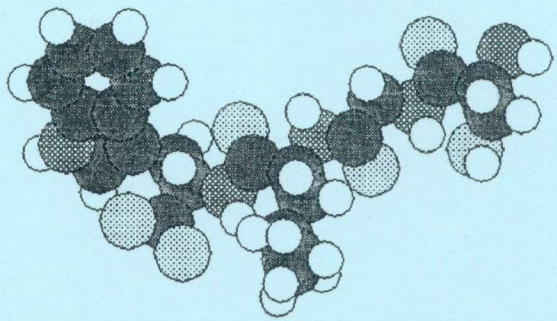
Molgraph

Dr. M Forster

ST Club

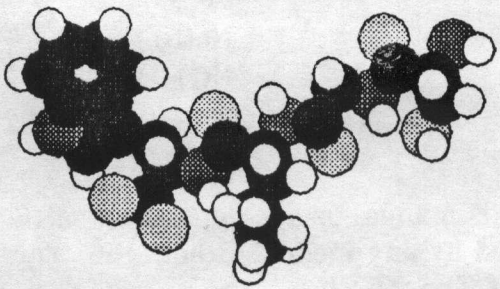
Molgraph



A Molecular Graphics Program
for the Atari ST

Dr. M J Forster

Molgraph



A Molecular Graphics Program
for the Atari ST

Dr. M J Forster

First Printing January 1990

Published by:
The ST Club
49 Stoney Street
Nottingham
NG1 1LX
Tel: (0602) 410241

Manual typeset on an Atari Mega ST computer and Atari SLM804 laser printer with Timeworks Desktop Publisher.

Table of Contents

1) Introduction

Molecular graphics is the use of computer graphics to display and manipulate three dimensional representations of the chemical structures of molecules. Such systems are of increasing use in the fields of chemistry, biology, biochemistry and drug design. However, molecular graphics programs are normally only available for large and expensive minicomputers and workstations.

MOLGRAPH represents an attempt to develop a viable system for displaying, rotating and scaling molecular models on the ATARI ST range of microcomputers. The program is GEM driven making use of drop-down menus and instruction boxes to guide the user through the operation of the program. It is surprisingly simple to use but allows the structure of even very large and complicated molecules to be handled. MOLGRAPH works in the high resolution monochrome and low resolution (16 colour) ATARI ST screen modes.

A molecule is a stable structure consisting of a number of atoms held together by chemical forces called bonds. An atom is the smallest individual unit of an element (examples of elements are hydrogen, carbon, oxygen, phosphorus, nitrogen etc). Please forgive me if you find this too trivial! Thus a molecule typically consists of a few dozen to several thousand carbon, hydrogen, oxygen and other atoms bonded together. Atoms are small having a radius of around 1E-10 metres (a unit called 1 Angstrom), while molecules can be several dozen times larger. Chemical structures can be written in a two-dimensional form on paper by

lines (representing the bonds) joining letters (representing the atoms) where different atoms are denoted by different letters. This representation is extremely useful and has served many generations of chemists but does not allow a true feeling or appreciation of the three dimensional nature of molecules.

Hence a molecular graphics program will be an invaluable aid to those wishing to visualise molecular structures, and those with an interest in the subject areas mentioned above. The latest release of MOLGRAPH allows the construction of molecular structures from fragments of previously saved structures.

2) How To Start Molgraph

The program is supplied as an executable file MOLGRAPH.PRG, and double clicking on this icon will start the program. After entering the desired information in the startup window it is necessary to select a file containing the molecular structure from those available on the file selector; valid files have a .CRD file extension. If you are a new user then you should answer no to the question "use build option YES/NO" since this is for advanced use only (see page 11).

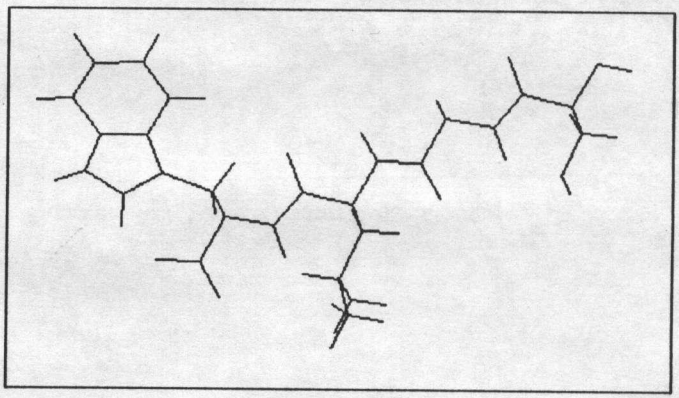
Certain information is displayed as the structure is loaded from disk and when this is complete the screen will clear, the work screen with menu bar will appear, and the selected structure will be drawn. The utility program PDB_CONV reads atomic coordinate data from an ASCII file and converts this into a .CRD file that can be read by the MOLGRAPH program. The ASCII file format used is the PDB (Brookhaven Protein Data Bank) format.

