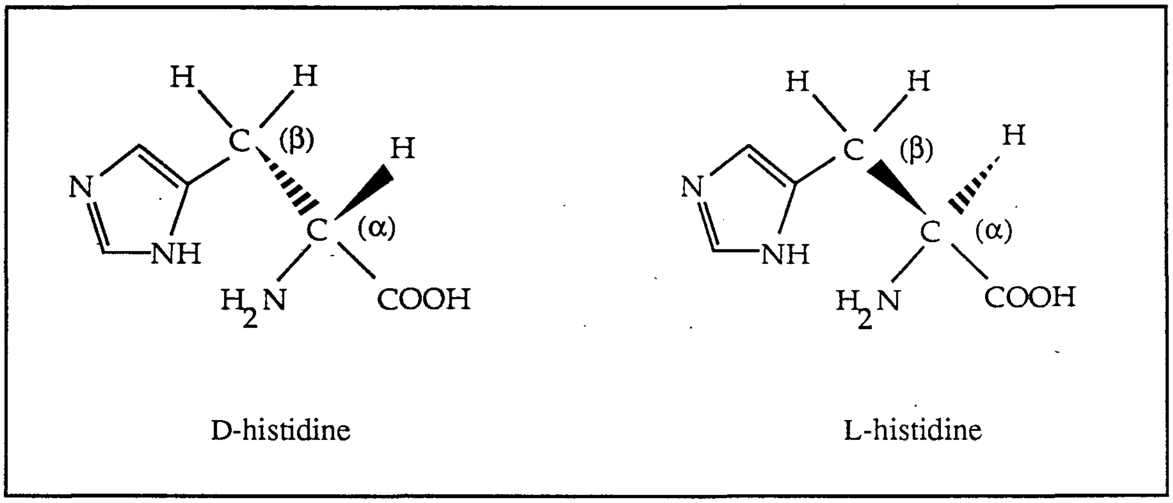


Figure 3.13: Comparison of D and L forms of histidine

1. Orient the molecule so that the amino and carboxy groups are both in the plane of the graphics display.

2. Note whether the side-chain coming off the α carbon is projecting forward. (If you already have the correct stereochemistry, go through the **invert** operation to see how it is done and then **invert** the group a second time at the end.)

3. Pick **select**, then pick **continous.** This allows us to invert about selected chiral atoms.

4. Pick **invert** on the Edit menu. BIOGRAF puts a red border around the graphics window and prompts us to pick an atom.

5. Pick the α carbon. A red cross marks the picked atom, then BIOGRAF inverts the structure by repositioning the amino group.

6. Pick off the molecule.

7. Orient the molecule such that the carboxy and amino groups are again in the plane of the graphics display. If the side-chain was originally projecting out, it will now be pointing backward; whereas if it was originally pointing back, it will now project forward.

8. If the inversion positioned the side-chain so that it is pointing backward, **invert** it a second time to put it back in the correct stereochemistry (the L form).

9. **Return** to the Organic menu.

Don't worry if your L-histidine looks distorted. We can regularize its structure by finding a low energy conformation. This is described in the next section.

3.12 Minimizing the Energy

The best way to regularize the structure of a molecule is to do an energy minimization. In this section, we minimize the energy of L-histidine to find a conformation that corresponds to the nearest local minimum in the potential energy surface.

1. Pick **mechanics** on the Organic menu. This brings up the Mechanics menu shown in Figure 3.14 and automatically sets up an energy expression which includes every atom in the molecule and allows every atom to move during the calculation. The conformation of the molecule will be updated on the display at each step of the minimization. If we had picked **simulate** instead of **mechanics** we would be required to set up the energy expression, choosing which atoms were movable, which were fixed, and which were updated.

   Each time BIOGRAF enters the Mechanics menu, it recalculates the lists of hydrogen-bond and nonbond interactions and brings up the Energy results menu as shown in Figure 3.15. A description of the items on this menu is given in the adjacent table. No values are listed since no calculations have been done yet.

2. Pick **one energy** on the Mechanics menu to tell BIOGRAF to evaluate the energy of the histidine molecule in its present conformation.

After BIOGRAF computes the energies, it lists the results in the text window and in the Energy results menu. All energies are in kcal/mol. If an energy is too large to fit into the menu box, the program prints a row of asterisks: *********.

1. Make a list of the initial energies on the Energy results menu.
2. Pick **minimize** on the Mechanics menu. The message *Minimizing....* appears during the calculations. The energies displayed on the Energy results menu show a rapid decrease early in the calculation.
3. When the calculation finishes, check **delta energy** on the Energy results menu. If it is non-negligible compared to the other energies, then the calculation has not yet converged.

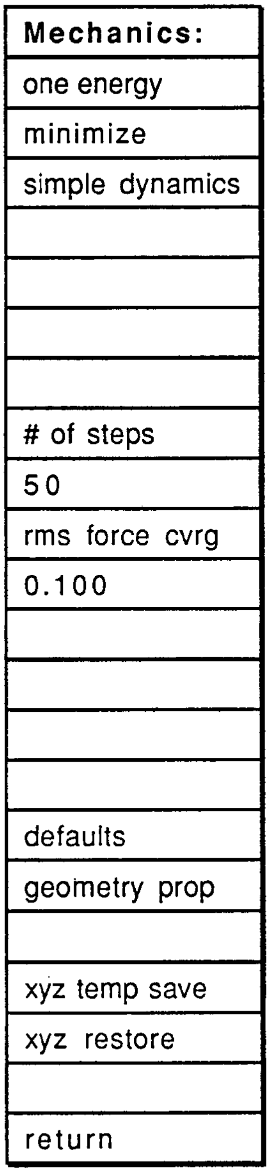
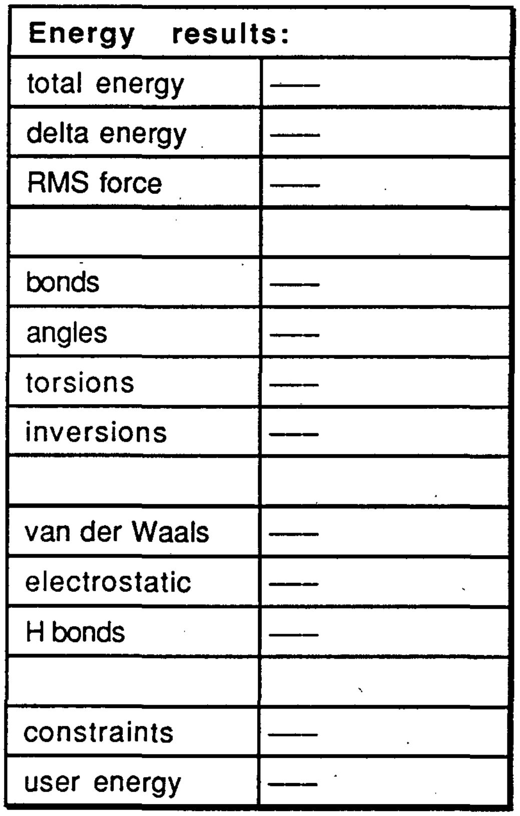


Figure 3.14: Mechanics menu

4. If **delta energy** is still comparatively large, pick **# of steps**. BIOGRAF prompts you for the number of steps.
5. Type in the number of steps you think might be needed to complete the minimization. Select the OK button to complete the entry. If this number of steps is not sufficient, you can always simply pick **minimize** again.



Energy results:	
total energy	—
delta energy	—
RMS force	—
bonds	—
angles	—
torsions	—
inversions	—
van der Waals	—
electrostatic	—
H bonds	—
constraints	—
user energy	—

Figure 3.15: Energy results menu

Energy	Description
total energy	Current total energy
delta energy	Change in total energy from previous step
RMS force	Root mean square force (derivative of potential energy)
bonds	Bond stretching
angles	Bond angle bending
torsions	Dihedral angle torsions
inversions	Inversions
van der Waals	van der Waals forces
electrostatic	Electrostatic forces
H bonds	Hydrogen bonds
constraints	Constraints
user energy	User energy

Table 3.3: Description of Energy results menu items

6. Repeat steps 2-5 above as many times as necessary until you are satisfied that the calculation has converged. The abort key (F12, F10 on the Titan) may be used at any time to terminate the calculation.

7. Compare the final energies to the initial energies. After minimization, the **total energy** should be on the order of 25 kcal/mol.

   L-histidine has a number of favorable conformations, each of which arises from a local minimum in its potential energy surface. Chapter 6 describes how to use molecular dynamics to find lower energy conformations of a molecule, and Chapter 7 describes how to do a conformational search for better conformations.

3.13 Saving Information

After you finish building a molecule, it is a good idea to write it to disk. Saving information protects you from losing work and in subsequent sessions allows you to resume where you left off without repeating all of the steps of the previous session.

1. Click the TOP MENU button to get back to the Top menu, then pick **in-out** to display the In-Out menu (see Figure 3.16).

2. Make sure that the **by file** toggle is highlighted next to the write defaults.

3. Pick **write**. The prompt *Pick the file type to be output* is displayed and the File types menu appears (see Figure 3.17).

   In order to reconstruct a structure, BIOGRAF needs the information about the coordinates of the atoms in the structure. The program saves coordinate information in coordinate files. Ten different coordinate file formats are available. These are listed at the top of the menu.

4. Pick **BioDesign** to indicate the BioDesign General Format file type. This type is recommended for internal use with BIOGRAF since it contains all pertinent information.

   Since there is only one file in the system (**lhis**), it is automatically choosen as the source of coordinate information. A dialog box comes up and you are asked to *Enter a name for the output file:*

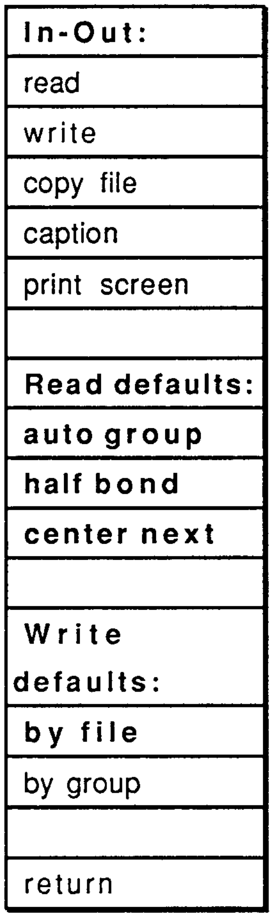


Figure 3.16: In-Out menu

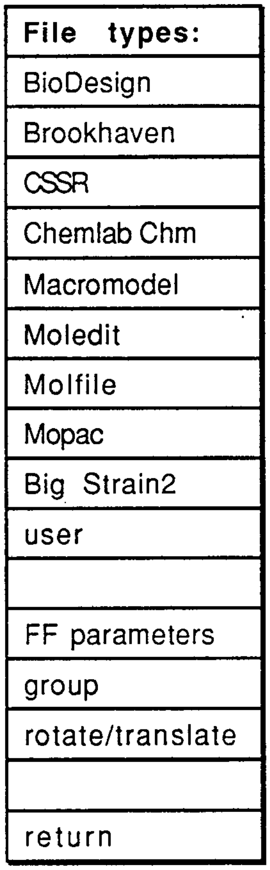


Figure 3.17: File types menu

5. Type ***lhis.bgf*** in the dialog box. Include disk and directory information if not saving it in your working directory. Verify the entry by clicking OK. BIOGRAF prompts you to enter remarks for the output file.

6. Type in remarks that will identify the file the next time you use it. Terminate the remarks with a blank line (Click OK or press <Enter>).

   Messages appear in the text window and in the Modules/Messages window while the program writes the data file.

3.14 Conclusions

The L-histidine molecule provides an example of the building and drawing capabilities of BIOGRAF. You can make any molecule using the Organic Builder; one of the other builders may be used to create more specialized molecules.

After you build a molecule, BIOGRAF can optimize its structure by

- doing a minimization,
- performing a search of conformational space (see Chapter 5), or
- running molecular dynamics calculations (see Chapter 6).

The program makes it possible to predict the likely conformation of a molecule and whether one conformation is more stable than another.

To conclude this tutorial, complete the following exercise:

1. Read in the *intro.bgf* file from your distribution data directory. Typically the full specifications for the distribution data directory are */usr/msi/biogv300/data.* The distribution data files are always located in the */biogv300/data* directory; however, the first part of the path specifications may differ on your system. To determine the location of the */biogv300/data* directory on your system, type *echo $MSIUSR* from a systems window. The directory listed in response to this command should precede */biogv300/data* when giving directory specifications.

   The morphine molecule appears in the graphics window.

2. Go into the Organic Builder and build a replica of the morphine group. To check your work, toggle on the group you created and the **morphine** group you read in simultaneously to see if they match.

3.15 Advanced Notes

BIOGRAF recognizes the standard files from the Brookhaven Protein Data Bank and the Cambridge crystallographic data base, as well as CHEMLAB CHM, CHEMLAB MOL, Macromodel, Moledit, and MOPAC files. The program also has its own internal file format which is similar to Brookhaven but contains additional force field information. (See Appendix D of the *BIOGRAF Reference Manual* for further details.)

To specify your own fragment library to use with the Organic Builder, proceed as follows:

1. Before going into BIOGRAF, make the BioDesign General Format coordinate files for your library and store them in their own directory.
2. In BIOGRAF:
   a) Pick **read library** on the Organic menu. This reads in the directory information for the libraries.
   b) Pick **user base frag** or **user funct grp** on the adjacent menu.
   c) Enter the name of the directory that contains your library.
   d) Pick **user base frag** or **user funct grp** on the Organic menu. A menu appears displaying your user-defined base fragment or functional group library.

The residue names and numbers used by the Organic Builder are a manifestation of the fact that the BIOGRAF system of storing atoms is similar to the convention used by the Brookhaven Protein Data Bank.

This concludes Chapter 3. To leave BIOGRAF, pick **exit** on the Top menu and then confirm by picking OK in the alert box.

Chapter 4

Creating Surfaces

This chapter describes how to construct and display dotted molecular surfaces with BIOGRAF. Both van der Waals and solvent accessible surfaces can be generated. The generated surface may be color-coded by atom type or to show hydrogen bond donors and acceptors. Electrostatic surfaces can also be displayed, showing electrostatic potentials and color-coded representations of the distribution of charge. These depictions of molecular surfaces can be useful in describing the steric extent of a molecule, in studying its accessibility to various solvents, and in looking at its interaction with other molecules.

In this chapter, we will look at the interaction between an enzyme and its inhibitor. The enzyme, thermolysin, contains zinc and catalyzes the hydrolysis of peptides. It is inhibited by the peptide analogue CLT (*N*-[1-carboxy-3-phenylpropyl]-L-leucyl-L-tryptophan). We will generate a van der Waals dotted molecular surface for CLT. This CLT surface will then be successively color-coded to show atom type, charge, and electrostatic potential. Files containing a solvent accessible surface describing the active site for thermolysin will then be read in and displayed so that we can examine the enzyme/inhibitor complex. We shall see how well the CLT inhibitor fits into the active site cavity.

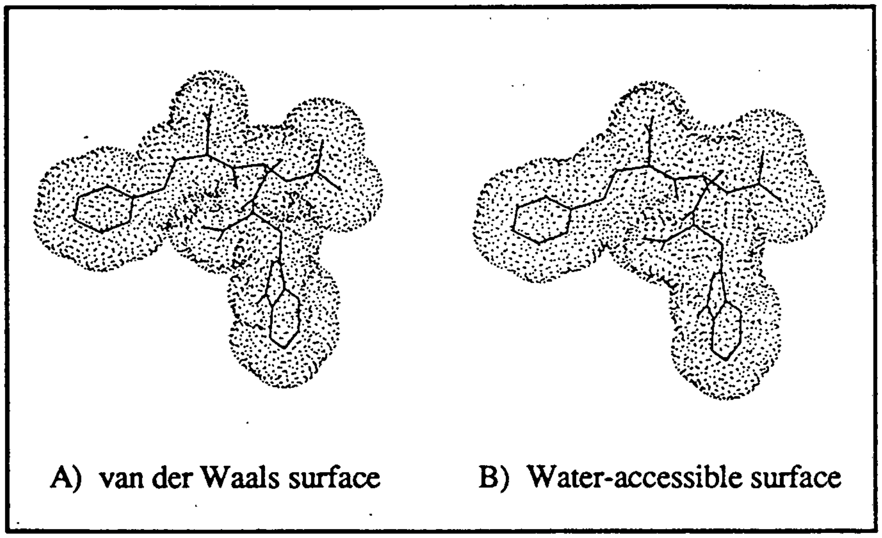
It is assumed that you have already completed Chapter 1.

4.1 How Molecular Surfaces are Created

A van der Waals surface for an atom consists of colored dots which form a sphere around the nucleus of the atom at a distance equal to the standard van der Waals radius for that atom. In a molecule, because the atoms are connected, the spheres around each atom would overlap. In order to create a *molecular* van der Waals surface, therefore, only the non-overlapping parts of the spheres are used. This is done by calculating the dot positions for a probe sphere rolling along the van der Waals spheres of the outer (accessible) atoms of the molecule.

If a probe radius of zero is used, then a van der Waals surface is generated. However, if the probe radius is set to 1.4 Å, the van der Waals radius of water, then a water-accessible surface is generated. This surface would just show the parts of the molecule that water can touch. Accessibility to other solvents can similarly be displayed by appropriately adjusting the probe radius. A van der Waals surface and a water-accessible surface for CLT are shown in Figure 4.1.

Figure 4.1: A van der Waals surface for CLT generated with a probe radius set to zero (A) is compared to a water-accessible surface for this same molecule where a probe radius of 1.4 Å was used (B).


A) van der Waals surface B) Water-accessible surface

4.2 The Surface Menu

The surface functions are found on the Visualize menu. The Surface menu, shown in Figure 4.2, includes surface operations and display operations. The surface operations are used for calculating the positions of the dots and for reading and writing surface files (dot files). The dot coordinates for the surface last calculated or read in through a dot file are stored in memory. Display operations control the computation and display of surface properties. This allows the dot surfaces previously generated to be color-coded in order to illustrate certain features. The color-coding can be done by group or according to other selected properties including atom type, electrostatic potential, atom charge or hydrogen bond donor/acceptor capability. This splitting of the surface and display functions has the advantage of allowing the displayed property to be changed without having to recompute the positions of the surface dots.

Surfaces can be calculated and surface files can be written or read without showing the surface on the graphics display. In order to see the dotted surface, the display functions *must* be executed.

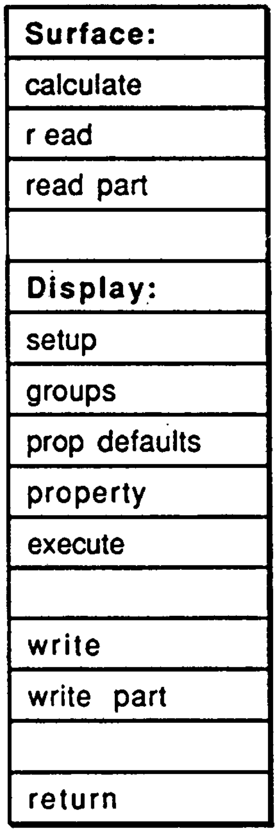


Figure 4.2: Surface menu

4.3 Surface Operations

Dot surfaces are generated using the **calculate** function. A surface can be generated for the entire structure on display or just for selected groups in the structure. The user can similarly decide to save the entire surface or he can specify certain groups to be saved using the **write** and **write part** functions. The **read** and **read part** functions read in existing surface files. **Read part** makes it possible to select groups on the Groups menu and input the parts of an existing surface file that correspond to the atoms in these groups. (This can also be done from the Top menu by selecting **in-out**, then choosing the **read** option and picking **dot** for the file type. See Section 2.1 in the *BIOGRAF Reference Manual.*)

4.3.1 Calculate

After choosing the **calculate** function, the Calculate menu shown in Figure 4.3 appears and the user is prompted to select which groups are to be included when the surface is calculated. It is wise to limit the surface computations to areas of the molecule that you are currently interested in, since calculations of a dot molecular surface for hundreds of atoms can require a substantial amount of computer time.

Calculating a surface for a portion of a structure can lead to a problem, however. Dots which are actually internal to the structure can be generated if neighboring atoms are not included in the surface calculation. One would like to be able to eliminate unwanted surface dots from adjacent groups, but still retain the true external surface shape. BIOGRAF provides a way to do this by allowing the user to select two sets of groups: those to be included and those to be excluded. One may initially choose a larger section of the molecule in the groups to be included in the surface calculations, and then eliminate the unwanted surface dots due to adjacent portions by specifying these in the groups to be excluded.

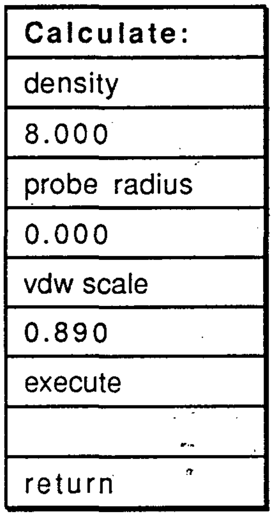


Figure 4.3:
Calculate menu

The **density** option defines the surface density of dots, in dots per square Å. The **probe radius** option allows the probe radius to be set. The default value of zero produces van der Waals surfaces. The **vdw scale** value provides the scaling factor in computing the van der Waals radii of atoms to be surfaced. The initial value of 0.89 yields the standard CPK radii. **Execute** starts the calculations.

4.4 Creating a van der Waals Surface for CLT

In this section, we shall read in the data for the structure of the CLT molecule and then calculate a van der Waals surface for it.

4.4.1 Reading the CLT Data

1. Enter the command ***biograf*** to start up BIOGRAF (this command may be different at your site).

2. Pick **in-out** from the Top menu to display the In-Out menu shown in Figure 4.4.

   Three read defaults appear on the menu which allow the user to automatically create a group for the file that is being read in (**auto group**), to have the structure drawn in a single color or with the half-bond colors (**half bond**), and to center it on the display (**center next**). These apply only when reading in coordinate files (our CLT data is in a coordinate file). We want to create a group for the file, display the molecule with the half-bond colors, and center it, so we'll leave these options at their default settings (**auto group** and **half bond** are toggled on and **center next** is displayed).

3. Pick **read** from the In-Out menu. The File types menu appears displaying the list of file types as shown in Figure 4.5. BIOGRAF prompts *Pick the file type to be input:*

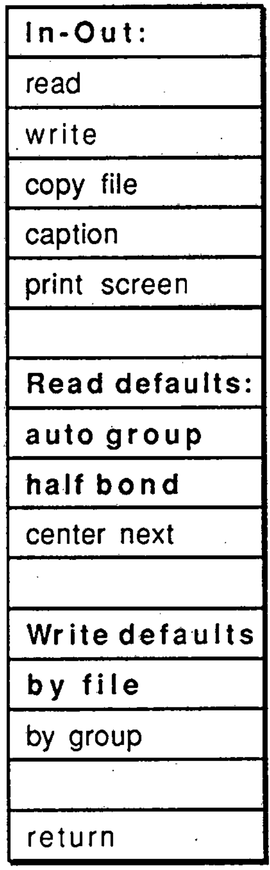


Figure 4.4:
In-Out menu

4. The data describing the structure of the CLT molecule is contained in a BioDesign General Format file called *cltxray.bgf*, so pick **BioDesign** as the file type to be input. The program then brings up the file name entry dialog box and prompts *Choose the file to be input.*

   The *cltxray.bgf* file is located in your distribution directory. Typically the full specifications for the distribution directory are */usr/msi/biogv300/data.* The distribution data files are always located in the */biogv300/data* directory; however, the first part of the path specifications may differ on your system. To determine the location of the */biogv300/data* directory on your system, type *echo $MSIUSR* from a systems window. The directory listed in response to this command should precede */biogv300/data* when giving directory specifications.

5. Use the browser box to locate your distribution data directory, then select the ***cltxray.bgf*** file from the listings. Verify the entry by clicking the OK button.

   When BIOGRAF has finished reading in the data, the CLT molecule is displayed. For BioDesign General Format and Brookhaven file types, a file descriptor is automatically supplied. A group of the same name is also created.

6. Rotate the structure about its different axes to examine the CLT molecule.

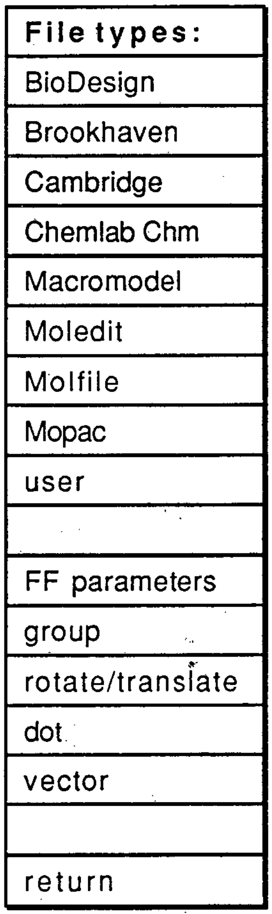


Figure 4.5:
File types menu

4.4.2 Calculating the van der Waals Surface

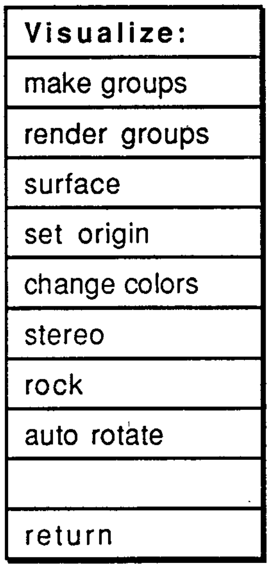


Figure 4.6: Visualize menu

Now that we have the CLT molecule on the display, we shall calculate a van der Waals surface for it. We'll therefore leave the probe radius at its default setting of zero. The defaults are suitable for all the other options as well. Since there is only one group, it is not necessary to specify the groups to be included and excluded in the surfacing calculation, so these prompts are skipped.

The surface operations are found on the Visualize menu.

1. **Return** to the Top menu, then pick **visualize**. The Visualize menu shown in Figure 4.6 appears.
2. Pick **surface** to display the Surface menu.
3. Pick **calculate** in the Surface menu to display the Calculate menu.
4. Pick **execute** to calculate the positions of the dots.

Messages appear in the Modules/Messages window while the program calculates the surface. When the calculation is completed, the Modules and other buttons reappear. Remember the dots will not yet appear on the display.

4.5 Display Operations

The display operations on the Surface menu are used to control the computation and display of surface properties. This allows the dot surfaces previously generated to be color-coded in order to illustrate certain features. The color-coding can be done by group or according to other selected properties including atom types, electrostatic potential, atom charges or hydrogen bond donors and acceptors. The dot coordinates for the surface to be displayed must have already been calculated or read in through a dot file (using the **calculate, read** or **read part** options) before any display functions can be executed. Once the dot coordinates have been generated, the displayed property can be changed without having to recompute the positions of the surface dots.

Typically, the display options are picked sequentially, from top to bottom starting with **setup** and working through to **execute**. If the whole calculated surface is to be displayed, then the **groups** option may be omitted. Likewise, if the default property parameters are to be used in computing the dot surface colors, then the **prop defaults** option may be skipped. Solid color dot surfaces can be displayed by picking **setup** and **execute** only. However, if color-coding to show surface properties is desired, then the kind of property to be used must be selected using the **property** option. Note that the **change colors** operations (accessed via the Visualize menu) can only be performed on solid color dot surfaces.

In the following sections, the CLT molecular surface will be successively color-coded to show atom type, charge, and electrostatic potential.

4.5.1 Setting Up Groups for Display

A group must always be set up for displaying the dot surfaces of a structure. Since we are initially color-coding the dotted molecular surface by atom type, the first display group to be set up will be named "ATOM".

