

KCP-613-5736

Models of the Solvent-Accessible Surface of Biopolymers

Federal Manufacturing & Technologies

R. E. Smith

KCP-613-5736

RECEIVED

OCT 02 1996

OSTI

Published September 1996

Topical Report

Approved for public release; distribution is unlimited.



Prepared Under Contract Number DE-AC04-76-DP00613 for the
United States Department of Energy

AlliedSignal
AEROSPACE

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade names, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

Printed in the United States of America.

This report has been reproduced from the best available copy.

Available to DOE and DOE contractors from the Office of Scientific and Technical Information, P. O. Box 62, Oak Ridge, Tennessee 37831; prices available from (615) 576-8401, FTS 626-8401.

Available to the public from the National Technical Information Service, U. S. Department of Commerce, 5285 Port Royal Rd., Springfield, Virginia 22161.

A prime contractor with the United States
Department of Energy under Contract Number
DE-AC04-76-DP00613.

AlliedSignal Inc.
Federal Manufacturing & Technologies
P. O. Box 419159
Kansas City, Missouri
64141-6159

KCP-613-5736
Distribution Category UC-706

Approved for public release; distribution is unlimited.

MODELS OF THE SOLVENT-ACCESSIBLE SURFACE OF BIOPOLYMERS

R. E. Smith

Published September 1996

Topical Report
R. E. Smith, Project Leader



DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

CONTENTS

<i>Section</i>	<i>Page</i>
Abstract.....	1
Summary	1
Discussion.....	2
Scope and Purpose	2
Prior Work	2
Activity	2
Description of a Solvent-Accessible Surface Model	2
Description of the Calculations.....	3
Stereolithography.....	5
The Biopolymers.....	7
Accomplishments.....	33
Future Work.....	34
References.....	35

ILLUSTRATIONS

<i>Figure</i>		<i>Page</i>
1	Ribbon Model of Ras Turned Off, From Brookhaven PDB File 1q21	11
2	Solvent-Accessible Surfaces of Ras Off (Left) and On (Right), From Brookhaven PDB Files 1q21 and 6q21	13
3	Solvent-Accessible Surface of Hemoglobin, From Brookhaven PDB File 1hga.....	14
4	Computer-Generated Picture of Ribbon Model of Hemoglobin With Space-Filling Model of Heme Groups. Each Color Represents a Different Polypeptide Subunit.....	15
5	Computer-Generated Picture of Space-Filling Model of Hemoglobin	16
6	Solvent-Accessible Surface of Malate Dehydrogenase, From Brookhaven PDB File 2cmd.....	18
7	Computer-Generated Picture of Ribbon Model of Malate Dehydrogenase Monomer With Space-Filling Model of Methionine 1 and Lysine 312	19
8	Computer-Generated Picture of Space-Filling Model of Malate Dehydrogenase Monomer.....	20
9	Solvent-Accessible Surface of Alcohol Dehydrogenase, From Brookhaven PDB File 5adh	22
10	Computer-Generated Picture of Ribbon Model of Alcohol Dehydrogenase Monomer With Space-Filling Models of Two Zn^{+2} , ADPR, 2-methyl-2,4-pentanediol, Serine 1, and Phenylalanine 374.....	23
11	Computer-Generated Picture of Space-Filling Model of Alcohol Dehydrogenase Monomer.....	24
12	Solvent-Accessible Surface of ZIF268 Bound to DNA, From Brookhaven PDB File 1zaa.....	26
13	Space-Filling Model of ZIF268 Bound to DNA	27

TABLES

<i>Number</i>		<i>Page</i>
1	Atomic Radii	4
2	Stereolithographic Files (.stl) Put on CD-ROM	6
3	Solvent-Accessible Surface of Proteins From PDB	8

ABSTRACT

This report describes how plastic models of the solvent-accessible surface of biopolymers were made. Computer files containing sets of triangles were calculated, then used on a stereolithography machine to make the models. Small (2 in.) models were made to test whether the computer calculations were done correctly. Also, files of the type (.stl) required by any ISO 9001 rapid prototyping machine were written onto a CD-ROM for distribution to American companies.

SUMMARY

Many biopolymers such as proteins, DNA, and RNA have been studied because they have important biomedical roles and may be good targets for therapeutic action in treating diseases. Scientists have determined the structures of many such biopolymers using X-ray crystallography and nuclear magnetic resonance (NMR). The structures were then submitted to Brookhaven National Laboratory, who put them on a protein data bank (PDB). Files on the Brookhaven PDB contain the x, y, and z coordinates of the centers of the atoms in the biopolymers. The files also contain text, providing a brief description of the biopolymers and literature references. Any of several computer programs can be used to visualize the structures of the biopolymers. One program, called RasMol,¹ was used to visualize some biopolymers. One of the most realistic ways to visualize them is to represent each atom in the biopolymer as a sphere, so that a set of overlapping spheres are produced. This shows how the atoms fill space. However, RasMol cannot depict the more realistic solvent accessible surface, which includes a monolayer of tightly bound water.

In this report, computer methods are described that were used to calculate the solvent-accessible surface of proteins, DNA, and RNA. These include programs to read a PDB file, assign a radius to each atom, draw a molecular surface, then a piece-wise quartic molecular surface, then roll a sphere (the size of one water molecule) over the surface of the biopolymer. The piece-wise quartic surface is then converted to a set of triangles, which is then converted into the ISO 9001 file type (called .stl) required by stereolithography machines such as the one at AlliedSignal Inc., Federal Manufacturing & Technologies (FM&T)--(SLA 250, from 3D Systems, Valencia, CA).

The .stl files were used to make several small models. A brief description of each biopolymer is presented.

Finally, several .stl files were written onto a CD-ROM. This CD-ROM will be made available to American industry. The first CD-ROM was sent to scientists at Procter & Gamble (Cincinnati, OH), and was used on their SLA-250 machine. Like other pharmaceutical companies, Procter & Gamble does rational drug design, using computer images. The plastic (stereolithographic) models will enable scientists to see a real three-dimensional model, instead of just a computer simulation. This should help them to produce more medicines for treating a variety of diseases.

DISCUSSION

SCOPE AND PURPOSE

The project objective is to develop software to make plastic models of biopolymers, such as proteins, DNA, and RNA, and then transfer this technology to American industry. In this phase of the project, solvent-accessible surfaces of biopolymers were calculated and fabricated.

PRIOR WORK

A description of stereolithographic models made previously under this project has been published. Also, a detailed description of the computer programs and the mathematical principles involved has been published.

ACTIVITY

Description of a Solvent-Accessible Surface Model

Molecules can be depicted in many ways. In text books, the one-letter abbreviations for hydrogen (H), carbon (C), oxygen (O), sulfur (S), nitrogen (N) and phosphorus (P), are joined by lines, or sticks, i.e., $\text{CH}_3\text{-CH}_2\text{-NH}_2$. Another way to depict this molecule would be to have each group represented by a ball, or sphere. When biopolymer structure is determined using X-ray crystallography, the data are depicted by RasMol¹ and other programs, so that the CH_3 , CH_2 , NH_2 , and similar groups are one sphere. RasMol can depict a PDB file as a ball and stick model, in which the groups are represented as small spheres, joined by sticks which represent chemical bonds. Another way to represent a molecule is called a space filling model, in which each group is represented as a larger sphere such that each sphere overlaps. This is a more realistic model of how the groups fill space.

When a biopolymer is dissolved in water, as it is in the cell, it is surrounded by a tightly held monolayer of water. This monolayer of water has a lower dielectric strength and other physical properties that distinguish it from bulk water. The bulk water dissolves this monolayer-covered biopolymer. This is called the solvent (water) accessible surface of the biopolymer.

This becomes significant in doing rational drug design. Pharmaceutical companies can save money and bring important new medicines to the market faster if their computer calculations can predict which small molecules (drugs) are most likely to be effective. Drugs usually act by binding to a protein or other biopolymer that is defective and causing a disease. The energy of binding can be calculated and used to predict a drug's efficacy. For the drug to bind to the protein, it must fit tightly into a binding pocket and the monolayer of water at the binding site must be displaced. Since two-dimensional computer-generated images of such binding have proven somewhat successful, it is anticipated that three-dimensional (stereolithographic) models produced in this project will be even more useful.

Description of the Calculations

The first step is to calculate a molecular surface from a PDB file. If the PDB file contains less than 4000 atoms, a program called pdb2ms is used. The computer is told to read the PDB file, which contains the x, y, and z coordinates of not only the atoms in the biopolymer, but also those water molecules that were bound tight enough to the biopolymer to be seen by X-ray crystallography.² These waters are not wanted at this time, so the pdb2ms program is instructed to ignore them. Then, a molecular surface file is written with the extension .ms, i.e., ras.ms for the protein called ras. The .ms files contain four columns of numbers, corresponding to the x, y, and z coordinates for the atom centers and the atomic numbers.

If the PDB file has more than 4000 atoms, the pdb2ms program cannot be used. Instead, a program (written by Jim Emery, Department 871) is used. Like pdb2ms, it reads the PDB file, then writes a .ms file, but the waters and any ligands (small molecules) present in the file are ignored. These are labeled HETATOM in PDB files.

Once a .ms file is produced, a program called pqms³ is used to read the .ms file, assign a radius for each of the atoms, then calculate a piece-wise quartic molecular surface. The surface is generated by rotating a sphere around two adjacent spheres to produce a torus, which is described by a 4th degree (quartic) polynomial. The sizes of the spheres are read from the radius file. Table 1 lists the possible atomic (spherical) radii. There are two files of atomic radii. They are elements2.rad, ms.rad and msexpanded.rad. Depending on the experimental method used to obtain the particular PDB file, one would choose one of these files to use in assigning atomic radii. For this report, most of the PDB files were based on X-ray crystallographic data. Such files do not contain hydrogen atoms because they are not observed by X-ray diffraction. However, the locations of the hydrogens are apparent from the known structures of the amino acids that make up the protein biopolymer. Thus, when such files are visualized, groups like -CH₂-, -CH₃, -NH₂, etc. are shown as one sphere depicting the entire functional group. The radii of such groups are available in the msexpanded.rad file. Thus, the first command given the computer in the pqms program is:

```
:molecule name.ms msexpanded.rad
```

The command molecule is used to read the .ms file. The radii (listed in Table 2 under the column heading msexpanded.rad) are assigned to the appropriate atom types in the .ms file. Since carbons and nitrogens have attached hydrogens, larger radii are assigned to them in the msexpanded.rad file.

There were a few PDB files based on NMR data. For these, the command given in the pqms program is:

```
:molecule name.ms elements2.rad
```

Table 1. Atomic Radii

Atom	Number	Elements2.rad Radius (Å)	Msexpanded.rad Radius (Å)
H	1	1.000	1.40
C	6	1.800	2.00
N	7	1.700	2.00
O	8	1.600	2.00
P	15	1.900	1.20
S	16	1.900	1.50

The radii in Table 1 under elements2.rad are used. Smaller radii for carbons and nitrogens are used because they do not represent -CH₂-, -NH₂, -CH₃, or other such groups but only the C and N atoms themselves.

Some of the data in the Brookhaven PDB are based on NMR data, in which hydrogens are observed. In the near future, models will be made from them. At that time, either the elements2.rad or the ms.rad files will be used to assign radii, and groups like -CH₂- and -NH₂ will be depicted as three smaller spheres, rather than the one larger sphere, as was done for this report, because data were generated by X-ray crystallography.

After completing the pqms program, a file that ends in .pqms is written. The piece-wise quartic molecular surface is then triangulated using a program called trb.³ The resolution of the triangulation is controlled by the parameter angle. This is the angle formed by an atom center and an edge of a triangle. In the end, a file containing vertices, edges, and triangles (.vet) is produced.

This .vet file is then read into a program called vet2tri, which calculates a set of triangles. This set of triangles can be visualized on the computer's CRT by using programs written by Jim Emery. They are called tri2t and vt. They create a triangle file for display, then display it in on the Silicon Graphics Iris Indigo's CRT. This represents the solvent-accessible surface of the biopolymer. The vt program displays it and slowly rotates it. As a quality control, the RasMol program can be used to display the same biopolymer as a space filling model. The space filling model represents each atom as a sphere, so the model looks almost like the solvent-accessible surface model, except that the final monolayer of water has not been added.

Finally, the .tri file is converted into a .stl file using the program tri2stl. This program allows the analyst to scale and translate the model. The default scale (maximum dimension) is 4.7 in., and the translation (location) is 0.3 inches for the x, y, and z minima. That is, the lower left front corner of the model is placed up, to the right, and in 0.3 inches from the lowest left corner of the vat of photoreactive resin (Exactomer 2202, see reference 2) in the SLA-250. In this phase of the project, four 2-inch models were made at a time. The first was set at 0.3, 0.3, 0.3 in. The second was set at 2.4, 0.3, 0.3, that is, 2.1 inches to the right of the first model. The third was set at 4.5, 0.3, 0.3, and the fourth was set at 6.6, 0.3, 0.3. Thus, at least 0.1 inch separates the models from each other.

For the CD-ROM, however, the .stl files were not scaled or translated. Different customers can decide for themselves how big to build their models. A list of the .stl files put onto CD-ROM is shown in Table 2.

