



Spring 5-1999

# Use of Computational Docking Procedure to Identify Potential Inhibitors of R67 Dihydrofolate Reductase

Ushma Jyotindra Shukla  
*University of Tennessee - Knoxville*

Follow this and additional works at: [https://trace.tennessee.edu/utk\\_chanhonoproj](https://trace.tennessee.edu/utk_chanhonoproj)

---

## Recommended Citation

Shukla, Ushma Jyotindra, "Use of Computational Docking Procedure to Identify Potential Inhibitors of R67 Dihydrofolate Reductase" (1999). *University of Tennessee Honors Thesis Projects*.  
[https://trace.tennessee.edu/utk\\_chanhonoproj/341](https://trace.tennessee.edu/utk_chanhonoproj/341)

This is brought to you for free and open access by the University of Tennessee Honors Program at Trace: Tennessee Research and Creative Exchange. It has been accepted for inclusion in University of Tennessee Honors Thesis Projects by an authorized administrator of Trace: Tennessee Research and Creative Exchange. For more information, please contact [trace@utk.edu](mailto:trace@utk.edu).

# UNIVERSITY HONORS PROGRAM

## SENIOR PROJECT - APPROVAL

Name: Ushma Shukla

College: LA& Sciences Department: Biochemistry

Faculty Mentor: Dr. Elizabeth Howell

PROJECT TITLE: use of computational Docking  
Procedure to Identify Potential Inhibitors  
of R67 Dihydrofolate Reductase

I have reviewed this completed senior honors thesis with this student and certify that it is a project commensurate with honors level undergraduate research in this field.

Signed: Elizabeth Howell, Faculty Mentor

Date: 5-12-99

Comments (Optional):

**Use of Computational Docking Procedures to  
Identify Potential Inhibitors of R67  
Dihydrofolate Reductase**

Ushma Shukla  
Senior Thesis  
Faculty Mentor: Dr. Elizabeth Howell  
May 1999

## **Abstract**

Indiscriminate use of antibiotics has created a selective advantage for resistant strains of bacteria. Resistance is spread through bacterial conjugation, whereby plasmids that contain genes for alternate enzymes are transmitted directly between bacteria. The R-plasmid encoded enzyme R67 dihydrofolate reductase (DHFR) confers resistance to the antibacterial drug trimethoprim. Typically, chemical screening for new drugs involves modification of the known ligands for a given enzyme. On the other hand, novel structures can be identified through computer screening. The program suite DOCK developed by Irwin Kuntz, et al. at UCSF docks ligands into an enzyme of known structure based on Van der Waals interactions. First, the program, MS, developed by Michael Connolly, creates a surface by calculating the solvent accessible surface over the crystal structure (obtained from the Protein Data Bank). Then SPHGEN fills the invaginations of the active site with a set of overlapping spheres. A negative image of the site is created from the set of sphere centers. The ligand interactions are evaluated by constructing a grid over the protein structure using GRID. Ligands from the Cambridge Structural Databank (CSD) were scored for Van der Waals interactions with R67 using DOCK and the top 1000 scoring molecules were run again considering more orientations. The top scoring compounds were identified for testing in the laboratory as potential inhibitors of R67 DHFR.

## **Introduction**

### Antibiotic Resistance

In 1929, Alexander Fleming discovered penicillin, the first antibiotic. Initial optimism about the universal efficacy of antibiotics led to their overuse. Antibiotics are used as a common addition to animal feed, in order to expedite their marketability. The indiscriminate use of antibiotics has resulted in a selective advantage for bacterial resistance. High levels of resistance usually indicate the presence of a Resistance plasmid or R-plasmid. Resistance plasmids are spread through the phenomenon of bacterial

conjugation, which involves the transfer of genetic material. Though it is impossible to determine when these plasmids first evolved, there is evidence to suggest that they predate the use of antibiotics. For example, resistance plasmids have been found in non-pathogenic soil bacteria. A selective pressure may have existed in the soil because many antimicrobial agents such as penicillin are soil organisms. Also a strain of *E. coli* that was frozen in 1946, was found to contain plasmids conferring resistance to tetracycline and streptomycin. Neither drug was used clinically until many years later. Thus the indiscriminate use of antibiotics has not sped the evolution of novel resistance mechanisms, but has simply created a selective advantage for the transmission of the R-plasmids. R-plasmids contain genes that code for new proteins that either inactivate the antibacterial drug, actively pump it out of the cell, or simply prevent its uptake. Once these proteins are identified they become the targets for new drugs.