1. **Return** to the Surface menu, then pick **setup** (in the Display section of the menu). The Groups menu is displayed and the prompt *Pick the group to be created* appears.
2. Pick **add group** from the Groups menu. The new group will contain the dotted surface. A dialog box appears and BIOGRAF prompts *Enter the name of the group to be created.*
3. Type ***ATOM*** as the name of the group. Click OK or press <Enter> to verify the entry.

4.5.2 Color-Coding the Surface to Show Properties

Now that we have a group to contain the dotted surface, we must select which property we want to illustrate. Color-coding of the dotted surface is available for five properties: atom type, electrostatic potential, atomic charge, gross atomic charge (-,0,+), and hydrogen bond donors and acceptors (see Figure 4.7). A "current" selection is also available which uses the property values currently stored in memory. This option would be used when you want to create a new surface group with the same property values as the most recently generated group, or when a dot surface file containing the property colors to be displayed has been read from disk.

The parameters used in computing dot surface colors can be changed using the property defaults (**prop defaults**) option. This option applies only to the electrostatic potential and charge properties. Since our first group will be color-coded according to atom type, the **prop defaults** option doesn't apply here.

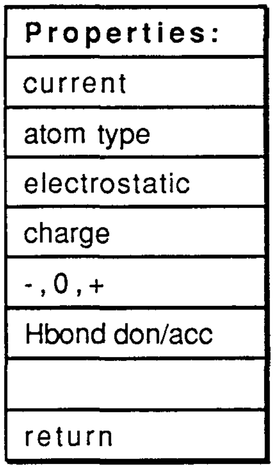


Figure 4.7: Properties menu

A. Color-Coding the Surface to Show Atom Type

1. Pick **property** in the display section of the Surface menu. BIOGRAF responds with the prompt *Select the property for the surface: Pick RETURN to continue.* The Properties menu appears as shown in Figure 4.7.
2. Pick **atom type** in the Properties menu. The message *Working...* appears briefly while BIOGRAF calculates the colors of the dotted molecular surface.

   When the calculation of surface properties is complete, the prompt to select the property reappears.

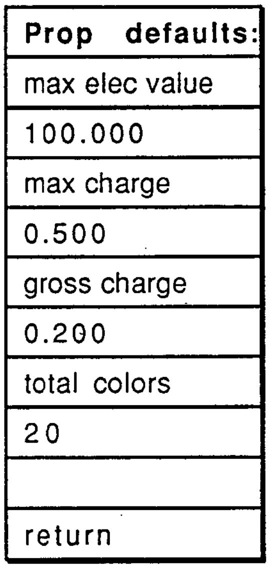
3. **Return** to the Surface menu, then pick **execute** to display the dotted molecular surface. Messages appear while the program generates the surface display. When the calculation is done, the van der Waals surface is displayed and a new group, **ATOM**, is created. The color of the surface varies according to the BIOGRAF half-bond color of the nearest atom: carbons are aqua, oxygens are red, nitrogens are blue, and hydrogens are white.

4. Manipulate the image in the graphics window to study the surface group **ATOM**. When finished, click the DISPLAY button and toggle the **ATOM** group off. Pick **return** to complete the operation.

B. Color-Coding the Surface to Show Charge

We shall now calculate and display a surface color-coded to show charge. The color of the surface will vary according to the charge of the nearest atom, using hues from red (negative) to blue (positive). The CLT molecule is a dianion with two free carboxylate groups. The dotted surface should be red in the region of the carboxylate groups (-0.24), blue-green in the region of the positively charged nitrogens (+0.14), and chartreuse (almost neutral) over the rest of the molecule.

The displayed properties can be changed without recomputing the positions of the dots. We just need to set up a new display group and change the property selected to **charge**. This time we'll also be changing one of the property default parameters using the **prop defaults** option. The Property defaults menu is shown in Figure 4.8.



Prop defaults:
max elec value
100.000
max charge
0.500
gross charge
0.200
total colors
20
return

Figure 4.8: Property defaults menu

1. **Setup** a new display group as we did for **ATOM** above (Section 4.5.1), but name it ***CHARGE***.

2. Pick **prop defaults** to display the Property defaults menu.

3. Pick **max charge**. This parameter defines the mapping of atomic charges into colors. BIOGRAF responds *Enter the desired maximum charge.*

4. Enter ***0.24*** as the maximum absolute value of the charge. Atomic charges between -0.24 and +0.24 will be interpolated between red and blue, and charges outside this region will be displayed as red (negative) or blue (positive). Select OK to complete the entry.

   The **max elec value** and **gross charge** parameters similarly define the mapping of electrostatic potential and gross atomic charges into colors, while **total colors** sets the total number of colors used. The range of total colors is from 3 to 80.

5. **Return** to the Surface menu and color-code the surface as we did for **ATOM** above (Section 4.5.2-A, steps 1-3), but specify charge for the property (pick **property**, then **charge**, then **return**, then **execute**).

6. Manipulate the image on the display to examine the surface group **CHARGE**. The dotted surface should be red in the region of the carboxylate groups (-0.24), blue-green in the region of the positively charged nitrogens (+0.14), and chartreuse (almost neutral) over the rest of the molecule. Click the DISPLAY button and toggle **CHARGE** off in the Display menu when finished. Complete by picking **return**.

C. Color-Coding the Surface to Show Electrostatic Potential

In this section we will create and display a surface of CLT color-coded to show electrostatic potential. The color of this surface will vary according to the electrical potential generated by the presence of nearby charged atoms. We will also be using the **prop defaults** option to change the maximum electrostatic potential parameter (**max elec value**) to 200. This will define the mapping of colors so that for a unit charge, a calculated potential of 200 kcal or greater will be displayed as blue, and potentials less than or equal to -200 kcal will be displayed as red; values in between will vary in hue from red to blue. This should create a CLT surface which is primarily yellow (electrostatically neutral), and red (electrostatically negative) in the region of the free carboxylate groups.

1. **Setup** a new display group as we did for **ATOM** above (Section 4.5.1), but name it ***ELEC***.

2. Pick **prop defaults** and change **max elec value** to *200*.

3. **Return** to the Surface menu, pick **property**, then choose **electrostatic**.

   Electrostatic potential calculations can be rather time consuming when done on very large molecules, so BIOGRAF allows the user to restrict the calculations to smaller portions of the molecule. This is done by specifying which groups are to be considered. The Groups menu comes up as shown in Figure 4.9 and we are prompted *Pick the groups to be considered during the calculation of electrostatic potential.*

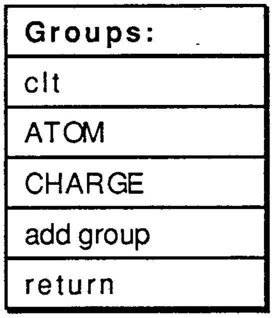


Figure 4.9: Groups menu

4. Pick **clt** from the Groups menu, since this is the only group with atoms in it. Pick **return** to complete the selection.

   While BIOGRAF calculates the colors of the dotted molecular surface, the message *Working...* appears in the Modules/Messages window. When the calculations are complete, the message to pick another property reappears.

5. **Return** to the Surface menu, then pick **execute** to display the newly surfaced group. BIOGRAF prints the message *Color-coding by property....* When the calculations are completed, the newly colored surface group is displayed and the message disappears.

5. Manipulate the image on the display to examine the surface group **ELEC**. The surface should be primarily yellow (electrostatically neutral) except for the region of the free carboxylate groups, where it will be red.

4.6 Saving the Dot Surface

In this section we shall save the dot surface in a dot file. Both the dot positions and the last calculated property, in this case electrostatic potential, will be saved.

1. Pick **write**. BIOGRAF brings up a dialog box and prompts *Enter the name of the data file to be created.*
2. Enter ***clt.dot*** (if not saving the file in your current working directory, include disk and directory information). While BIOGRAF writes the dot file to disk, messages appear in the Modules/Messages window and in the text window.

You may not always want to save the dot surface information for the entire structure. BIOGRAF therefore also provides a **write part** option which saves this information only for selected groups.

4.7 Using Dot Surfaces to Study An Enzyme/Inhibitor Complex

One of the advantages of surfacing is that it describes the steric extent of a molecule and allows us to study its interaction with other molecules. For example, surfacing could be used to describe the active site of an enzyme and to see how well different suspected inhibitors fit into the active site cavity. Color-coding adds to this process by illustrating areas of positive or negative electrostatic potential or charge that may contribute to the way the molecules interact.

In this section, we shall illustrate the usefulness of surfacing in studying the interaction of the active site of the enzyme thermolysin with the inhibitor CLT. Existing files containing structural data for both the thermolysin and CLT molecules will be read in and displayed. We shall then read in two dot files: the one for CLT that we saved above, and another dot file which describes the active site for thermolysin. We shall use these surfaced structures to examine the enzyme/inhibitor complex and see how well the CLT inhibitor fits into the active site cavity.

4.7.1 Reading in the Thermolysin and CLT Data

We shall first examine the unsurfaced structures of thermolysin and CLT. First we'll reset the program to erase all the current information, then we'll read in two coordinate files containing the structural data for thermolysin and CLT. A group file will also be read in which partitions the thermolysin/CLT inhibitor complex into various groups.

1. Click the TOP MENU button to return to the Top menu, pick **utilities**, then pick **program reset**. Verify the reset by clicking OK. When BIOGRAF finishes the program reset, it returns to the Top menu.

2. Pick **in-out** to bring up the In-Out menu (leave **auto group** and **half bond** toggled on).

3. Pick **read** from the In-Out menu, then pick **BioDesign** as the file type to be input.

4. Use the browser box to locate your distribution data directory, then select ***therm.bgf*** from the file names listed. Click OK to verify the selection. The thermolysin molecule appears in the graphics window.

5. Repeat steps 3 and 4 to read in the CLT molecule (pick **read**, pick **BioDesign** for the file type, then select the ***cltxray.bgf*** file from the listings in your distribution directory.

6. To read in the group file for the thermolysin/CLT inhibitor complex, pick **read**, pick **group** for the file type, then select ***therm.grp*** from the file names listed in the distribution directory.

   Messages appear while the group data is being read in. The original thermolysin and CLT groups disappear when the new group data is displayed.

4.7.2 Looking at the Molecular Structures of Thermolysin and CLT

The thermolysin/CLT inhibitor complex has been partitioned into various groups as shown in Table 4.1. The first group contains CLT; the remaining groups contain thermolysin data.

1. Click the DISPLAY button to bring up the Display menu.

2. Zoom out if necessary and toggle the various groups on and off. Rotate and translate the structure about the different axes to examine the enzyme/inhibitor complex.

Group	Description
CLT	The entire CLT molecule
BACK	The thermolysin backbone or main chain
SIDE	The peptide residues or side chains
CALCIUM	The four Ca^{+2} ions
WATER	The crystallographic water molecules
ZINC	The Zn^{+2} ion
CAVITY	The active site side chains

Table 4.1: Groups making up the thermolysin/CLT inhibitor complex

3. Toggle off all the groups except **ZINC** and **BACK**. The zinc ion is coordinated to one water, two histidine residues and one glutamic acid residue. Note how the zinc ion (a magenta sphere) is bound inside a cleft in the enzyme molecule. The zinc ion is necessary for catalytic activity.

4. Toggle on **CALCIUM**. The four calcium ions (blue spheres) contribute to the unusual heat stability of the enzyme.

5. Toggle on **CAVITY**. This local group was created based on the distance of the side chains from the CLT inhibitor; it is our working definition of the thermolysin active site. (See Chapter 7, Section 7.7 for the procedure to create a local group.)

6. Pick **return** from the Display menu when finished.

4.7.3 Reading and Displaying the CLT Dot File

In this section we will read in and display the CLT dot file saved in Section 4.6 above.

1. Click the TOP MENU button to get back to the Top menu, pick **visualize**, then **surface**.

2. Pick **read** in the Surface menu. BIOGRAF brings up the file name entry dialog box and prompts *Enter the name of the dot file to be input.*

3. Locate the directory you saved your CLT dot file in, then select ***clt.dot*** from the listings. While BIOGRAF reads the dot file, messages appear in the Modules/Messages window and in the text window. When the program has finished reading in the file, the Modules and other buttons reappear.

   Remember, the dotted molecular surface will not be visible until after the display functions have been executed.

   Note that dot files can also be read in using **in-out** from the Top menu followed by **read**, but the display functions on the Surface menu must be used to display them.

4. **Setup** a display group to hold the dotted CLT surface (pick **setup**, pick **add group** from the Groups menu, and enter ***INHIB*** as the name of the group.

5. Color-code the surface: pick **property**, then select **current**. **Current** will color-code the surface according to the last calculated property. In this case, the last calculated property was the electrostatic potential in the absence of the enzyme.

6. **Return** to the Surface menu, then pick **execute** to display the surface group color-coded for electrostatic potential.

7. Use the rotate and translate functions to examine the surface group **INHIB**. When finished, click the DISPLAY button to bring up the Display menu then toggle **INHIB** off. Pick **return** to complete the operation and remove the Display menu.

4.7.4 Reading and Displaying the Active Site Dot File

In this section, we shall read in a dot file describing the active site for thermolysin. The surface is a *water-accessible* surface; it shows the parts of the molecule that water can touch. This surface was created by changing the **probe radius** default in the Calculate menu from 0.000 to 1.4, the diameter of water in Å, and surfacing the **CAVITY** and **ZINC** groups. The dotted molecular surface of the active site was then color-coded to show electrostatic potential in the absence of CLT inhibitor.

1. Read in and display the active site dot file as we did for the CLT dot file in Section 4.7.3 above (steps 2 - 6), but select the ***thermcav.dot*** file from the distribution data directory. Name the new group ***CAV***.

   When color-coding the surface, be sure to pick **current** from the Properties menu before picking **execute**. If you were to accidentally pick **electrostatic** instead of **current**, the molecular surface would be recalculated to show the electrostatic potential of the active site in the *presence* of the inhibitor. (We want it to show the electrostatic potential in the *absence* of the inhibitor.) This **electrostatic** calculation would take about 10 minutes on the Silicon Graphics Iris 4D/70 workstation.

   After completing step 6 (**execute**), the water-accessible surface for the active site is shown in the graphics window and the group **CAV** is created.

2. Click the DISPLAY button again and toggle everything off except for **CAV** and **CLT**. Note how the dotted molecular surface describing the active site is primarily yellow (slightly negative electrostatic potential). There are also a few regions of strongly negative electrostatic potential (red) and one area of positive electrostatic potential (greenish-blue) in the region of the zinc ion.

   The dotted surface describing the active site for thermolysin is shown in Figure 4.10. Shown next to it in the cavity is the CLT molecule (unsurfaced). To see how well CLT fits, we'll go to step 3 and use the surfaced CLT molecule that we saved above.

3. Toggle **CLT** off. Toggle **INHIB** on (the surfaced CLT molecule). Manipulate the image on the display system to get a good view of the CLT/thermolysin active site complex. Rotate them until the cavity becomes clearly visible (see Figure 4.10). It might be helpful to toggle **INHIB** on and off a few times during this process. Note how well the CLT inhibitor fits in the cavity.

4.8 Conclusion

This completes our study of the creation and display of dotted molecular surfaces. Van der Waals surfaces have been generated and several properties calculated and displayed for the peptide analogue CLT. Dot coordinates and dot colors were saved in a dot file. A solvent-accessible surface describing the active site of an enzyme was read in and displayed.

You may wish to dock the CLT inhibitor with thermolysin. The docking functions are described under Build Operations in Section 3.10 of the *BIOGRAF Reference Manual* and are illustrated in Chapter 5 of this tutorial. Please note that although the **INHIB** dot surface group can be docked, the **modify coords** function on the Dock menu will *not* update the coordinates of the surface group; only coordinates of real atoms and bonds can be updated (*e.g.*, the **CLT** group). If you need a more detailed description of the defaults in the Calculate and Properties menus, please see surface operations in Section 4.3 of the *BIOGRAF Reference Manual.*

To leave BIOGRAF, pick **exit** on the Top menu, then verify the exit by clicking the OK button.

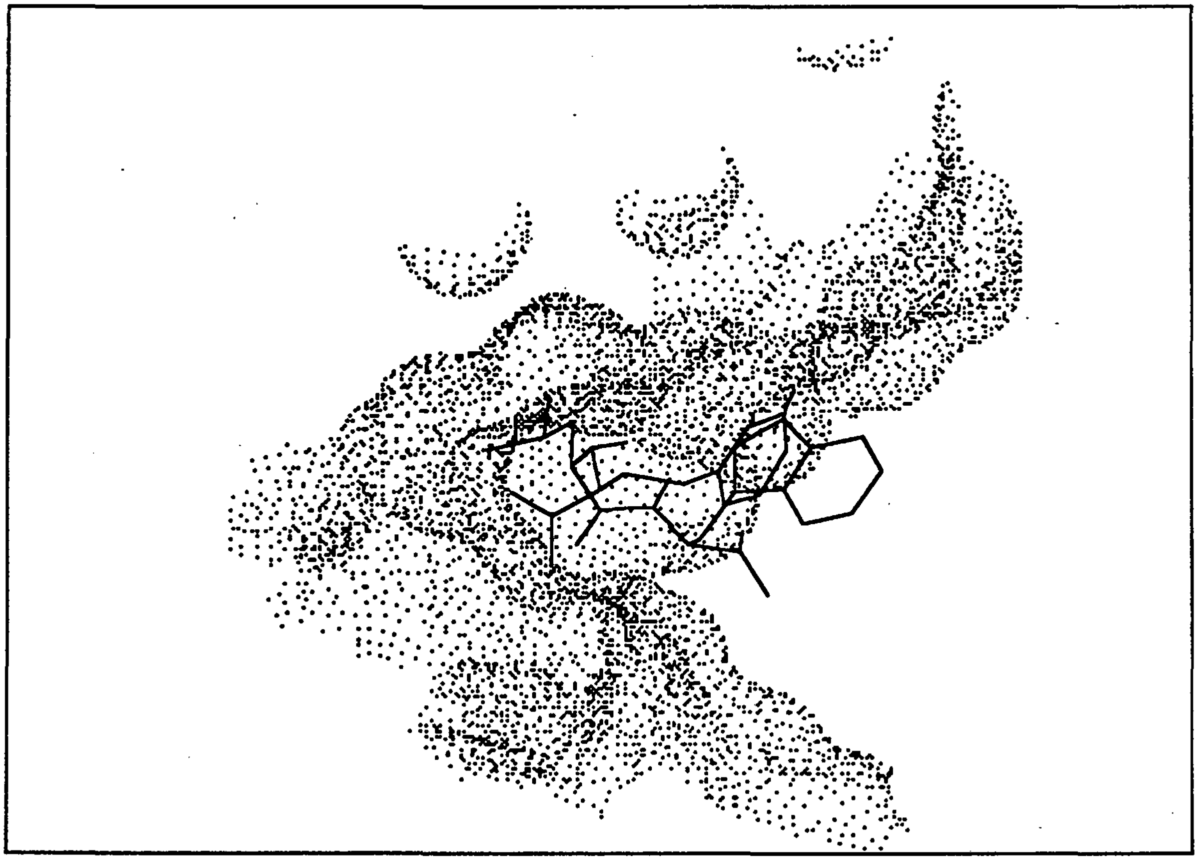


Figure 4.10: Dotted molecular surface describing the active site for the enzyme thermolysin. Shown next to it is the inhibitor CLT (unsurfaced molecule).

Chapter 5

Docking

This chapter describes how to move molecules on a display relative to each other. Such a process is called *docking*. We will read in data for the anti-viral drug netropsin and a segment of DNA, and then dock the netropsin on the DNA. It is assumed that you have completed Chapter 1.

5.1 Starting the Program and Reading the Data

1. Enter the command ***biograf***. This starts up BIOGRAF without reading in information about any molecule. As explained in the introductory chapter, this command may be different at your site. The session starts out with the Top menu shown in Figure 5.1.

   Next, we'll read in the data that describes the netropsin and DNA molecules. This information is stored in two files: a *bgf* file (*dna.bgf*) which contains coordinate information about the molecules and a group file (*dna.grp*) which contains the group information. We could read in each of these files using the ***read*** option on the In-Out menu or we could retrieve this data by executing a macro which has already been set up to do this. Since the program comes with just such a macro (*dna.macro*), we'll read the files in using the macro. Macros can only be executed in command mode.

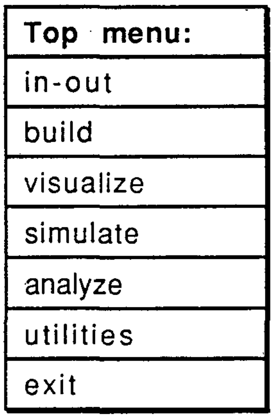


Figure 5.1:
Top menu

2. Click the TEXT button in the lower left window to bring the text window forward (see Figure 5.2). The commands executed when in command mode appear in the text window, so we want to be able to see this window clearly. If necessary enlarge or

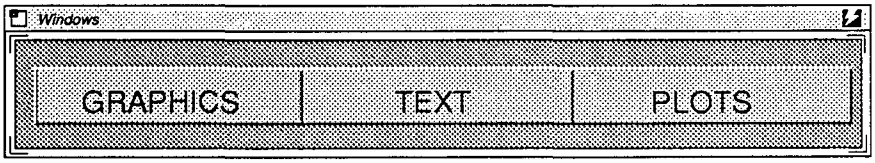


Figure 5.2: GRAPHICS, TEXT, and PLOTS Buttons

reposition the text window to make it more visible. Make sure it doesn't obscure the Menu Pad. A good place to put it is toward the bottom of the graphics window.

3. Click the COMMAND button in the Modules/Messages window to get into command mode (see Figure 5.3). The text window displays the command mode prompt with the current menu name (Top menu:) and a message appears in the Modules/Messages window reminding us that the cursor must be in the graphics window to use command mode.

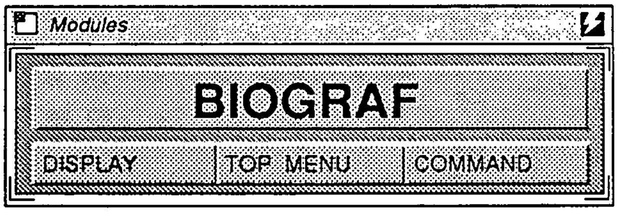


Figure 5.3: Modules/Messages Window with DISPLAY, TOP MENU and COMMAND Buttons

4. Move the cursor into the graphics window. Macros are executed by typing in the character "@" followed by the macro file name. If the macro is not located in the current working directory, then directory specifications must also be included.

   The *dna.macro* file is located in your distribution macro directory. Typically the full specifications for the distribution macro directory are */usr/msi/biogv300/macro*. The distribution macro files are always located in the */biogv300/macro* directory; however, the first part of the path specifications may differ on your system. To determine the location of the */biogv300/macro* directory on your system, type *echo $MSIUSR* from a systems window. The directory listed in response to this command should precede */biogv300/macro* when giving directory specifications.

5. Type @ followed by the macro file name including full directory specifications if needed (*e.g.*, ***@/usr/msi/biogv300/macro/dna***). Do *not* include the *.macro* extension at the end of the file name. The program quickly executes the series of commands in the *dna.macro* file. The coordinate and group files for DNA/netropsin are read in and the molecules appear in the graphics window. The last command in the macro file ("menu") returns us to graphics mode.

6. Bring the graphics window forward to get a good view of the structures. This can be done by clicking in the graphics window border or by clicking the GRAPHICS button.

The four groups which are read in are **CHAIN A**, **CHAIN B**, **NETROP**, and **WATER**. The **NETROP** group is the netropsin molecule, which nestles in a groove of the DNA double helix. The **WATER** group is toggled off, so the water molecules do not appear in the graphics window. The structures are oriented so that we are looking down through the top of the DNA double helix.

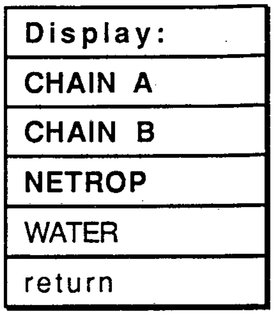


Figure 5.4:
Display menu

6. Click the DISPLAY button then toggle on the **WATER** group in the Display menu (see Figure 5.4). The water molecules appear as pyramids scattered around the DNA and netropsin molecules.

5.2 Changing the Group Color

In order to see things more clearly, we'll first turn off the **WATER** group, then we'll change the color of the **NETROP** group to purple and reorient the display.

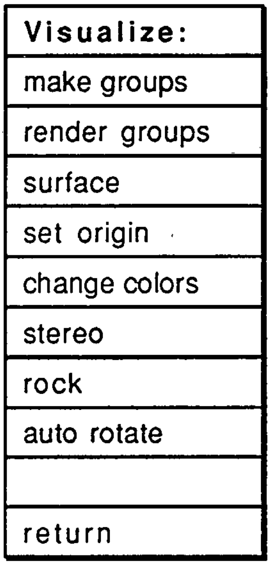


Figure 5.5:
Visualize menu

1. Toggle **WATER** off. The pyramids disappear from the display. Pick **return.** The Display menu is replaced with the Top menu.

2. Pick **visualize** from the Top menu. The Visualize menu shown in Figure 5.5 is displayed.

3. Pick **change colors** on the Visualize menu. The Groups menu appears, and BIOGRAF prompts: *Pick the desired group.*

4. Pick **NETROP** from the Groups menu. The Colors menu is then displayed listing the colors as shown in Figure 5.6 and we are prompted to *Pick the desired group color.*

5. Pick **purple.** The netropsin molecule turns purple on the display. If you wish, continue picking other colors on the menu to see how they look.

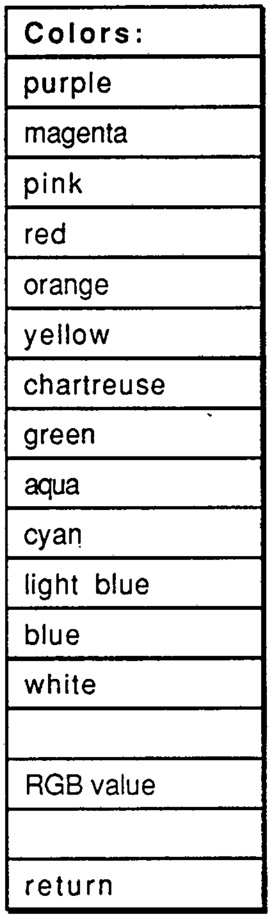


Figure 5.6:
Colors menu

6. Pick **return** when satisfied with the color.

7. Rotate the DNA around the y axis until it is lengthwise and horizontal in the graphics window.

8. Rotate the DNA around the z axis until it is vertical in the graphics window, then rotate around the y axis until the netropsin molecule is on its right side.

9. Center the image. Make sure that neither the DNA nor the netropsin molecule is clipped.

5.3 Creating Groups

BIOGRAF expects all data to be specified in terms of groups. A group is a collection of atoms and bonds chosen in such a manner as to reduce a molecular system into recognizable parts. Groups may be defined in either of two ways:

- with standard structural elements or *primitives* (*e.g.*, an amino acid residue), or
- with standard structural descriptors or *selectors* (*e.g.*, hydrophilic side chains).

The Standard Group Element Appendix of the *Biograf Reference Manual* lists the primitives and selectors that may be used to describe biological macromolecules.

Groups are created using the operations on the Make groups menu.