Stereolithography

Stereolithography (SL) is a process that grew out of computer-aided design (CAD). The SL process uses a computer-directed laser to map out the topology of the CAD object at the surface of a photosensitive liquid resin, such as Exactomer 2202. The SL workstation requires the .stl file format. The .stl file contains the topology of the CAD model represented as a set of triangles. Thus, the surface of the biopolymer is represented by a set of triangles, or facets. Each facet is defined by a normal and three vertices. The facets define the surface of the three-dimensional biopolymer (or any other CAD object).

The direction of the normal must be outward and the vertices listed in counterclockwise order when looking at the object from the outside. Each triangle must share two vertices with each of its adjacent triangles. The model's surface must be closed, that is, no open or unconnected regions. These .stl data are used by a slicing algorithm to determine the cross sections of the model to be built. All this is done under the network called MDMA (Mechanical Design and Manufacturing Automation). This way, the .stl files can be imported into Part Man, then into ProEngineer. The files are sent over the MDMA network to the SLA-250 computer (workstation) for the SL process.

At the point of laser irradiation, the liquid resin is converted into a solid plastic. The laser only penetrates about 0.1-0.3 mm. The newly formed solid layer is then lowered to let fresh resin cover it, and the laser mapping process is continued for the next layer of the object. By drawing out the biopolymer's image layer by layer in the liquid resin, the three-dimensional SL model is produced.. The model is made from bottom to top.

Support structures are required and are built first. The problem is that the polymerized solid would just float away if it were not stuck onto an already existing piece of polymer. Thus, a small rectangular base with a small post is built as support. The rectangular base is made first, so the biopolymer model has something to adhere to as it is being made. Once the model is polymerized, it is removed from the vat of unreacted resin, and the liquid resin is allowed to drain off. The rectangular base support is then removed. The biopolymeric model is then washed with isopropyl alcohol and post-cured by heating at 120-160°F. During the summer, when the humidity is higher, 160°F for several (4-6) hours is needed, but 120°F for only 1 hour is needed when the humidity is low. Post-curing is done until the model no longer feels very sticky. A clear coating is applied after the post-curing.

Table 2. Stereolithographic Files (.stl) Put on CD-ROM

Protein	.stl File Name	File Size (Mbyte)	Atoms Deleted
ras on with GDP	rasgdp.stl	6.16	Mg ⁺²
ras on without GDP	rasoff.stl	6.50	Mg ⁺² & GDP
ras off with GDPCP	rasgdpcp.stl	6.16	Mg ⁺²
ras off without GTP	rason.stl	6.44	Mg ⁺² & GDPCP
Deoxyhemoglobin	1hga.stl	22.2	4 Fe ⁺²
DNA + ZIF258	1zaa.stl	6.4	3 Zn ⁺²
Chymotrypsin	1cho.stl	9.6	none
HIV Protease	9hvp.stl	7.7	none
Oxy Myoglobin	1mbo.stl	6.0	Fe ⁺³ & SO ₄ ²⁻
Cytochrome P450	1pha.stl	13.1	4 Ca ⁺²
Calmodulin	1osa.stl	7.4	none
Alcohol Dehydrogenase (DH)	5adh.stl	12.3	2 Zn ⁺²
Alcohol DH Dimer w/o NAD	6adh.stl	23.2	4 Zn ⁺² & 2 NAD ⁺ & DMSO
DNA Decamer	1d56.stl	3.0	Ca ⁺²
tRNA-alanyl	1tra.stl	9.3	5 Mg ⁺²
Anti-Fluorescein Antibody	4fab.stl	15.4	none
Aspartate Amino Transferase	1aam.stl	15.3	SO ₄ ²⁻
tRNA	1tra.stl	9.31	5 Mg ⁺²
Malate DH with Citrate	2cmd.stl	10.9	none
Protein Kinase C	1atp.stl	12.4	2 Mn ⁺²
HLA-A w/Hepatitis Peptide	1hhh.stl	15.0	none
Ubiquitin Conjugating Enzyme	1aak.stl	6.5	none
Di-Ubiquitin	1aar.stl	5.6	none
Antifreeze Protein	1atf.stl	3.4	none
c-myb Proto-oncogene	1pom.stl	5.5	none
Alpha-1 Amphiphilic Helix	1al1.stl	0.95	none
Deoxyhemoglobin	2hhb.stl	22.2	4 Fe ⁺²
Malate DH with NAD ⁺	1emd.stl	10.7	none
Cytochrome c	1ccr.stl	4.3	Fe ⁺²
CD4	1cd4.stl	7.4	none

Table 2 Continued. Stereolithographic Files (.stl) Put on CD-ROM

Protein	stl File Name	File Size (Mbyte)	Atoms Deleted
Amicyanin	1aa.j.stl	3.8	none
Oxymyoglobin	1mbo.stl	6.0	Fe ⁺²
cAMP-dependent Protein Kinase C with cAMP Attached:			
Amino Acids 143-260	1apk.stl	5.7	none
Amino Acids 261-379	Problem faces found in trb		
Amino Acids 143-264	2apk.stl	6.4	none
Trimethyl Calmodulin with Trifluoperazine Attached	2cln.stl	7.4	4 Ca ⁺² and Trifluoperazine
Cytotoxic T-lymphocyte Proteinase	2cp1.stl	8.6	none
Avidin-Biotin	2avi.stl	10.0	none
Alpha-Lytic Protease	4lpr.stl	5.5	5 SO ₄ ²⁻
Acetylcholine Esterase	1ace.stl	14.9	none
D-Xylose Isomerase	8xia.stl	15.5	2 Mn ⁺²
BDS Anti-hypertensive, Anti-viral Protein	1bds.stl	3.6	none
Bacteriorhodopsin Amino Acids 8-63, 74-100, 106-127, 137-191 and 202-225	1brd.stl	8.1	none
E. Coli Capsular Polysaccharide	1cap.stl	1.6	none
Arabinose Binding Protein	6abp.stl	9.8	none
Arabinose Binding Protein	7abp.stl	9.7	none
α-Amylase	6taa.stl	13.7	none

The Biopolymers

In Table 3 is a list of biopolymers whose solvent-accessible surfaces were calculated or attempted to be calculated. The first few files were chosen from those selected by previous FM&T customers at DOE and the University of California at Berkeley. Other types of models of some of these biopolymers were also made. The number of amino acids in the proteins is given. DNA and RNA have nucleic acids, not amino acids. Nucleic acids are also called base pairs. For each biopolymer, the number of atoms and the calculated volume of water solvent excluded by the biopolymer, that is, the volume of the triangulated surface with a monolayer of attached water, is also given.

Some of the files had problems. PDB file 1emd had problem faces that prevented a .stl file from being calculated (the tri2stl program wouldn't accept the problem faces). PDB file 1d56 had an open loop that prevented the slicing algorithm from working so a model was not made. File 7aat had a triangle error. Calculations on files 1pes and 1olg, both based on NMR data, failed. File 1pom caused a memory fault (core dump). In some other cases, the trb program found omega less than zero, so it changed the angle for some atoms so that the set of triangles could be calculated. One file, 110d, was successfully converted into a .stl file, but when we tried to make a model on the SLA-250, the slicing algorithm did not work because there was an open loop. Some files had trivial problems. File 1mbo had some atoms listed more than once. They simply had to be deleted for the .stl file to be calculated.

Following Table 3 is a brief description of each biopolymer. Also, the type of small molecule, or ligand, that is bound to the biopolymer is given. In some cases, metal cations such as Mg^{+2} , Zn^{+2} , Ca^{+2} , and Fe^{+3} were deleted. In others, ligands were deleted. These are detailed in the descriptions following the table.

Table 3. Solvent-Accessible Surface of Proteins From PDB

Biopolymer	PDB File	Amino Acids	Atoms	Solvent Excluded Volume (Å ³)	stl File Name
ras off with GDP	1q21	171	1396	25220	rasgdp.stl
ras off without GDP	edited 1q21	171	1320	25220	rasoff.stl
ras on with GDPCP	6q21	198	1400	25220	rasgdpcp.stl
ras on without GTP	6q21	198	1400	24742	rason.stl
Deoxyhemoglobin T	1hga	564	4384	78983	1hga.stl
Deoxyhemoglobin T	2hhb	564	4384	80255	2hhb.stl
DNA+Adriamycin	110d	Slicing algorithm found an open loop			
Alcohol Dehydrogenase	5adh	374	2831	52355	5adh.stl
Alcohol Dehydrogenase	6adh	2x374	5569	103951	6adh.stl
dna + ZIF258	1zaa	85 + 11 bp	1161	19887	1zaa.stl
Chymotrypsin	1cho	245+56	2150	39386	1cho.stl
Cytochrome P450	1pha	414	3254	60324	1pha.stl
Malate Dehydrogenase with NAD ⁺	1emd	312	2335	43510	1emd.stl
Malate Dehydrogenase with Citrate	2cmd	312	2797	43556	2cmd.stl
HIV Protease	9hvp	198+ Inhibitor	1574	29453	9hvp.stl

Table 3 Continued. Solvent-Accessible Surface of Proteins From PDB

Biopolymer	PDB File	Amino Acids	Atoms	Solvent Excluded Volume (Å ³)	stl File Name
Oxyhemoglobin T	1hgc	564	4384	79033	Can't do problem faces
Hemoglobin-deoxy, without Fe-Heme	edited 2hhb	564	4384	80255	2hhb.stl
tRNA-alanyl	1tra	1	1652	26209	trna.stl
Fab of Antibody	4fab	219L +216H	3395	59610	4fab.stl
HLA-A with Hepatitis B Capsid Decameric Peptide	1hhh	275+100+ 10	3167	56902	1hhh.stl
Oxy Myoglobin	1mbo	153	1269	23104	1mbo.stl
(had to remove redundant atoms (i.e., 714&716& others) from 1mbo)					
DNA 10 Base Pairs	1d56	10 bp	404	6309	1d56.stl
Aspartate Amino Transferase without SO ₄ ²⁻ , but with PLP	1aam	396	3081	56518	1aam.stl
Mutant with arg386 Replaced by phe	3aat	396	3083	56130	3aat.stl
(Same as above)	7aat	Triangle Error - can't run trb			
Alcohol Dehydrogenase Homo Dimer without NAD ⁺	6adh	2x374	5569	103951	6adh.stl
cAMP Dependent Protein Kinase with ATP and Peptide Inhibitor (Mn ⁺² edited out)	1atp	350+24	2965	54394	pkc.stl
CD4	1cd4	173	1348	24962	cd4.stl
Calmodulin	1osa	148	1337	27016	calm.stl
Cytochrome P450	1pha	414 + heme	3254	60324	p450.stl
Cytochrome c	1ccr	111 + heme	952	17419	cytc.stl
Amicyanin	1aaj	105	807	14635	1aaj.stl
Di-Ubiquitin	1aar	2x76	1203	22360	1aar.stl
Ubiquitin Conjugating Enzyme	1aak	152	1203	21814	ubiq.stl
Antifreeze Protein	1atf	37	452	3165	1atf.stl
c-myb Proto-oncogene	1pom	31	781	6807	Memory fault
Alpha-1 Amphiphilic Helix	1al1	12	95	1695	1al1.stl

Table 3 Continued. Solvent-Accessible Surface of Proteins From PDB

Biopolymer	PDB File	Amino Acids	Atoms	Solvent Excluded Volume (Å ³)	stl File Name
p53 Quaternization Domain	1pes	2088	20372	trb fails, check vertex 6765 lfn=0	
p53 Quaternization Domain	1olg	4x42	2729	27205	1olg.stl
Euler characteristic = 4 by trb					
cAMP-dependent Protein Kinase C with cAMP Attached:					
Amino Acids 143-260	1apk	117	958	17530	1apk.stl
Amino Acids 261-379	1bpk	118	958	Problem faces found in trb	
Amino Acids 143-264	2apk	121	1018	18144	2apk.stl
Trimethyl Calmodulin with Tri-fluoperazine attached - Delete 4 Ca ²⁺ and Trifluoperazine	2cln	148	1166	20543	2cln.stl
Cytotoxic T-lymphocyte Proteinase	2cp1	227	1775	32989	2cp1.stl
Avidin-Biotin	2avi	2x121	1952	34935	2avi.stl
Alpha-Lytic Protease	4lpr	183	1405	25352	4lpr.stl
Acetylcholine Esterase	1ace	508	4115	74082	1ace.stl
D-Xylose Isomerase	8xia	387	3054	55496	8xia.stl
BDS Anti-hypertensive, Anti-viral Protein	1bds	43	629	4975	1bds.stl
Bacteriorhodopsin Amino Acids 8-63, 74-100, 106-127, 137-191 and 202-225	1brd	177	1363	24856	1brd.stl
E. Coli Capsular Polysaccharide	1cap	No amino acids	230	1865	1cap.stl
Arabinose Binding Protein	6abp	305	2316	42383	6abp.stl
Arabinose Binding Protein	7abp	305	2316	42305	7abp.stl
Alpha Amylase	6taa	476	3686	67384	6taa.stl

Catalytic Domain of Ras. The protein called ras plays a pivotal role in the signaling of growth and differentiation of plant and animal cells. Probably for this reason, other types of models (called ribbon models) of ras on and off were made under an integrated contractor order (ICO) with the Lawrence Berkeley Laboratory. A ribbon model of ras turned off is shown in Figure 1. Ribbon Model of Ras Turned Off, from Brookhaven PDB File 1q21 When the ras protein binds a modified form of the nucleic acid called guanine, or G, it is turned on and the cell can grow or differentiate into a specialized cell. The modified form of G is a form containing three phosphates and is usually called GTP (for guanosine triphosphate).