Traditional drug screening involves modification of the known ligands through the addition or removal of functional groups. Though effective, this mechanism feeds the evolution of the enzymes through mutation of the residues that interact with the ligands. Inhibitors developed in this manner do not require a complete overhaul of the enzyme structure to be rendered ineffective. Thus, it becomes a race between the development of mutations within the bacteria, and the development of new drugs by the researcher (1).

#### A Resistance Enzyme: R67 Dihydrofolate Reductase

The enzyme dihydrofolate reductase (DHFR) converts dihydrofolate (DHF) to tetrahydrofolate using nicotinamide adenine dinucleotide phosphate (NADPH) as a cofactor. It thus replenishes folate cofactors required for nucleotide synthesis (2). The

chromosomal enzyme is inhibited by trimethoprim, an antibiotic administered with sulfonamides. R67 dihydrofolate reductase, an R-plasmid encoded protein, bears no homology to the chromosomal DHFR either in sequence or structure. Thus it confers resistance to the antibiotic.

R67 DHFR is a roughly doughnut shaped homotetramer  $30 \times 35 \times 40 \text{ \AA}^3$  (Figure 1). A  $25 \text{ \AA}$  pore with 2,2,2 symmetry is found along the length of the molecule (3). The active site pore narrows to  $12 \text{ \AA} \times 9 \text{ \AA}$  in the center. Since R67 is a homotetramer, four identical folate binding sites are expected. However, for steric reasons, only two ligands are able to bind asymmetrically to symmetry-related binding sites (4). R67 DHFR will bind either two folate molecules or two NADPH molecules in an unproductive complex or one NADPH plus one folate in a productive complex. However, the binding of NADPH inhibits the binding of a second NADPH molecule and thus facilitates the binding of DHF to form a productive complex. There is no evidence of an induced-fit mechanism in the binding of the ligands. Thus, R67 DHFR is a good choice for the rigid approximation of a computational binding study.

#### The DOCK Program Suite

The DOCK program suite characterizes the active site of the enzyme of interest, matches potential ligand molecules to that site, and then scores the complementarity of the ligand to the active site (5)(6). As input, DOCK requires crystal structures of the protein and ligand molecules, and a molecular surface file generated by Michael Connolly's program Molecular Surface (MS).