1. Pick **make groups** on the Visualize menu to display the Make groups menu shown in Figure 5.5.

The top half of the menu contains the proper sequence of steps:

setup ⇨ **files** ⇨ **chains** ⇨ **residues** ⇨ **limits** ⇨ **create**

The primitives and selectors used to define the groups are chosen under **limits**. The sequence given above *must* be performed in the prescribed order. However, it may not be necessary to repeat every step each time you define a group because **files**, **chains**, and **residues** always correspond to their last setting.

You can make changes in the middle of the sequence by redoing the appropriate step and then repeating the subsequent steps in the proper order. The defaults for **files**, **chains**, and **residues** are often appropriate, so that constructing groups involves only the steps

setup ⇨ **limits** ⇨ **create**

Primitives and **selectors** also retain their previous settings, but the residues picked or keyed in under **limits** are lost each time you toggle **setup**, so that the limits include the entire molecule.

To create a group containing all the atoms of any one coordinate file, toggle the **full file** switch on. If the **full file** switch is on, the group created will contain all the atoms in the file. You must **setup** the group and select the file whose atoms will be used, but the **chains**, **residues** and **limits** toggles are not active. In order to create a group containing only part of a file, the **full file** switch must be toggled off. **Full file** is toggled off when the program starts.

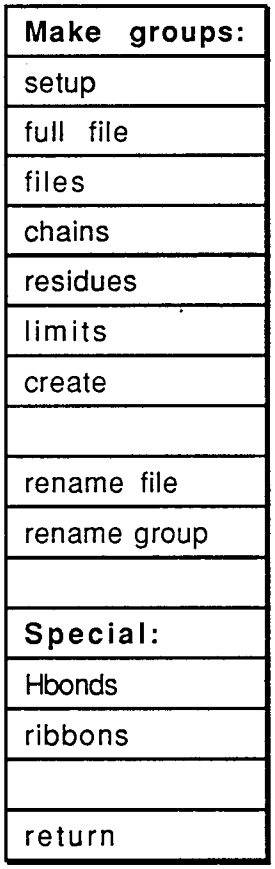


Figure 5.5: MAKE GROUPS menu

5.3.1 Making the AT Base Pairs Group

The DNA/netropsin system is broken up into the groups **CHAIN A**, **CHAIN B**, **NETROP**, and **WATER**. To see how the netropsin molecule preferentially binds to sequences of four or more adenosine-thymine (AT) base pairs, we will make a new group that contains all of the AT base pairs. Then we will save the original conformation of the DNA/netropsin system by creating another group. This group will allow us to compare the position of the netropsin molecule after docking with its original position prior to docking.

1. Pick **setup**. The Groups menu appears and we are prompted *Pick the group to be defined.*
2. Pick **add group**. A dialog box comes up and we are asked to *Enter a name for the group to be defined.*
3. Type ***AT*** as the group name in the dialog box. To verify the entry click the OK button or simply press <Enter>.
4. Pick **files** from the Make groups menu. The Files menu appears with the file descriptor **DNA** highlighted. This is the file we want, so we'll leave it toggled on. Verify the file selection by picking **return**.
5. Pick **chains** from the Make groups menu. A new menu appears listing the chain designators currently in the system as shown here in Figure 5.8.

   Four chain designators are listed: **?**, which indicates that the program will accept any chain designator, **A** and **B** for the A and B chains of the DNA helix, and _ (underscore), a blank chain designator for water and netropsin. The first time the Chains menu is entered, **?** is toggled on. If it is not on, toggle it on. Pick **return** to verify the selection.

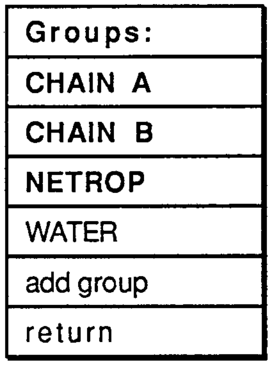


Figure 5.6: Groups menu

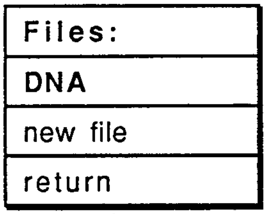


Figure 5.7: Files menu

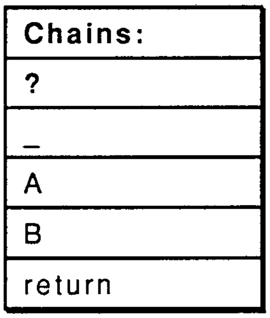


Figure 5.8: Chains menu

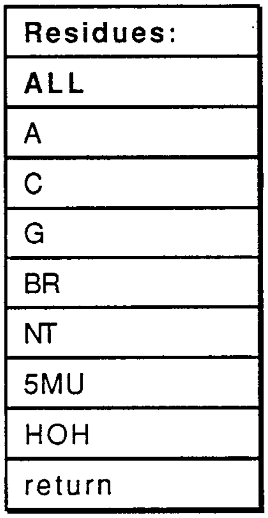


Figure 5.9: Residues menu

6. Pick **residues** from the Make groups menu. The Residues menu appears as shown in Figure 5.9. **ALL** should be toggled on, indicating that any residue will be accepted.
7. Pick **A** and **5MU** from the list of residues; complete the selection by picking **return**. Note that **5MU** stands for 5-methyluridine, which is the same as thymidine. Brookhaven Protein Data Bank coordinate files generally use **5MU** instead of **T**.

Defining the Primitives

We next consider the primitives that describe the residues. Figure 5.11 shows the Primitives menu of structural elements. The primitives that have names starting with "H" (*e.g,.* HMAIN) refer to the hydrogen atoms and bonds attached to the corresponding primitive (*e.g.,* MAIN). Very few X-ray crystallographic structures include data for H atom positions (H atoms are poor diffractors of X-rays), so BIOGRAF often needs to generate the positions. In general, it is a good idea to include the H primitives when setting up a group to make it easier to add H atoms later.

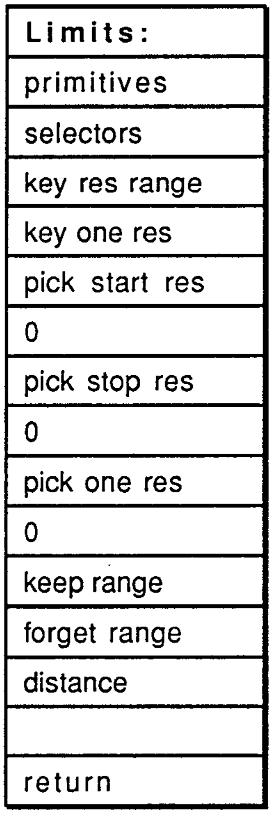


Figure 5.10:
Limits menu

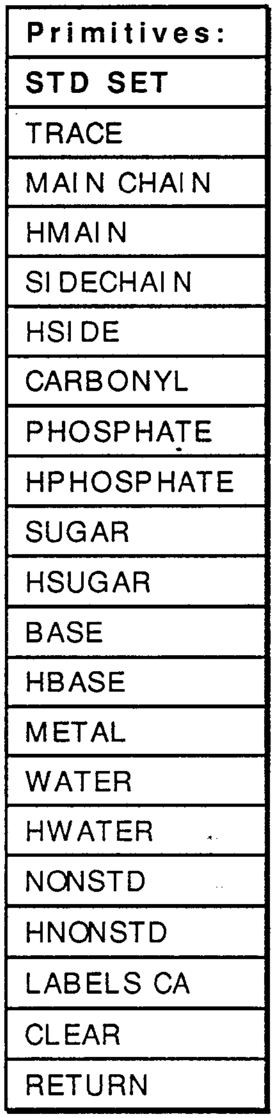


Figure 5.11:
Primitives menu

1. Pick **limits** on the Make groups menu to display the Limits menu shown in Figure 5.10.

2. Pick **primitives** on the Limits menu. The Primitives menu is then displayed listing the primitive structural elements as shown in Figure 5.11. **STD SET** should be toggled on, indicating that the standard set of primitive elements (all the elements listed except **TRACE** and **LABELS CA**) are included in the group.

3. Toggle on the primitives **PHOSPHATE, HPHOSPHATE, SUGAR, HSUGAR, BASE** and **HBASE** followed by **RETURN.**

4. Pick **return** on the Limits menu to return to the Make groups menu, then pick **create**. The message *Working...*is displayed in the Modules/Messages window. When the program finishes, the **AT** group appears in the graphics window.

5. Click the DISPLAY button, then toggle off the groups **CHAIN A** and **CHAIN B** and note how the netropsin recognizes the **AT** rich region of the DNA molecule.

6. Toggle off the **AT** group and toggle **CHAIN A** and **CHAIN B** back on. Complete by picking **return.**

5.3.2 Making a Group to Hold the Original Conformation

The structure of netropsin displayed in the graphics window derives from the X-ray data and analysis by Dickerson, *et al.* We will create another netropsin group called **XRAY** to save the original conformation of the DNA/netropsin system so that after we move the netropsin around we can compare its new position with that of the original conformation.

1. Pick **setup** on the Make groups menu, add a new group and define it as ***XRAY***.
2. Verify that files and chains remain unchanged from when you created the **AT** group.
3. Pick **residues**.
4. Pick **ALL** from the Residues menu, followed by **return**.
5. Pick **limits**, then **primitives**.
6. Pick **CLEAR** to turn the primitives off, then pick **NONSTD** and **HNONSTD**, followed by **RETURN**.
7. Pick **return** on the Limits menu to get back to the Make groups menu and **create** the group **XRAY**. The **XRAY** group appears in the graphics window in the same position as the **NETROP** group. The color of the last drawn group (**XRAY**) is displayed. It will be half-bonded.
8. Click the DISPLAY button and toggle **XRAY** off. The color changes back to purple. Pick **return** to remove the Display menu.
9. Click the TOP MENU button in the Modules/Messages window to get back to the top level menu.

5.4 Docking

At the moment, the mouse and dials connect to every group in the graphics window; when you use the mouse or dials to manipulate the objects in the graphics window, all of the groups move. The **dock** operation on the Build menu makes it possible to connect the mouse or dials to a selected subset of groups and move that subset while the rest of the display remains fixed.

5.4.1 Selecting the Groups to be Docked

In order for BIOGRAF to perform a docking operation, it must know which groups to move and which to keep fixed.

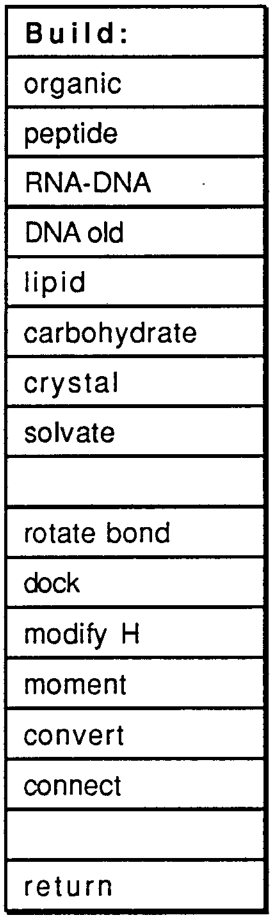


Figure 5.12: Build menu

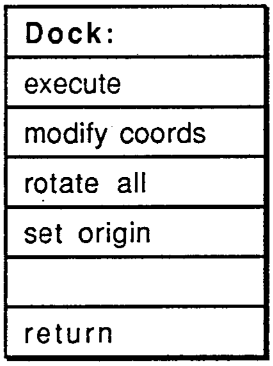


Figure 5.13: Dock menu

1. Pick **build** on the Top menu to display the Build menu shown in Figure 5.12.
2. Pick **dock** to bring up the Dock menu shown in Figure 5.13.
3. Pick **execute**. BIOGRAF prompts *Pick the groups containing atoms to be docked:* and displays the Groups menu.
4. Pick **NETROP** on the Groups menu. Terminate the list of groups to be docked by picking **return**.
5. BIOGRAF then prompts *Pick additional groups containing atoms to be used in analysis (if desired).* If any groups are selected, the program goes into interactive docking mode which allows the user to monitor interatomic distances, force vectors, and total energy as one of the structures is moved. We won't be doing interactive docking, however, so just pick **return**.
6. Translate the image along the x axis and note how only the netropsin group moves.
7. Use the other rotate and translate functions to manipulate the netropsin.

5.4.2 Rotating the Image

At the moment, all of the groups in the graphics window rotate about the center of the DNA/netropsin system. We intend to move only the netropsin, so it is convenient to change the center of rotation to the **NETROP** group.

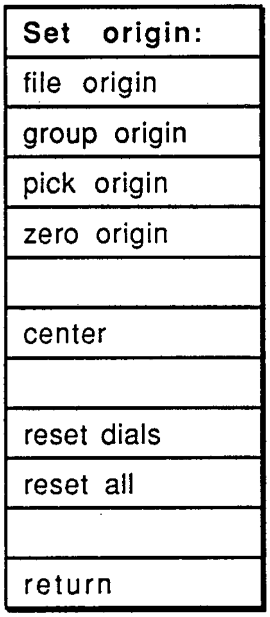


Figure 5.14:
Set Origin menu

1. Pick **set origin** to display the Set Origin menu shown in Figure 5.14.

2. Pick **group origin**. BIOGRAF displays the Groups menu and prompts *Pick the groups to center rotation on:*

3. Pick **NETROP** on the Groups menu to change the origin of rotation to the netropsin molecule. Complete the selection by picking **return**.

4. Rotate the netropsin molecule. Note how the change in rotation origin makes it easier to control the **NETROP** group.

   The **file origin, pick origin** and **zero origin** operations work in a similar manner to the **group origin** operation. Section 3.10 of the *BIOGRAF Reference Manual* describes how the program handles rotations and translations in more detail.

5. Pick **reset dials**. The reset dials operation resets the dials or mouse to zero rotation, zero translation, and zero zoom, and restores the docked group to the orientation it had before being last connected, *i.e.,* the last time **rotate all** or **execute** (on the Dock menu) was toggled.

6. If resetting the dials (or mouse) caused the image to shrink, zoom in until the image fills the display.

The **rotate all** operation allows you to temporarily reconnect all of the groups to the dials (or mouse), making it possible to change the orientation of the entire system during a docking operation.

1. **Return** to the Dock menu, then toggle **rotate all** on.

2. Turn the dials or use the mouse and note that all of the groups now move.

3. Restore the structure to the vertical orientation with the netropsin on the right side of the DNA molecule.

4. To resume docking, toggle **rotate all** off.

5.4.3 Docking the Netropsin

All systems except those using GKS graphics have several features which enhance the graphics window display. To invoke the features, you use function keys F1 - F5 located at the top or left side of the keyboard.

1. Press the F1 key to enlarge the graphics window to fill the entire display.
2. Press the F2 key to display the structure as two images for split screen stereo viewing. Cross your eyes to see the image in three dimensions.
3. Press the F3 key to rotate the image on the right by 90° about the x axis.
4. Translate the netropsin along the x axis and note how the two orthogonal views make it easier to see how the netropsin moves relative to the DNA.
5. To view the image in three dimensions, either
    - press the F3 key again to return the rotated image on the right back to its original orientation. The two images appear again as a stereo pair for split screen stereo viewing; cross your eyes to see the three-dimensional image.
    - if you have an IRIS equipped with special glasses (Personal Stereo Viewer), press the F5 key to enable the glasses. Consult the appropriate SGI manual for details on how to use the glasses.
    - if you have a Titan equipped with a special stereo monitor, make sure the alias command to execute the stereo version of the program has been set up and the procedure to get into stereo mode has been followed. Use the F5 key to activate stereo mode.
6. Try to dock the netropsin in its original position using the function keys and the **rotate all** operation as aids.
7. Press the F2 and F1 keys again to restore the molecules to a single image and to return the graphics window to its normal size.

5.4.4 Storing the New Coordinates

After docking a molecule or group, BIOGRAF can update the molecular coordinates it stores in memory.

1. Pick **modify coords** from the Dock menu. BIOGRAF prints the message *Working...* while it saves the new coordinates in memory.

   At the moment, the new coordinates are saved only in BIOGRAF's memory. If you exit the program at this point, you will lose the modifications. To save the new coordinates permanently, you must write them out to disk using the **write** operation under **in-out**.

We can compare the new orientation of **NETROP** in the DNA molecule to its old orientation.

2. Pick **set origin**, then pick **file origin** on the Set Origin menu to restore the origin of rotation to the center of the combined DNA/netropsin structure.
3. Click the DISPLAY button, then toggle on the **XRAY** group to see how close you came to docking the netropsin in its original position. Toggle **XRAY** off again. Pick **return** when finished.
4. Click the TOP MENU button to return to the top level menu.

Now that the netropsin is docked, we will look for a favorable conformation of the DNA/netropsin structure.

5.5 Calculating Energies

The best way to regularize the structure of a molecule is to do an energy calculation which takes all forces on the system into account. In this section we minimize the energy of the DNA/netropsin system to find the conformation that corresponds to the nearest energy minimum of the potential energy surface.

5.5.1 Setting up an Energy Expression

The first step is to set up an energy expression for the problem.

1. Pick **simulate** on the Top menu to display the Simulate menu as shown in Figure 5.15.

   Any atom can be included or excluded from a molecular mechanics calculation, and if included it can be movable or fixed. In setting up the energy expression, we must define those groups which are allowed to move and those which will remain fixed. In order to keep the run short enough for an interactive tutorial, we'll minimize the energy of the netropsin group while keeping the DNA helix fixed.

2. Pick **setup eex** to setup the energy expression. BIOGRAF displays the Groups menu and prompts *Pick the groups containing movable atoms:*

3. Pick **NETROP** to be movable, then pick **return**. BIOGRAF responds with the prompt *Pick the groups containing fixed atoms.*

4. Pick **CHAIN A** and **CHAIN B** to be fixed, followed by **return**. We are now asked to *Pick the groups to be updated during calculation.*

5. Pick **NETROP** to be updated, then pick **return**. BIOGRAF prints the message *Setting up energy expression* then lists a summary of the energy expression statistics in the text window (see Table 5.2).

6. Pick **show eex**. The movable atoms are indicated by small green crosses and the fixed atoms are marked with small blue crosses. Click anywhere to remove the crosses.

7. Click the DISPLAY button, toggle off the **NETROP**, **CHAIN A**, **CHAIN B** and **XRAY** groups, then pick **show eex** again. You should be able to see the crosses better. Make certain that the only green crosses in the molecule are those on **NETROP**.

8. Click anywhere to remove the crosses from the display.

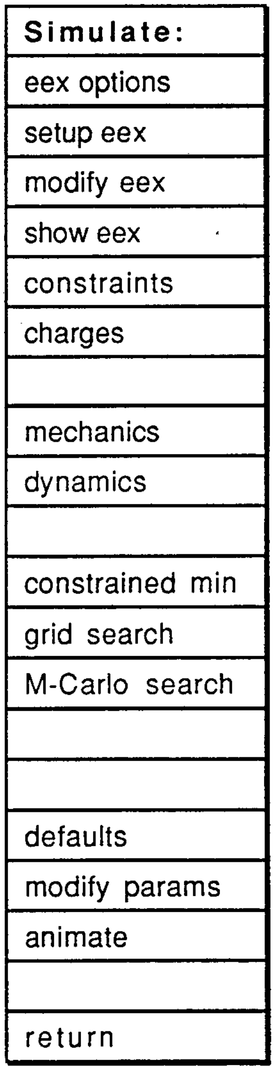


Figure 5.15: Simulate menu

9. Click the DISPLAY button and toggle on **NETROP**, **CHAIN** A and **CHAIN B** again, followed by **return**.

```
Movable atoms      :    31        Fixed atoms          :488
Bonds              :    32        Angles               : 44
Torsions           :    51        Scaled torsions      : 12
Pi Twist           :     0
Inversions         :    33        Angle-angle (1C)     :  0
Bond cross angle   :     0
Nbond exclusions   :    76
Hbond acceptors    :   245        Hbond donors         :  0
User energy terms  :     0
Total charge       :   -23.734    movable atoms only   :  0.000
```

Table 5.2: Energy expression statistics (in text window)

5.5.2 Minimizing the Energy

To do energy minimizations, we use the operations on the Mechanics menu.

1. Pick **Mechanics** to display the Mechanics menu shown in Figure 5.16.

   Figure 5.17 shows the Energy results menu that BIOGRAF displays whenever it goes into the **mechanics** function. A description of the items on this menu is given in Table 5.2. No values are given for the energies at this point because we haven't done any energy calculations yet.

2. Pick **one energy** on the Mechanics menu to calculate the energy of the DNA/netropsin system in its present conformation.

   After BIOGRAF computes the energy, it displays the results in the text window and on the Energy results menu. All energies are in kcal/mol. If an energy is too large to fit into the menu box, the program prints a row of asterisks for it: *********.

3. Make a list of the initial energies (or asterisks) on the Energy results menu.

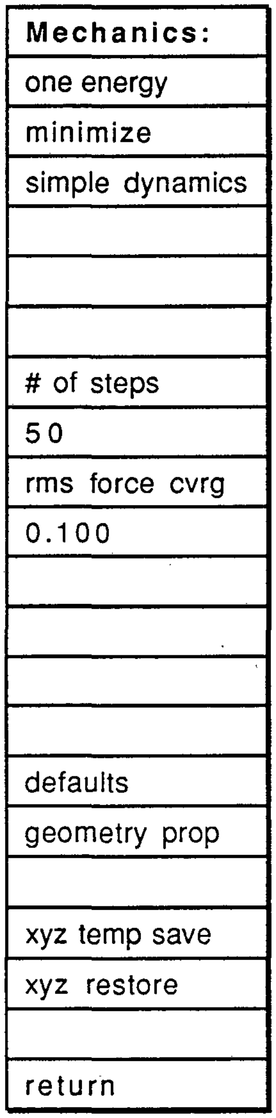


Figure 5.16: Mechanics menu

We are now ready to optimize the DNA/netropsin system. The energies at each step will appear in both the Energy results menu and the text window, and the conformation of the system for the corresponding step will appear in the graphics window.

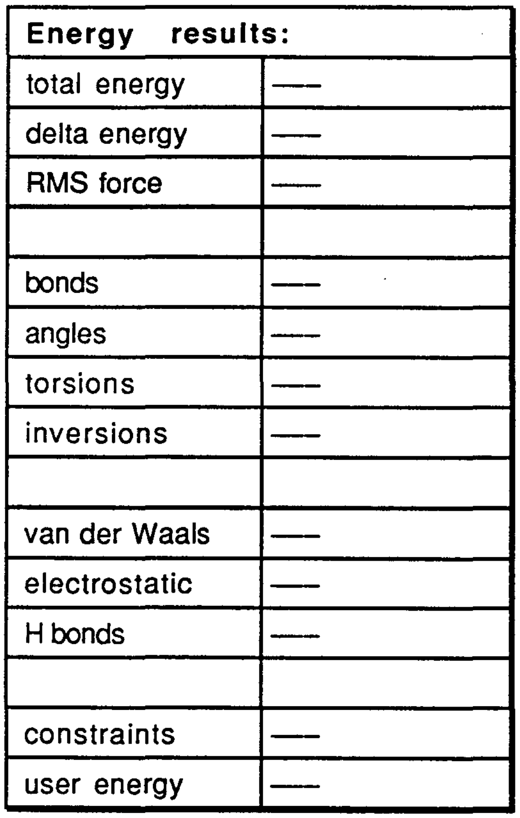
1. Center and zoom in on the **NETROP** group until you have a good view of it.

2. Pick **minimize** on the Mechanics menu. BIOGRAF prints the message *Working...* in the Modules/Messages window. The energies should show a rapid decrease early in the calculation.

3. When the minimization finishes, check **delta energy**. If it is non-negligible compared to the other energies, then the calculation has not yet converged.

4. If **delta energy** is still comparatively large, toggle **# of steps**. BIOGRAF then prompts you for the number of steps.

5. Type in the number of steps you think are needed to complete the minimization. Select the OK button to complete the entry.



Energy results:	
total energy	—
delta energy	—
RMS force	—
bonds	—
angles	—
torsions	—
inversions	—
van der Waals	—
electrostatic	—
H bonds	—
constraints	—
user energy	—

Figure 5.17: Energy results menu

Energy	Description
total energy	Current total energy
delta energy	Change in total energy from previous step
RMS force	Root mean square force (derivative of potential energy)
bonds	Bond stretching
angles	Bond angle bending
torsions	Dihedral angle torsions
inversions	Inversions
van der Waals	van der Waals forces
electrostatic	Electrostatic forces
H bonds	Hydrogen bonds
constraints	Constraints
user	User energy

Table 5.2: Description of Energy results menu items

6. Repeat steps 2-5 above as many times as necessary until you are satisfied that the calculation has converged. Pressing the abort key (F12, F10 on the Titan) aborts any mechanics calculation after the next complete step.

7. Compare the minimized results to the energies of the original docked system. The final energies should be on the order of 10 kcal or less.

8. Click the DISPLAY button, then toggle on **XRAY**. Compare the optimized position of the netropsin molecule to its position prior to being docked. Pick **return** when finished.

You now have a docked netropsin molecule nestled in a favorable conformation within the DNA structure. However, large flexible systems have many favorable conformations, each of which arises from a *local minimum* in the potential energy surface. Energy minimization yields only the nearest local minimum, so it is likely that the conformations of the **XRAY** group and the **NETROP** group differ even though both are optimized structures.

Chapter 6 describes how to use molecular dynamics to find lower energy conformations for a system and Chapter 7 describes how to do a conformational search. You can also redock the netropsin to search for other favorable binding sites.

5.6 Saving Information

After you find a good conformation for a system, it is a good idea to save it. This section describes how to save information in *group* files and *coordinate* files.

5.6.1 Group Files

A group is a collection of atoms and bonds chosen in such a manner as to reduce a molecular structure into recognizable parts. There are two types of files that save group information:

Type 1	saves all the sequences of menu picks that defined the various groups. When BIOGRAF reads a type 1 group file, it regenerates the groups based on the saved sequences of menu picks. This allows the same group file to work for different structures.
Type 2	saves the list of atoms and bonds for each group as well as the sequence of menu picks used to define them. A type 2 file regenerates groups faster than type 1 because it reconstructs them directly from the atom lists rather than from the menu picks.

Both types of group file contain the display status information that determines the color of the group and whether or not the group is toggled on or off.

Note: Type 2 files should only be used for coördinate files with the same set of atoms and connections. When in doubt, make a type 1 file.

To save group information about the molecule, proceed as follows.

1. Click the TOP MENU button to return to the Top menu, then pick **in-out** to bring up the In-Out menu (see Figure 5.18).

2. Pick **write**. The prompt *Pick the file type to be output:* is displayed and the File types menu appears in the Menu Pad.

3. Pick **group** from the File types menu to let BIOGRAF know that the type of file you want to write contains group information. BIOGRAF brings up a dialog box and asks you to *Enter a name for the groupings file.*

4. Type ***dock.grp*** (include disk and directory information if not saving the file in your working directory), then select OK to verify the entry. The *grp* extension indicates that the file is a group file. BIOGRAF prompts *Enter output file type [1 (normal) or 2].*

5. Enter the number ***1*** to indicate type 1 group information, then select OK. BIOGRAF then prompts you to enter remarks for the output file.

6. Type in remarks that will identify the file the next time you use it. BIOGRAF will print the remarks in the text window when it reads the file. Terminate the remarks with a blank line.