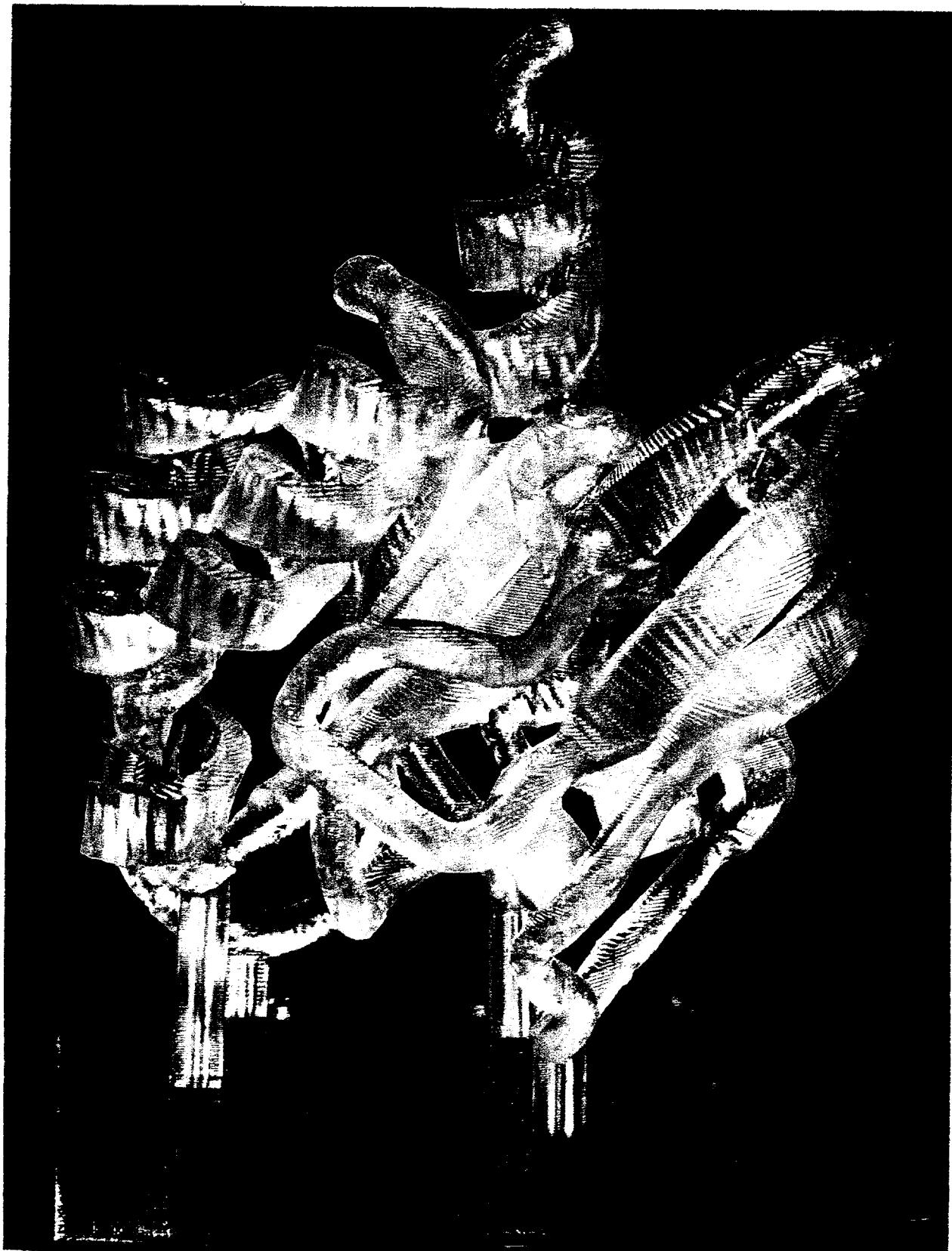


Figure 1. Ribbon Model of Ras Turned Off, From Brookhaven PDB File 1q21

In a healthy cell, growth and differentiation are carefully controlled. Once the cell divides or differentiates, the ras protein should be turned off. This is done by breaking off, or hydrolyzing, one of the phosphates off GTP to produce GDP, or guanosine diphosphate. Thus, ras off with attached GDP is stable, but ras on with GTP attached is not very stable. It takes many hours to do X-ray crystallography, so the ras cannot be turned on with its natural switch, GTP. Instead, an artificial switch, that is, a guanosine with three phosphates, but with an additional $-\text{CH}_2-$ between the second and third phosphates (abbreviated GDPCP), was used.

One Mg^{+2} was in each of the PDB files (1q21 and 6q21). The Mg^{+2} was deleted before beginning the calculations of the solvent accessible surface for the first two models. It is also interesting to note the part of the ras proteins that bind GDP (off) or GDPCP (on). For the third and fourth models, the PDB files were edited, removing the atoms due to GDP or GDPCP.

Switching between ras on and off causes a change in its structure,^{4,5} so models of ras on and ras off were made and are shown in Figure 2. They are in about the same orientation as the ribbon model in Figure 1. The highest point on the ribbon model, which is the $-\text{COOH}$ terminus of the ras (sometimes called p21) protein, is also the highest point on the solvent accessible surface models (Figure 2). The solvent accessible surface models show that this (as most proteins) have no hollow portions (as appears in the ribbon model). It should be noted also that the GDP and GDPCP ligands bind at the part of the model that is on the bottom as viewed in these pictures. The ligand binding sites appear as grooves in those models in which the GDP or GDPCP were deleted.

Hemoglobin. Several PDB files exist for this important protein which carries oxygen in the blood. Hemoglobin consists of four polypeptide subunits ($\alpha_2\beta_2$), two having 141 amino acids and the other two having 146 amino acids. Each of the four polypeptides have one iron and one heme group. The four irons and hemes were deleted from PDB file 2hhb before calculating the solvent-accessible surface. This file, contributed by Perutz and coworkers,⁶ depicts deoxy-hemoglobin with the four polypeptides arranged symmetrically. Other PDB files of hemoglobin show the polypeptides in other configurations. Perutz's original analysis (file 2hhb) was done on hemoglobin crystals grown from an aqueous solution containing high salt concentration. Liddington and coworkers⁷ grew deoxy-hemoglobin crystals from an aqueous solution containing poly(ethylene glycol), and deposited their data in PDB file 1hgc. It proved impossible to calculate a .stl file from 1hgc because there were problem faces that the computer could not fit into the model. However, in the same paper, they described the structure of oxygenated, or met-hemoglobin; it is in PDB file 1hga. Four Fe^{+2} and four hemes were deleted from PDB file 1hga before preparing the .ms file. There were no problem faces, and a .stl file was successfully calculated and a model built, as shown in Figure 3. To get a better idea of the hemoglobin structure, RasMol¹ was used to generate a picture of a ribbon model with the four hemes as space-filling models, in Figure 4. The four different polypeptides with their hemes are shown in four different colors. Hemoglobin is viewed from approximately the same position as in Figure 3. RasMol was also used to generate a space-filling model of hemoglobin viewed from this position in Figure 5. Remember, the solvent-accessible surface includes an extra monolayer of water.

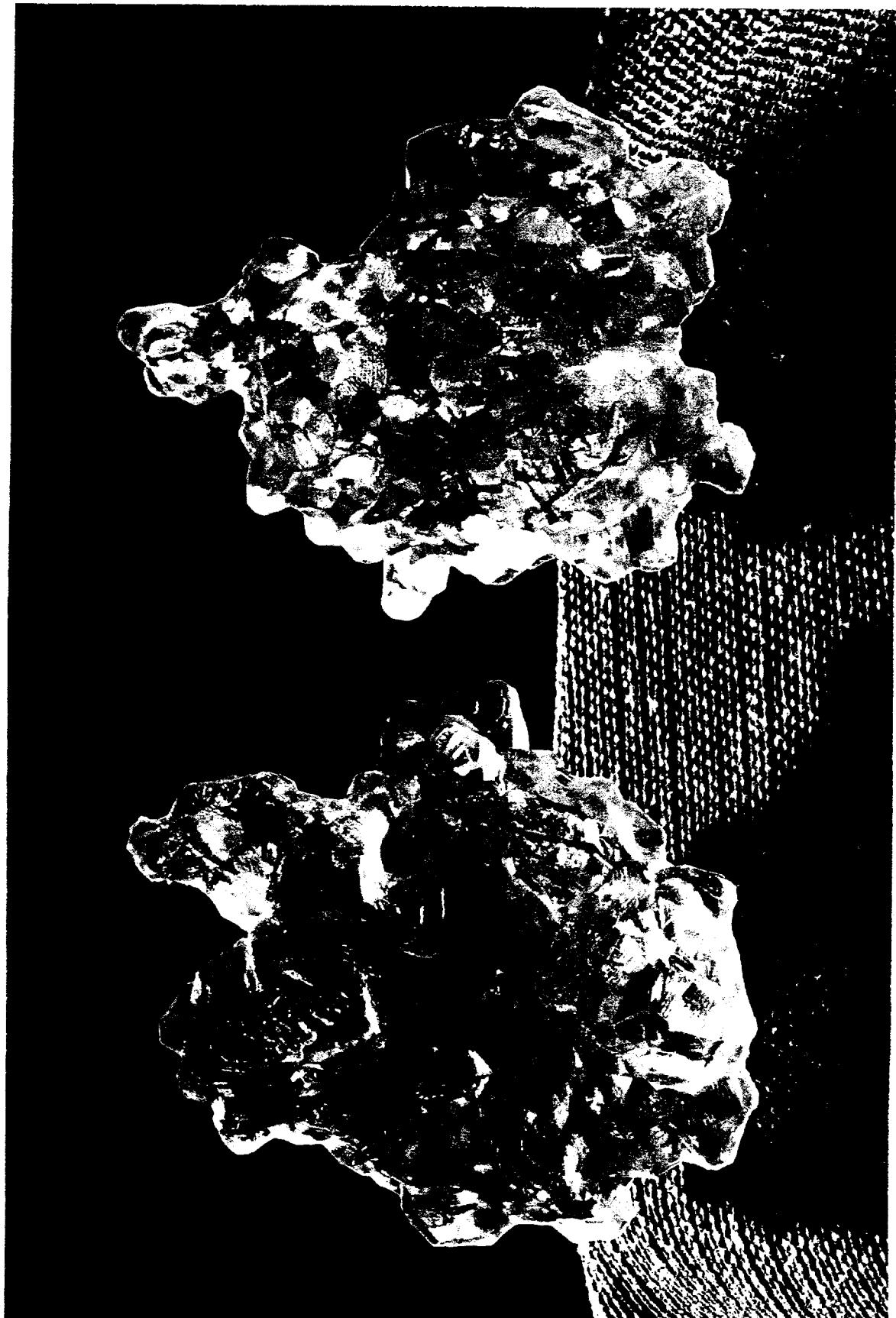


Figure 2. Solvent-Accessible Surfaces of Ras Off (Left) and On (Right), From Brookhaven PDB Files 1q21 and 6q21