If ASCII files in this format can be downloaded to the ST using KERMIT or a similar program, then any structure thus obtained can be converted for display by MOLGRAPH. I have access to PDB databank structures, but I am unsure whether the structures are public domain and can be freely copied and passed around. The MOLGRAPH build routine can be used (if you have a knowledge of bond lengths, bond angles etc) to overcome this barrier and construct new molecules from bits of old ones - a kind of molecular LEGO.

3) Types Of Display Used

MOLGRAPH has three different methods of displaying molecules that can be selected:

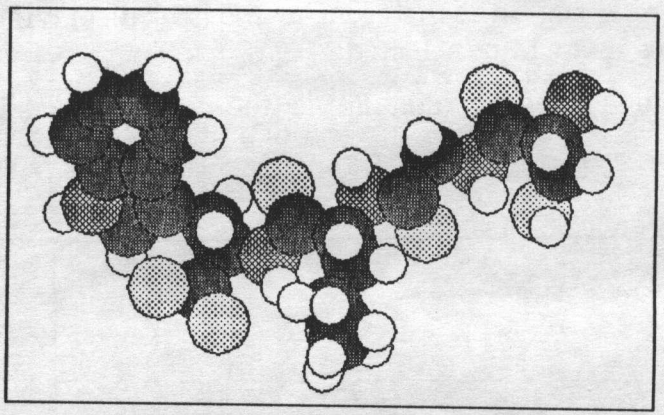
a) The one used on startup is the **vector** model.



Example of Vector Display

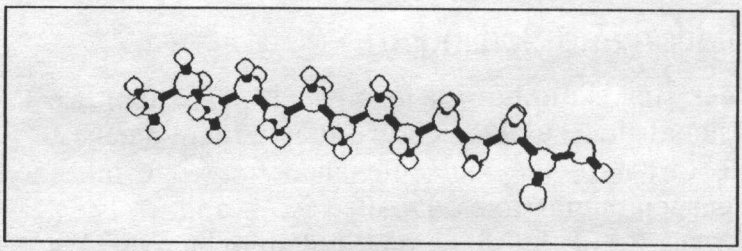
In this model bonds are drawn as straight lines joining the positions of two atoms. Thus the atoms are positioned at the end of a line or at the bend where two bonds join. The vector model is the quickest to draw and is used by default after any rotation of the molecule has taken place. Two other types of display are available from the 'MODEL' menu. These are the sphere and ball-stick models.

b) In the **sphere** model each atom is represented by a circle drawn at the atom position, the circle size being an indication of the different sizes of atoms. Bonds are not drawn in the sphere model.

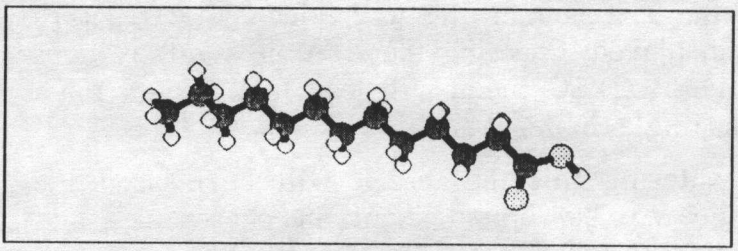


Example of Sphere Display

c) Finally the **ball-stick** displays available from the model menu show atoms as small spheres intersected by short sticks that imply a bond between two atoms. In ball-stick 1 atoms are unshaded or coloured white, while in the sphere and ball-stick 2 displays, the shading or colour used depends upon the type of atom and enables atoms of a given type to be simply identified.



Example of Ball-Stick 1 Display



Example of Ball-Stick 2 Display

The sphere and ball-stick models of large molecules may involve a short wait while the atom coordinates are properly arranged prior to displaying the correct

3D structure. The initial size of the spheres and bonds in the ball-stick display can be reduced by using the b-st radius menu item and entering a scaling factor between 0 and 1. The initial value being 1, lower values allow greater visibility of atoms at the rear of large complex structures. After entering this ball-stick scaling factor choose one of the ball-stick displays.