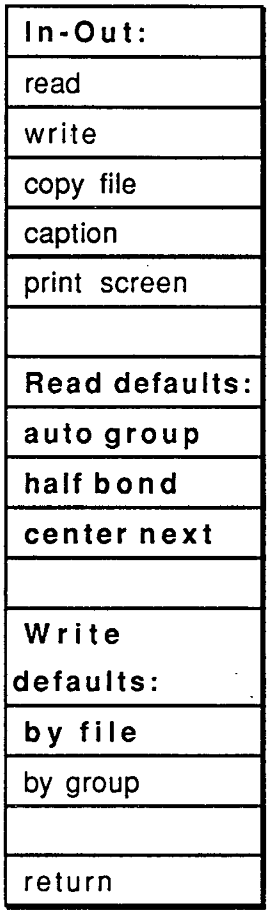


Figure 5.18: In-Out menu

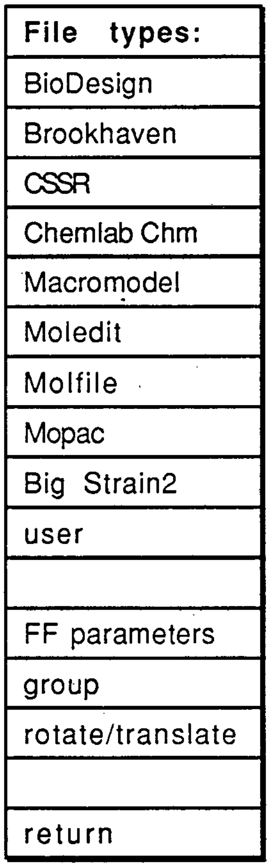


Figure 5.19: File types menu

5.6.2 Coordinate Files

In order to reconstruct a group, BIOGRAF needs the information about the coordinates of the atoms in that group. The program saves coordinate information in coordinate files. Here we make a coordinate file of the BioDesign General Format type.

1. Make sure that the **by file** toggle is highlighted under the write defaults. The source of the coordinate information that we will be saving is in the DNA file that we read in.

2. Pick **write**. The prompt *Pick the file type to be output:* appears and the File types menu shown in Figure 5.19 again displays the file types. Ten different coordinate file formats are available.

3. Pick **BioDesign** to indicate the BioDesign General Format file type. This type is recommended for internal use with BIOGRAF since it contains all pertinent information.

Since there is only one file in the system (DNA), it is automatically chosen as the source of coordinate information. BIOGRAF now asks you to *Enter a name for the output file.*

4. Type ***dock.bgf*** in the dialog box (include disk and directory information if needed), then verify the entry by selecting OK. The *bgf* extension indicates that the file contains BioDesign General Format type data. BIOGRAF prompts you to enter remarks for the output file.
5. Type in remarks that will identify the file the next time you use it. Terminate the remarks with a blank line (Click OK or press <Enter>).

5.6.3 Testing Our Work

As an exercise, we test the new group and coordinate files by resetting the program and having BIOGRAF read them in again.

1. **Return** to the Top menu and pick **utilities**. The Utilities menu shown in Figure 5.20 is displayed.
2. Pick **program reset**. An alert box appears asking you to verify the reset.
3. Click the OK button. When BIOGRAF finishes the program reset it no longer remembers anything about the molecule. The Top menu is again displayed.
4. Pick **in-out** to bring up the In-Out menu.
5. Leave the **auto group** and **half bond** options under the read defaults toggled on. The **auto group** option treats the molecule as a single group and includes every atom and bond in the file. The **half bond** option uses the half-bond colors to display the molecule. Leave the centering option with **center next** displayed. This will cause the next file that we read in to be centered in the graphics window.

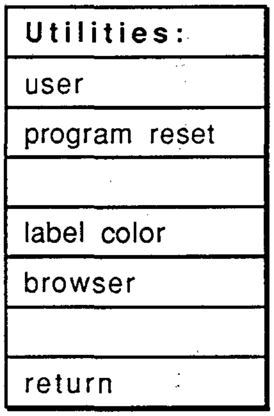


Figure 5.20: Utilities menu

6. Pick **read**. The File types menu appears and we are prompted to *Pick the file type to be input.*
7. Pick **BioDesign**, since we saved our data in a *bgf* file. BIOGRAF brings up the file name entry dialog box and prompts *Choose the file to be input.*
8. Use the browser box to bring up the directory that you saved the files in, then select ***dock.bgf*** from the file names listed. Click the OK button to verify the selection. The message *Working* ... is displayed as the coordinate file that we created in the previous section is read in. When finished, the DNA and netropsin structures appear in the graphics window.

Next we will read in the group file that we created to test whether it correctly saved the groups.

9. Pick **read** again from the In-Out menu and then pick **group** from the File types menu. BIOGRAF prompts *Enter the name of the groupings file.*

10. Select ***dock.grp*** from the files listed in the browser box, then click the OK button. The groups that had been toggled on will appear in the graphics window as the group file is read in.

5.7 Advanced Notes

BIOGRAF recognizes the standard files from the Brookhaven Protein Data Bank and the Cambridge crystallographic data base, as well as CHEMLAB CHM, CHEMLAB MOL, Macromodel, Moledit, and MOPAC files. The program also has its own internal file format which is similar to Brookhaven but contains additional force field information. (See the File Formats Appendix of the *BIOGRAF Reference Manual* for further details.)

This concludes Chapter 5. To leave BIOGRAF, **return** to the Top menu, pick **exit**, then verify the exit by clicking the OK button in the alert box.

Chapter 6

Molecular Dynamics of Small Peptides

In this chapter we will build a small peptide (leuenkephalin) and minimize it to yield a favorable conformation. However, flexible molecules such as peptides have many favorable conformations, each of which arises from a *local minimum* in the potential energy surface.

One way to look for favorable geometries is to do a conformational search as discussed in Chapter 7. However, a full search of the available conformations takes a prohibitive amount of time. Consider a small protein with 15 freely rotating bonds. Each bond rotates 360°, so doing a calculation for every 10° of rotation requires 36 calculations per bond. This gives a total of 36^{15} possible conformations. If it takes one second to find the energy of a conformation, then it takes 36^{15} seconds, or 7000 trillion years, to do the calculation.

The equations of molecular dynamics solve Newton's equations of motion for the atoms in a molecular structure. Given any starting geometry of the system, BIOGRAF can evaluate the forces acting on every atom to determine how the velocity and position of the atoms change with time. This provides an efficient means to search for low energy conformations of a system without going through every one of its conformations.

This chapter describes how to do a molecular dynamics calculation on leuenkephalin. The results of the dynamics run are then viewed by animating the trajectory file generated. The best (lowest energy) conformer is saved in a separate file. A macro to perform a longer dynamics run is also created and executed in the background.

6.1 Building the Peptide

To run a dynamics calculation, we need a molecule. We will build the peptide leuenkephalin, which is a chain of five amino acids with the sequence tyr-gly-gly-phe-leu.

1. Enter the command ***biograf*** (this command may be different at your site). This starts up BIOGRAF without reading in the information for any molecule. The Top menu shown in Figure 6.1 appears in the Menu Pad.

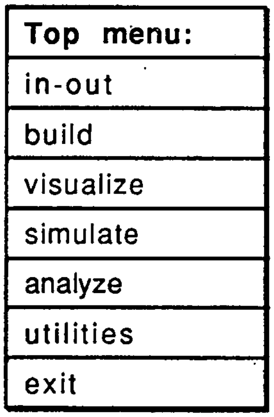


Figure 6.1: Top menu

2. Pick **build** on the Top menu to display the Build menu shown in Figure 6.2.

   BIOGRAF supports several builders: Organic, Peptide, RNA-DNA, DNA-old, Lipid, Carbohydrate, Crystal and Solvate. The purpose of having different builders is to take advantage of the fact that specialized molecules such as peptides, DNA, and lipids are put together in a predefined way, *e.g.*, peptides are single chains of amino acids. BIOGRAF provides short-cuts for making them. The Organic Builder is the most flexible of the builders and can be used to make any molecule. In this tutorial, we use the Peptide Builder.

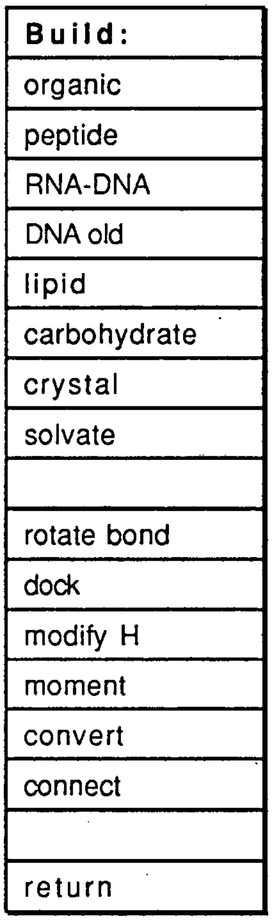


Figure 6.2: Build menu

3. Pick **peptide**. Before we begin building, we must set up an internal file which will hold information about the new molecule. The **setup** option is therefore automatically invoked, a dialog box comes up, and we are prompted to *Enter a descriptor for the new file*.

4. Enter ***pep*** as the name for the file descriptor, then select OK (click the OK button or press <Enter>). In general, a descriptor name can be anything you want up to 14 characters. For the rest of the session, whenever the Files menu comes up, the first box will read **pep**.

   The Peptide menu is displayed as shown in Figure 6.3.

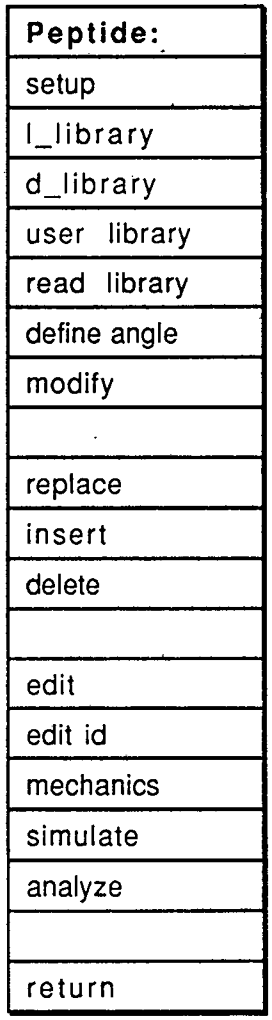


Figure 6.3: Peptide menu

Now that we have a file to contain the peptide, we can begin building. The first time you enter the Peptide Builder during a session, the program reads in its library of amino acid residues. BIOGRAF has libraries for both the L- and D-amino acids. L-amino acids are the normal form found in nature, whereas D-amino acids are very rare.

5. Pick **l_library** on the Peptide menu. The list of L-amino acids shown in Figure 6.4 appears and BIOGRAF prompts *Pick next peptide unit to be added.*

6. Pick **tyr_l**. A tyrosine residue in half-bond colors appears in the graphics window. The molecule is automatically centered on the display. We are prompted to pick another peptide unit.

7. Pick the rest of the sequence: **gly-gly-phe-leu**.

   As you pick the residues they appear in the graphics window as a chain, adding on to the carboxyl terminus of the peptide according to the default values for the ϕ, ψ, and ω angles. The default values are 180° for ϕ and ψ and 0° for ω.

8. Pick **return** to terminate the addition of residues.

   At this point any necessary hydrogens or lone pairs are automatically added and the charges on the amino and carboxyl termini are corrected.

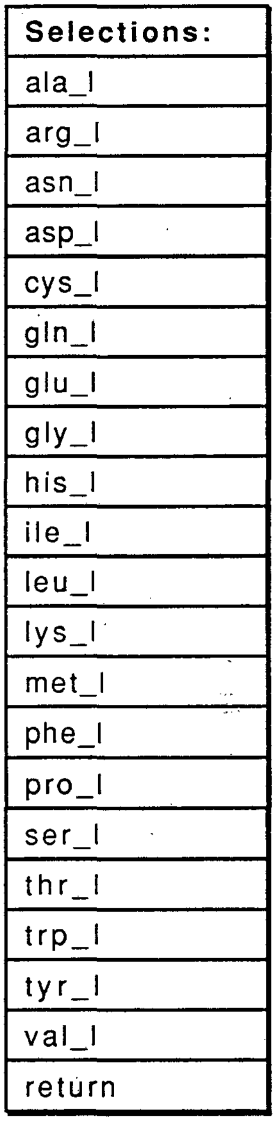


Figure 6.4:
L-Amino acid
residues menu

6.2 Minimizing the Energy

In this section we minimize the energy of the leuenkephalin molecule to find an optimal conformation in the local region of the potential energy surface. The options on the Mechanics menu allow us to do energy minimizations and/or molecular dynamics.

1. Pick **mechanics**. This brings up the Mechanics menu shown in Figure 6.5 and the Energy results menu shown on the next page (Figures 6.7). BIOGRAF prints the message *Setting up energy expression...* in the Modules/Messages window until it finishes, after which the summary of energy expression statistics shown in Table 6.1 appears in the text window.

   When **mechanics** is picked from any of the Builders the energy expression for the structure being built is automatically set up so that *every* atom in the molecule is included and allowed to move during the calculation. If you had picked **simulate** instead of **mechanics**, then you would have had to set up the energy expression yourself and specify which atoms are to be movable, which are to be fixed, and which are to be updated on the display.

   We can check to see how the energy expression is set up at any time by using the **show eex** option on the Simulate menu.

2. **Return** to the Peptide menu, then pick **simulate**. The Simulate menu comes up as shown in Figure 6.6.

3. Pick **show eex**. Movable atoms are marked with green crosses and fixed atoms are marked with blue crosses. Every atom in the peptide should be marked with a green cross. Since we have no fixed atoms, no blue crosses should be visible. Click anywhere to remove the crosses from the display.

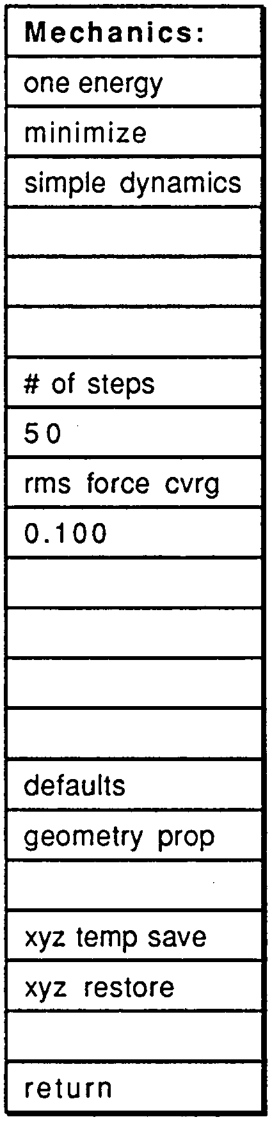


Figure 6.5: Mechanics menu

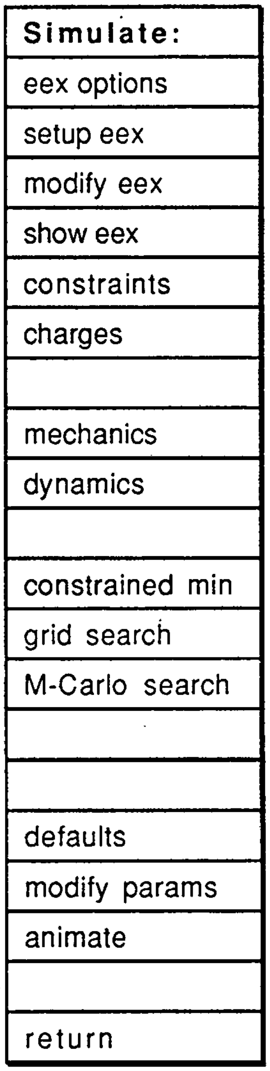
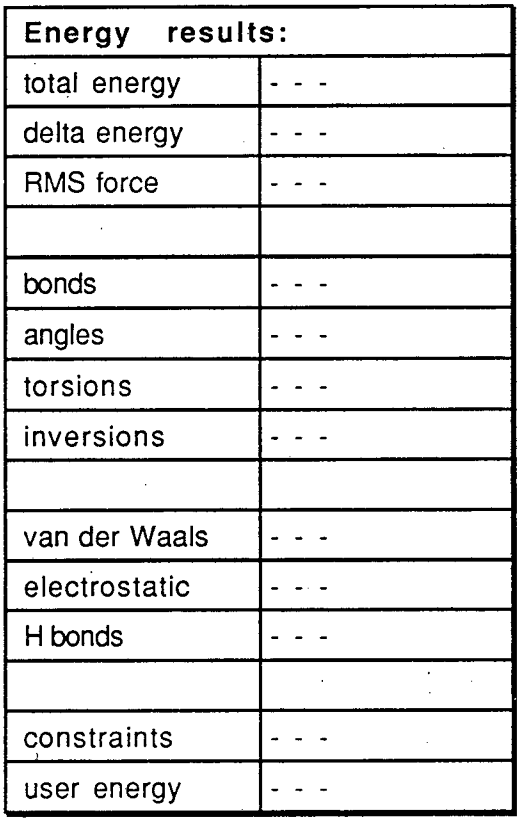


Figure 6.6: Simulate menu

```
Movable atoms       :    48      Fixed atoms         :     0
Bonds               :    49      Angles              :    69
Torsions            :    82      Scaled torsions     :     2
Pi Twist            :     0
Inversions          :    48      Angle-angle (1C)    :     0
Bond cross angle    :     0
Nbond exclusions    :   118
Hbond acceptors     :     7      Hbond donors        :     8
User energy terms   :     0
Total charge        :     0.000  movable atoms only  :     0.000
```

Table 6.1: Energy expression statistics (in text window)

4. Click the TEXT button to bring the text window forward so that you can view the energy expression statistics. Reposition or resize the windows if necessary so your views of the structure in the graphics window and the Menu Pad are not obscured.



Energy results:	
total energy	- - -
delta energy	- - -
RMS force	- - -
bonds	- - -
angles	- - -
torsions	- - -
inversions	- - -
van der Waals	- - -
electrostatic	- - -
H bonds	- - -
constraints	- - -
user energy	- - -

Figure 6.7: Energy results menu

5. Note that the energy expression statistics show 48 movable atoms (none are fixed). Pi twist, bond cross angle and scaled torsion terms are not included in the calculations (values given are always 0); they are listed here since they will be included in a future release of the program.

6. Pick **mechanics** again to bring up the Mechanics menu, then pick **one energy**. BIOGRAF now evaluates the energy of the peptide in its present conformation.

7. Make a list of the energies (or asterisks) on the Energy results menu (see Figure 6.7). All energies are in kcal/mol. If an energy is too large to fit into the menu box, BIOGRAF prints a row of asterisks: *********. The **total energy** should be about 170 kcal/mol and the van der Waals energy should be about 120 kcal/mol, indicating that the conformation is unfavorable.

8. Pick **minimize**. The message *Minimizing...* appears as the minimization proceeds.

   By default, BIOGRAF runs molecular mechanics calculations for 50 steps. During a minimization, the energies at each step appear in both the text window and the Energy results menu. The conformation of the peptide is also updated in the graphics window. The energies should show a rapid decrease early in the calculation.

9. When the calculation finishes, check **delta energy** on the Energy results menu. If it is non-negligible compared to the other energies, then the calculation has not yet converged. Convergence is reached when RMS Force is less than the value specified for **rms force cvrg** on the Mechanics menu (0.100 kcal/mol).
10. If the calculation has not converged, pick **# of steps**, then type in the number of steps you think are needed to complete the minimization. Select the OK button to complete the entry.
11. Continue to pick **minimize** until the calculation converges. Pressing the abort key (F12, F10 on the Titan) aborts any mechanics calculation after the next complete step.
12. Compare the final energies to the initial energies. After minimization, the **total energy** should be on the order of 90 kcal/mol.

6.3 Running Dynamics

The minimization yielded a favorable conformation for leuenkephalin. However, flexible molecules such as peptides have many favorable conformations, each of which arises from a *local minimum* in the potential energy surface. Molecular dynamics provides an efficient means to search for low energy geometries of a system without going through every one of its conformations.

BIOGRAF does several types of molecular dynamics. These can be divided into two groups: adiabatic (does not exchange heat), and isothermal (exchanges heat with a thermal bath). The adiabatic group includes the *adiabatic, quenched, annealed,* and *impulse* options, while isothermal dynamics is performed using the *canonical* and *temperature damping* options. These are briefly described below (see Section 5.8 and Appendix P of the *BIOGRAF Reference Manual* for more details).

Adiabatic Dynamics: The program does adiabatic dynamics where it starts with a user-supplied temperature T_i and calculates the dynamics for a total period t. The default step size equals 0.002 picoseconds, so there are 1000 steps of dynamics in a dynamics run that lasts 2 picoseconds. By default BIOGRAF checks the temperature every 0.1 picosecond. If T deviates too much from T_i, the program rescales the velocities to match T_i again.

Quenched Dynamics: The program does adiabatic dynamics as above, but it performs an energy minimization at intervals of 0.100 picoseconds and writes the minimized result to disk. This is useful for finding optimum geometries.

Annealed Dynamics: The temperature starts at T_i and increments by ΔT (*e.g.*, 10°) at intervals of 0.100 picoseconds until it reaches T_f. It increments back to T_i, then to T_f again and so on. Although annealed dynamics does not actually minimize the energy, if the low temperature is small (*e.g.* $\approx 0°$) and the number of steps large, then the system may cool slowly enough to find the best minimum in the potential energy surface without getting trapped in a high energy local minimum. Only those geometries that correspond to energies less than the previous result are written out.

Impulse Dynamics: The initial velocities of the atoms of the structure are assigned in a direction that the user has chosen, giving the structure an "impulse" in that direction. Then the program does adiabatic dynamics.

Canonical Dynamics: The program does constant temperature constant volume (TVN) dynamics via an implementation of the Nosé-Hoover[1,2] method. In order to maintain a constant thermodynamic temperature (not kinetic temperature) the velocities are scaled by an additional variable, s, which represents the interaction of the system with a heat bath. The rate of change of s is determined by the difference between the thermodynamic and the kinetic temperature at each step. The relaxation time constant and algorithm used may be specified.

---

[1] S. Nose, *J. Chem. Phys.*, **91** (1984) 511, and *Mol. Phys.*, **52** (1984) 255.

[2] W. H. Hoover, *Phys. Rev. A.*, **31** (1985) 1695.

Temerature Damping Dynamics: The program does constant temperature constant volume dynamics via weak coupling to a thermal bath using the Berendsen *et al*[3] approach. As in canonical dynamics, the strength of coupling is specified through a damping time constant.

This section describes how to do a simple adiabatic dynamics calculation on leuenkephalin. It was necessary to minimize the peptide first because if we try to do a dynamics run on a system with highly unfavorable interactions, the calculation may severely distort the molecule.

6.3.1 Setting the Defaults

The defaults for a dynamics run are as shown in Table 6.2:

T_i (initial temperature) = 300° K
T_f (final temperature) = 300° K
ΔT (anneal increment) = 10° K (annealed runs)
time = 0.500 picoseconds
ΔT (heat increment) = 20° K (straight dynamics)
time step = 0.001 picoseconds

Table 6.2: Dynamics defaults

Dynamics run faster at hotter temperatures because the atoms move faster, allowing the system to overcome conformational barriers more easily. However, if the temperature is too hot, then almost all the conformations are high in energy and the dynamics run no longer gives much useful information. For the calculations done in this chapter, reasonable temperatures lie between 300° K and 1200° K. In general, the floppier a molecule, the lower the temperature needed to run dynamics.

If the initial temperature (T_i) is not equal to the final temperature (T_f), then BIOGRAF increments the temperature by the value ΔT, where ΔT is specified by the heat increment. The temperature is incremented after each period τ where τ is specified by the heat equilibration frequency. The default for $\tau = 0.100$. The values for the initial and final temperatures and for the heat increment are specified on the Temperatur var (temperature variables) menu under **defaults**, while the heat equilibration frequency is specified on the Dynamics var (dynamics variables) menu (also under **defaults**).

---

[3] Berendsen, H. J., Postma, J. P. O., van Gunsteren, W. I., Di Niola, A. and Haak, J. R., *J. Chem. Phys.*, 81 (1984), 3684.

Changing the Temperature

We'll keep τ at its default setting, but we'll change T_i and T_f to 600° K.

1. Pick **defaults** to bring up the Defaults menu (see Figure 6.8).

2. Pick **temperatur var**. The menu shown in Figure 6.9 appears. The values of the parameters shown are the temperature defaults that BIOGRAF uses during a dynamics calculation. The temperatures are in degrees Kelvin.

3. Pick **initial temp** and enter ***600*** in the dialog box. Choose OK to complete the entry.

4. Pick **final temp** and enter ***600*** in the dialog box. Choose OK to complete the entry.

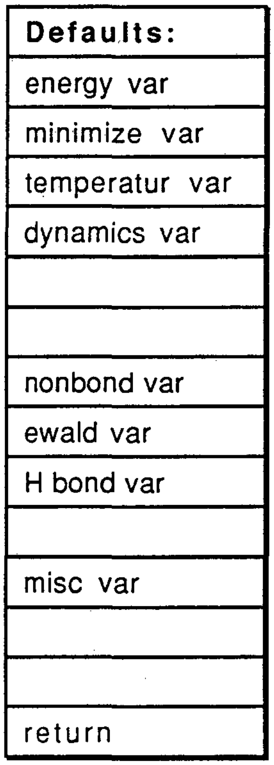
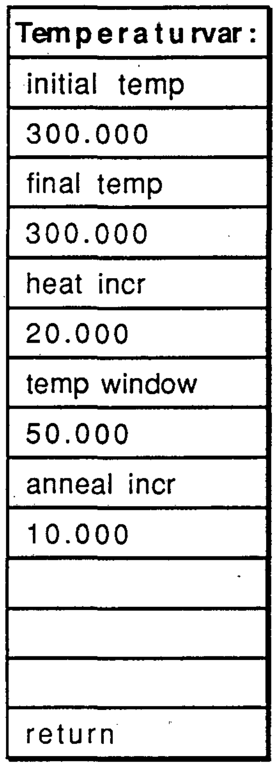


Figure 6.8: Defaults menu



initial temp
300.000
final temp
300.000
heat incr
20.000
temp window
50.000
anneal incr
10.000

Figure 6.9: Temperatur var menu

Changing the Frequency for Writing Data to the Trajectory File

We'll be saving the results of our dynamics run in a trajectory file. The rate in which the data is written to the trajectory file depends on the value specified for **write trj freq**. The default is 0.10 ps. We'll be doing a very short dynamics run (0.40 ps), so we'll change the write frequency to 0.01 ps. This will give us 40 snapshots (we'll be animating these trajectory snapshots later). The **write trj freq** option is on the Dynamics variables menu.

1. **Return** to the Defaults menu, then pick **dynamics var**. The menu shown in Figure 6.10 is displayed.

2. Pick **write trj freq**, then enter ***0.01*** in the dialog box. Select OK to complete the entry.

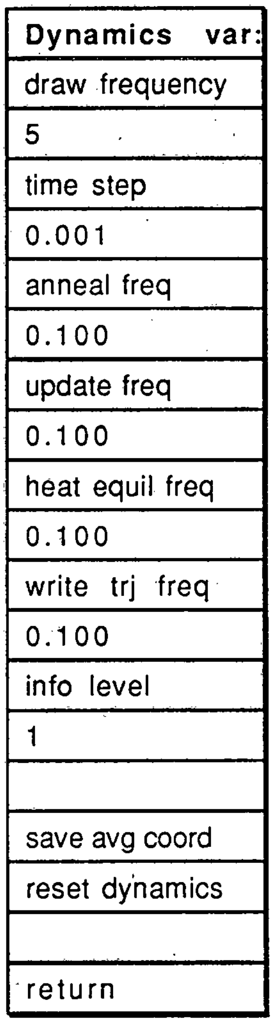


Figure 6.10: Dynamics var menu

6.3.2 Doing the Calculation

The default duration of a dynamics run is 0.500 picoseconds. Since the default for the length of each step is 0.001 picoseconds, a default run has 500 steps. We want to do a straight dynamics run of length 0.4 picoseconds (400 steps).