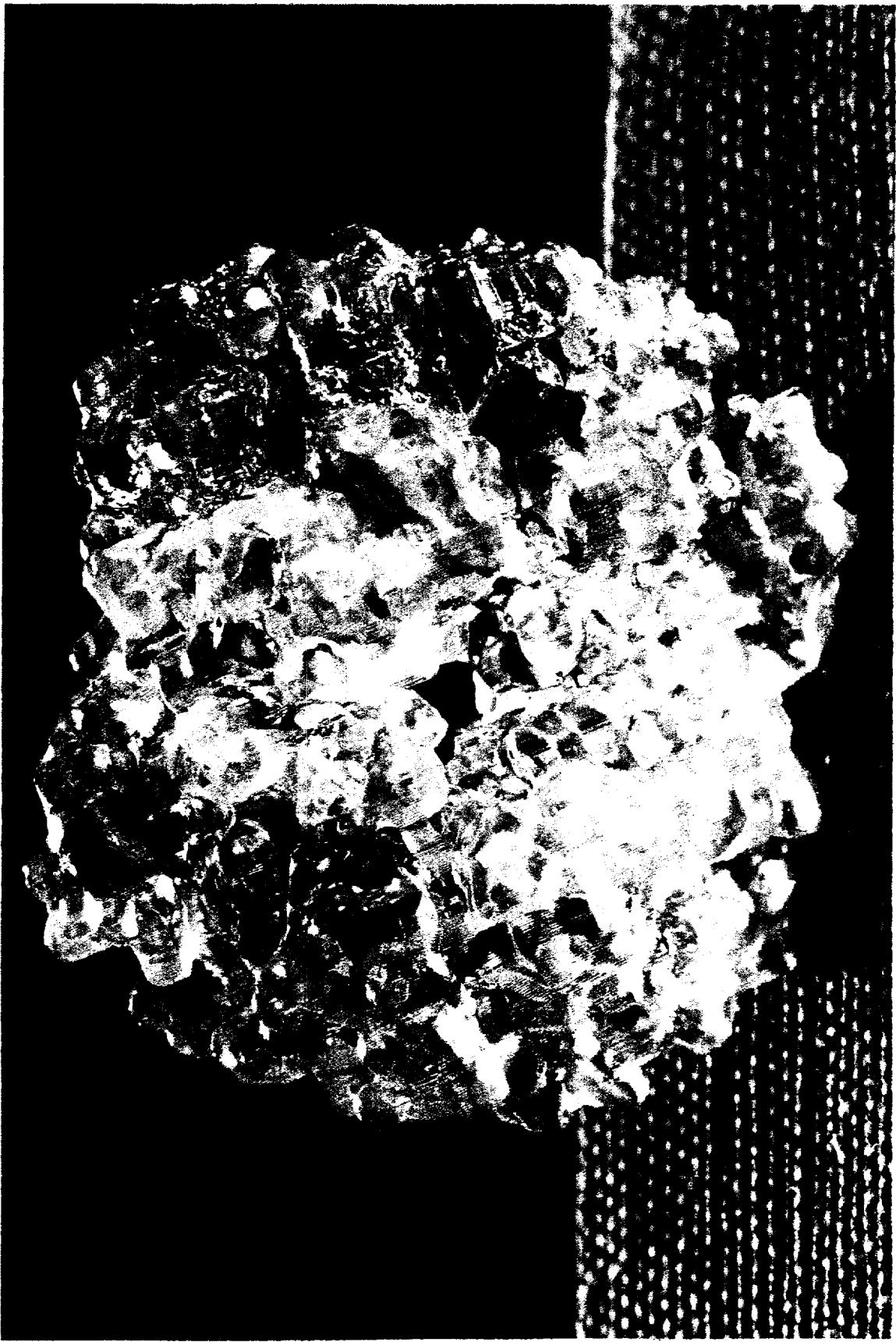


Figure 3. Solvent-Accessible Surface of Hemoglobin, From Brookhaven PDB File 1hga



Figure 4. Computer-Generated Picture of Ribbon Model of Hemoglobin With Space-Filling Model of Heme Groups. Each Color Represents a Different Polypeptide Subunit.

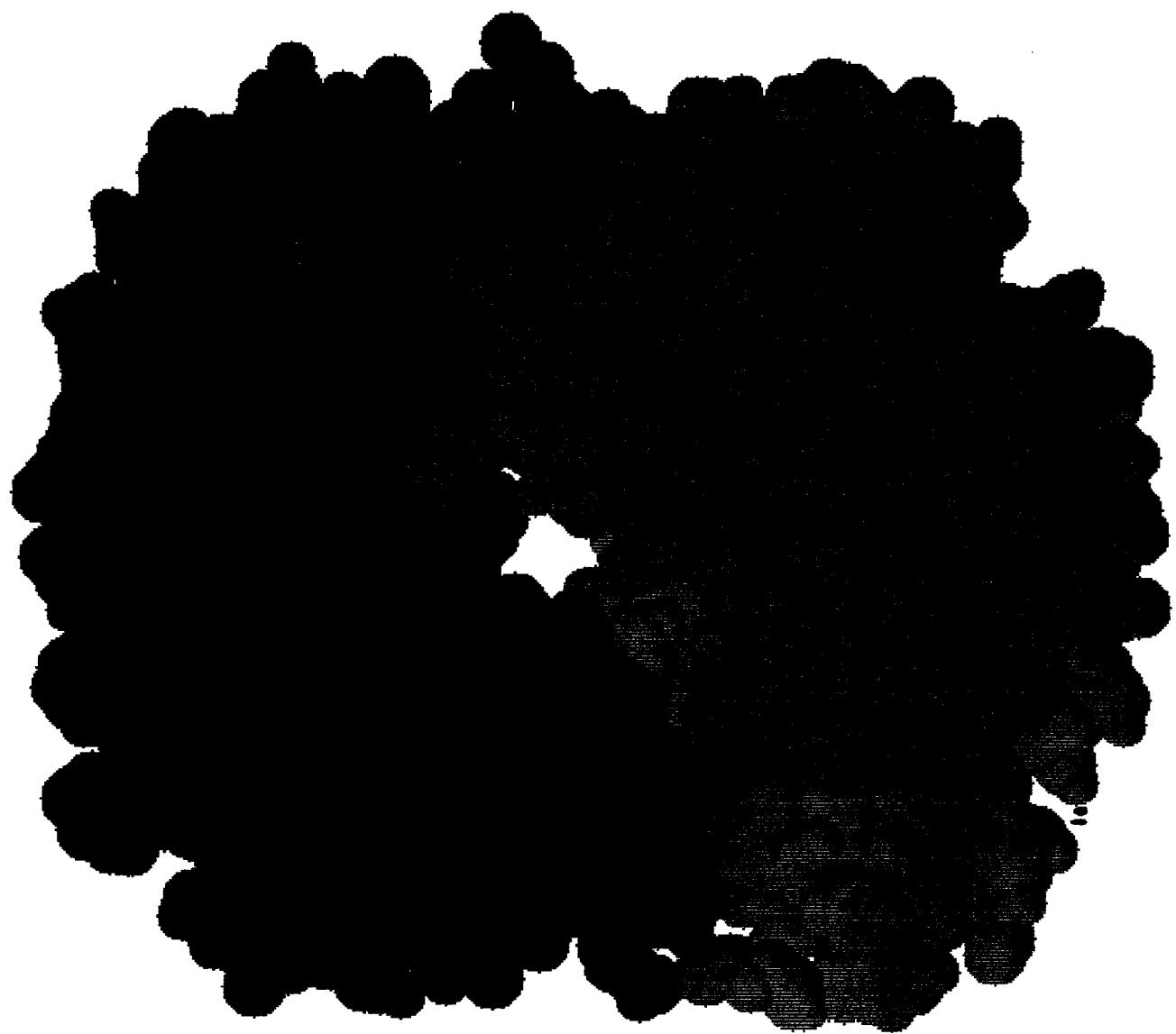


Figure 5. Computer-Generated Picture of Space-Filling Model of Hemoglobin

HIV Protease. This protein, required by the human immunodeficiency virus (HIV), has been a popular site for therapeutic action in treating AIDS. Other types of models of the HIV protease were described previously. In this report, the PDB file 9hvp⁸ was used. It depicts the HIV protease with an inhibitor attached. HIV protease is in the family called aspartate proteases because it contains two aspartates (as well as glycines and threonines) in its active site. The inhibitor binds to this active site, preventing the HIV protease from binding to its natural substrate, so the HIV will not mature and be able to spread to other white blood cells.

α -Chymotrypsin. This protein is similar to the HIV protease in that it also is a protease. That is, it catalyzes the breakdown of proteins (proteolysis). However, it is a serine protease. It uses serine, not aspartate, in its active site. Instead of breaking down large viral proteins made by HIV, it is not as specific. It is a digestive enzyme that our bodies use to digest proteins and is one of the most studied enzymes. There is an α and a β form, which are chemically identical but differ in their crystallization properties. The PDB file 1cho⁹ was used to calculate the .stl file. This file depicts chymotrypsin (from the pancreas) complexed with a protein that it is about to digest. This protein, from turkeys, is the third domain of Kazal-type ovomucoid.

DNA-Adriamycin. This model shows a complex formed between six base pairs of DNA and adriamycin, the most commonly used cancer chemotherapeutic drug. Adriamycin, which has saved many lives, kills cancer cells by binding to their DNA. The PDB file 110d was used. A .stl file was successfully calculated and sent to the SLA250, but the slicing algorithm did not work because there was an open loop. Since a closed surface is needed, no model could be made.

Malate Dehydrogenase. This protein catalyzes the removal of two hydrogens from malic acid to form oxaloacetic acid. In eukaryotic cells, the mitochondrial form is important in energy metabolism, as part of the tricarboxylic acid (or Kreb's) cycle. The cytosolic form participates in the malate/aspartate shuttle, which exchanges reducing equivalents across the mitochondrial membrane. The PDB file 1emd was used for one model. It depicts the complex formed between malate dehydrogenase and its coenzyme, called NAD⁺ (nicotinamide adenine dinucleotide) and a molecule of citric acid. Another PDB file, 2cmd,¹⁰ depicts malate dehydrogenase (from *E. coli*) complexed with citrate, but not NAD⁺. *E. coli* malate dehydrogenase is a homodimer, with 312 amino acids per subunit. This file depicts just the monomer, and it was made into a model as shown in Figure 6. To better visualize its structure, malate dehydrogenase is represented (using RasMol) as a ribbon model with Methionine 1 and Lysine 312 as space-filling models in Figure 7. In Figure 8 is a space-filling model of the entire monomer.

F_{ab} of an Antibody. Antibodies are made by some types of white blood cells to help fight diseases. Antibodies are large proteins. One part of an antibody has the same structure as other antibodies in its class (i.e., IgG, IgA, IgC). The other part of the antibody structure is unique for the particular antigen invading the body. This variable region of an antibody is called the F_{ab} region. This is the region depicted in this model, made from PDB file 4fab, which depicts an anti-fluorescein antibody with fluorescein and 2-methyl-2,4-pentanediol attached.¹¹ This portion of the antibody contains a light (L) chain with 219 amino acids and a heavy (H) chain with 216 amino acids. The .stl file was calculated from 4fab without any deletions.

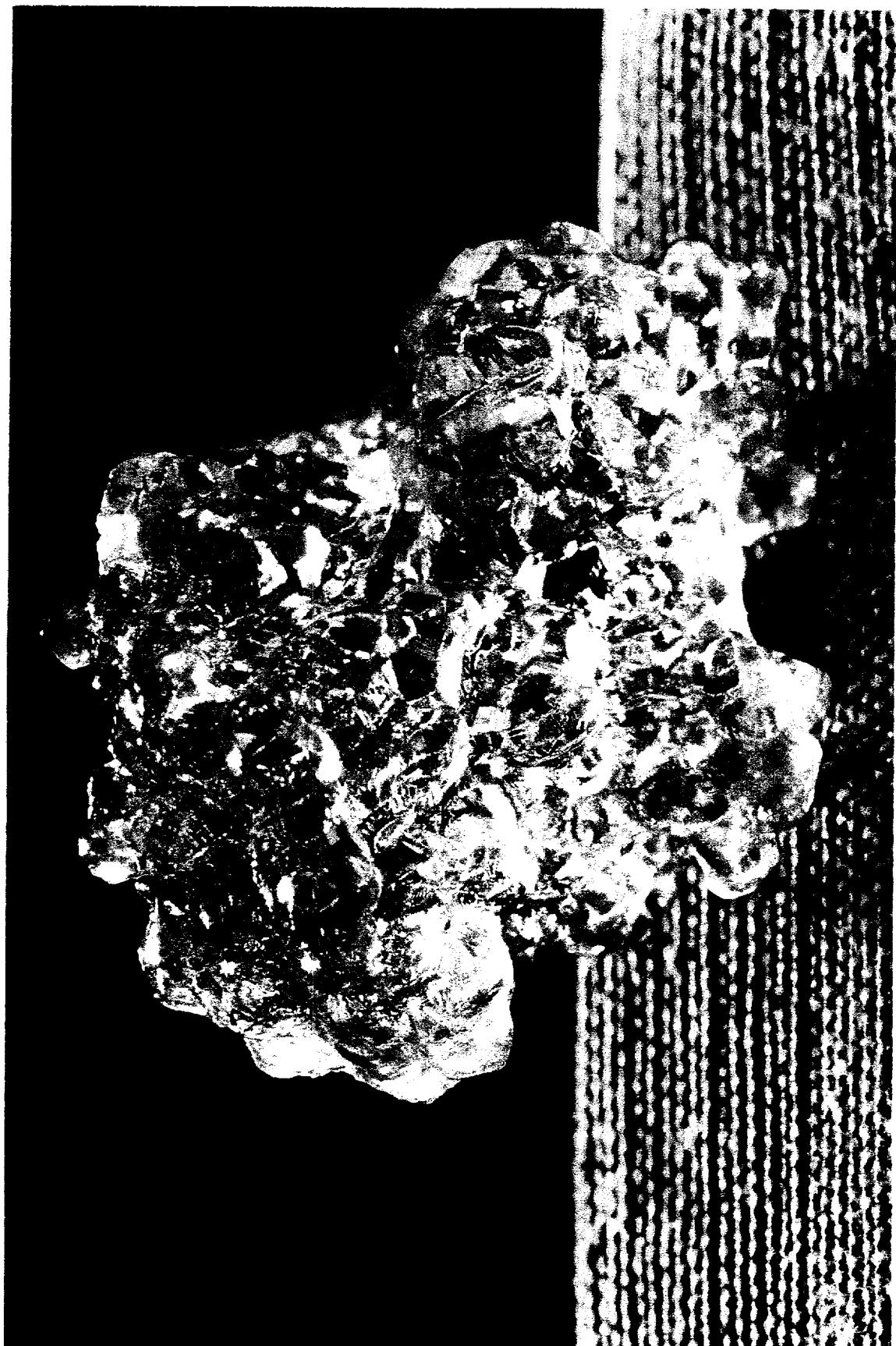


Figure 6. Solvent-Accessible Surface of Malate Dehydrogenase, From Brookhaven PDB File 2cmd