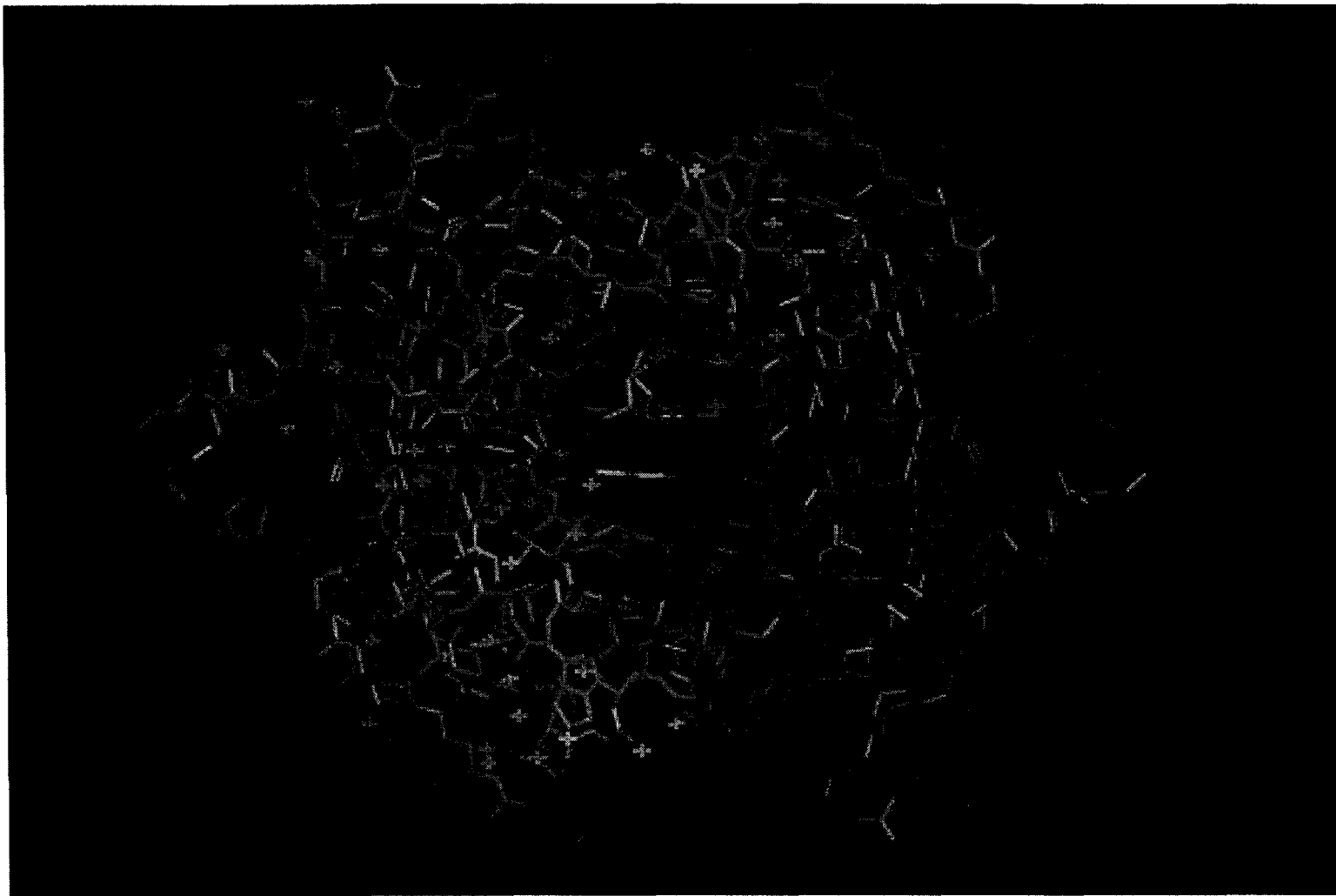


Figure 1: R67 DHFR with Folate (bound) and NADP+ (docked).

MS estimates the solvent accessible surface of the receptor by rolling a water molecule over the surface of the protein (7). The first DOCK program, SPHGEN, then estimates the grooves and invaginations of this surface with a set of overlapping spheres. It then groups these spheres into clusters. This creates a negative image of the active site (8). Next, a scoring grid is created from the original receptor file with the program GRID to provide a means of evaluating ligand complementarity. DOCK compares the structures of the potential ligands to the image of the active site by comparing internal distances based on van der Waals radii. This is known as contact scoring. The number of orientations to be tried, and other input parameters, are specified in the dock.in file. In addition, to contact scoring, DOCK is able to take into account chemical complementarity, flexibility of ligand and receptor, energetic considerations, and electrostatic interactions with the solvent and receptor molecule. With these considerations, DOCK is able to provide more accurate predictions of ligand interactions (9).

DOCK can be used to analyze the interaction of a known or potential ligand in detail, or it can be used to screen a database of molecules for potential drug targets. The following assumptions are made in the latter case: the receptor and ligands are rigid, water and counter-ions are neglected, interaction energies are simplified. Computing time increases enormously with the incorporation of these factors, therefore it is impractical to consider them while screening a large number of molecules (8).

## The Role of Electrostatics

Electrostatic interactions are known to play an important role in the binding of ligands to the chromosomal DHFR in *E.coli*. The molecule carries a net charge of -10, yet it binds NADPH and folate, both of which have net negative charges. An analysis of the electrostatic charge distribution revealed a positive charge distribution at the active site. This is thought to be a result of the presence of positively charged residues at the entrance to the ligand binding site (10). Electrostatic potential is calculated with the Finite Poisson Boltzmann Method.

R67 DHFR has an overall neutral charge and binds negative ligands, analysis of the electrostatic charge distribution could help elucidate the mechanism of ligand binding. Also, the charge distribution should be taken into account while identifying potential ligands for docking.