1. Pick **return** three times to get back to the Simulate menu, then pick **dynamics**. The menu shown in Figure 6.11 comes up.
2. Pick **adiabatic**. The Adiabatic dynamics menu appears as shown in Figure 6.12. EVN in the menu title indicates that constant energy, constant volume dynamics will be performed.
3. Pick **time**, enter ***0.4*** in the dialog box, then select OK.
4. Toggle **write traject** on. This tells the program to write the results of the dynamics run to a trajectory file.

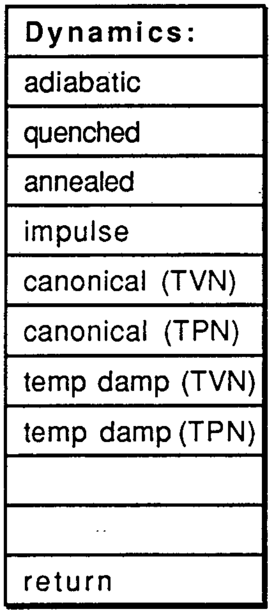


Figure 6.11: Dynamics menu

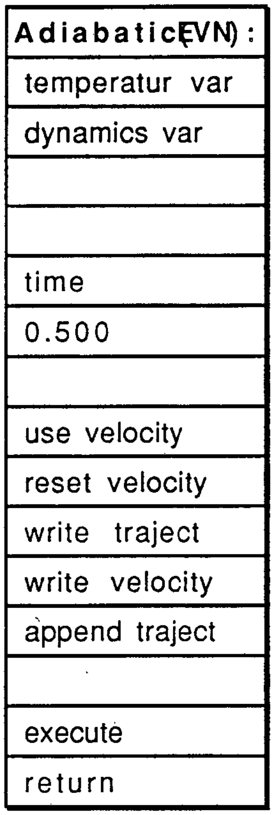


Figure 6.12: Adiabatic dynamics menu

5. Pick **execute**. A dialog box comes up and prompts *Enter trajectory file name*.
6. Enter ***pep.trj*** (*trj* is the extension used for trajectory files). Include disk and directory specifications if you do not want to save the trajectory file in your working directory. Select OK to complete the entry.
7. Type in remarks that will help identify the file when you use it. Terminate the remarks with a blank line (click OK or press <Enter>).

   BIOGRAF now performs a 400 step (0.4 picosecond) dynamics calculation with an initial and final temperature of 600° K. Results are sent to the text window and the Energy results menu for each step of the calculation. Time is in picoseconds, energy in kcal/mol, and the temperature (which is proportional to the kinetic energy), is in degrees Kelvin. The internal energy is the sum of the energies due to bond stretching, bond angle bending, dihedral angle torsions, and inversions. The nonbonding energy is the sum of the van der Waals, electrostatic, and hydrogen bond energies.

6.4 Minimizing the Structure Again

1. After the run finishes, pick **return** twice to get back to the Simulate menu, then pick **mechanics** to bring up the Mechanics menu.

2. Pick **minimize** to minimize the peptide again. The energies may start out relatively high compared to the previous minimization because dynamics tends to stretch out bonds. However, the energies should rapidly decrease. Continue to pick **minimize** until the calculation converges.

3. Compare the final energies to those of the minimized structure we started out with. The total energy should be slightly less than it was for the minimized peptide prior to doing dynamics. Thus even a short dynamics run may yield an improved energy.

   A full dynamics calculation on a small peptide requires a τ of at least several picoseconds. Such a run is too long for an interactive session. However, we can create a macro to do this and run it in the background. We'll do this at the end of the tutorial. The advantage of doing an interactive calculation first is that it provides an optimized conformation as a starting point for the full calculation.

6.5 Saving Information

Next, we'll save the minimized structure in an external file. Saving information protects you from losing work and in subsequent sessions allows you to resume where you left off without repeating all of the steps of the previous session.

1. Click the TOP MENU button to get back to the Top menu, then pick **in-out** to display the In-Out menu (see Figure 6.13).

2. Make sure that the **by file** toggle is highlighted next to the write defaults.

3. Pick **write**. The prompt *Pick the file type to be output:* is displayed and the File types menu appears (see Figure 6.14).

   In order to reconstruct a structure, information about the coordinates of the atoms in the structure is needed. The program saves coordinate information in coordinate files. Several different coordinate file formats are available. These are listed at the top of the menu.

4. Pick **BioDesign** to indicate the BioDesign General Format file type. This type is recommended for internal use with BIOGRAF since it contains all pertinent information.

   Since there is only one file in the system, the program automatically chooses **pep** as the file which is to be output. A dialog box comes up and you are asked to *Enter a name for the output file.*

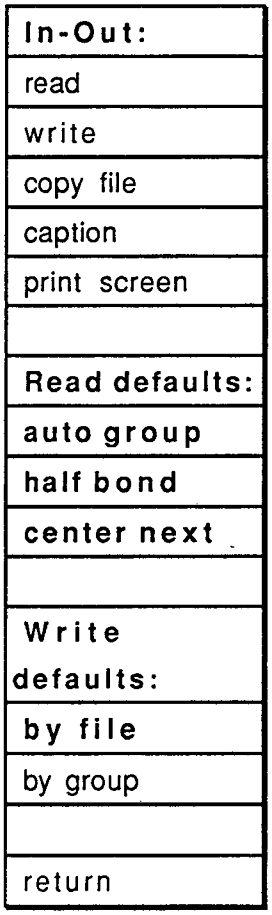


Figure 6.13:
In-Out menu

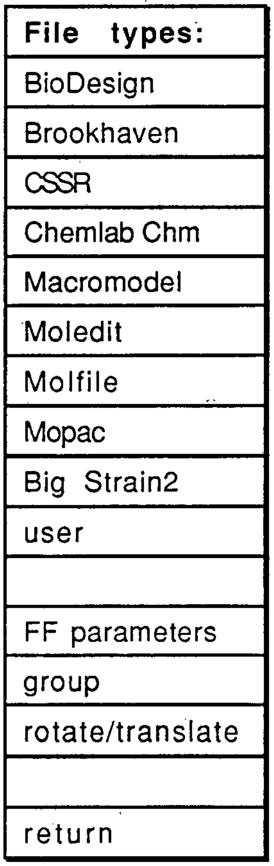


Figure 6.14:
File types menu

5. Type ***pep.bgf*** in the dialog box. Include disk and directory information if not saving it in your working directory. Verify the entry by clicking OK. BIOGRAF prompts you to enter remarks for the output file.

6. Type in remarks that will identify the file the next time you use it. Terminate the remarks with a blank line (Click OK or press <Enter>).

   Messages appear in the text window and in the Modules/Messages window while the program writes the data file.

6.6 Rendering a Group in Monochrome

We'll be animating the conformations from our dynamics calculation in the next section. When animating, we can choose to color-code the conformations according to sequence number or energy, thus making it easier to see how the peptide moved in space during the calculations. Color-coding requires that the group being animated is displayed in monochrome. However, our peptide group was created using the half-bond representation (each atom is colored according to its element and the half of the bond closest to the atom is made the same color as the atom). Before animating we must therefore render the peptide group in a single color. This is done using the **render groups** operation on the **visualize** menu.

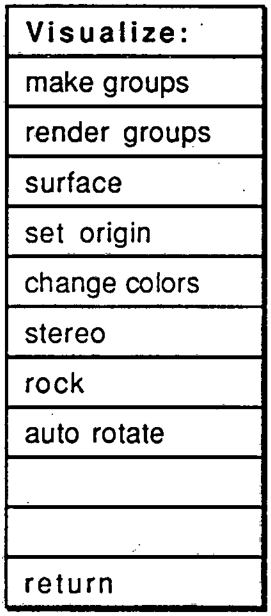


Figure 6.15: Visualize menu

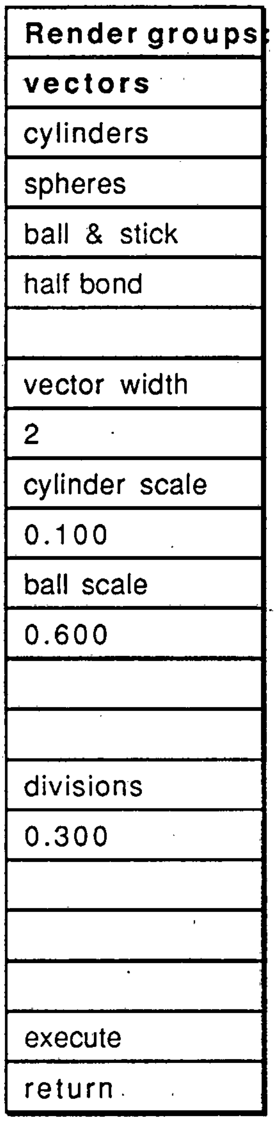


Figure 6.16: Render groups menu

1. **Return** to the top level menu, then pick **visualize**. The Visualize menu shown in Figure 6.15 is displayed.
2. Pick **render groups** to bring up the menu shown in Figure 6.16. Since there is only one group in the system (**peptide**), it is automatically chosen as the group to be rendered.
3. Make sure **half bond** is toggled off.
4. Pick **execute**.

   The leuenkephalin molecule is redisplayed in a single color. The color of the group could be changed if desired using the **change colors** option on the Visualize menu. We won't we doing that here, however.

6.7 Animating a Dynamics Trajectory

We now animate the leuenkephalin dynamics calculation.

6.7.1 Defining the Trajectory File and Listing the Conformers

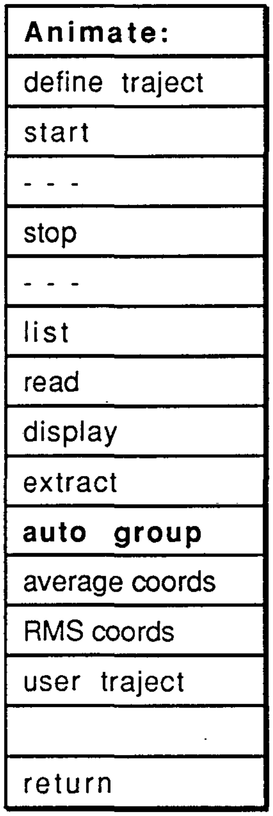


Figure 6.17: Animate menu

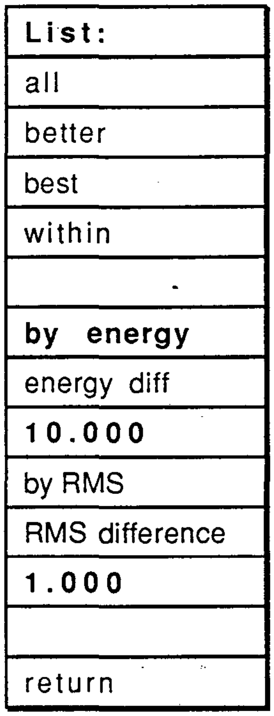


Figure 6.18: List menu

1. Click the TOP MENU button to get back to the top level menu, then pick **simulate**.
2. Pick **animate** to display the Animate menu shown in Figure 6.17.
3. Pick **define traject** to define the trajectory file to be animated. A browser box comes up and BIOGRAF prompts *Enter the name of the trajectory file.*
4. Locate the directory in which you saved the trajectory file, then select the ***pep.trj*** file name from the listings in the browser box. Click OK to complete the selection.

   When the file has been read in, the times for the first and last conformers in the file are displayed under the **start** and **stop** options on the menu. **Start** and **stop** can be used to restrict the sequence of conformations to be included in the animation. We want to look at the whole range of conformations, so we'll leave these values as they are (0.010 ps for **start** and 0.400 ps for **stop**).
5. Pick **list** to display the List menu shown in Figure 6.18.

   The conformers may be listed in two modes: using their associated energy values to make comparisons (**by energy**), or using the root mean square (RMS) values obtained from RMS matching to a reference conformation which is specified by the user (**by RMS**). We'll use **by energy** mode. This is the default, so it should be highlighted. If not, toggle it on.
6. Pick **all** to obtain a listing of all the conformers in the text window. Each conformation number is listed with its energy value. A listing of the best conformer (the one with the lowest energy value) can be obtained by picking **best** on the menu. **Better** lists the conformers that are consecutively better than the best one so far and **within** lists those that are within a specified energy difference (**energy diff**) from the best.

7. Pick **better, best** and **within** and note the listings obtained with each option. Note the conformer with the best (lowest) energy value.

8. Note the number of conformers within **energy diff** (10 kcal) of the best. Pick **energy diff,** enter a new value, and pick **within** until there are about 20 conformers within energy diff of the best. You'll be using this energy diff value later to look at the low energy conformers.

9. Pick **return** to return to the Animate menu.

6.7.2 Reading in the Positions of the Conformers

1. Pick **read** on the Animate menu to read in the positions of the conformers. (A **read** operation must be performed before any animations can be done.) After picking **read,** the group to be animated must be specified. Since there is only one group in the system, **pep** is automatically chosen.

   The Read menu will then appear as shown in Figure 6.19.

   As with the list operations, the conformers may be read in **by energy** or **by RMS** mode (we'll leave it in the default **by energy** mode); similarly, **all** the conformations may be read in or they may be restricted to those specified using the **better** or **within** options.

2. Pick **all** to read in all the conformations. BIOGRAF responds with the prompt *Enter the read increment (1 implies read every point):*

3. Enter *1* to read every conformer, then select OK. The message *Working...* appears as they are read in. Each of the conformers are listed in the text window along with their associated time and energy values.

4. **Return** to the Animate menu.

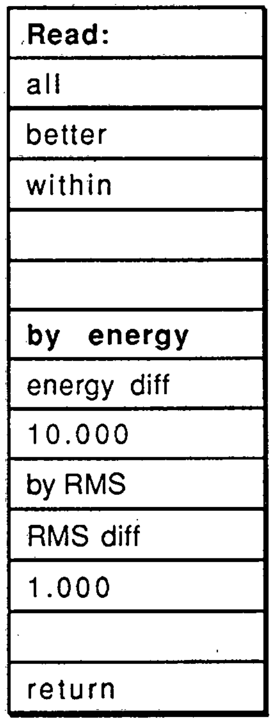


Figure 6.19: Read menu

6.7.3 Animating the Conformations

Now that we have read in the positions of all the conformers, we are ready to animate them.

1. Pick **display** to bring up the Display menu shown in Figure 6.20.

   The conformations may be colored by sequence number, by energy, or by RMS value. Conformations that were read in by energy mode are automatically color-coded by energy, so **color by energy** appears highlighted on the menu.

   The conformation with the lowest energy (the best) will be red, the conformation with the highest energy will be blue, and those in between will be interpolated between red and blue. **Energy diff** is automatically set so that the conformers that were read in span the whole range of available colors.

2. Pick **animate**. BIOGRAF displays the first conformation and prompts

   To animate the sequence, drag the cursor
   with the control key and the middle mouse
   button depressed.
   Click the left mouse button to stop.

3. Use the mouse to animate the sequences as described in the prompt (if using a dials unit, rotate dial 8 to cycle through the conformations). The time and energy value for each of the conformations are displayed in the lower left corner of the graphics window (labeled T and E respectively). Find the conformation with the lowest energy.

4. Click the left mouse button to stop the animation.

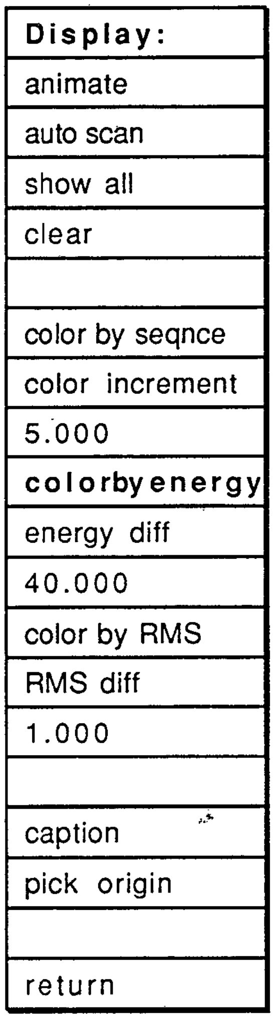


Figure 6.20:
Display menu

6.7.4 Using Auto Scan

We can also scan through the conformations automatically.

1. Pick **auto scan**. BIOGRAF prompts.

   Click anywhere to start the autoscanning.
   Use control-middle mouse button to adjust
   the speed (left is slower, right is faster).
   Click in the graphics window to stop.

2. Click anywhere to start scanning. Adjust the scan speed if desired (if you have a dials unit, rotate dial 8 to change the scan speed; rotate it clockwise to scan faster).

3. Click in the graphics window to stop.

6.7.5 Looking at the Low Energy Conformers

We may wish to restrict the conformations that we look at to the lower energy ones, that is, to those that fall within a specified energy difference from the best. To do this, we must change **energy diff** to the desired cutoff, read the trajectory file in again this time using the **within** option, and reanimate. We'll choose the **energy diff** value we found earlier when listing the conformers (the one which gave us about 20 conformers within energy diff of the best).

1. **Return** to the Animate menu, then pick **read.**
2. Pick **energy diff** and enter the value which produced about 20 conformers within energy diff of the best.
3. Pick **within**. Only the lower energy conformers will be read in.
4. **Return** to the Animate menu, then pick **display**.
5. Reanimate using **animate** or **auto scan**.
6. Stop the animation (click the left mouse button if animating or click in the graphics window if auto scanning).

6.7.6 Showing All the Conformations at Once

All the conformations may be displayed simultaneously (on top of each other) as well. This is useful in illustrating the degree to which different parts of the molecule move or remain stationary.

1. Pick **show all**. A dialog box comes up and we are successively prompted to enter values for the first and last conformation and for the increment.
2. Enter ***1*** for the first conformation to be displayed, ***999*** for the last and ***1*** for the increment; pick OK after each entry (entering a much higher number for the last makes sure we get them all).

   BIOGRAF will print the message *Working...* , then all the conformations that we read in above will be displayed at once. Note the best conformation displayed in red. An example of this is shown at the top of Figure 6.21. Since this tutorial is not in color, we show the best conformation at the bottom of the figure for comparison (this does not appear on your display). Note the movement of the side chains with respect to the backbone.

   Note: Your conformers may look somewhat different since the minimized structure you obtained was probably different (there are several local minima in the potential energy surface).

3. Pick **clear** to clear the display, then **return** to the Animate menu.

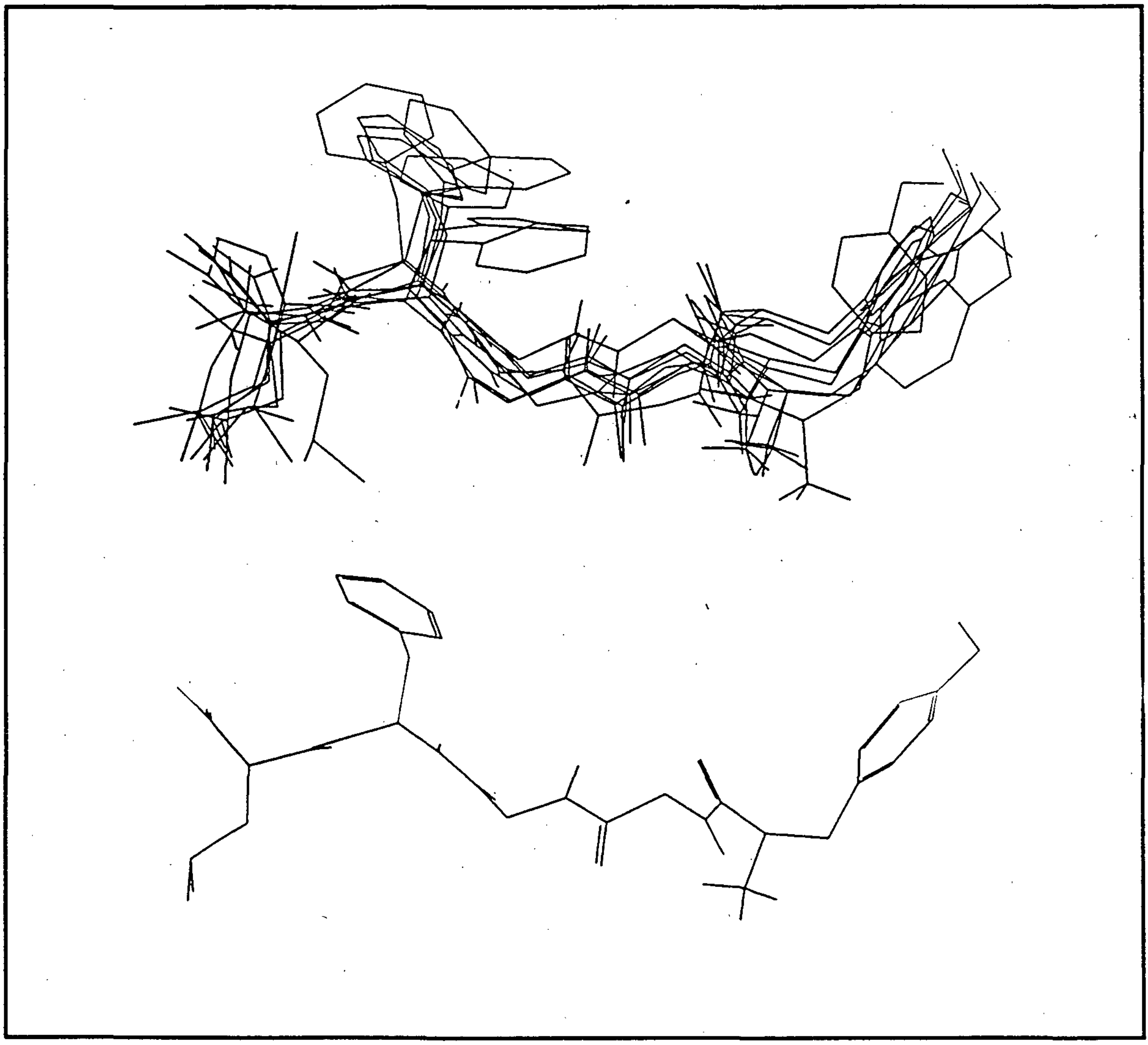


Figure 6.21: Top: Leuenkephalin conformations displayed using the **show all** option.
Bottom: Best conformation (lowest energy).

6.7.7 Extracting the Best Conformer

BIOGRAF allows you to extract the information for a particular conformation (*e.g.*, the best) and create a file for it so that energy minimizations or other analyses may then be performed on it. If the trajectory file was created using dynamics calculations, the conformation to be extracted is specified by giving its time value. Otherwise, the conformation's sequence number is used. We can get the time value for the best conformation using the list operations described earlier.

1. Pick **list,** then pick **best** to list the best conformer. Note the time value of the best conformer (be sure to note the exact value, all four decimal places).

2. **Return** to the Animate menu, and make sure the **auto group** option is toggled on. This will create a group for the extracted conformer.

3. Pick **extract**. BIOGRAF prompts *Enter the time value of the conformer to be extracted:*

4. Enter the time of the best conformer (include all four decimal places).

   A new file, named "CONF*n*" where *n* is the sequence number of the extracted conformer, is created. The best conformation will be displayed, and a new group, also named "CONF*n*", will be created. Other conformers may be extracted in a similar fashion.

   It might be a good idea to save the coordinates of the best conformation in a *bgf* file so that you can leave BIOGRAF if you want and then resume again later where you left off. (In order to reconstruct a group, BIOGRAF needs the information about the coordinates of the atoms in that group.)

   Warning: Once the Animate menu is exited, all the set up data is lost. You must redefine the trajectory file and read in the conformations again before any more animations can be done.

5. Click the TOP MENU button to get back to the Top menu.

6. Make a BioDesign General Format type coordinate file as we did in section 6.5, but choose **CONF*n*** as the file descriptor for which you want to output information, and name the new file ***CONFn.bgf***.

6.8 Creating a Macro and Running It in the Background

One of the easiest ways to run a job offline or in the background is to begin the job as an interactive session as we did in this tutorial. The program automatically creates a macro command file of the session called *logfile.macro* and places it in your current working directory. One can rename this file, edit it as desired, disable the graphics displays, and then run the macro in the background. It is usually a good idea to save the file information sent to the text window. This can be done by redirecting output to a file on the command line which is used to execute the program. The stored output file can be followed by running "tail -f" on the output file from another window.

In this section, we will demonstrate how to create a macro which performs a full dynamics calculation on the peptide leuenkephalin. We'll edit the *logfile.macro* generated from this tutorial to run dynamics for a longer time (10 ps) and change the initial and final temperatures to 400° K. We'll then illustrate how to run this macro in the background.

1. Exit the program (**return** to the Top menu, pick **exit**, then confirm by picking OK).

2. Use the ***mv*** command to rename the *logfile.macro* file generated from this session *dynamics.macro*.

3. Edit the *dynamics.macro* file so that it builds leuenkephalin, checks the energies, minimizes it, and runs dynamics as before, but change the dynamics time to 10 ps and the initial and final temperatures to 400° K. Delete the lines which change the write trajectory frequency (we'll leave this at the default value of 0.1). Change the name of the trajectory file created to *pepdyn10.trj* and alter the remarks appropriately. Remove the lines which minimize the structure again. The structure should still be saved but the name of the *bgf* file should be changed to *pepdyn10.bgf* and the remarks updated. All of the lines after this point (starting from rendering the peptide group in monochrome) should be deleted except for the exit and endmacro lines. An example of the edited macro is included on the following pages.

4. Disable the graphics and menu displays by adding the following line to your defaults file:

   ```
   L_BATCH          yes
   ```

5. Execute the macro in the background and redirect output to an output file by entering the following command:

   ```
   biograf   dynamics.macro   >&   dynamics.out   &
   ```

   The command used to start up the program (biograf) may be different at your site. The first "&" tells the program to include error messages in the output; the final "&" indicates that the job is to be run in the background.

6. Follow the progress of the job by entering the following command from another window:

   ```
   tail   -f   dynamics.out
   ```

7. When the job finishes, examine the results by viewing the *dynamics.out* file. You may want to minimize the structure or continue running more dynamics on it.

```
beginmacro
%
% version      : 3.00
% version date : 16:03:58 3/3/92
% link date    : 05:07:32 3/26/92
%
% Macro created on  6/04/92   12:39:45
%
% /usr/msi/johndoe/biograf.def
Top menu/build
      Build/peptide
              new file
              "pep"
            Peptide/l_library
            Selections/tyr_l
%           /biodesign/v300/lib/biograf/peptide_l/tyr_l.bgf
            Selections/gly_l
%           /biodesign/v300/lib/biograf/peptide_l/gly_l.bgf
            Selections/gly_l
%           /biodesign/v300/lib/biograf/peptide_l/gly_l.bgf
            Selections/phe_l
%           /biodesign/v300/lib/biograf/peptide_l/phe_l.bgf
            Selections/leu_l
%           /biodesign/v300/lib/biograf/peptide_l/leu_l.bgf
            Selections/return
            Peptide/mechanics
                  Mechanics/return
            Peptide/simulate
                  Simulate/show eex
                  Simulate/mechanics
                        Mechanics/one energy
                        Mechanics/minimize
                        Mechanics/# of steps
                          "300"
                        Mechanics/minimize
                        Mechanics/minimize
                        Mechanics/defaults
                              Defaults/temperatur var
                                    Temperatur var/initial temp
                                      "400"
                                    Temperatur var/final temp
                                      "400"
                                    Temperatur var/return
                              Defaults/return
                        Mechanics/return
                  Simulate/dynamics
                        Dynamics/adiabatic
                              Adiabatic(EVN)/time
                                "10"
                              Adiabatic(EVN)/write traject
                              Adiabatic(EVN)/execute
                                "pepdyn10.trj"
                                "10 ps dynamics on leuenkephalin at 400 K"
                                " "
                              Adiabatic(EVN)/return
```

```
                        Dynamics/return
                  Simulate/return
            Peptide/return
      Build/return
Top menu/in-out
      In-Out/write
      File types/BioDesign
        pep
        return
        "pepdyn10.bgf"
        "leuenkephalin after 10 ps dynamics at 400 K"
        " "
      In-Out/return
Top menu/exit
  "OK"
%
endmacro
```

6.9 Conclusion

This concludes our study of the dynamics of small peptides. The leuenkephalin calculations indicate that the peptide is a floppy system with many favorable conformations. Leuenkephalin also provided an example of the peptide building capabilities of BIOGRAF. You can make any molecule with the program using the Organic Builder, or for more specialized molecules, the RNA-DNA, Lipid, Carbohydrate, Crystal, or Solvate Builders.