Figure 7. Computer-Generated Picture of Ribbon Model of Malate Dehydrogenase Monomer With Space-Filling Model of Methionine 1 and Lysine 312

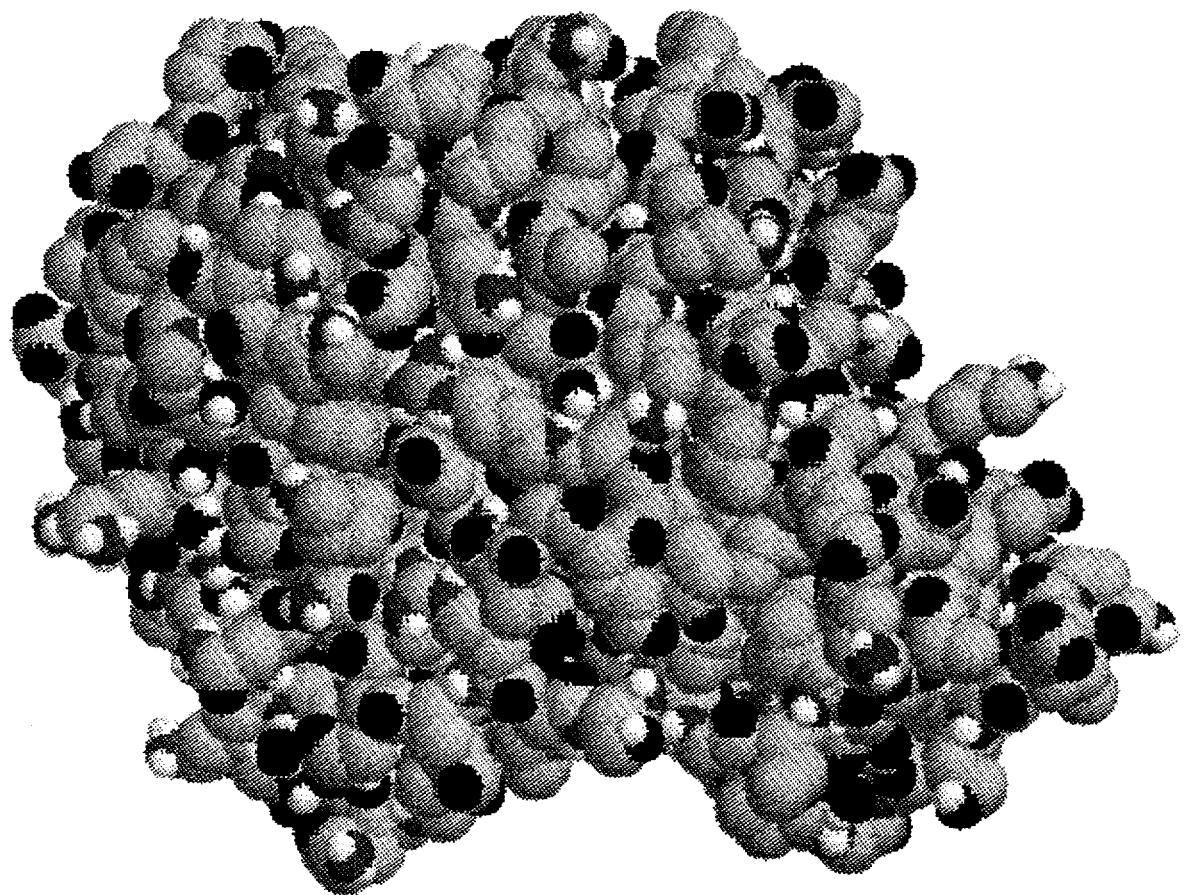


Figure 8. Computer-Generated Picture of Space-Filling Model of Malate Dehydrogenase Monomer

Transfer RNA (tRNA). To make a protein, a cell must assemble the proper amino acids. To do this, it requires that amino acids be attached to a tRNA. A model was calculated from the PDB file 1tra, which depicts phenylalanine-tRNA from yeast.¹² It contains the one amino acid (phenylalanine) and 76 ribonucleic acids. Five Mg⁺² were deleted from this file before calculating the .stl file.

Human Lymphocyte Antigen (HLA). Lymphocytes, or white blood cells, carry certain proteins on their membrane surface. One of these proteins, called HLA (for human lymphocyte antigen) by some and histocompatibility antigen by others, helps define a person's histocompatibility. It is also involved in presenting antigens to other white blood cells to stimulate an immune response. PDB file 1hhh¹³ depicts HLA bound to a piece of a hepatitis B virus (a protein). This file was used (without deletions) to calculate a .stl file.

Aspartate Amino Transferase. This protein uses a form of vitamin B₆, called pyridoxal-5'-phosphate (PLP), to catalyze the transfer of an amino group from the amino acid called aspartate to α -ketoglutarate to form glutamate plus oxaloacetate. Thus, it plays a central role in amino acid metabolism. This enzyme (a homodimer) is the most thoroughly studied of all the aminotransferases. There are many PDB files describing different aspartate amino transferases from various sources. The PDB file 1aam,¹⁴ depicts a mutant aspartate amino transferase monomer with arginine 292 being replaced by aspartate. The second order rate constant of this mutant is reduced by five orders of magnitude compared to the wild-type enzyme. However, the overall structure is quite similar to the wild-type enzyme. There are atomic coordinates for one SO₄²⁻ which was deleted before calculating the .stl file. The PDB file 3aat was also used to calculate an .stl file (after deleting the SO₄²⁻ but keeping the attached PLP). However, it should be noted that there is a portion of the solvent-accessible surface of aspartate amino transferase (amino acids met 5 to ala 11) that sticks out and would require a support to be added to keep it from falling off during the stereolithographic production. Other PDB files exist for this enzyme. All but one of them (on the October 1993 release) were successfully converted into .stl files. File 7aat, though, had triangle errors in the trb program, so no tri or .stl files could be calculated.

Liver Alcohol Dehydrogenase. This protein catalyzes the removal of two hydrogens off ethanol and other alcohols. It is important in energy and in alcohol metabolism. It consists of two identical polypeptides, so it is a homodimer. It uses two Zn⁺² and the coenzyme NAD⁺ for each polypeptide. There are several PDB files for alcohol dehydrogenase. Files 5adh and 6adh^{15,16} were used to calculate .stl files. The file 5adh depicts horse liver alcohol dehydrogenase monomer with an attached adenosine diphosphoribose (ADPR). It was made into a solvent-accessible surface model as shown in Figure 9. To help visualize the structure, a ribbon model is shown in Figure 10 with two Zn⁺², ADPR, 2-methyl-2,4-pentanediol, Serine 1, and Phenylalanine 374 as space-filling models. A computer-generated space-filling model is shown in Figure 11. File 6adh depicts the homodimer. The .stl file was calculated and put on a CD-ROM, but a model has not been made.

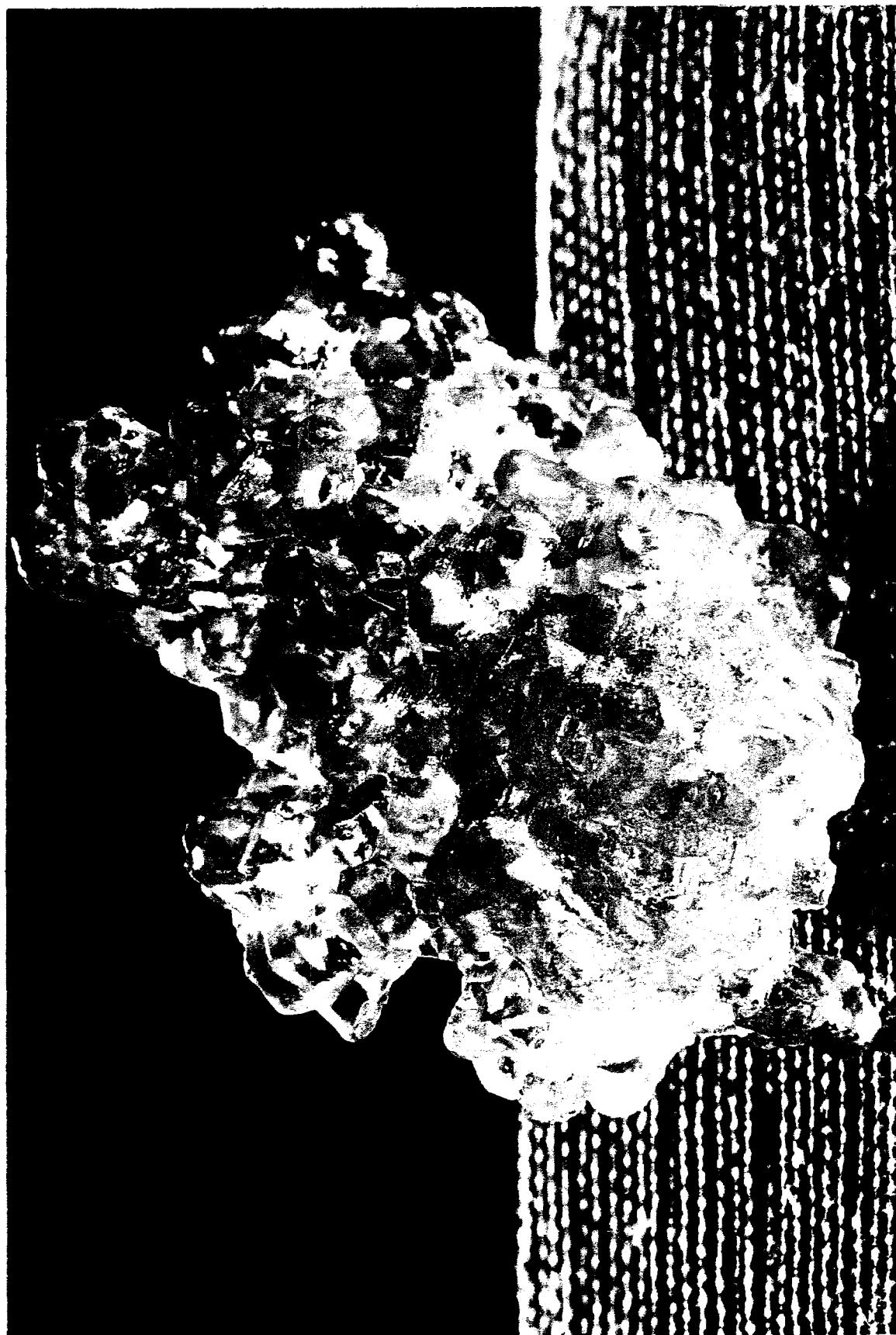


Figure 9. Solvent-Accessible Surface of Alcohol Dehydrogenase, From Brookhaven PDB File 5adh

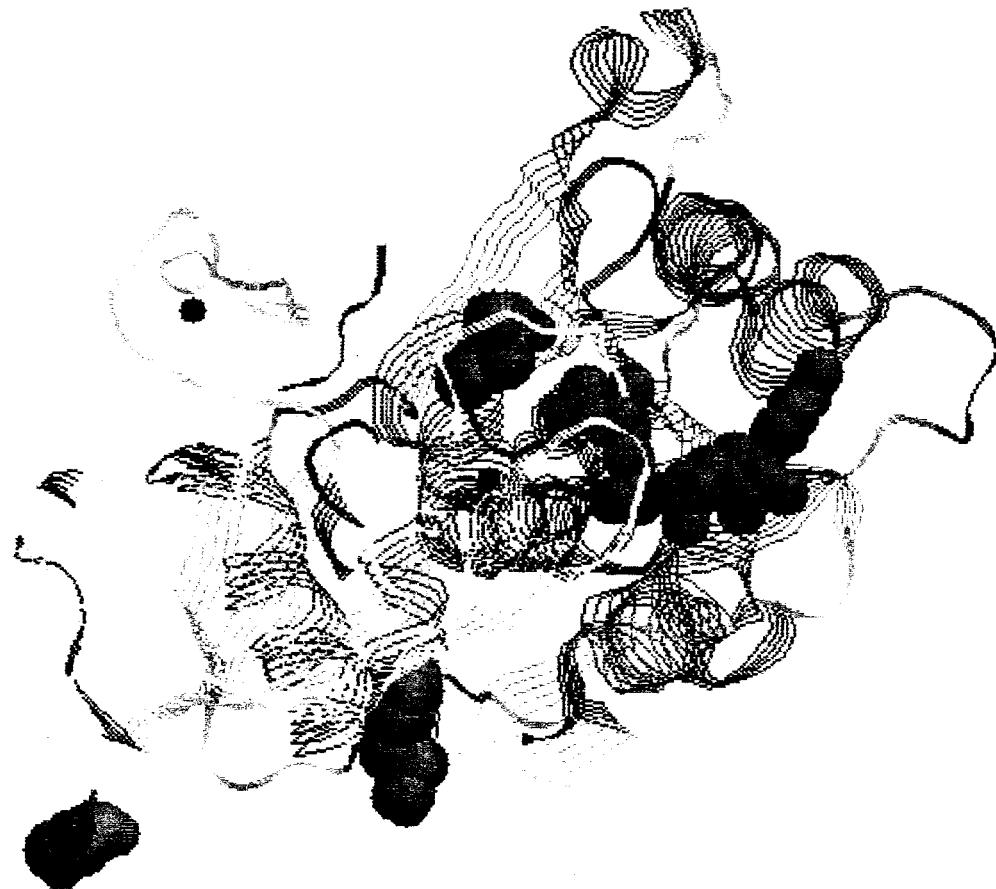


Figure 10. Computer-Generated Picture of Ribbon Model of Alcohol Dehydrogenase Monomer With Space-Filling Models of Two Zn^{+2} , ADPR, 2-methyl-2,4-pentanediol, Serine 1, and Phenylalanine 374