4) Rotation and Scaling

In order to examine various parts of the three dimensional molecular structure it is advantageous to be able to rotate the molecule about any one of the three perpendicular axes, labelled X, Y and Z. The X and Y axes lie in the plane of the display screen, X is the horizontal axis while Y is the vertical axis. The Z axis is at right angles to both the other axes and hence projects into and out of the plane of the screen. It is along the Z axis that the user views the structure being displayed. Choosing the X, Y or Z rotate items from the 'ROT/SC' menu will lead to a request for a rotation angle (in degrees).

After entering this the screen will be cleared the molecule will be rotated about the chosen axis and re-drawn using the vector model. It is important to remember that the axes remain fixed relative to the screen and are unchanged by any rotation of the molecule. The SCALE item on the menu allows the overall displayed size of the molecule to be altered; the present value is shown and a new value is requested.

The NEW ORIGIN feature allows any atom to be chosen as the coordinate origin, just pick the atom with the cursor. This is very useful for examining large structures where a large SCALE factor may be necessary, and the atom of interest would otherwise be outside the display window. The CENTRE item returns the coordinate origin to the approximate centre of mass of the molecule.

Rotation about chemical bonds can be achieved by using the BOND ROTATE menu item; pick two bonded atoms and enter the required rotation angle. The first picked atom is at the stationary end of the bond, atoms on the side of the second picked atom are those rotated.

NB: Some funny effects can occur if the atoms picked are part of a ring structure - which is not a sensible bond rotation to attempt anyway, but this error is not currently trapped out.

In order to store particular views of a molecule while searching for the better ones the STORE item allows the current coordinates to be stored in a memory buffer. After several rotations, translations etc the stored view can be regained by using the RECALL menu item. SWAP simply exchanges the current and stored coordinates and redisplays the structure.

5) Atom Identification and Structural Data

The PICK ATOM item on the 'MODEL' menu allows individual atoms to be identified. Their atom numbers, atom names C1, H2, O4 etc are displayed along with any group name (e.g. CYT 1 for cytosine 1 in the DNA

structure CGTACG.CRD) they may also possess. This information was loaded from the file selected at program startup.

An atom is picked by placing the mouse on it and clicking the left button. After picking an atom with this option PICK ATOM should again be selected to pick further atoms and display their names. In later versions the x,y,z coordinates of the picked atom are also displayed. This is of great use in deciding how much to translate molecules for use in merging stored structures to build new ones.

The DISTANCE option on the 'MODEL' menu allows two separate atoms to be picked using the mouse and the distance between them to be found. The distance is given in units of Angstroms.

To find a second distance simply choose this menu item once again.

The BOND ANGLE menu option calculates and displays the angle between two connected bonds, the three atoms forming the two bonds are picked using the cursor.

The TORSION option allows four connected atoms to be picked with the cursor. If these are bonded together in an A-B-C-D pattern then the angle calculated and displayed is that between the A-B and C-D bonds when viewed along B-C. Because it is sometimes difficult to keep track of a group of atoms after a rotation operation the MARK ATOM feature has been added. This allows an atom to be picked with the cursor, following this the atom name for any marked atoms are displayed in the vector and the ball-stick displays.

Atoms can have their mark removed by the UNMARK menu item.

Since a simple distance algorithm is used to test for atom-atom connectivities (atoms are considered bonded if they are closer than 1.9 Angstroms) then spurious bonds can sometimes be generated. To remove these a BREAK BOND feature removes the bond between two picked atoms and redraws the display. The converse operation of forming a bond between two picked atoms is performed by the MAKE BOND menu item. This is of special use for forming bonds between the current molecule and a (correctly oriented and positioned) fragment stored on disk as a .CRD file; this is the basis of the BUILD feature of MOLGRAPH. This feature will be explained later and examples of its use in molecular construction are given.

The DEL ATOM allows an atom to be picked with the cursor and then to be deleted from the structure.

6) Methods of Obtaining a Hard Copy

The simplest method of obtaining a printed copy of the display is to choose the SCREEN menu item from the dump menu when the desired view is obtained. This option works at any stage of the program execution. The number of pixels per line can be set by the INSTALL PRINTER accessory - assuming you booted the ST with this accessory present of course.

The COORDS item on this menu allows the set of atom names and coordinates to be listed on screen and optionally on the printer as well. The DEGAS item

dumps the screen to disk in a format compatible with the DEGAS, and DEGAS ELITE design programs. This opens the way for molecular structures to be fully incorporated into graphs, text and other work. The saved screen is not corrupted by the file selector. This feature works in both monochrome and colour modes, and the colour palette is saved correctly together with the screen image. The DEGAS colour images have a .PI1 extension and can be loaded into other art packages such as QUANTUM PAINT. The SNAPSHOT accessory can be used to convert molecular images in DEGAS format into a format (GEM image or .IMG format) compatible with the 1st Word Plus word processing package. Thus MOLGRAPH can be used to incorporate molecular models into chemical and scientific documents. Examples of this are provided on disk.