After you build a molecule, BIOGRAF can optimize its structure by

- doing a minimization,
- performing a search of conformational space (see Chapter 7), or
- running molecular dynamics calculations.

This concludes Chapter 6. To leave BIOGRAF, pick **exit** from the Top menu and then confirm by picking OK.

Chapter 7

Site-Directed Mutagenesis

This chapter introduces the operations which allow you to replace one side chain with another in a molecule. Such a substitution of amino acids in a protein is called *site-directed mutagenesis*, and is a key element in the development of protein engineering. BIOGRAF can model the effect of site-directed mutagenesis, making it possible to interpret experimental results and suggest new experiments. This tutorial goes through the replacement of a phenylalanine with a tyrosine residue in the Cytochrome C molecule found in tuna.

In addition to modeling side chain replacement, BIOGRAF also predicts optimal conformations for the altered molecule. The program uses molecular mechanics to model the forces between atoms and so predict structures. It is assumed that you have completed Chapter 1.

7.1 Executing a Macro on Start Up

Macros are often used to perform routine operations, thus speeding up the process. For example, a macro might be written to read in the data for a molecule that one is working with on a regular basis. Macros are executed by getting into command mode and typing in the character "@" followed by the macro file name. The commands specified in the macro file are executed until the macro finishes or until control is passed to the menus. Macros may also be specified in the command line used to start up and run the program (*e.g., biograf filename.macro*). If the macro is not located in the current working directory, then directory specifications must also be included.

We have a macro file which reads in the coordinate and group data for the Cytochrome C molecule. It is called *test.macro* and is stored in your distribution macro directory. We'll specify this macro as part of the start up command.

Typically the full specifications for the distribution macro directory are */usr/msi/biogv300/macro* The distribution macro files are always located in the */biogv300/macro* directory; however, the first part of the path specifications may differ on your system. To determine the location of the */biogv300/macro* directory on your system, type *echo $MSIUSR* from a systems window. The directory listed in response to this command should precede */biogv300/macro* when giving directory specifications.

1. Type in the start up command followed by the specifications for your distribution macro directory and the macro file name *test.macro*.

   Example: ***biograf /usr/msi/biogv300/macro/test.macro***

   As explained in Chapter 1, the command used to start up and run BIOGRAF may vary from site to site. Consult your systems supervisor for the command used at your site.

   The program starts up then quickly executes the series of commands in the *test.macro* file. The coordinate and group files for Cytochrome C are read in and the molecule appears in the graphics window. The last command in the macro file ("menu") returns us to interactive mode. The Top menu is displayed in the Menu Pad as shown in Figure 7.1.

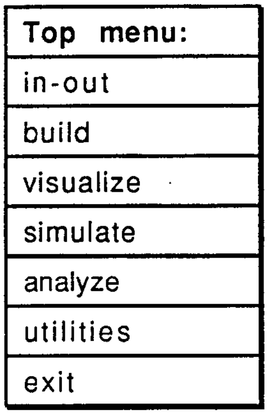


Figure 7.1: Top menu

7.2 Looking at the Groups

BIOGRAF handles molecular information as *groups*. A group is a collection of atoms and bonds chosen in such a manner as to reduce a molecular structure into recognizable parts. There are various ways to partition the Cytochrome C molecule into meaningful groups. We chose the ten groups shown in Table 7.1.

Group	Description
All	The entire molecule
Trace	A trace through all the α-carbons (C_α)
Main	The peptide backbone or main chain (-N-C_α-C-)
Side	The peptide residues or side chains
Carbonyl	The peptide C=O groups
Heme	The heme group
Ligands	The histidine (HIS 18) and methionine (MET 80) residues which constitute the fifth and sixth ligands to the heme iron
Thio E	The two thioether bonds which attach the heme to the protein
Water	The crystallographic water molecules
Labels	The residue labels and numbers

Table 7.1: Groups in Cytochrome C molecule

When BIOGRAF finishes reading the data, it displays four groups: Trace, Heme, Ligands, and Thio E. These groups are turned on, while the remaining groups (All, Main, Side, Carbonyl, Water, and Labels) are turned off and thus do not appear on the display. Groups may be turned on or off (displayed or not) by clicking the DISPLAY button and toggling on the group name which appears in the Display menu.

1. Click the DISPLAY button located in the Modules/Messages window (see Figure 7.2). The Display menu shown in Figure 7.3 comes up and lists all the groups currently in the system. Note that the four groups which are currently displayed in the graphics window appear highlighted in the menu.

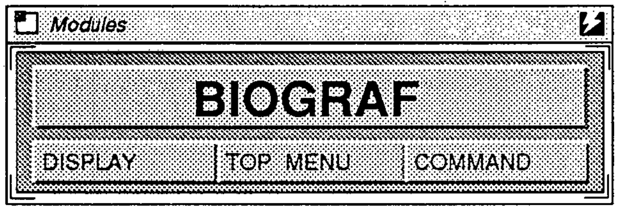


Figure 7.2 Modules/Messages Window with DISPLAY, TOP MENU and COMMAND Buttons

2. Toggle the various groups on and off to look at the molecule.
3. Set the toggles so that the groups **Main**, **Heme**, **Ligands**, and **Thio E** are shown in the graphics window.
4. Pick **return** when finished.

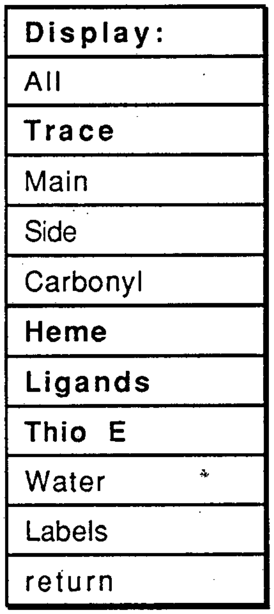


Figure 7.3: Display menu

7.3 Creating Groups

In this section, we create a group which initially will correspond to a phenylalanine residue. After the replacement it will correspond to a tyrosine residue.

Groups may be defined in either of two ways:

- with standard structural elements or *primitives* (*e.g.*, an amino acid residue), or
- with standard structural descriptors or *selectors* (*e.g.*, hydrophilic side chains).

The Standard Group Element Appendix of the *Biograf Reference Manual* lists the primitives and selectors that may be used to describe biological macromolecules.

Groups are created using the operations on the Make groups menu found under **visualize**.

1. Pick **visualize** from the Top menu to bring up the Visualize menu shown in Figure 7.4.
2. Pick **make groups** on the Visualize menu to display the Make groups menu shown in Figure 7.5.

The top half of the menu contains the proper sequence of steps:

setup ⇨ files ⇨ chains ⇨ residues ⇨ limits ⇨ create

The primitives and selectors used to define the groups are chosen under **limits**. The sequence given above *must* be performed in the prescribed order. However, it may not be necessary to repeat every step each time you define a group because **files, chains**, and **residues** always correspond to their last setting.

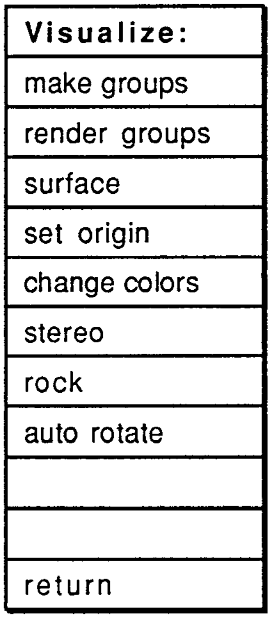


Figure 7.4: Visualize menu

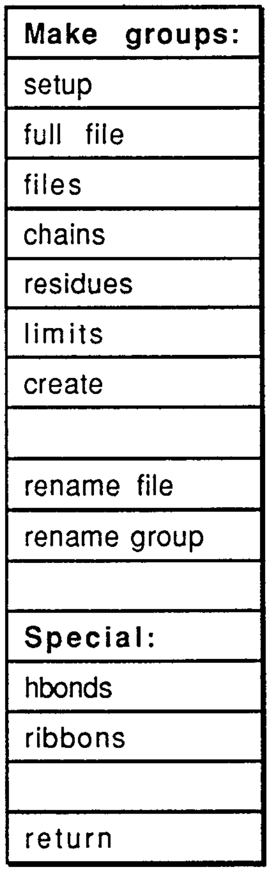


Figure 7.5: Make groups menu

You can make changes in the middle of the sequence by redoing the appropriate step and then repeating the subsequent steps in the proper order. The defaults for **files, chains**, and **residues** are often appropriate, so that constructing groups involves only the steps

setup ⇨ limits ⇨ create

Primitives and **selectors** also retain their previous settings, but the residues picked or keyed in under **limits** are lost each time you toggle **setup**, so that the limits include the entire molecule. To create a group containing all the atoms of any one coordinate file, clude the entire molecule. To create a group containing all the atoms of any one coordinate file, toggle the **full file** switch on. If the **full file** switch is on, the group created will contain all the atoms in the file. The user must **setup** the group and select the file whose atoms will be used, but the **chains, residues** and **limits** toggles are not active. In order to create a group containing only part of a file, the **full file** switch must be toggled off. **Full file** is toggled off when the program starts.

7.3.1 Defining the Tyrosine 82 Group

1. Pick **setup** on the Make groups menu. The Groups menu appears as shown in Figure 7.6 and we are prompted *Pick the group to be defined.*

2. Pick **add group**. A dialog box comes up and we are asked to *Enter a name for the group to be defined.*

3. Type ***TYR 82*** as the group name in the dialog box. To verify the entry click the OK button or simply press <Enter>.

4. Pick **files** from the Make groups menu. The Files menu appears with the file descriptor **Tuna** highlighted as shown in Figure 7.7 and we are prompted to *Pick the data file designators.* This is the file we want, so we'll leave it toggled on. Verify the file selection by picking **return**.

5. Pick **chains** from the Make groups menu. A new menu appears listing the chain designators currently in the system as shown here in Figure 7.8.

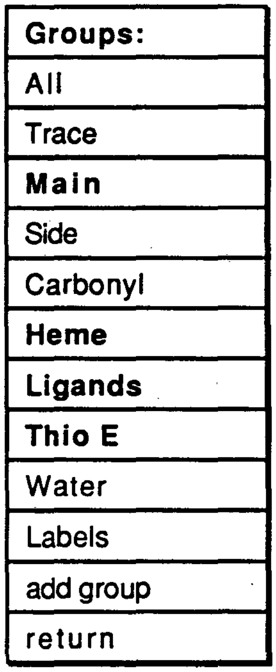


Figure 7.6:
Groups menu

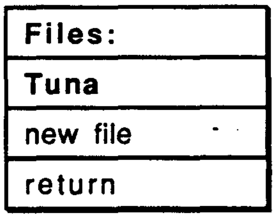


Figure 7.7:
Files menu

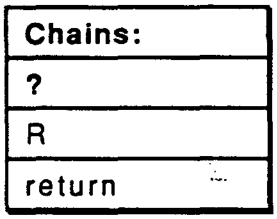
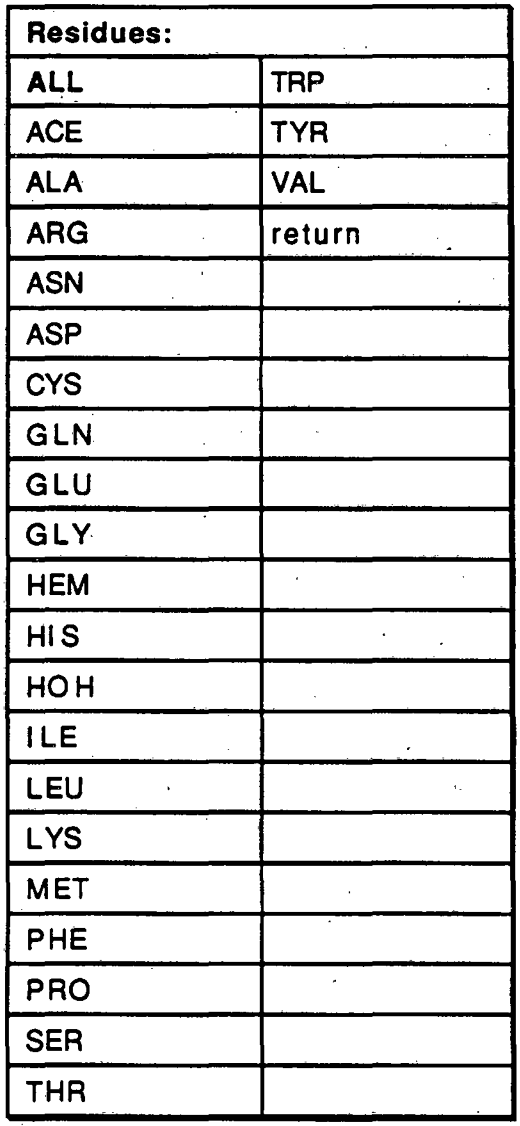


Figure 7.8:
Chains menu

Two chain designators are listed: ?, which indicates that the program will accept any chain designator, and **R** (for reduced Cytochrome C). The first time the Chains menu is entered, ? is toggled on. If it is not on, toggle it on. Pick **return** to verify the selection.

6. Pick **residues** from the Make groups menu. The Residues menu appears as shown in Figure 7.9. **ALL** should be toggled on, indicating that any residue will be accepted. If **ALL** is not on, toggle it on. Complete the selection by picking **return**.



Residues:	
ALL	TRP
ACE	TYR
ALA	VAL
ARG	return
ASN	
ASP	
CYS	
GLN	
GLU	
GLY	
HEM	
HIS	
HOH	
ILE	
LEU	
LYS	
MET	
PHE	
PRO	
SER	
THR	

Figure 7.9: Residues menu

7.3.2 Choosing the Primitives

We next consider the primitives that describe tyrosine 82. A number of the primitives shown in Figure 7.11 have names that start with "H" (*e.g.*, HMAIN). These refer to the hydrogen atoms and bonds attached to the corresponding primitive (*e.g.*, MAIN). Very few X-ray crystallographic structures include data for H atom positions, so BIOGRAF often needs to generate the positions. In general, it is a good idea to include the H primitives when setting up a group to make it easier to add H atoms later.

1. Pick **limits** on the Make groups menu to display the Limits menu (see Figure 7.10).
2. Pick **primitives** on the Limits menu. The Primitives menu is then displayed listing the primitive structural elements as shown in Figure 7.11. **STD SET** should be toggled on, indicating that the standard set of primitive elements are included in the group. The standard set includes all the elements listed except the α carbons (**TRACE**) and the labels for the α carbons (**LABELS CA**).
3. Toggle on the primitives **SIDE CHAIN** and **HSIDE**, then pick **RETURN**.

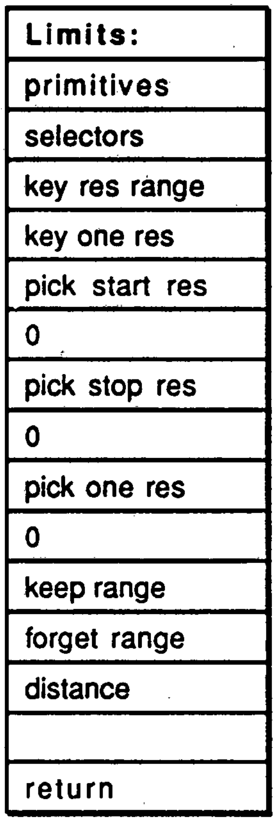


Figure 7.10:
Limits menu

4. Pick **key one res** on the Limits menu. BIOGRAF prompts *Enter the residue number*. Enter *82* in the dialog box, then select OK.
5. Pick **return** on the Limits menu to return to the Make groups menu, then pick **create**. *Working*...is displayed in the Modules/Messages window. When the program finishes, a group corresponding to the phenylalanine residue at position 82 in the carbon backbone of Cytochrome C appears in the graphics window.

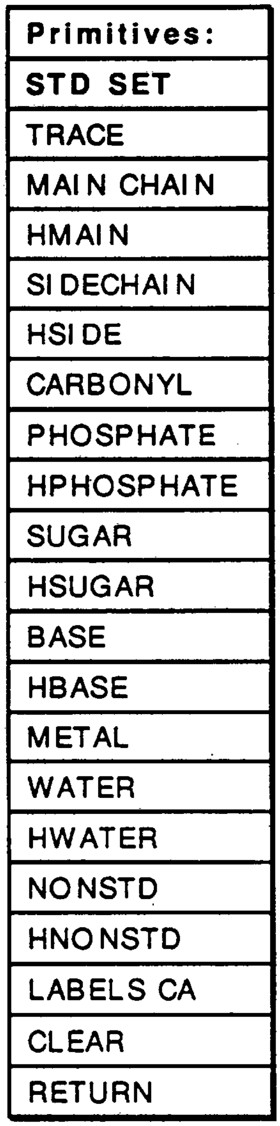


Figure 7.11:
Primitives menu

6. If it is difficult to see this group, move it to the front of the molecule or decrease the depth cueing to make the distant residues brighter.

   We have defined a separate group to hold the residue that will be replaced so that later we can do calculations on it alone. However, in general you are not required to define separate groups for the replacement residue.

7.4 Replacing the Side Chain

The **replace** operation in the Peptide Builder allows you to substitute one side chain group with another. To replace the phenylalanine with tyrosine, we proceed as follows:

1. Click the TOP MENU button in the Modules/ Messages window to get back to the top level menu.
2. Pick **build** to bring up the Build menu shown in Figure 7.12.
3. Pick **peptide**. The Peptide menu is displayed as shown in Figure 7.13. The **setup** option is automatically invoked and we are prompted to pick the file on which to operate from the Files menu.
4. Pick **Tuna**. The **Tuna** group appears in the graphics window; all other groups that had been on are turned off. We want to keep the groups as they were, so we'll turn them back on.

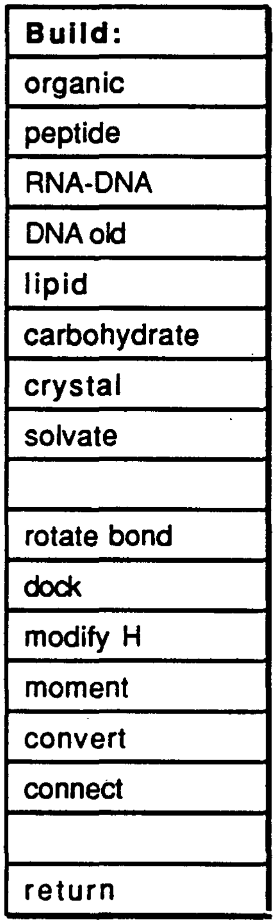


Figure 7.12: Build menu

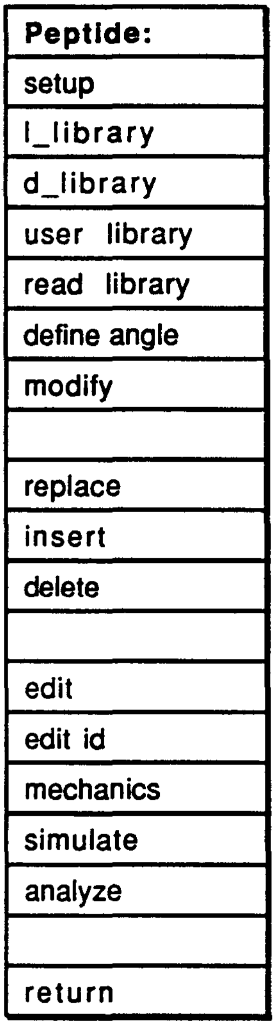


Figure 7.13: Peptide menu

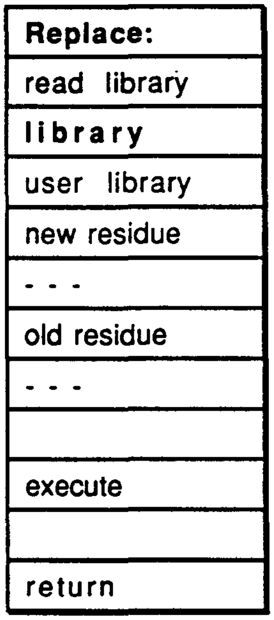


Figure 7.14: Replace menu

5. Click the DISPLAY button. Toggle off the **Tuna** group, then toggle on the **Main, Heme, Ligands, Thio E**, and **TYR82** groups as before. Pick **return** when finished.
6. Pick **replace** on the Peptide menu. The Replace menu comes up as shown in Figure 7.14.

   The first time that BIOGRAF enters the Replace menu, it reads in the BIOGRAF library of side chain fragments. When the fragments have all been read in, the **library** option is highlighted, indicating that this library is available.

7. Pick **new residue**. BIOGRAF prompts *Pick the new side chain* and displays the list of amino acids as shown in Figure 7.15.
8. Pick **tyr** for the replacement residue. The residue name (**TYR**) appears on the Replace menu beneath the **new residue** box.
9. Click the DISPLAY button, then toggle on the **Labels** group from the Display menu. Labels appear next to the residues in the graphics window. Pick **return**.
10. Pick **old residue** on the Replace menu. BIOGRAF responds with the prompt *Pick the residue to be replaced* and outlines the graphics window with a red border.
11. Zoom in on the **TYR 82** group (currently it is a phenylalanine residue and is labeled **PHE 82**) and rotate it until the residue is centered in the display and not overly blocked by the rest of the molecule.
12. Pick the **PHE 82** residue off the display by toggling an atom on it the same way that you toggle a menu function. A red cross marks the picked atom and information about the residue is listed in the text window. The residue label (**PHE82**) appears on the Replace menu below the **old residue** box.
13. If you picked the wrong amino acid, repeat the **old residue** operation.
14. Pick **execute**. BIOGRAF recalculates the lists of atom positions for both the primitives and the groups. After the final recalculations, a tyrosine residue replaces the phenylalanine residue and the label on carbon 82 of the MAIN chain changes to **TYR 82**. Note that the tyrosine differs from phenylalanine only by the hydroxyl group at the *para* position.

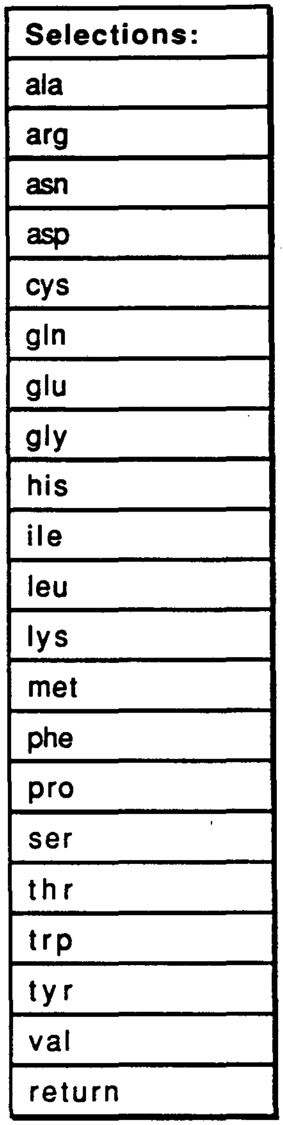


Figure 7.15: Amino acid residues menu

7.5 Changing the Center of Rotation

In a replacement operation, BIOGRAF inserts the new residue at an arbitrary orientation. As an aid to optimizing the position of the new amino acid, it is useful to be able to rotate it relative to the rest of the molecule. This is done using the **rotate bond** operation. For convenience, we first change the origin of rotation for the *entire* display to the center of **TYR 82**. This is done using the **group origin** option which is found on the Rotate bond menu.

1. **Return** to the Peptide menu, pick **edit** to bring up the Edit menu (see Figure 7.16), then pick **rotate bond**. The Rotate bond menu shown in Figure 7.17 appears.

   The **setup** function is automatically invoked and we are prompted to define the bond vector. We'll come back to this later.

2. Pick off the molecule (a non-atom). The prompt asking us to define the bond vector disappears.

3. Zoom in and center the **TYR 82** group until it fills the graphics window.

4. Rotate the Cytochrome C. Note that it takes very little rotation to move **TYR 82** out of view. This results from the fact that the Cytochrome C rotates around its own center.

   There are four options on the Rotate bond menu which can be used to change the center of rotation (**group origin, file origin, pick origin,** and **zero origin**). These options let you specify the point around which the *entire* molecule rotates. The **group origin** option changes the origin of rotation to the center of a specified group. We'll specify the **TYR 82** group.

5. Pick **group origin**. BIOGRAF brings up the Groups menu and prompts *Pick the groups to center rotation on:*

6. Pick **TYR 82**, then pick **return**.

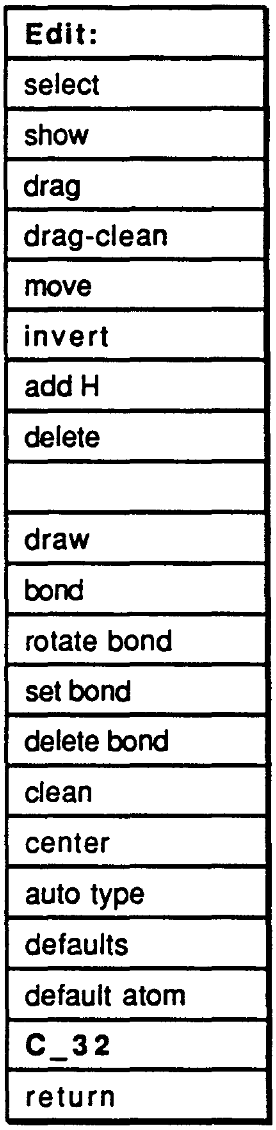


Figure 7.16: Edit menu

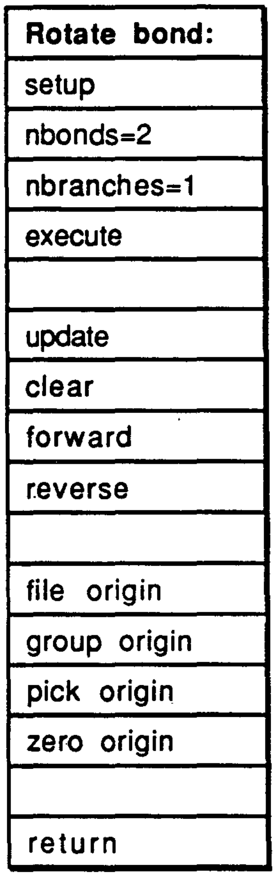


Figure 7.17: Rotate bond menu

7. Rotate the molecule. Note how the **TYR 82** no longer moves off the display.

7.6 Rotating the Side Chain

We'll now go back and define the bond vector so that we can rotate the TYR 82 side chain *relative* to the rest of the Cytochrome C molecule. That is, we'll rotate around the bond connecting the tyrosine residue to the rest of the molecule. The bond rotation operations are at the top of the menu.

1. Pick **setup**. We are again prompted to define the bond vector (to select the two atoms of the bond that is to be rotated). The portion of the molecule to be rotated is located past the second atom picked. Rotations are done using the dials or mouse; up to eight bonds may be rotated simultaneously. The number of bonds to be rotated is set with the **nbonds** option. We'll leave it at the default (two bonds).