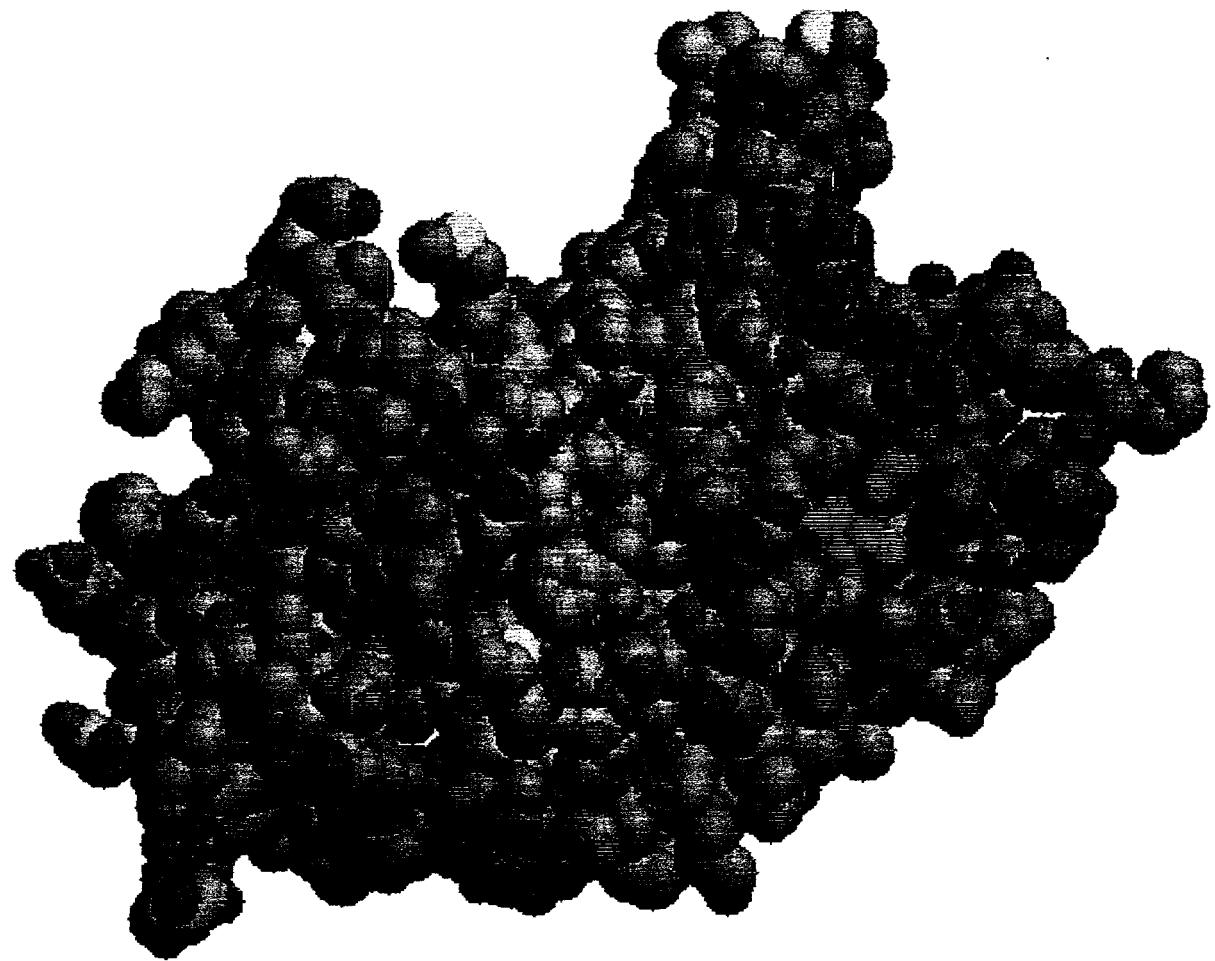


Figure 11. Computer-Generated Picture of Space-Filling Model of Alcohol Dehydrogenase Monomer

Protein Kinase C. This protein catalyzes the addition of phosphate to proteins that regulate metabolism. The family of protein kinases (with about 200 different members identified so far by molecular biologists) play key roles in metabolism. This includes the pathways in which hormones, growth factors, neurotransmitters, and toxins—such as the cholera toxin—exert their effects. Four different PDB files, 1atp,¹⁶ 1apk,¹⁷ 2apk,¹⁷ and 1bpk,¹⁷ were used to try to calculate .stl files. File 1atp depicts the catalytic subunit of a cyclic adenosine monophosphate (cAMP) dependent protein kinase C with two Mn⁺², adenosine triphosphate (ATP), and a polypeptide (20 amino acids) inhibitor. Amino acids threonine 197 and serine 338 are phosphorylated in the protein kinase C. The two Mn⁺² were deleted before calculating the .stl file. File 1apk depicts only amino acids 143 to 260 and cAMP. Nothing was deleted from this file. File 2apk depicts amino acids 143-264 plus cAMP. Nothing was deleted from this file. The trb program found omega to be negative, so it changed the angle for atom 121 from 0.800 to 0.500. With this correction, an .stl file was calculated and was put on CD-ROM. File 1bpk depicts amino acids 261 to 379 and cAMP. Nothing was deleted from this file. However, there were two problem faces discovered when running the trb program on 1bpk. As a result, trb was not able to triangulate the surface, and no .stl file could be calculated from PDB file 1bpk. There were no problem faces in files 1atp or 1apk, though. No models have been made of protein kinase C, but the .stl files from 1atp and 1apk are being copied onto a CD-ROM.

Zinc Finger. This protein, coded by a gene called ZIF268 immediate early gene, binds to DNA causing it to be transcribed into a messenger RNA (mRNA) which is subsequently translated into a protein. The name “zinc finger” indicates that there is a catalytic zinc and that the protein has a shape that resembles a finger. Such fingers occur in many proteins that are turned on by growth and differentiation signals. The term zinc finger has been used in many different ways, so it is important to realize that this model is different from those of cysteine-rich motifs in the steroid receptors, yeast transcription factor GAL4, or the Cys-Cys-His-Cys motif in some retroviral proteins. PDB file 1zaa¹⁸ depicts the portion of the ZIF268 gene product (zinc finger) containing three zinc fingers, corresponding to amino acids 349 to 421, bound to 11 base pairs of DNA. The three Zn⁺² atoms were deleted from the PDB file before calculating the .stl file. There is one part of this model, corresponding to arginine 87, which sticks out somewhat and was not supported during the building (SL) process. That is, there was no photocured polymer for that part of the model to stick to, so it floated off. As a result, there is a small part of this model that is missing. On any future attempts to build it, the .stl file should be rotated so that the part sticking out is on a support. The stereolithographic model of the solvent-accessible surface is shown in Figure 12. A space-filling model viewed from about the same position is shown in Figure 13. The zinc finger protein is in yellowish-green at the upper left and lower right, while the two strands of DNA are in navy blue and blue-green.

DNA Decamer. A model of 10 (decamer) base pairs (C-G-A-T-A-T-A-T-C-G) of DNA was made from PDB file 1d56.¹⁹ Sequence-dependent differences in DNA structures probably have functional significance in controlling gene expression. For example, the homopolymer poly(dA)-poly(dT) has both a bend in the helix axis and stiffness that prevent it from winding around the nucleosome core. In contrast, the alternating heteropolymer poly(dA-dT)-poly(dA-dT) is more like normal B-DNA and is capable of winding around a nucleosome core.

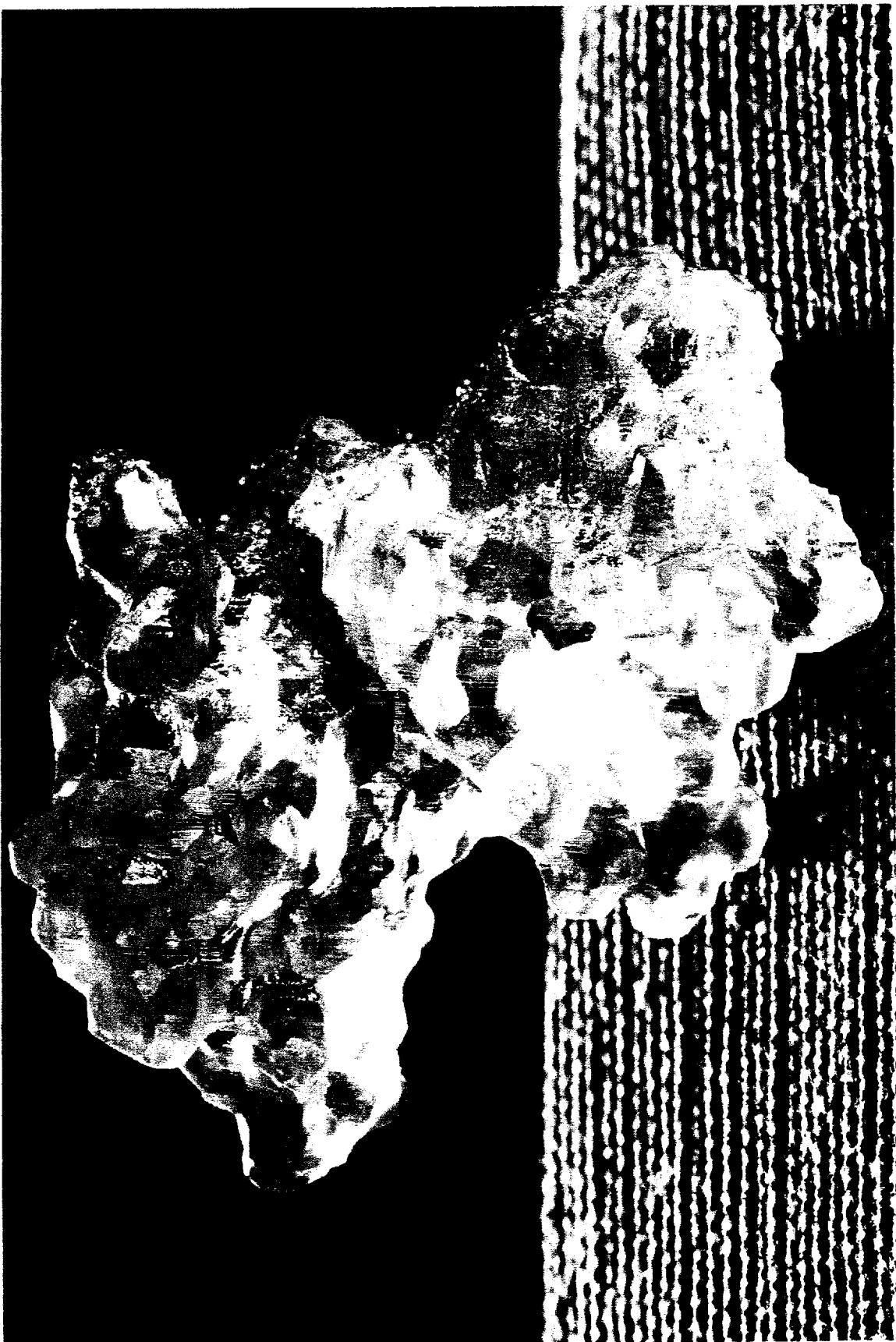


Figure 12. Solvent-Accessible Surface of ZIF268 Bound to DNA, From Brookhaven PDB File 1zaa

