

MacMolecule Image Files © 1990 - 92.
All rights reserved. Molecular Image Database.

Richard B. Hallick
Departments of Biochemistry
University of Arizona
Tucson, Arizona 85721

Portions ©1990-92
(Crystal Lattice Images)
Peter Hardman
Surface Science Centre
Manchester University
Oxford Road, M13 9PL

Portions ©1992
(Small Inorganics, VESPR)
C. Eugene Burchill, Adamant Graphics
Department of Chemistry
The University of Manitoba
Winnipeg, MB
Canada R3T 2N2

Conditions of Use of this Program.

The University of Arizona grants you an non-exclusive license to use this program on the following terms: You may use the library directly or on a computer network. This library may be freely distributed by any means, including bulletin boards, provided (1) that there is no fee or charge for consideration of any kind which is in excess of \$8 (aside from on-line charges) in distributing the software, (2) All documentation and example files are included. However, we retain all the rights to this image library.

You may NOT: Remove any proprietary notices, labels or marks on the library and accompanying documentation.

If you wish to use files from this image library for commercial purposes, they are covered by the University of Arizona copyright, and require prior consent.

The library is provided on an "as is" basis, and without warranty of any kind, including but not limited to any implied warranty of merchantability or fitness for a particular purpose.

In no event shall the authors or the University of Arizona be liable for any direct, indirect, incidental, special, or consequential damages arising from use or distribution of this library. The University of Arizona also shall not be liable for any claim against any user of this library by any third party.

File Format

MacMolecule 1.7 data files have been modified from MacMolecule 1.0 to accommodate information about covalent bond locations, the desired default image form (space-filling, ball-and-stick, wire-frame), and the atom diameters for ball-and-stick models. MacMolecule 1.5 files can be read directly by MacMolecule 1.7, or updated by inclusion of text information for the text pull-down menu. MacMolecule 1.7 will display MacMolecule 1.0 images, but only in space-filling mode.

MacMolecule 1.7 data files are in ASCII text format. They can be created and modified with any text editor or word processor that can save files in text only format. A file for the amino acid glycine is shown as an example.

```
; glycine H+
2
;model
BS
C = 1.5, 0.6 (.5,.5,.5)
O = 1.4, 0.58 (1.,0.,0.)
N = 1.5, 0.6 (0.,0.,1.)
H = 0.90, 0.36 (1.0, 1.0, 1.0)
C1 :    -3.49503  1.79605  2.15052
C11 :    -4.56421  0.74904  1.99995
H1 :    -6.12061  1.89748  2.79283
H2 :    -6.34809  0.23019  2.87950
H21 :    -1.79454  2.28689  1.44555
H3 :    -5.30122  0.99524  3.90571
H31 :    -4.14040 -0.23229  2.19028
H32 :    -4.95760  0.75616  0.99245
N1 :    -5.64483  0.98849  2.95291
O1 :    -3.56196  2.69996  2.94043
O2 :    -2.50337  1.59972  1.31374
;BONDS
C1, C11
C1, O1
C1, O2
C11, H31
C11, H32
C11, N1
H1, N1
H2, N1
H21, O2
H3, N1
glycine
```

The amino acid glycine is the smallest of the 20 common amino acids found in proteins. The color scheme used to designate atoms is as follows: gray = carbon, blue = nitrogen, red = oxygen, white = hydrogen.

Explanation of File Format Line by Line

EXAMPLE 1:

; glycine H⁺

Example 1 is a Comment Line. Comment lines are optional, and are not required for image display. Comment lines begin with a semicolon (;), and are ignored by the program. Comments may also be placed at the end of data lines. They are ignored by MacMolecule 1.7, and may be placed anywhere within the file. It is often useful to include comment lines in image files for large macromolecules. For example, with comment lines for each residue in an enzyme image file, it is possible to create new files showing key features of the enzyme structure, such as active site residues, alpha-helix or beta sheet secondary structure, polypeptide backbone stripped of side chains, etc. We recommend that users include in comment lines the relevant literature or data base citations for the coordinates used to create a new image file. Note that comment lines are not preserved by the 'save as' option of the program.

EXAMPLE 2:

2
;model
BS

Example 2 shows the first 2 required data lines, separated by a comment line. The first data line is a "2" to distinguish a MacMolecule 1.7 or 1.5 file from a MacMolecule 1.0 file. The second data line designates the default molecular model, WF for wire-frame, SF for space-filling, and BS for ball-and-stick. Both data lines are required.

EXAMPLE 3:

C = 1.5, 0.6 (.5,.5,.5)

Example 3 is a definition line for a type of atom that occurs in the image. The four atom data lines define the diameters and color for the atoms used in the image, carbon (C), oxygen (O), nitrogen (N), and hydrogen (H). Each atom used in the image must be in the file header. The first two numbers after the equal (=) sign, separated by a comma, are the user-specified diameters of the atom in Angstroms for the space-filling, and ball-and-stick models, respectively. We have set the ball-and-stick radii at 0.4 of the covalent radii in the image library examples. Other radii may be chosen. The three numbers in parentheses are the (red, green, and blue) color values chosen to represent the atoms on the Macintosh color screen, with numbers ranging from 0 to 1. Note that red is (1, 0, 0), green is (0, 1, 0), blue is (0, 0, 1) and white is (1, 1, 1). Custom colors can be created with fractional mixes of primary colors, e.g., carbon above is gray. Some of the images in the image library have custom colors for substrates or ligands bound to enzymes. (Hint: If you know the custom color you want based on the (red, green, and blue) color values within the 'color' window of the Macintosh control panel under the Apple pull-down menu, divide the values by 65535 to get the corresponding colors in MacMolecule 1.5).