   Figure 7.18 shows the labels used to identify the atoms on the tyrosine residue. The nomenclature comes from the Brookhaven Protein Data Bank and is based on the IUPAC-IUB Commission on Biochemical Nomenclature ["Abbreviations and Symbols for the Description of the Conformation of Polypeptide Chains. Tentative Rules (1969)", *J. Biol. Chem.* **245**, 6489 (1970)].

2. Pick the **TYR 82** Cα as the base atom of the bond vector (where TYR 82 attaches to the main chain). A red cross appears at the α position.

   If you make a mistake at any time while specifying the atoms, pick **setup** again and start over.

3. Pick the **TYR 82** Cβ atom. A yellow cross appears at the β position. The red border surrounding the graphics window disappears.

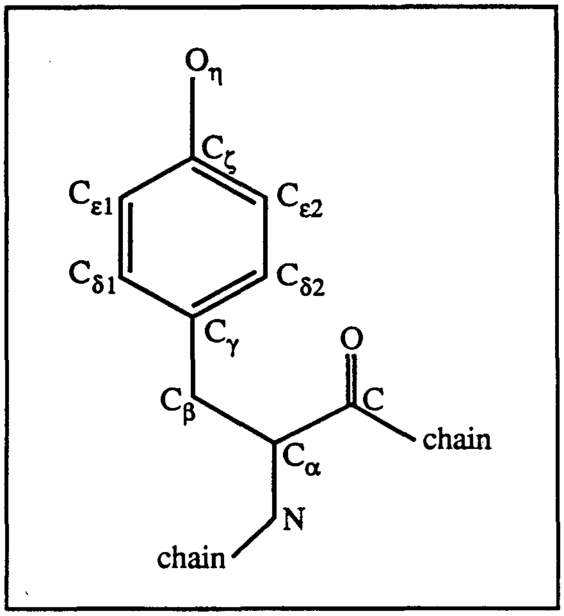


Figure 7.18: Labeling of atoms on the tyrosine residue

4. Pick **execute** on the Rotate bond menu. This generates a ghost image for the region being rotated and enables the dials or mouse. The dihedral angle values for the first two bonds are displayed and labels appear on the structure to indicate which dial affects which bond. As shown in Table 7.2, dial 8 (D8) which is usually the depth cueing dial, is for the first bond (the Cα-Cβ bond we defined); dial 7 (D7) which is usually the z translate dial is for the next bond down (the Cβ-Cγ bond). If using the mouse, the first bond is rotated by pressing the middle mouse button and the control key and moving the mouse; the second bond is rotated by pressing the middle mouse button and the shift key and moving the mouse.

Bond Rotated	Dials	Mouse Button and Key	Normal Function
bond #1	dial 8	middle button + control key	depth cueing
bond #2	dial 7	middle button + shift key	z translate

Table 7.2: Using the dials or mouse to rotate bonds

5. Use the dials or mouse to rotate the ghost tyrosine residue about the first bond (see Table 7.2).

   The ghost image of the tyrosine residue is shown in red. The actual tyrosine remains displayed in its original position and colors. The other dials or mouse manipulations retain their regular functions, so you can rotate about the x, y, and z axes to get a better view of the differences between the positions of the ghost tyrosine and the true image.

6. Rotate the ghost into a position that looks reasonably well accommodated by the rest of the molecule.

7. Pick **update** on the Rotate bond menu. The tyrosine residue then updates to the position of its ghost.

See the description of **rotate bond** in Section 3.9 of the *BIOGRAF Reference Manual* for more details on these operations.

7.7 Adding Hydrogens

In order to determine an optimal conformation for the TYR 82 residue, at least some of the hydrogen atoms should be included in the minimization of the energy.

7.7.1 Hydrogen Atom Types

There are two standard approaches used to generate hydrogen atom positions: the *extended atom* model and the *all atom* model. The extended atom model treats H atoms implicitly by adjusting the van der Waals radius of the atoms to which the H atoms are covalently bonded to account for the hydrogen atoms. The model thus represents a methyl group by a single "CH3 atom" with a van der Waals radius slightly larger than the radius of an isolated carbon atom and a mass equal to that of the methyl group. In contrast, the all atom model generates the actual H atoms.

BIOGRAF recognizes two explicit hydrogen atom types, H___A and H_. The H___A type represents the hydrogen bonding hydrogen atoms, whereas the H_ type represents all other H atoms used with the all atom representation. The types determine whether or not nonbond interactions (van der Waals and electrostatic) are included in calculations:

H___A	nonbond interactions excluded
H_	nonbond interactions included

An explicit hydrogen bond potential is used for H___A hydrogens. More details on the hydrogen atom types can be found in the appendices of the *BIOGRAF Reference Manual.*

7.7.2 Heteroatoms

An H atom bonded to an electronegative or *heteroatom* such as O and N acquires a fractional charge which makes it a good candidate for hydrogen bonding. BIOGRAF gives you the option of adding H atoms to heteroatoms only. Cytochrome C has many hydrogens, so to decrease the time needed for the calculations we add H atoms only to the heteroatoms.

1. Click the TOP MENU button to return to the top level menu.
2. Pick **build,** then pick **modify H.** The Modify H menu appears as shown in Figure 7.19.
3. Pick **add het H.**

   BIOGRAF selects for optimal hydrogen bonding by pointing donor-hydrogen bonds towards the nearest acceptor. If we intended to have Cytochrome C interact with other molecules, we could tell BIOGRAF to take the other molecules into account when it generates the hydrogen atom positions for Cytochrome C. BIOGRAF would then add the H atoms such that they point towards hydrogen bond acceptors on the other molecules. At the moment, we just have the Cytochrome C molecule, so we need not specify any other molecules.

   While BIOGRAF adds the hydrogen atoms, it prints informative messages in the Modules/Messages window. When it finishes, the number of hydrogens added is listed in the text window.

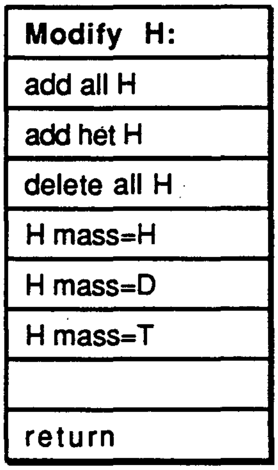


Figure 7.19: Modify H menu

The Cytochrome C groups are automatically updated to include the added hydrogen atoms.

7.8 Creating Groups Based on Distance Criteria

Although at this point we could use molecular mechanics to optimize the structure of the mutant Cytochrome C, the calculation could take several hours (depending on the system you're using). However, the substitution of TYR 82 for PHE 82 primarily affects atoms in the environment local to TYR 82, so it is reasonable to do a shorter calculation that includes only that portion of the molecule. To create such a local group, we use the **distance** operation which is found under **limits** on the **Make groups** menu.

1. Click the TOP MENU button to return to the top level menu.
2. Pick **visualize**, then pick **make groups**.
3. Pick **set up**, choose **add group**, and give it the name ***LOCAL***.
4. Pick **limits**, then **primitives**.
5. Turn on the **MAIN CHAIN, HMAIN, CARBONYL, NONSTD, HNONSTD**, and **LABELS CA** options in the Primitives menu. Leave **SIDE CHAIN** and **HSIDE** on. **RETURN** to the Limits menu.
6. Pick **distance**. BIOGRAF prompts you to pick the groups containing the base atoms and brings up the Groups menu.
7. Pick **TYR 82** from the Groups menu, then pick **return** to terminate the list.
8. Enter *5* Å for the cutoff distance in the dialog box. Verify the entry by clicking the OK button or pressing <Enter>.
9. **Return** to the Make groups menu and **create** the group. A **LOCAL** group is created which contains all primitive structural elements within 5 Å of any atom in the **TYR 82** group.
10. Click the DISPLAY button and turn off all groups except **TYR 82** and **LOCAL** to display the **LOCAL** group more clearly. Pick **return** to complete the operation.

7.9 Saving Information

At this point it is a good idea to save the work you have done so far in the session. This allows you to resume from this point instead of repeating the entire tutorial if you exit from BIOGRAF. This section describes how to save information in *group* files and *coordinate* files.

7.9.1 Group Files

A group is a collection of atoms and bonds chosen in such a manner as to reduce a molecular structure into recognizable parts. There are two types of files that save group information:

Type 1	saves all the sequences of toggles that defined the various groups. When BIOGRAF reads a type 1 group file, it regenerates the groups based on the saved sequences of toggles. This allows the same group file to work for different structures.
Type 2	saves the list of atoms and bonds for each group as well as the sequence of toggles used to define them. A type 2 file regenerates groups faster than type 1 because it reconstructs them directly from the atom lists rather than from the toggles.

Both types of group file contain the display status information that determines the color of the group and whether or not the group is toggled on or off.

Note: Type 2 files should only be used for coordinate files with the same set of atoms and connections. When in doubt, make a type 1 file.

To save group information about the molecule, proceed as follows:

1. Click the TOP MENU button to return to the Top menu, then pick **in-out** to bring up the In-Out menu (see Figure 7.20).

2. Make sure the **by file** option under the write defaults is on, then pick **write**. The prompt *Pick the file type to be output:* is displayed and the File types menu appears in the Menu Pad as shown in Figure 7.21.

3. Pick **group** from the File types menu to let BIOGRAF know that the type of file you want to write contains group information. BIOGRAF brings up a dialog box and asks you to *Enter a name for the groupings file.*

4. Type ***cytc.grp*** (include disk and directory information if needed), then select OK to verify the entry. The *grp* extension indicates that the file is a group file. BIOGRAF prompts *Enter output file type [1 (normal) or 2].*

5. Enter the number *1* to indicate type 1 group information, then pick OK. BIOGRAF then prompts you to enter remarks for the output file.

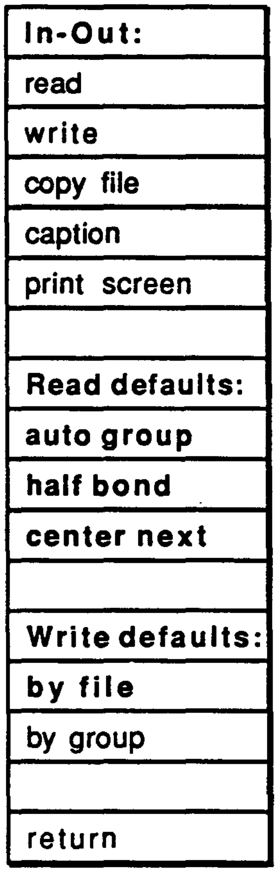


Figure 7.20: In-Out menu

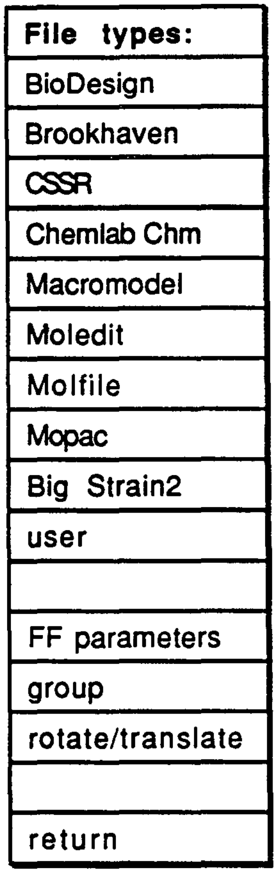


Figure 7.21: File types menu

6. Type in remarks that will identify the file the next time you use it. BIOGRAF will print the remarks in the text window when it reads the file. Terminate the remarks with a blank line (click the OK button or press <Enter>).

7.9.2 Coordinate Files

In order to reconstruct a group, BIOGRAF needs the information about the coordinates of the atoms in that group. The program saves coordinate information in coordinate files. Here we make a coordinate file of the BioDesign General Format type.

1. Pick **write**. The prompt *Pick the file type to be output:* appears and the File types menu shown in Figure 7.21 again displays the file types. Ten different coordinate file formats are available.

2. Pick **BioDesign** to indicate the BioDesign General Format file type. This type is recommended for internal use with BIOGRAF since it contains all pertinent information.

Since there is only one file in the system (Tuna), it is automatically chosen as the source of coordinate information. BIOGRAF now asks you to *Enter a name for the output file.*

3. Type *cytc.bgf* in the dialog box, then verify the entry by clicking OK. The *bgf* extension indicates that the file contains BioDesign General Format type data. BIOGRAF prompts you to enter remarks for the output file.

4. Type in remarks that will identify the file the next time you use it. Terminate the remarks with a blank line (Click OK or press <Enter>).

7.9.3 Testing Our Work

As an exercise, we test the new group and coordinate files by resetting the program and having BIOGRAF read them in again.

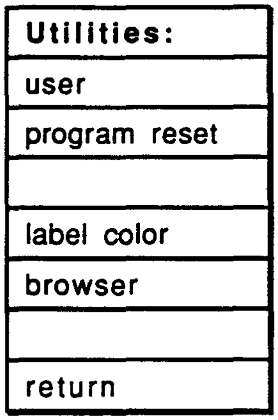


Figure 7.22: Utilities menu

1. **Return** to the Top menu and pick **utilities**. The Utilities menu shown in Figure 7.22 is displayed.

2. Pick **program reset**. An alert box appears asking you to verify the reset.

3. Click the OK button. When BIOGRAF finishes the program reset it no longer remembers anything about the molecule. The Top menu is again displayed.

4. Pick **in-out** to bring up the In-Out menu.

5. Make sure the **auto group** and **half bond** options under the read defaults are toggled on. The **auto group** option treats the molecule as a single group and includes every atom and bond in the file. The **half bond** option uses the half-bond colors to display the molecule. Leave the centering option with **center next** displayed. This will cause the next file that we read in to be centered in the graphics window.

6. Pick **read**. The File types menu appears and we are prompted to *Pick the file type to be input:*

7. Pick **BioDesign**, since we saved our data in a *bgf* file. BIOGRAF brings up the file name entry dialog box and prompts *Choose the file to be input.*

8. Use the browser box to bring up the directory that you saved the files in, then select *cytc.bgf* from the file names listed. Click the OK button to verify the selection. The message *Working* ... is displayed as the coordinate file that we created in the previous section is read in. When finished, the Cytochrome C structure appears in the graphics window.

Next we will read in the group file that we created to test whether it correctly saved the groups.

9. Pick **read** again from the In-Out menu and then pick **group** from the File types menu. BIOGRAF prompts *Enter the name of the groupings file.*
10. Select ***cytc.grp*** from the files listed in the browser box, then click the OK button. The groups that had been toggled on will appear in the graphics window after the group file is read in. Verify that the groups were saved as expected.

7.10 Calculating Energies

We now predict an optimal position for TYR 82 within the Cytochrome C molecule. The best way to regularize the structure of a molecule is to do an energy calculation, which takes all forces on the structure into account. In this section we minimize the energy of the mutant Cytochrome C to find the conformation that corresponds to the nearest minimum of the potential energy surface.

7.10.1 Setting Up an Energy Expression

The first step is to set up an energy expression for the system.

1. **Return** to the Top menu and pick **simulate** to display the Simulate menu shown in Figure 7.23.

   Any atom can be included or excluded from a molecular mechanics calculation, and if included it can be movable or fixed. In setting up the energy expression, we must define those groups which are allowed to move and those which will remain fixed. In order to keep the run short enough for an interactive tutorial, we find the optimal position of **TYR 82** (movable) in the presence of the **LOCAL** group (fixed).

2. Pick **setup eex** to setup the energy expression. BIOGRAF displays the Groups menu and prompts *Pick the groups containing movable atoms:*

3. Pick **TYR 82** to be movable, then pick **return**. BIOGRAF responds with the prompt *Pick the groups containing fixed atoms:*

4. Pick **LOCAL** to be fixed, followed by **return**. We are now asked to *Pick the groups to be updated during the calculation.*

5. Pick **TYR 82** to be updated, then pick **return**. BIOGRAF prints the message *Setting up energy expression* then lists a summary of the energy expression statistics in the text window (see Table 7.3).

6. Click the TEXT button to bring the text window forward so that you can view the energy expression statistics. Reposition or resize the windows if necessary so your views of the structure in the graphics window and the Menu Pad are not obscured.

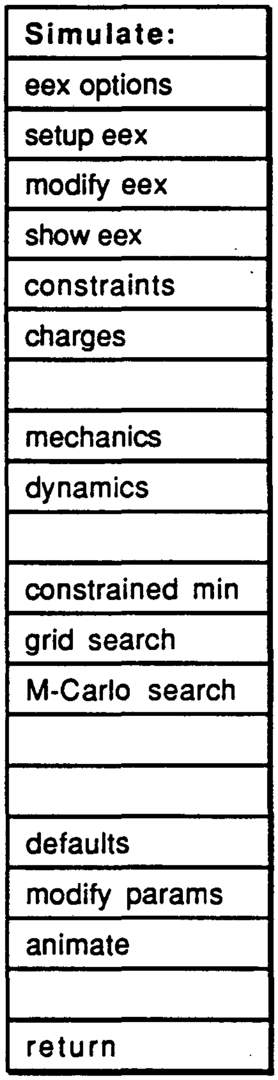


Figure 7.23:
Simulate menu

```
Movable atoms      :  10       Fixed atoms          : 115
Bonds              :  12       Angles               :  19
Torsions           :  28       Scaled torsions      :   0
Pi Twist           :   0
Inversions         :  15       Angle-angle (1C)     :   0
Bond cross angle   :   0
Nbond exclusions   :  30
Hbond acceptors    :  29       Hbond donors         :  10
User energy terms  :   0
Total charge       :  -1.000   movable atoms only   :   0.101
```

Table 7.3: Energy expression statistics (in text window)

The statistics may vary somewhat from run to run depending on the exact position of the tyrosine residue relative to the rest of the molecule. However, at this point there should *always* be 10 movable atoms.

7. Pick **show eex.** The movable atoms are indicated by small green crosses and the fixed atoms are marked with small blue crosses. Click anywhere to remove the crosses.

7.10.2 Fixing the Cytochrome C Backbone

The disposition of the atoms can also be changed on an atom-by-atom basis. We fix the position of the α–carbon of TYR 82 so that the entire backbone of the Cytochrome C remains stationary during the calculation.

1. Pick **modify eex** on the Simulate menu to display the Modify eex menu shown in Figure 7.24.
2. Pick **fix single** to fix single atoms. BIOGRAF outlines the graphics window in red and prompts you to pick the atoms you want fixed: *Toggle atoms to be included in the list. Pick off the molecule to terminate. Pick EXECUTE to execute function or RETURN to abort function.*
3. Pick the C_α atom on TYR 82. A red cross marks the picked atom.
4. Pick off the molecule (a non-atom) to terminate atom selection. The red border is removed and a new menu appears listing options to continue, execute or return.
5. Pick **execute.** Pick **return** on the Modify eex menu to get back to the Simulate menu.

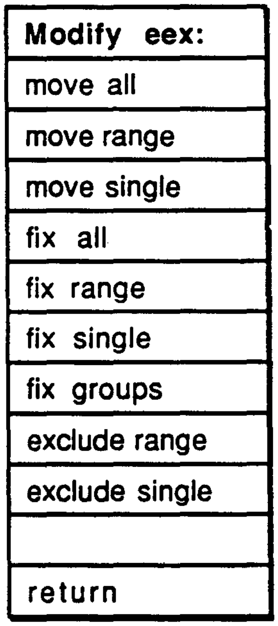


Figure 7.24: Modify eex menu

The program now updates the energy expression with the changes made under **modify eex**. The energy expression statistics again appear in the text window. Note that the number of movable atoms has changed to 9, while the number of fixed atoms has increased by one.

6. Pick **show eex**. The cross on TYR 82 C_α is now blue, indicating that it can no longer move during the calculation.
7. Zoom in and out to make certain that there are no green crosses in the molecule other than those on the side chain portion of the tyrosine residue. Zoom in until **TYR 82** fills the graphics window. Click anywhere to remove the crosses from the display.

7.10.3 Partitioning the Energy

It is now possible to look at the energies of the Cytochrome C system.

1. Pick **Mechanics** to display the Mechanics menu shown in Figure 7.25.

Each time BIOGRAF enters the Mechanics menu, it recalculates the nonbond and hydrogen bond potentials and displays the Energy results menu as shown in Figure 7.26. The values are updated for each step of the calculations.

The items listed include the current total energy, the change in the total energy from the previous step (delta energy), and the root mean square force which is the derivative of the potential energy (RMS force).

The energy is also broken down into its components. Values are listed for internal forces: bond stretching, angle bending, dihedral angle torsions and inversions. Nonbond forces include van der Waals forces, electrostatic forces, and hydrogen bonds. The energy due to constraints and user defined forces are also listed.

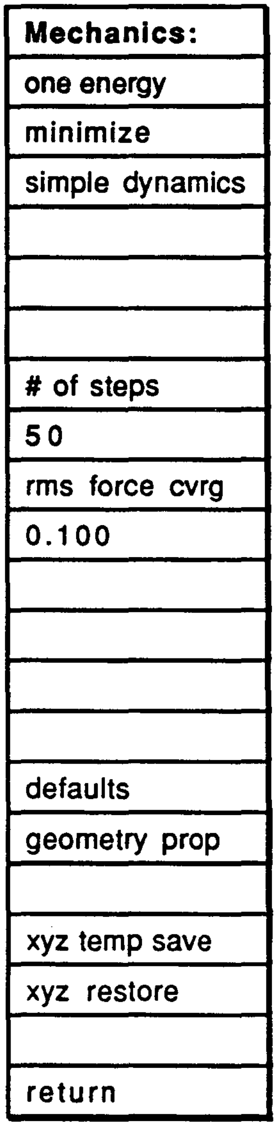


Figure 7.25: Mechanics menu

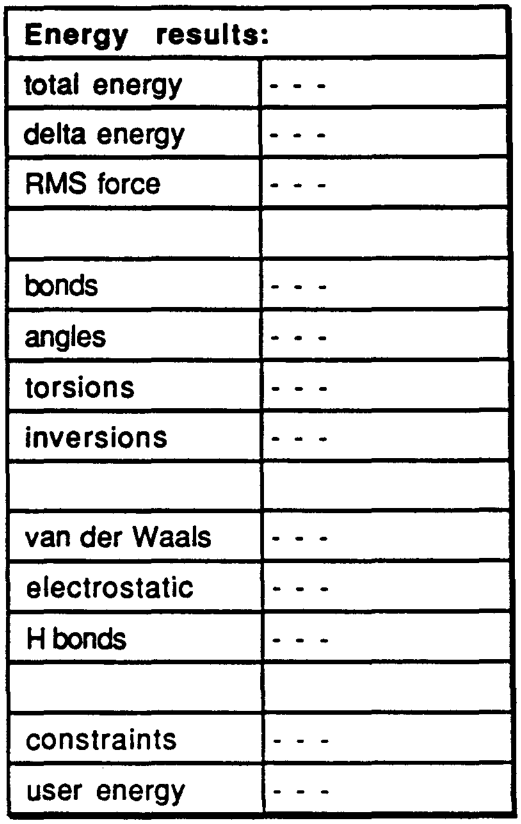


Figure 7.26: Energy results menu

2. Pick **one energy** on the Mechanics menu to calculate the energy of the Cytochrome C molecule in its present conformation.

After BIOGRAF computes the energy, it displays the results in the text window and on the Energy results menu. All energies are in kcal/mol. If an energy is too large to fit into the menu box, the program prints a row of asterisks for it: *********.

3. Make a list of the energies (or asterisks) on the Energy results menu.

   If the van der Waals energy makes a large contribution to the repulsive interactions, the conformation is unfavorable. To determine how much of the van der Waals repulsion arises from the hydroxyl group, we go back and exclude it from the energy calculation.

4. **Return** to the Simulate menu and pick **modify eex.**

5. Pick **exclude single**, then pick the oxygen and hydrogen atoms on the hydroxyl group of the tyrosine. Pick off the molecule to terminate.

6. Pick **execute** from the menu to execute the function, then **return** to the Simulate menu. The energy expression is automatically updated so that the excluded atoms are no longer movable.

7. Pick **show eex**. The picked atoms are no longer marked with green crosses. Pick anywhere to remove the crosses from the display.

8. Pick **mechanics** on the Simulate menu, then pick **one energy** on the Mechanics menu to calculate the new energy.

9. Compare the new results with the list you made before; note the differences in the energies.

10. **Return** to the Simulate menu and restore the original energy expression by using the **move single** operation under **modify eex** to make the oxygen and hydrogen atoms movable again.

7.10.4 Minimizing the Energy

We are now ready to minimize the energy of the substituted tyrosine residue in the field of the local group. By default, BIOGRAF runs the calculation for 50 steps.

The energies at each step appear in both the text window and on the Energy results menu, and the graphics window shows the corresponding conformation of the molecule.

1. **Return** to the Simulate menu again and pick **mechanics** to bring up the Mechanics menu.

2. Pick **minimize** on the Mechanics menu. BIOGRAF prints the message *Minimizing...* in the Modules/Messages window. The energies should show a rapid decrease early in the calculation.

3. When the minimization finishes, check **delta energy**. If it is non-negligible compared to the other energies, then the calculation has not yet converged.

4. If **delta energy** is still comparatively large, toggle **# of steps**. BIOGRAF then prompts you for the number of steps.

5. Type in the number of steps you think are needed to complete the minimization. Select the OK button to complete the entry.

6. Repeat steps 2-5 above as many times as necessary until you are satisfied that the calculation has converged. Pressing the abort key (F12, F10 on the Titan) aborts any mechanics calculation after the next complete step.

7. Make a list of the energies on the Energy results menu and compare it to the previous list. The energies should show a significant decrease.

7.11 Doing a Conformational Energy Search

Energy minimization yields a favorable conformation for TYR 82. However, large flexible molecules such as proteins have many favorable conformations, each of which arises from a *local minimum* in the potential energy surface. One way to look for other good conformations is to use molecular dynamics as discussed in Chapter 6. In this chapter, we do a conformational energy search by varying key dihedral angles of the system.

A complete search of all the conformations available to the mutant Cytochrome C would take a prohibitive amount of time. Each bond rotates 360°, so if we look at a conformation for every 10° of rotation, there are 36 calculations per bond. Even if we considered a much smaller molecule with just 15 freely rotating bonds, it would still have 36^{15} possible conformations. If it takes one second to calculate the energy of a conformation, then it takes 36^{15} seconds, or 7000 trillion years, to do the calculation.

Fortunately, the portion of conformational space most affected by the side chain replacement is relatively small. Changes in the conformation of TYR 82 relative to the rest of Cytochrome C are made by rotating the residue about its $C\alpha$-$C\beta$ and $C\beta$-$C\gamma$ bonds. To examine the conformational space spanned by these two dihedral angles we use the **grid search** function under the Simulate menu. **Grid search** will display up to 100 of the best conformations obtained from either a systematic or random variation of the dihedral angles. **Grid search** can also do a specified number of steps of minimization for each set of dihedral angles. Although the minimizations yield a better picture of the relative energies by allowing the conformations to relax, minimizing a large number of structures is too time consuming for an interactive calculation, so we won't be doing any minimizations here.

First, we'll create a group to hold the results of the search, setup the energy expression with this group as the only movable group and again fix the $C\alpha$ so that the backbone of the molecule remains stationary. Then we'll set up the search so that the two dihedral angles are systematically varied in 20° increments, with the first one rotating about the $C\alpha$-$C\beta$ bond and spanning the whole range (360°) and the second dihedral rotating about the $C\beta$-$C\gamma$ bond and spanning 180° (since the structure being rotated is symmetrical). We'll also specify that the ten best conformations be saved (so that they may be viewed with the **best** option) and that the coordinates of the structures generated be saved in a trajectory file so that we can animate them later.