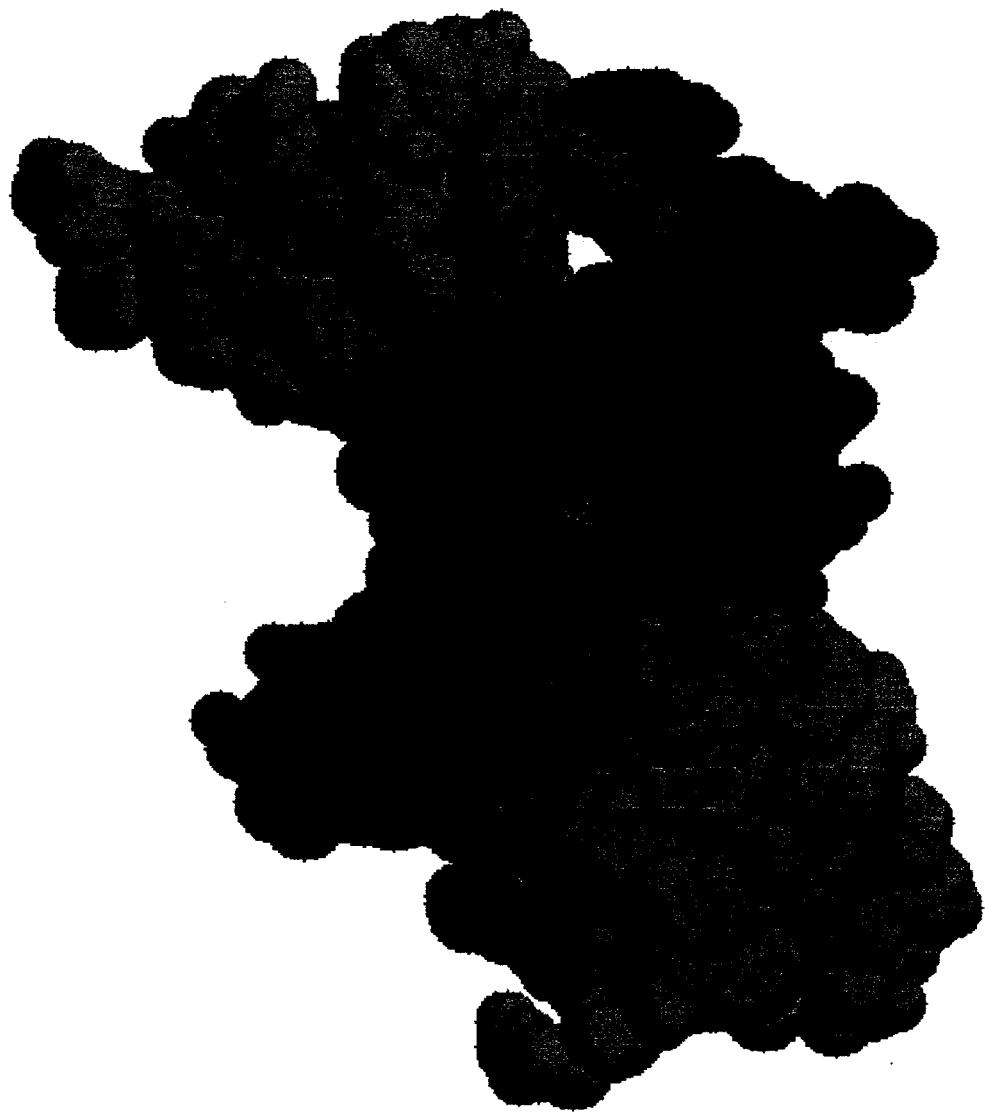


Figure 13. Space-Filling Model of ZIF268 Bound to DNA

The DNA decamer was prepared as part of a systematic study of the sequence C-G-A-T-C-G-A-T-C-G (sometimes called the KK sequence). In this case, the central two bases were inverted to yield a short alternating tract, sometimes called an AT helix. The crystal environment is the same as the KK, but the twist, or propeller, angle is larger, causing the minor groove to be wider at the center.¹⁹ Two Ca^{+2} atoms in this file were deleted before calculating the .stl file. No model has been made yet, but the stl file was put on a CD-ROM.

Oxymyoglobin. PDB file 1mbo,²⁰ which depicts sperm whale myoglobin at pH 8.4, was used. Myoglobin is similar to hemoglobin in that it contains an iron-heme group that binds oxygen. Myoglobin is in muscle tissue. It binds oxygen when the oxygen content of blood is high and releases it when blood oxygen content is low. However, myoglobin consists of only one polypeptide, unlike hemoglobin, which has four polypeptides. This file contains some redundant atomic coordinates, that is, two sets of coordinates for the same atom (e.g., 714 and 716). Redundant atoms had to be deleted before the pqms program would run. Also, the atoms due to one iron and one SO_4^{2-} were deleted before calculating the .stl file. Myoglobin is similar to hemoglobin in that it has a heme group and binds oxygen. No model has been made yet.

CD4. The glycoprotein called CD4 is on the surface of those white blood cells (T helper cells) to which HIV binds. HIV eventually kills those white blood cells, severely weakening the patient's immune system, causing AIDS. In healthy people, CD4 belongs to the immunoglobulin super family which serves in recognition processes. CD4 consists of a large (about 370 amino acids) extracellular segment composed of four tandem immunoglobulin-like domains (D1, D2, D3, and D4), a single transmembrane span, and a short (38 amino acids) COOH-terminal tail. Crystals of whole soluble CD4 have been grown, but they have limited diffraction order.²¹ The PDB file 1cd4,²¹ which depicts the N-terminal 183 amino acids (called the D1D2 structure) of CD4 plus two missense amino acids, aspartate and threonine, was used. This N-terminal fragment is the HIV binding fragment. No model has been made yet, but the .stl file is on CD-ROM.

Calmodulin. This is a protein that binds calcium and regulates many different cellular activities. It is a small (molecular weight 16,700) acidic protein with four Ca^{+2} binding sites. It plays a pivotal role in the regulation of several enzymes. Calmodulin belongs to the family of calcium-binding proteins (including troponin-C, parvalbumin, and the intestinal Ca^{+2} -binding protein). These proteins have a helix-loop-helix structural motif often referred to as the EF-hand. The EF-hand comprises about 33 amino acids in a linear sequence. The central Ca^{+2} -binding loop has 12 amino acids, and the two flanking α -helices have from 10 to 14 amino acids. The angle between the helical axes, E-F, varies from about 110° in the Ca^{+2} -filled protein to between 133° to 150° in the Ca^{+2} -free protein. PDB file 1osa,²² which depicts calmodulin from paramecium, was used. There are 148 amino acids in this calmodulin, but only residues 5-147 are included in this PDB file. The positions of amino acids 1-4 and 148 are uncertain at this time. Also, the four Ca^{+2} in this file were deleted. No model has been made yet.

Cytochrome P-450. This protein is important in liver metabolism. It helps degrade and detoxify substances that reach the liver. P-450 enzymes participate in steroid hormone and bile acid biosynthesis, vitamin activation, fatty acid hydroxylation, and detoxification of various solvents,

antibiotics, anesthetics, insecticides, and polycyclic aromatic hydrocarbons. Unfortunately, the detoxification is sometimes associated with the carcinogenic activation of some compounds. A better understanding of P-450 structure should help inhibitor design for harmful P-450 isozymes as well as the engineering of new P-450-like isozymes capable of metabolizing environmentally hazardous organic compounds.²³ P-450 has an iron and a heme group. PDB file 1pha,²³ which depicts a cytochrome P-450 (from a bacterium called *Pseudomonas putida*) with a small molecule called camphor bound to it, was used. This protein is sometimes called camphor 5-monoxygenase. The one iron atom in this file was deleted before calculating the .stl file. No model has been made yet.

Ferricytochrome C. This protein is involved in photosynthesis and respiration in prokaryotes and eukaryotes. It contains an iron (Fe^{+3}) and a protoporphyrin IX (similar to a heme group). PDB file 1ccr,²⁴ which depicts cytochrome c from rice, was used, but the Fe^{+3} was deleted. The rice cytochrome c was studied to help characterize the differences between plant and animal forms. There are some unusual (or modified) amino acids in this cytochrome c. There are trimethyllysines at amino acids 2 and 3. Lysine 80 and trimethyllysine 2 form an ϵ -N-trimethyllysine as do lysine 94 and trimethyllysine 3. No model has been made yet.

Ubiquitin. Ubiquitin is a small, highly conserved protein that functions in selective protein degradation as well as direct modification of the structure and function of proteins.²⁵ The formation of ubiquitin conjugates requires ATP and requires three steps. The first step is catalyzed by a specific ubiquitin-conjugating enzyme. In ubiquitin-mediated protein degradation, multiple molecules of ubiquitin are attached to the target protein and these conjugates are recognized and degraded by specific proteases. PDB file 1aak²⁵ was used. A .stl file was calculated for the CD-ROM, but no model has been made yet.

Antifreeze Protein. PDB file 1atf²⁶ was used. It is based on NMR data. This protein is present in flounder that live in water at temperatures below 0°C. To keep the water in them from freezing, their cells contain antifreeze proteins. The structure of the antifreeze protein is like a zipper that can bind to ice crystals as they start to form. The hydroxyl groups of four threonines (numbers 2, 13, 24, and 35) form hydrogen bonds with the water, stopping or retarding the growth of ice crystals.²⁶ The pqms file was made using the atomic radii in elements2.rad. No problem faces or problem atoms were found in pqms. An .stl file was successfully calculated and put on CD-ROM. No model has been made yet.

p53. The protein coded by the p53 gene is mutated in about half of all human cancers.²⁷ It is a tumor suppressor, regulator of the cell cycle, and DNA repair enzyme. It binds to specific sites of DNA (trans, or distant, from the p53 gene itself), turning on transcription (of DNA to RNA), and turning off transcription at other specific sites. Also, when p53 is over expressed (present at higher concentration), it may interact with cytomegalovirus to cause coronary restenosis, or blood vessel narrowing, common after angioplasty.²⁸

The p53 protein is a dimer of dimers. It has several structural domains. There is an amino-terminal transactivating domain, a DNA-binding core domain, an inhibitory domain, and a tetramerization domain. PDB file 1pes²⁹ contains the tetramerization domain of p53. The DNA

binding core is made up of amino acids 102-292. Most cancer-causing mutations occur in this region. The tetramerization domain is comprised of amino acids 319-360. Its structure was solved by NMR, so hydrogens are included in the structure. Thus, when calculating the molecular surface with the program pqms, elements2.rad was used to provide the atomic radii. The program pqms found six problem faces and 14 problem atoms on its first try. On its second try, it increased the atomic radii of arg 335 NH2, arg 337 NH1 and NH2, leu344 CG (i.e., γ -carbon) and HG (i.e., γ -hydrogen), glu 326 CD (i.e., δ -carbon) and OE1 (i.e., ϵ -oxygen #1), lys 351 CD and HD, arg 337 NH1 and HH1, leu 344 HG, and lys 351 CD and HD. The atomic radii of the problem nitrogens (N) were increased from 1.700 to 1.800, carbons from 1.800 to 1.900, oxygen from 1.600 to 1.700, and hydrogens from 1.00 to 1.100. Once this was done, pqms calculated a piece-wise quartic molecular surface with no problem faces. However, the trb program was unable to triangulate the surface. It gave the message check vertex (9625) lfn = 0.

Next, the pqms program was re-run, using the msexpanded.rad table for atomic radii, but the same error occurred. There is also a PDB file 1olg,³⁰ which has the p53 tetramerization domain. The program pqms found three problem faces and nine problem atoms. It increased the radii of arg 342 HG and HE, glu 346 HB, arg 342 HG and HE, glu 346 HB, arg 342 HG and HE, and glu 346 HB. The trb program found two angles that were negative. It changed the angle for atom 367 and atom 2461 from 0.800 to 0.500. Then it stated that there was a memory fault (core dump). The pqms program was re-run using the msexpanded.rad table of atomic radii. This time the trb program ran, after changing several omega angles, but it calculated an Euler characteristic of 4 for the molecular surface. However, the vet2tri, tri2t, vt and tri2stl programs ran and an .stl file was calculated.

c-myb Proto-oncogene. Oncogenes are genes that code for proteins that help control cell growth and/or differentiation. When they are mutated, they can cause cancer (oncology is the study of cancer, thus the name oncogene). The c-myb proto-oncogene is the cellular counterpart to two avian (bird) retroviruses (AMC and E26) that can cause cancer in myeloid cells.³¹ Related genes have been found in all eukaryotic species from yeast to humans. The region most similar in the different myb gene-coded proteins is a sequence at the N-terminal end which consists of imperfect repeats of 50-53 amino acids and which correspond to the DNA-binding domain. Proteins containing the c-myb-related repeat domain act as transcriptional (transcribing DNA to RNA) regulators.

Secondary structure predictions suggested that there is an arrangement of α -helices homologous to that seen in the homeodomains, members of the helix-turn-helix family of DNA-binding proteins. Frampton and coworkers³¹ used molecular modeling and site-directed mutagenesis to test the feasibility of this structure. They produced a theoretical model, not based on NMR or X-ray crystallography, and deposited the atomic coordinates in PDB file 1pom³¹ which was used without any deletions to try to calculate the .stl file. Because this file includes hydrogens, atomic radii from elements2.rad were used. No problem faces were found, but a memory fault (core dump) occurred while running the trb program. Thus, the surface could not be triangulated and no .stl file was calculated.

Trimethylcalmodulin With Four Ca^{+2} and Trifluoperazine. Strydnadka and James³² developed a molecular model of calmodulin complexed with the anti-psychotic drug, trifluoperazine. The model was based on the prediction that calmodulin's structure would closely resemble that of troponin-C, whose structure had already been solved by X-ray diffraction. The atomic coordinates of the modeled structure were deposited in PDB file 2cln,³² which was used after deleting the four Ca^{+2} and trifluoperazine. The .stl file will be put on CD-ROM, but no model will be made.

Cytotoxic T-Lymphocyte Proteinase. Cytotoxic T-lymphocytes bind to cells that express foreign molecules (antigens) and kill them (induce them to lyse). This class of lymphocytes (white blood cells) is one of the main effectors of cell-mediated immune responses that includes the destruction of cells infected with viruses and cancer cells. These cells are killed, in part, by destroying the cells' proteins, using proteinases. Murphy and coworkers³³ cloned a gene called CCP1 (cytotoxic cell protease 1) that is expressed when precursor cytotoxic T-lymphocytes are transformed to T killer cells. They solved the structure using comparative molecular modeling and deposited the atomic coordinates in PDB file 2cp1,³³ which was used with no deletions to calculate an .stl file. The .stl file was put on a CD-ROM, but no model will be made.