The MOVIE option rotates the displayed molecule about an axis saving a series of screen images to disk. These can later be prestored in memory to form the frames in a short movie sequence showing molecular rotation.

A separate program called MOLDEMO2 is available to demonstrate this feature, this replays a 10 frame movie using either monochrome or low resolution colour mode. To use this feature select the MOVIE item from the menu, enter a disk with enough space to save the frames (320k for 10 frames) and allow the program to generate the frames and save them to disk. Then run the MOLDEMO2 program and enter the disk of saved screen images when required by the program (note MOLDEMO2 and the frames need to be placed in the same folder). The DEGAS screen images that

constitute the MOVIE frames have filenames PIC.001, PIC.002,... etc. Individual frames will need to be renamed to have either .PI1 (colour) or .PI3 (mono) file extension if they are to be loaded in DEGAS. The FILE item under the dump menu allows MOLGRAPH coordinate files that have been modified by rotations, bond rotations, bond deletions etc to be saved in either PDB or CRD file formats. The current SCALE factor is not saved and should be noted and reset if a desired view of a structure is to be recreated.

7) Building New Molecules

The BUILD feature of MOLGRAPH is an advanced and versatile method of combining together 3 dimensional fragments of molecules to build new molecular structures. A molecular fragment on display is merged with a fragment stored on disk (as a CRD file).

The DEL ATOM routine can be used to deplete a molecule until a desired fragment such as a methyl group is left and this can be stored to disk as a CRD file. The essence of molecule building with MOLGRAPH is to add stored coordinate files to the structure already displayed, make the necessary new bonds thus generating the new (larger) structure.

This feature is only available if you answer yes to the build feature alert box when the program is started. If you do answer yes then sufficient memory must be reserved for you to add new atoms to the existing structure, hence a maximum number of atoms to allow for building must be entered. Having done this the program will still function as normal, but will allow additional CRD files to be merged with the current

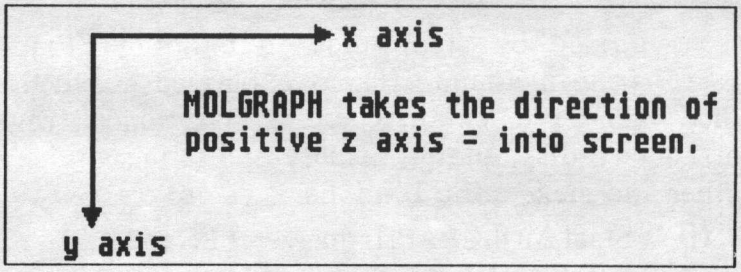
structure until the number of atoms merged in exceeds the maximum number that you allowed for building. Even for merging small fragments I normally allow enough space for about 100 extra atoms - this is not a large number even on a 520 STFM and allowing extra space does not cause any harm.

Naturally the stored fragment and the currently displayed structure must have the correct relative orientation and position, otherwise incorrect bond lengths and angles will result. MOLGRAPH allows precise control of the orientation and position of structures using the X-ALIGN and TRANSLATE menu options respectively.

The function of X-ALIGN is to allow two atoms to be picked, the first is placed at origin and the second is oriented along the positive X axis. Thus the orientation of the structure is fixed to a known reference position. Other orientations can be achieved by the relevant x,y or z axis rotations by 90 or 180 degrees (or any angle you like!).

Having obtained the desired orientation in space, the atoms in the current structure that will not be required in the new structure are deleted using the DEL ATOM menu item.

The TRANSLATE menu item should then be used to define the position of the current structure in space. Enter the desired translation distance in Angstroms, positive values lead to movement along the positive x, y, z axes, negative values lead to movement in the opposite direction. The axis orientations are those defined by GEM.



Having obtained a correctly positioned and oriented fragment in the display the BUILD item is used to select a CRD file for merging. If enough space for the extra atoms is present then the file will be loaded and the two distinct fragments displayed.

The necessary bonds should then be formed using the MAKE BOND item and picking the two desired atoms with the cursor. The new molecular structure should now have been obtained. The MOLGRAPH build routine does require a knowledge of atom-atom bond lengths (and maybe bond angles), it requires careful planning and storing of suitably positioned fragments as CRD files on disk, but it is very flexible and not limited to any particular type of molecular structure. With practice it is quite quick and easy to use, examples of how to build a few simple small organic compounds from currently available fragments are presented in the hope that this will assist you in learning the technique.

a) Building Ethane from Methane.