EXAMPLE 4:

C1 : -3.49503 1.79605 2.15052

Example 4 is a data line for atomic coordinates. These data lines are the 3-space coordinates

for each atom in the image file. The first field is the atom designation, followed by a space, a colon, and one or more spaces (or a tab). The three numbers are the absolute X, Y, and Z molecular coordinates in angstroms, separated from one another by one or more spaces (or a tab). There must be a data line with X, Y, and Z coordinates for each atom to be displayed. The total number of atoms and bonds that can be displayed will be dependent on the available memory of the computer.

EXAMPLE 5:

C1, C11

Example 5 is a data line for a covalent bond. The covalent bonds, identified as a list following the X, Y, Z atomic coordinates. "C1, C11" specifies a covalent bond in the wire-frame and ball-and-stick models between the two carbon atoms designated C1 and C11 in the coordinate list.

EXAMPLE 6:

glycine

The amino acid glycine is the smallest of the 20 common amino acids found in proteins. The color scheme used to designate atoms is as follows: gray = carbon, blue = nitrogen, red = oxygen, white = hydrogen.

Example 6 is information to appear in the text window during image display. The first line is the title for the window. The following lines appear in the window itself. Text window titles and information are identified by (option-8, option-8). This information is best placed at the end of the image file. Text window information is optional, and not required for image formation.

Source of Molecular Coordinates

The X, Y, and Z coordinates of atoms in a molecule are normally determined by means of x-ray crystallographic analysis. The structures of several hundred macromolecules, and thousands of small molecules have been solved using this technique. Coordinates are obtained from the primary scientific literature, or from one of two structural Databases. The Cambridge Structural Database (CSD), University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, United Kingdom is the source for small molecule coordinates, and the Brookhaven Protein Data Bank, Chemistry Department, Brookhaven National Laboratory, Upton, New York, USA is the source of large molecule coordinates.

Image Library

This Image Library, which was previously released with MacMolecule 1.5, was designed to give representative examples of space-filling models of small molecules, proteins, and nucleic acids. A few new images are added to the MacMolecule 1.7 release, including the complete structure of a transfer RNA molecule and a zinc finger protein-DNA complex. We have used the following color scheme for the image library: Red is oxygen, gray is carbon, blue is nitrogen, white is hydrogen, and yellow is sulfur. Other atoms are given special colors.

Correspondence with the Authors

Please address correspondence about MacMolecule Image File (c) 1990-1992 to Dr. Richard

B. Hallick, Department of Biochemistry, Bioscience West 524, University of Arizona, Tucson, Arizona, 85721, U.S.A. FAX: 602-621-9288. Email address: Internet, hallick@arizona.edu.

We would appreciate learning about your experiences using MacMolecule, and receiving copies of user-created image files that can be incorporated into the image library distributed with the software. Future updates to MacMolecule will be released by anonymous FTP at joplin.biosci.arizona.edu. User contributed image files will also be made available at this address, and bundled with future releases.

Acknowledgments

The crystal lattice images were created by Peter Hardman, Surface Science Centre, Manchester University, Oxford Road, M13 9PL. (Hardman@uk.ac.mcc.cgu.v2). He has written a program called "Lattice Maker" to automate file creation. Lattice Maker is available by anonymous FTP at sumex-aim.stanford.edu.

Image Revision History

Images Included with MacMolecule 1.7 (not included in this file) (Compatible with Ball-and Stick, Wire-Frame, and Space Filling Imaging)

A Few New Images:

5'-yGMP in yeast phenylalanyl tRNA Phe, biotin, glycine (sample file from 'About MacMolecule' and 'About MacMolecule Image Files', yeast phenylalanyl tRNA Phe complete, Zinc finger-DNA complex, triple nucleotide base pairing from a tRNA, apolipoprotein 3B alpha carbon trace, triose-phosphate isomerase backbone and beta carbons

VSPER Models (Valence Shell Electron Pair Repulsion. Image Files provided courtesy of Dr. C. Eugene Burchill, Univ. of Manitoba, ceburch@ccu.umanitoba.ca):
Electron pair conformations: linear, trigonal planar, tetrahedral, trigonal bipyramid, octahedral

Small Inorganics (Image Files provided courtesy of Dr. C. Eugene Burchill, Univ. of Manitoba):
amide ion, ammonium ion, beryllium chloride, boron trichloride, bromine pentafluoride, carbon dioxide, chlorate ion, HCN, hydronium ion, iodine trichloride, nitrate, nitrite, phosgene, phosphorus pentachloride, phosphorus trichloride, silicon tetrachloride, stannous chloride, sulfate ion, sulfur dichloride, sulfur dioxide, sulfur hexafluoride, sulfur tetrafluoride, triiodide ion, water, xenon difluoride, xenon tetrafluoride