Amicyanin. Thermobacillus versutus is a chemolithotropic bacterium that uses sulfur compounds for its energy metabolism. This bacterium can grow on methylamine and uses a copper-containing protein called amicyanin to act as an electron transport acceptor from methylamine dehydrogenase. Van Beeumen and coworkers used X-ray diffraction to solve the structure of amicyanin, and they deposited the atomic coordinates in PDB file 1aaj,³⁴ which was used with no deletions to calculate an .stl file. The .stl file will be put on a CD-ROM, but no model will be made.

Alpha-1 Amphiphilic Helix. This protein was studied as part of a project to better understand how proteins fold into their active conformation upon being made in the cell. The specific aim was to design a four α -helical bundle by synthesizing a peptide that would self-associate. A 12-amino acid chain that self-associated through hydrophobic leucine residues was prepared, and the data was entered into PDB file 1a11.³⁵ It was used to calculate an .stl file after deleting the SO_4^{2-} . There is an acetyl group attached to the N-terminal amino acid. Also, the PDB file does not have the C₈, C_ε or N for lys 5 and lys 11. The .stl file was put on CD-ROM, but no model will be made.

Avidin-Biotin Complex. PDB file 2avi³⁶ was used with no deletions. Avidin is a homodimer, containing two biotins and two N-acetylglucosamines in this file. The .stl file was put on a CD-ROM, but no model will be made.

α -Lytic Protease. α -lytic protease is an extracellular protease from Lysobacter enzymogenes. That is, it uses the amino acid serine in its active site to break up, or hydrolyze, proteins. α -lytic protease is specific for small proteins (peptides) containing succinyl-Ala-Ala-Pro.³⁷ The PDB file 4lpr,³⁷ which depicts a mutant α -lytic protease with met 192 changed to ala, complexed with

methoxysuccinyl-ala-ala-pro-leu-boronic acid, was used after deleting a SO_4^{2-} and the boronic acid. This boronic acid-containing peptide is an analogue of a high energy reaction intermediate. Mutations like this one produce an altered α -lytic protease with broad specificity.³⁷ The detailed structure helped researchers understand the basis for this change to broad specificity. The .stl file was put on CD-ROM, but no model will be made.

Acetylcholine Esterase. This enzyme catalyzes the hydrolysis of the neurotransmitter, acetylcholine. It is important that nerve impulses be terminated rapidly, so this enzyme has a high specific activity and functions at a rate approaching that of a diffusion-controlled reaction. By hydrolyzing this neurotransmitter, the body ensures that the neurological message is short-lived, preparing the receptor cell for more instructions, or just letting it relax to its unstimulated state. The powerful acute toxicity of organophosphorous poisons is due to their potent inhibition of this enzyme. Some acetylcholine esterase inhibitors are used in the treatment of glaucoma and myasthenia gravis, and their use has been proposed as a treatment for Alzheimer's disease.

Knowledge of the three-dimensional structure of acetylcholine esterase is important for understanding its remarkable enzyme activity, for rational drug design, and for developing treatments for organophosphate poisoning. This is an unusually difficult enzyme to crystallize, since it is a membrane-bound enzyme. It is a homodimer, attached to the plasma membrane through covalently attached phosphatidylinositol, attached to the COOH-terminus of each monomer through an intervening oligosaccharide. Sussman et. al.³⁸ solubilized the enzyme using a bacterial phosphatidylinositol-specific phospholipase C, which broke the acetylcholine esterase off its glycolipid anchor. PDB file 1ace,³⁸ which depicts acetylcholine esterase from an electric ray, *Torpedo Californica*, with acetyl choline attached, was used. The enzyme belongs to a class of α/β proteins, consisting of a 12-stranded central mixed β sheet surrounded by 14 α helices.³⁸

A remarkable feature of this enzyme is a deep, narrow gorge about 20 Å long that penetrates halfway into the enzyme and widens out close to its base. Sussman et. al.³⁸ named this the active site gorge because it contains the catalytic triad (Serine 200, Histidine 440 and Glutamate 327). This gorge is apparent in the solvent-accessible surface model.

D-Xylose Isomerase. D-xylose isomerase catalyzes the conversion of D-xylose to D-xylulose, and, more slowly, D-glucose to D-fructose. Only α -anomers are used. During catalysis a proton migrates in a completely stereospecific way (no exchange with water solvent) between two adjacent carbons, C-1 and C-2 of the substrate.³⁹ The structure of D-xylose isomerase from *Streptomyces rubiginosus* and of its complexes with substrate and an active site-directed inhibitor were determined by X-ray diffraction.³⁹ The active site was identified, along with two metal ion binding sites. PDB file 8xia,³⁹ which depicts the xylose isomerase from *Streptomyces rubiginosus*, with D-xylose attached, was used. Two Mn^{+2} were deleted from the file before calculations began. The .stl file will be put on CD-ROM, but no model will be made.

BDS, Anti-hypertensive, Anti-viral Protein. Sea anemones contain several pharmacologically active polypeptides of molecular weight around 5000.⁴⁰ Some are cardiotoxic and neurotoxic. One small polypeptide, BDS-I, is neither cardiotoxic nor neurotoxic. Instead, it lowers blood

pressure and has antiviral activity, but its mode of action is unknown.⁴⁰ PDB file 1bds,⁴⁰ which depicts this protein from sea anemone (*Anemonia sulcata*), was used without deletions. As a primary reference, this file cites G.M. Clore and others as experimental data. It then cites the article in Biochemistry⁴⁰ as the second reference. The results described in this reference indicate that the BDS-I protein contains a triple-stranded antiparallel β -sheet, similar to that found in the partially homologous sea anemone polypeptide toxins. The data were obtained by NMR, so the atomic radii from the table elements2.rad were used. No problems were encountered during the calculations, which produced an .stl file for CD-ROM.

Bacteriorhodopsin Amino Acids 8-63, 74-100, 106-127, 131-191 and 202-225. Bacteriorhodopsin from the purple membrane of *Halobacterium halobium* acts as a light-driven H⁺ pump. The purple membrane is almost 75% bacteriorhodopsin, with lipid making up the remaining 25%. PDB file 1brd,⁴¹ which depicts several fragments of bacteriorhodopsin from *Halobacterium halobium* with attached rhodopsin, was used. The data were obtained by cryo-electron microscopy.

E. Coli (Mutant Strain M41 of Serotype K29) Capsular Polysaccharide. Polysaccharides are present on the cell wall of bacteria as a capsule, or extracellularly as a slime. The capsular polysaccharide from *E. coli* strain Bi161/42(09:K29(A):H-) displays serotype 29 and consists of hexasaccharide repeating units.⁴² This *E. coli* K29 polysaccharide is the receptor of *E. coli* phage (virus) 29, which is highly specific for this capsule. Its structure was solved by X-ray diffraction and atomic coordinates deposited in PDB file 1cap,⁴² which was used to calculate the .stl file. The PDB file contains no amino acids. It contains only the atoms in the *E. Coli* polysaccharide.

Arabinose Binding Protein Mutant With Met 108 Replaced by Leu. L-arabinose binding protein is a member of the family of proteins found in the periplasm of Gram negative bacteria.⁴³ These proteins are essential components of the osmotic shock-sensitive active transport (energy-requiring) systems for carbohydrates, amino acids, and ions.⁴³ L-arabinose binding protein consists of a single polypeptide chain with a high affinity binding site that accommodates both α - and β -anomers of L-arabinose. PDB files 6abp and 7abp⁴³ were used without any deletions.

α -Amylase. Neutral α -amylase from *Aspergillus oryzae* (TAKA amylase) is a glycoprotein which catalyzes the hydrolysis of internal α (1-4) glycosidic linkages in various polysaccharides.⁴⁴ This enzyme consists of a single polypeptide chain of 478 amino acid residues. PDB file 6taa⁴⁴ was used without any deletions to successfully obtain an .stl file, which was put on CD-ROM.

ACCOMPLISHMENTS

Eight stereolithographic models of the solvent-accessible surfaces of various biopolymers were made from ISO 9001 standard .stl files. A CD-ROM with four .stl files, depicting the ras protein in different shapes, was sent to Procter & Gamble Pharmaceutical Research (Cincinnati, OH),

who used them to make their own stereolithographic models. More .stl files ⁴⁰were calculated and put on CD-ROMs, to be made available to American industry.

FUTURE WORK

More CD-ROMs will be made, containing .stl files. As many .stl files as possible will be calculated from the remaining Brookhaven PDB files. More stereolithographic models will be made. Collaborations with X-ray crystallographers will be established.

REFERENCES

- ¹ R. Sayle, "RasMol Molecular Renderer, Version 2.3," Feb. 1994.
- ² F. C. Bernstein et. al., **J. Mol. Biol.**, 112, 535 (1977).
- ³ M. L. Connolly, **J. Appl. Crystallogr.**, 18, 499 (1985).
- ⁴ M. V. Milburn, L. Tong, A.M. de Vos, A. Brunger, Z. Yamaizumi, S. Nishimura, and S.-H. Kim, **Science**, 247, 939 (1990).
- ⁵ L. Tong, A.M. de Vos, M.V. Milburn, and S.-H. Kim, **J. Mol. Biol.**, 217, 503 (1991).
- ⁶ G. Fermi, M. F. Perutz, B. Shaanan, and R. Fourme, **J. Mol. Biol.**, 175, 159 (1984).
- ⁷ R. Liddington, Z. Derewenda, E. Dodson, and R. Hubbard, **J. Mol. Biol.**, 228, 551 (1992).
- ⁸ J. Erickson, **Science**, 249, 527 (1990).
- ⁹ M. Fujinaga, A. R. et. al., **J. Mol. Biol.**, 195, 397 (1987).
- ¹⁰ M. D. Hall et. Al., **J. Mol. Biol.**, 226, 867 (1992).
- ¹¹ J. N. Herron, **Proteins Struct. Funct.**, 5, 271 (1987).
- ¹² C. D. Stout et. al., **Acta Crystallogr. B.**, 34, 1529 (1978).
- ¹³ D. N. Garboczi, D. T. Hung, and D. C. White, **Proc. Natl. Acad. Sci.**, 89, 3429 (1992).
- ¹⁴ A. T. Danishefsky, et. al., **Biochemistry**, 30, 1980 (1991).
- ¹⁵ F. Colonna et. al., **J. Biol. Chem.**, 261, 15273 (1986)
- ¹⁶ D. R. Knighton et. al., **Science**, 253, 407 (1991).
- ¹⁷ I. T. Weber, et. al., **Biochemistry**, 26, 343 (1987).
- ¹⁸ N. P. Pavletich and C. O. Pabo, **Science**, 252, 809 (1991).
- ¹⁹ H. Yuan, J. Quinatana, and R. E. Dickerson, **Biochem.**, 31, 8009 (1992).
- ²⁰ S. E. V. Phillips and B. P. Schoenborn, **Nature**, 292, 81 (1981).
- ²¹ M. Rosenberg et. al., **Nature**, 348, 419 (1990).

- ²² S. T. Rao et. al., **Protein Sci.**, 2, 436 (1993).
- ²³ R. Raag et. al., **Biochem.**, 32 4571 (1993).
- ²⁴ H. Ochi et. al., **J. Mol. Biol.**, 166, 407 (1983).
- ²⁵ W. J. Cook et. al., **J. Biol. Chem.**, 267, 15116 (1992).
- ²⁶ K. C. Chou, **J. Mol. Biol.**, 223, 509 (1992).
- ²⁷ Y. Cho et. al., **Science**, 265, 346 (1994).
- ²⁸ E. Speir et. al. **Science**, 265, 391 (1994).
- ²⁹ W. Lee et. al., **Nature Struct. Biol.**, 1, 877 (1994).
- ³⁰ G. M. Clore et. al., **Science**, 265, 386 (1994).
- ³¹ J. Frampton et. al., **Protein Eng.**, 4, 891 (1991).
- ³² N. C. J. Strynadka and M. N. G. James, **Protein Struct. Funct.**, 3, 1 (1988).
- ³³ M. E. P. Murphy et. al., **Proteins Struct. Funct.**, 4, 190 (1988).
- ³⁴ R. C. Durley et. al. **Protein Sci.** 2, 739 (1993).
- ³⁵ C. P. Hill et. al., **Science**, 249, 543 (1990).
- ³⁶ O. Livnah et. al., **Proc. Natl. Acad. Sci.**, 90, 5076 (1993).
- ³⁷ R. Bone et. al., **Biochem.**, 30, 10388 (1991).
- ³⁸ J. L. Sussman et. al., **Science**, 253, 872 (1991).
- ³⁹ H. L. Carrell et. al., **Proc. Natl. Acad. Sci.**, 86, 4440 (1989).
- ⁴⁰ P. C. Driscoll et. al., **Biochem.**, 28, 2188 (1989).
- ⁴¹ R. Henderson et. al., **Nature**, 257, 28 (1975).
- ⁴² R. Moorehouse et. al., **J. Mol. Biol.**, 109, 373 (1977).
- ⁴³ F. A. Quiocio et. al., **Nature**, 310, 381 (1984).
- ⁴⁴ H. J. Swift et. al., **Acta Crystallogr. B.**, 47, 535 (1991).