Here is a step by step guide to building Ethane CH3-CH3 from methane CH4, as explained above the method will be to obtain correctly positioned fragments, one on disk, one in memory and to merge them together and make a C-C bond, here we go.

(i) Start MOLGRAPH, answer YES to build query and allow for 100 extra atoms.

(ii) Set SCALE to 80 and perform an x rotation of 10 degrees to allow you to see all three H atoms.

(iii) Select X-ALIGN, pick a C atom first and any H atom second. Use PICK ATOM to confirm that the C atom is at the coordinate origin (x=0, y=0, z=0).

(iv) Delete the H atom that is along the x axis (ie the right hand one!) using DEL ATOM. Do a y rotation by 180 degrees and using the FILE item save the fragment as a CRD file called METHYL.CRD. This is your stored fragment.

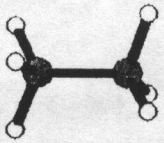
(vi) Perform another y rotation by 180 degrees and translate along x by -1.54 units (angstroms); this distance is the C-C bond distance. This is your displayed fragment

(vii) All that remains is to merge the two together. Select BUILD and pick the fragment METHYL.CRD, the two fragments should be displayed on screen. MAKE BOND between the two C atoms, CENTRE molecule and that is it. Save new structure as ETHANE.CRD.

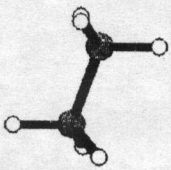
b) Building Propane CH3-CH2-CH3.

If you passed the last hurdle just carry on and build propane.

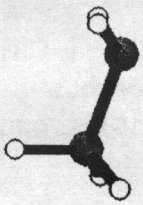
(i) Start with ethane



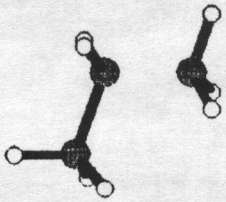
(ii) Using X-ALIGN pick a C atom then one of its bonded H atoms



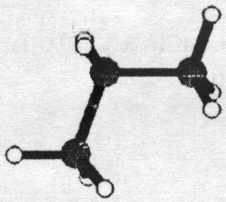
(iii) Delete H atom on axis (x=1.096A), translate x by -1.54



(iv) Select BUILD and choose file METHYL.CRD



(v) Make C-C bond, CENTRE structure and save as PROPANE.CRD using FILE menu.



c) Building Propionic Acid. CH3-CH2-CH2-CO2H.

This is pretty easy:

(i) Start with propane. Then using X-ALIGN pick C first then a bonded H atom.

(ii) Delete the H atom and translate along x by -1.54.

(iii) Merge in the file CARBOXYL.CRD which is a CO2H group with the carbon atom at the origin, make the relevant C-C bond and save your handywork.

A knowledge of bond lengths and bond angles will be invaluable in building sensible molecular structures, bond lengths can be *approximately* taken as the sum of the covalent radii of the two atoms forming the bond. Here are the covalent radii of most of the common atoms you are likely to need in building small organic molecules with MOLGRAPH.

Atom	**Radius (Å)**	**Atom**	**Radius (Å)**
H	0.37	N(i)	0.74
C(i)	0.77	N(ii)	0.65
C(ii)	0.67	P(i)	1.10
C(iii)	0.60	As(i)	1.21
C(ar)	0.695	Sb(i)	1.41
O(i)	0.74	Br	1.14
O(ii)	0.57	I	1.33
S(i)	1.04	F	0.72
S(ii)	0.95	Cl	0.99

Where (i), (ii) and (iii) refer to single, double and triple bonded atoms and (ar) refers to Carbons in aromatic rings.

As an example we expect a C-C single bond length to be 0.77+0.77 or 1.54 Angstroms, the C=O double bond should be 0.77+0.57 or 1.34 Angstroms. Please note these radii cannot account exactly for all bonding situations; for example the radii above lead to a predicted C-H bond length of 1.14Å, whereas the accepted value is nearer to 1.10Å.

I hope these examples provide sufficient explanation to allow you to build further structures. It should be

pretty easy with a knowledge of bond lengths and bond angles to develop small functional groups (carboxyl CO2H and amino NH2 are already on the disk). If you develop structures I hope you will make them available to other MOLGRAPH users. If you are not inclined to try to build up your own structures I hope that a large library of available structures will be developed that you can obtain.