7.11.1 Creating a Search Group and Setting up the Energy Expression

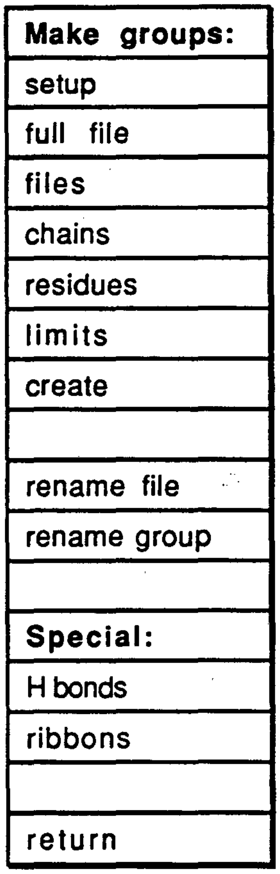


Figure 7.27: Make groups menu

1. Click the DISPLAY button, toggle off **TYR 82**, then pick **return.**
2. Click the TOP MENU button to bring up the Top menu, pick **visualize,** then pick **make groups.** The Make groups menu is displayed (see Figure 7.27).
3. Pick **setup**, then select **add group** from the Groups menu.
4. Enter ***SEARCH*** as the name of the new group in the dialog box. Select the OK button to complete the entry.
5. Leave **files, chains** and **residues** at their current settings.
6. Pick **limits**, then **primitives**. Choose **SIDECHAIN** and **HSIDE** from the Primitives menu which comes up, then pick **RETURN** to get back to the Limits menu.
7. Pick **key one res** and enter *82* in the dialog box to define the TYR 82 residue. Complete the entry by clicking the OK button.
8. **Return** to the Make groups menu and pick **create.** The **SEARCH** group created appears in the graphics window. It will be used to display the results of the conformational search.
9. Click the DISPLAY button and toggle the **SEARCH** group off and on. Complete by picking **return.**
10. Click the TOP MENU button and pick **simulate.**
11. Pick **setup eex** and set up the energy expression. Select **SEARCH** as the only movable group, **LOCAL** as the only fixed group, and **SEARCH** as the only group to be updated.
12. Pick **modify eex**, then **fix single** and fix the C_α atom of TYR 82 as you did before.
13. **Return** to the Simulate menu.

7.11.2 Defining the Search Space

1. Pick **grid search** to display the menu shown in Figure 7.28.

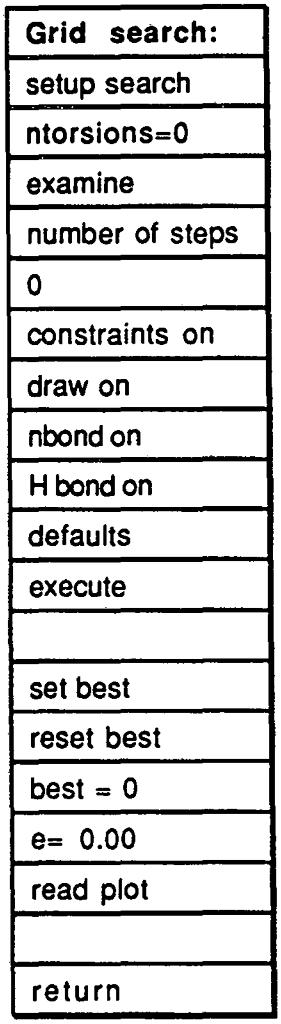


Figure 7.28: Grid search menu

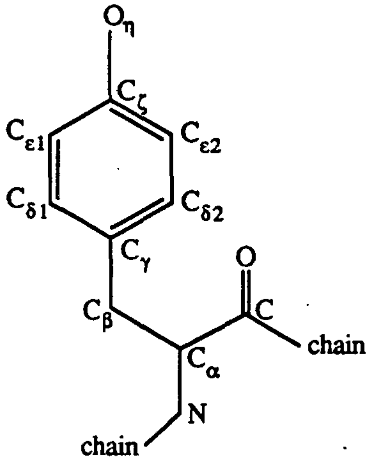
2. Pick **setup search.** A dialog box comes up and prompts you to enter the number of best conformations to be saved.

3. Enter ***10***, then click the OK button. BIOGRAF now prompts you for the method to be used in generating the conformations and displays a menu with two options, **Sequential** and **M-Carlo** (Monte Carlo). The **Sequential** method generates the conformations sequentially within a specified range of angles, while the **M-Carlo** method generates random angles within the whole range (0° to 360°).

4. Pick **Sequential.** BIOGRAF prompts you to define the dihedral angle to be varied by successively picking the first, second, third and last atoms. They must be picked in the proper order; the second and third atoms specify the central bond.

5. Define the dihedral angle about $C\alpha$-$C\beta$ by picking the following atoms (see diagram):
   - either of the two atoms next to the $C\alpha$ in the main chain (carbonyl carbon or N),
   - $C\alpha$,
   - $C\beta$, and
   - $C\gamma$.

   The first three atoms picked are successively marked with a red, yellow, and blue cross to indicate the order selected. If you make a mistake, pick off the display and start again with **setup search.**



Labeling of atoms on tyrosine

6. Enter ***0*** for the starting angle, ***360*** for the stopping angle, and ***18*** for the number of divisions in response to the prompts. This corresponds to doing a calculation every 20°.

7. Define the next dihedral angle about $C\beta$-$C\gamma$ by picking:
   - $C\alpha$,
   - $C\beta$,
   - $C\gamma$, and
   - either $C\delta_1$ or $C\delta_2$.

8. Enter ***0*** for the starting angle, ***180*** for the stopping angle, and ***9*** for the number of divisions (again giving an increment of 20°).

9. Pick off the molecule to terminate the operation. The number of dihedrals defined (2) is indicated in the **ntorsions** box on the Grid search menu.

10. Pick **execute** to initiate the search.

11. Pick **no** in response to the prompt to create a plot file.

12. Pick **yes** to save the generated structures as a trajectory (*trj*) file and give ***tyr.trj*** as the file name. Be sure to include appropriate directory specifications if not saving the file in you working directory.

13. Type in remarks that will help identify the file when you use it. Terminate the remarks with a blank line (click OK or press <Enter>).

    BIOGRAF now runs the conformational search. If the **draw on/draw off** toggle is set to **draw on** (the default), then each of the 190 conformations searched will appear in the graphics window. The energies for each of these are also displayed in both the Energy results menu and in the text window. All energies are in kcal/mol. If an energy is too large to fit into the display, BIOGRAF prints a row of asterisks: *********.

    Depending on your system, it may take several minutes to calculate all the conformations.

7.12 Displaying the Search Results

There are three ways to view the results of the conformational search: displaying the best conformations (10 in this case), animating the trajectory generated by the search, and displaying the potential energy surface. We'll use the first two methods here.

7.12.1 The Best Conformations Method

The easiest way to view the results of the search is to look at each of the best conformations individually.

1. Toggle **best=0**. The box cycles to **best=1** and the display shows the best conformation. The energy of the conformation appears in the box below **best=1**.

2. Toggle **best** repeatedly and observe that the program cycles through the 10 best conformations. Toggling **reset best** resets **SEARCH** to the original optimized conformation, while toggling **set best** lets you select which of the 10 best conformations you want to display.

7.12.2 Rendering a Group in Monochrome

We'll be animating the conformations from our conformational search in the next section. When animating, we can choose to color-code the conformations according to sequence number or energy, thus making it easier to see how the structure moved in space during the calculations. Color-coding requires that the group being animated is displayed in monochrome. However, our **SEARCH** group was created using the half-bond representation (each atom is colored according to its element and the half of the bond closest to the atom is made the same color as the atom). Before animating we must therefore render the **SEARCH** group in a single color. This is done using the **render groups** operation on the **visualize** menu.

1. Click the TOP MENU button to return to the top level menu, then pick **visualize.** The Visualize menu shown in Figure 7.29 is displayed.

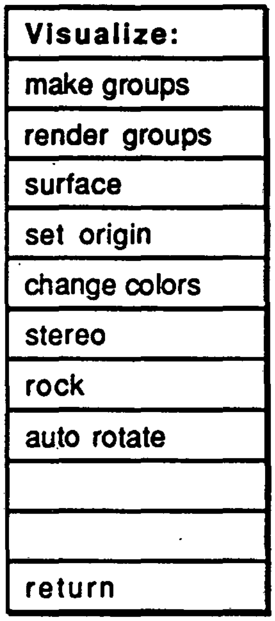


Figure 7.29: Visualize menu

2. Pick **render groups** to bring up the menu shown in Figure 7.30. The Groups menu is displayed and we are prompted to pick the group to be rendered.

3. Pick **SEARCH**, then pick **return.**

4. Make sure **half bond** is toggled off.

5. Pick **execute.**

   The leuenkephalin molecule is redisplayed in a single color. The color of the group could be changed if desired using the **change colors** option on the Visualize menu. We won't we doing that here, however.

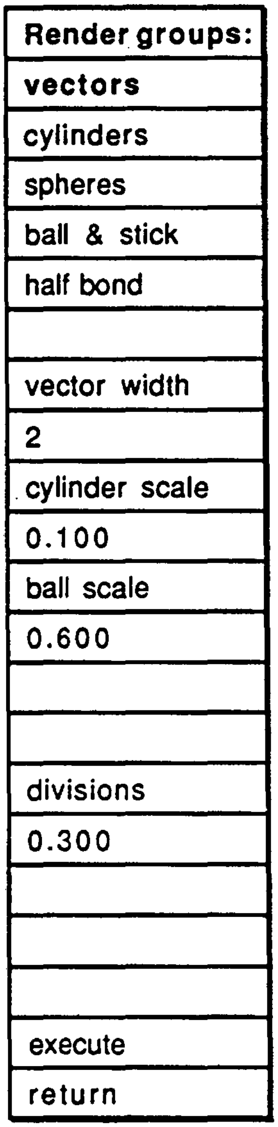


Figure 7.30: Render groups menu

7.12.3 Animating a Conformational Energy Surface

At the onset of the search, we told BIOGRAF to save all of the conformations in a trajectory file. Those conformations can be read in to animate the results of the calculation.

A) Defining the Trajectory File and Listing the Conformers

1. **Return** to the Simulate menu, then pick **animate** to display the Animate menu shown in Figure 7.31.

2. Pick **define traject** to define the trajectory file to be animated. A browser box comes up and BIOGRAF prompts *Enter the name of the trajectory file:*

3. Locate the directory in which you saved the trajectory file, then select the ***tyr.trj*** file name from the listings in the browser box. Click OK to complete the selection.

   When the file has been read in, the sequence number for the first conformer in the file is displayed under the **start** option on the menu; the sequence number for the last conformer is similarly shown under the **stop** option. **Start** and **stop** are options which can be used to restrict the sequence of conformations to be included in the animation. We want to look at the whole range of conformations, so we'll leave these values as they are.

4. Pick **list** to display the List menu shown in Figure 7.32.

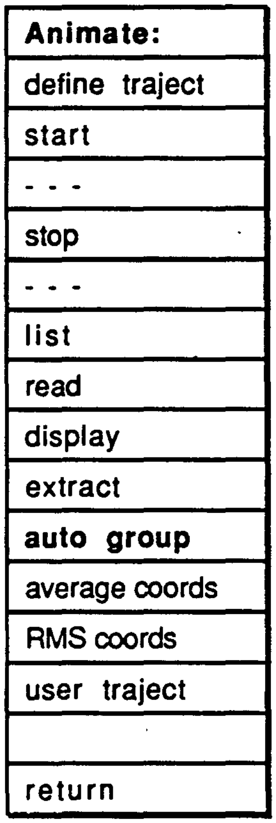


Figure 7.31: Animate menu

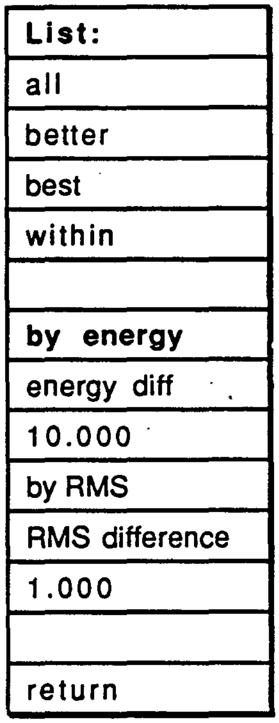


Figure 7.32: List menu

The conformers may be listed in two modes: using their associated energy values to make comparisons (**by energy**), or using the root mean square (RMS) values obtained from RMS matching to a reference conformation which is specified by the user (**by RMS**). We'll use **by energy** mode. This is the default, so it should be highlighted. If not, toggle it on.

5. Pick **all** to obtain a listing of all the conformers in the text window. Each conformation number is listed with its energy value. A listing of the best conformer (the one with the lowest energy value) can be obtained by picking **best** on the menu. **Better** lists the conformers that are consecutively better than the best one so far and **within** lists those that are within a specified energy difference (**energy diff**) from the best.

6. Pick **better**, **best** and **within** and note the listings obtained with each option. Note the conformer with the best (lowest) energy value. Note the number of conformers that are within **energy diff** (10 kcal) of the best.

7. Pick **energy diff**, enter ***1000*** for the energy difference and pick **within** again. Note the number of conformers that are now within 1000 kcal of the best. Continue to change **energy diff** until you obtain about 50 conformers within **energy diff** of the best.

B) Reading in the Positions of the Low Energy Conformers

We don't need to animate all the conformers. Instead we will restrict the conformations that we look at to the lower energy ones, that is, to those that fall within a specified energy difference (**energy diff**) from the best. To do this, we must change **energy diff** to the desired cutoff, read in the positions of the low energy conformers using the **within** option, then **animate**. From our listings in step 7 above, we found an **energy diff** value that produced about 50 conformers. This is a reasonable number to animate, so we'll set **energy diff** to this value.

1. **Return** to the Animate menu.

2. Pick **read** (a **read** operation must be performed before any animations can be done). A group to contain the trajectory file coordinates must be specified. BIOGRAF prompts *Pick the group to be animated:*

3. Pick the **SEARCH** group from the Groups menu. Terminate the selection by picking **return**. We want to color-code our conformers by energy when we animate them; this requires that the group being animated be displayed in monochrome. The **SEARCH** group is therefore chosen here rather than **TYR 82** since the **SEARCH** group is displayed in monochrome, while **TYR 82** is half-bonded. The Read menu will then appear as shown in Figure 7.31.

   As with the **list** operations, the conformers may be read in **by energy** or **by RMS** mode (we'll leave it in the default **by energy** mode); similarly, **all** the conformations may be read in or they may be restricted to those specified using the **better** or **within** options.

4. Make sure **energy diff** is a value that will produce about 50 conformers within **energy diff** of the best (see listings, step 7 above).

5. Pick **within**. Only the lower energy conformations are read in. Each of the conformers is listed in the text window along with its associated energy value.

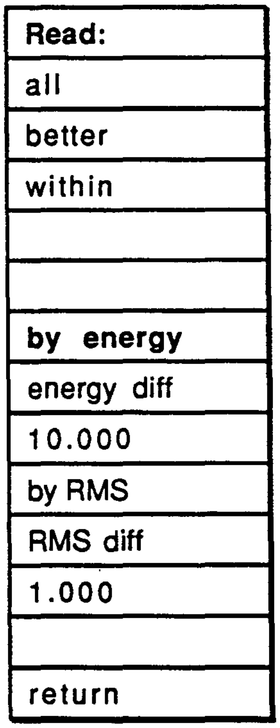


Figure 7.33: Read menu

6. **Return** to the Animate menu.

C) Animating the Conformations

Now that we have read in the positions of the conformers, we are ready to animate them.

1. Pick **display** to bring up the Display menu shown in Figure 7.34.

   The conformations may be colored by sequence number, by energy, or by RMS value. Conformations that were read in by energy mode are automatically color-coded by energy, so **color by energy** appears highlighted on the menu.

   The conformation with the lowest energy (the best) will be red, and the conformation with the highest energy will be blue. The conformations in between will be interpolated between red and blue. **Energy diff** is automatically set so that the conformers that were read in span the whole range of available colors.

2. Pick **animate**. BIOGRAF displays the first conformation and prompts:

   > *To animate the sequence, drag the cursor with the control key and the middle mouse button depressed. Click the left mouse button to stop.*

3. Use the mouse to animate the sequences as described in the prompt (if using a dials unit, rotate dial 8 to cycle through the conformations). The current conformation number and its energy value are displayed in the lower left corner of the graphics window (labeled C and E respectively). The higher energies of some of the conformations may be due to bad van der Waals (steric) interactions.
4. Find the conformation with the lowest energy (red).
5. Click the left mouse button to stop the animation.

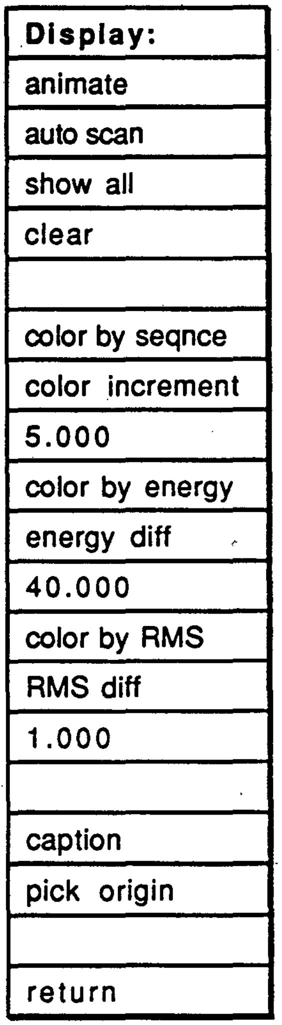


Figure 7.34:
Display menu

D) Using Auto Scan

We can also scan through the conformations automatically.

1. Pick **auto scan**. BIOGRAF prompts.

   Click anywhere to start the auto scanning.
   Use control-middle mouse button to adjust
   the speed (left is slower, right is faster).
   Click in the graphics window to stop.

2. Click anywhere to start scanning. Adjust the scan speed if desired (if you have a dials unit, rotate dial 8 to change the scan speed; rotate it clockwise to scan faster).

3. Click in the graphics window to stop.

E) Showing All the Conformations at Once

All the conformations may be displayed simultaneously (on top of each other) as well. This is useful in illustrating the degree to which different parts of the molecule move or remain stationary.

1. Pick **show all**. A dialog box comes up and we are successively prompted to enter values for the first and last conformation and for the increment.

2. Enter *1* for the first conformation to be displayed, *999* for the last and *1* for the increment (entering a much higher number for the last makes sure we get them all).

   BIOGRAF prints the message *Working...*, then all the conformations that we read in above are displayed at once.

3. Note the best conformation displayed in red.

4. Pick **clear** to clear the display.

5. To make things a little clearer, repeat steps 1 and 2, but this time use *2* for the increment This will give us a total of about 25 conformers instead of 50 (see left side of Figure 7.35). The best conformation is also shown at the right of Figure 7.35 for comparison.

   Note: Your conformations may be slightly different since the minimized structure you began with (prior to doing the conformational search) was probably different.

6. Pick **clear** to clear the display.

7. Pick **return** to return to the Animate menu.

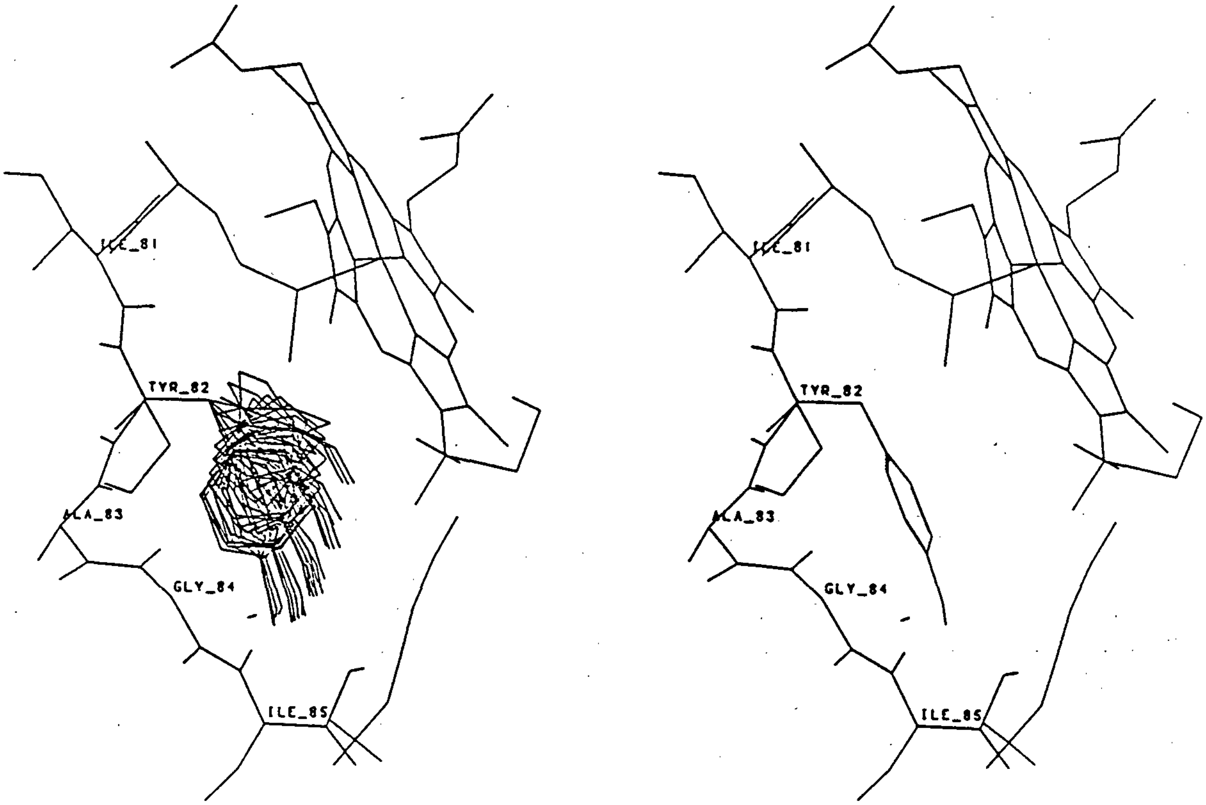


Figure 7.35 Left: Low energy tyrosine conformations displayed using the **show all** option; every 2nd conformation is displayed.
Right: Best conformation (lowest energy)(**extract** option).

F) Extracting the Best Conformer

BIOGRAF allows you to extract the information for a particular conformation (*e.g.*, the best) and create a file for it so that energy minimizations or other analyses may then be performed on it. If the trajectory file was created using conformational search calculations, the conformation to be extracted is specified by giving its sequence number. We can get the sequence number for the best conformation using the list operations described earlier.

1. Pick **list** to display the List menu.
2. Pick **best** to list the best conformer. Note the sequence number of the best conformer.
3. **Return** to the Animate menu, and make sure the **auto group** option is toggled on. This will create a group for the extracted conformer.
4. Pick **extract**. BIOGRAF prompts *Enter the number of the conformer to be extracted.*
5. Enter the number of the best conformer in the dialog box, then select OK. BIOGRAF then prompts *Pick the group to be redrawn using the extracted conformation.*
6. Pick **SEARCH** from the Groups menu. A new file, named "CONFn" where n is the sequence number is created (this file contains all the atoms in the molecule). The best conformation (similar to that shown on the right side of Figure 7.35) will be displayed, and a new group, also named "CONFn" will be created.

   It might be a good idea to save the coordinates of the best conformation in a *bgf* file so that you can leave BIOGRAF if you want and then resume again later where you left off. (In order to reconstruct a group, BIOGRAF needs the information about the coordinates of the atoms in that group.)

7. Click the TOP MENU button to get back to the Top menu.

   Warning: Once the Animate menu is exited, all the set up data is lost. You must redefine the trajectory file and read in the conformations again before any more animations can be done.

8. Make a BioDesign General Format type coordinate file as we did in section 7.9.2, but choose CONFn as the file descriptor for which you want to output information, and name the new file ***CONFn.bgf***.

7.13 Conclusion

This completes our modeling study of the replacement of the PHE 82 residue on Cytochrome C with a tyrosine residue. The results of the BIOGRAF calculations indicate that Cytochrome C accomodates the tyrosine residue well and suggest that its presence causes negligible changes in the tertiary structure and activity of Cytochrome C. Further confirmation can be obtained by doing a calculation that allows the atoms in the LOCAL group or the entire protein to relax.

See the *BIOGRAF Reference Manual* for more detailed explanations of the animate operations.

7.14 Advanced Notes

BIOGRAF recognizes the standard files from the Brookhaven Protein Data Bank and the Cambridge crystallographic data base, as well as CHEMLAB CHM, Macromodel, Moledit, CHEMLAB MOL, and MOPAC files. The program also has its own internal file format which is similar to Brookhaven but contains additional force field information. See the *BIOGRAF Reference Manual* for further details.

This concludes Chapter 7. To leave BIOGRAF, pick **exit** from the Top menu and then confirm by picking OK.

Appendix A

Details of Molecular Dynamics

In molecular dynamics, Newton's equations are used to describe how the *acceleration* of the atoms in a molecular structure changes with respect to time *t*, where acceleration is the change of velocity with time *t*. To find the change of acceleration for particle *s*, BIOGRAF solves the equation

$$\frac{\delta v_{s_j}}{\delta t} = \frac{1}{M_s} F_{s_j} = -\frac{1}{M_s} \frac{\delta E}{\delta R_{s_j}} \qquad \text{(A.1)}$$

where E is the total energy. The subscript j refers to x, y, or z, the components of the position vector for particle s. The symbol v_{sj} refers to the j-th component of the velocity of particle s, F_{sj} to the j-th component of the force, R_{sj} to the magnitude of the j-th component of the position vector, and M_s to the mass.

The 3N values of R_{sj} (there are N atoms, and each atom has an x, y, and z component of its position vector) define the starting geometry or *conformation* of the system. Given any starting geometry, BIOGRAF evaluates the 3N forces given by $\delta E/\delta R_{sj}$. The program then solves equation A.1 to find the change in velocity and position at regular time increments for *every* atom in the molecular structure. BIOGRAF thus determines how the geometry of the structure changes with time. This is molecular dynamics.

An important concept in molecular dynamics is that of temperature. For a system of N particles in equilibrium with a heat reservoir at temperature T, the velocities of the particles satisfy the relationship

$$\sum_{s_j} \frac{1}{2} M_s \left\langle v_{s_j}^2 \right\rangle = \frac{3}{2} NkT \qquad \text{(A.2)}$$

where T is the temperature, k is the Boltzman constant, N is the number of moles, and the brackets denote an average. The Maxwell-Boltzman distribution gives the distribution of velocities. When

BIOGRAF starts a calculation at specified temperature T_i, it generates an initial velocity for each particle such that the overall distribution is Maxwell-Boltzman at temperature T_i. As the dynamics proceeds, velocity fluctuates and the instantaneous values for

$$\sum_{s_j} \left(\frac{M_s}{3k}\right) v_{s_j}^2$$

differ from T_i. The total energy remains the same, so that when the system is at a low energy (*i.e.*, a favorable conformation) the value of T in equation A.2 is high, and when the system is at a high energy (*i.e.*, an unfavorable conformation) the value of T in equation A.2 is low. These instantaneous values of T are not the true temperature, which is defined as an average over time. BIOGRAF calculates the temperature by averaging equation A.2 over a number of steps in the dynamics calculation (*e.g.*, 50 steps).