Images Included with MacMolecule 1.5/1.5.1 (Compatible with Ball and Stick, Wire-Frame, and Space Filling Imaging)

Amino Acids:

D-alanine, D-tryptophan, D-tryptophan(H⁺), di L-leucine(H⁺), glycine(H⁺), L-alanine, L-arginine(H⁺), L-asparagine, L-aspartate(H⁺), L-cysteine(H⁺), L-glutamate(H⁺), L-glutamine(H⁺), L-histidine(H⁺), L-isoleucine(H⁺), L-leucine(H⁺), L-lysine(H⁺), L-methionine(H⁺), L-phenylalanine, L-proline(H⁺), L-serine(H⁺), L-threonine, L-tryptophan, L-

tyrosine(H⁺), L-valine(H⁺)

Peptides:

2(L-ala-L-ala-L-ala), L-ala-L-ala-L-ala (in beta sheet conformation), 3 anti-parallel beta sheets, alpha-L-glu-L-glu, gly-tyr, L-ala-gly-gly, L-alanyl glycine, L-alanyl-L-serine, L-cys-L-cys-cyclic disulfide, L-seryl-glycine, L-trp-L-met-L-asp-L-phe-, L-tyr-L-phe (H⁺), L-tyr-L-tyr-L-tyr

Nucleotides:

deoxycytidine 3'-monophosphate, deoxyguanosine 3'-monophosphate, adenosine 5'-monophosphate, cb3717 (anti-cancer, tetrahydrofolate analog), cytosine 5'-monophosphate, deoxyguanosine-deoxycytidine base pair, deoxyguanosine 5'-monophosphate-deoxycytidine 5'-monophosphate base pair, guanosine 5'-monophosphate, GMP-GMP dinucleotide, nicotine adenine dinucleotide (reduced), S-adenosyl homocysteine, uridine 5'-monophosphate, uridyl-3'-vanadate (RNase A inhibitor)

Other Biologicals:

heme group from myoglobin, penicillin G, vitamin A, vitamin B12, chlorophyllide B with hydrogens, chlorophyllide B without hydrogens, alpha-D-glucopyranose, chlorophyllide A

Small Molecules:

oxygen, phosphate anion, picrate anion, water, cyclohexane in chair conformation, diethyl ether, formate anion, methanol, buckminsterfullerene, carbon dioxide, cyclohexane in boat conformation, benzyloxy proline (H⁺)

Ribonuclease Images:

Entire enzyme with active site residues and bound inhibitor, active site cluster, residues 80-86 (beta sheet), residues 98-103 (beta sheet), residues 43-48 (beta sheet), residues 79-87 (beta sheet), residues 80-85 (beta sheet), residues 3-13 (alpha helix), residues 4-12(alpha helix)

DNA/RNA:

d(GGCC)-d(GGCC) 4 base pairs, DNA double helix of 12 base pairs, GCGG RNA tetranucleotide, yeast phenylalanyl tRNA AA stem and CCA, yeast phenylalanyl tRNA complete (not included in early releases of this image library)

Protein Chains:

myoglobin backbone without heme ligand, thymidylate synthase complexed with antifolate analog and substrate, myoglobin backbone with heme ligand, alpha carbon backbone of FAB L1 fragment, alpha carbon trace of myoglobin

Crystal Lattices: (Courtesy of Dr. Peter Hardman)

Bisco Model 2, Stepped TiO₂ (100)

**Images Included with MacMolecule 1.0
(Space Filling Imaging Only; Compatible with Later Versions of Program)**

Amino Acids:

All 20 common amino acids, with and without hydrogens.

Nucleotides:

3',5'-Y-GMPP of yeast tRNA Phe, 5'-CMP (2'-O-methyl), 5'-CMP, 5'-CMP (5-methyl), 5'-GMP, 5'-pseudoUMP, ADP, ATP, dUMP

Organics:

cesium crown ether, cyclohexane boat, cyclohexane chair, helicene

Small Biologicals:

carbon dioxide, chlorophyllide B, cholesterol, dUMP, folate analog, glucose, heme ligand of myoglobin, NAD, oxygen, penicillin G, phosphate anion

Peptides/Proteins:

alpha helix and turn backbone only, H-helix of myoglobin, example of beta sheet with 13 amino acids (with and without side chains), ferredoxin, chick lysozyme, myoglobin helices with heme ligand, oligosaccharide of NAM-NAG-NAM (lysozyme inhibitor), pig insulin (with and without side chains), RNase A helical segments 1-14 and 3-12, RNase A beta sheets, thymidylate synthase beta sheets

DNA/RNA:

(dCGCATATATGCG)₂, (dCGCATATATGCG)₂ with backbone, AU base pair in yeast tRNA Phe, CG base pair in yeast tRNA Phe, dGC base pair in DNA, dTA base pair in DNA, dTG-dAC base pairs stacked, yeast tRNA Phe nucleotides 27-43, yeast tRNA Phe backbone