The program DelPhi, developed by Barry Honig, is a finite Poisson Boltzmann difference solver (11). It takes not only charge interactions within the macromolecule into consideration, but also the interaction of those charges with the solvent. The presence of a charged residue on the surface of the molecule induces a dipole moment onto nearby solvent molecules. This induced dipole is strongest near the point charge, and decreases with distance. The induced dipole of the solvent serves to connect areas of similar potential.

To visualize this, DelPhi generates a potential map that can be viewed with the molecular modeling software Insight (Biosym Technologies, San Diego, CA), in which red is negative, white neutral, and blue positive (12). Analysis of electrostatic potential can refine a docking search by eliminating molecules with inhibited binding.

## **Materials and Methods.**

All programs were run on a Silicon Graphics IndigoII computer (Silicon Graphics, Inc. Mountain View, CA). A step by step protocol is included in Appendix 1.

### Preparation of Ligand Files

Potential ligand molecules from Cambridge Structural Database (13) were screened using Quest to exclude organometallic compounds. The remaining 20,000 molecules were saved as a \*.mol2 format file.

### Preparation of the Receptor

Docked ligands and crystallographic waters were removed from the Protein Databank (PDB) format receptor file in Insight. The PDB file was then run through MS, by a DOCK accessory program called AUTOMS. Unnecessary parts of the molecule were saved as an exclude.pdb file using another modeling program called Sybyl (Tripos, Inc. St. Louis, MO). The output MS file was fed into SPHGEN. SPHGEN parameters were specified in a file called INSPH (see appendix). Sphgen results were viewed using the DOCK accessory program, SHOWSPHERE.

Unwanted sphere clusters were deleted by hand in Sybyl, and the results were saved as a \*.sph file using PDBTOSPH. Next, a scoring grid was created using GRID. Input parameters for GRID were specified in a file called grid.in. Grid results were viewed using SHOWBOX.

### DOCK

DOCK input parameters were specified in dock.in file (see appendix).

Compounds from the CSD were then docked with R67. Fifty orientations were scored for each crystal structure. The top 1000 scoring compounds were rerun considering 400 orientations.

For a more realistic model of ligand binding, a new Sphere file was created with the crystallographic waters intact. Molecules from the CSD were then docked with this new sphere cluster. Also, a mol2 file of NADP<sup>+</sup> and of folate were created in INSIGHT and docked with R67 as a basis for comparison. The NADP<sup>+</sup> and folate coordinates were obtained from the chromosomal DHFR structure.

Results were viewed by dividing the output files into manageable portions, and viewing them in INSIGHT referenced to the R67 molecule (14). The top scoring molecules were alphabetized using a program written by Rod Bunn (15), to create a reference code list that could be viewed in the Cambridge Structural Database.

#### DelPhi Electrostatics

Electrostatic Potential Maps were created using DelPhi. The default charge and size files obtained with the program were used. The influence of Lysine residues on the electrostatic potential in the pore was then analyzed by removing the charge for the amino acid in the charge file and viewing the resulting potential map. The process was repeated for Histidine and Arginine. Alternatively, individual lysines were mutated on the computer and replaced by methionine residues.

## Results and Discussion

### SPHGEN

The negative image created with SPHGEN reflects the active site symmetry of R67 (Figures 2 & 3). Viewed from the front, the pore appears split into four chambers. When viewed from the side the negative surface looks like two flattened cones oriented perpendicular to one another (Figure 4). The negative surface provides a means of visualizing the asymmetric binding of the ligands. Also, it is clear that only two ligands can bind due to the van der Waals interactions between residues in the pore.

### DOCK

Some difficulties were encountered while running GRID. It proved complicated to limit the size of the scoring grid to only the active site of the enzyme. This may have contributed to the length of time required to run DOCK.

Due to computational constraints, the number of heavy atoms in the ligand molecules was limited to 50. This is unrealistic, since folate and NADP<sup>+</sup> possess 60-70 heavy atoms.

The top scoring compound BZANTC10 scored at -210. In comparison, folate, which was scored for 400 orientations scored -106, and NADP<sup>+</sup> scored -149. Many of the compounds scored higher than the natural ligands, based on a comparison distances to residues in the active site pore.

DOCK identified many modified purines and pyrimidines (Figures 5 & 6). These results were not surprising, since purines and pyrimidines are similar in structure to the natural ligands. However, DOCK also identified compounds of novel structure such as a crown ether (Figure 7). Upon viewing the compounds in the CSD, many were too

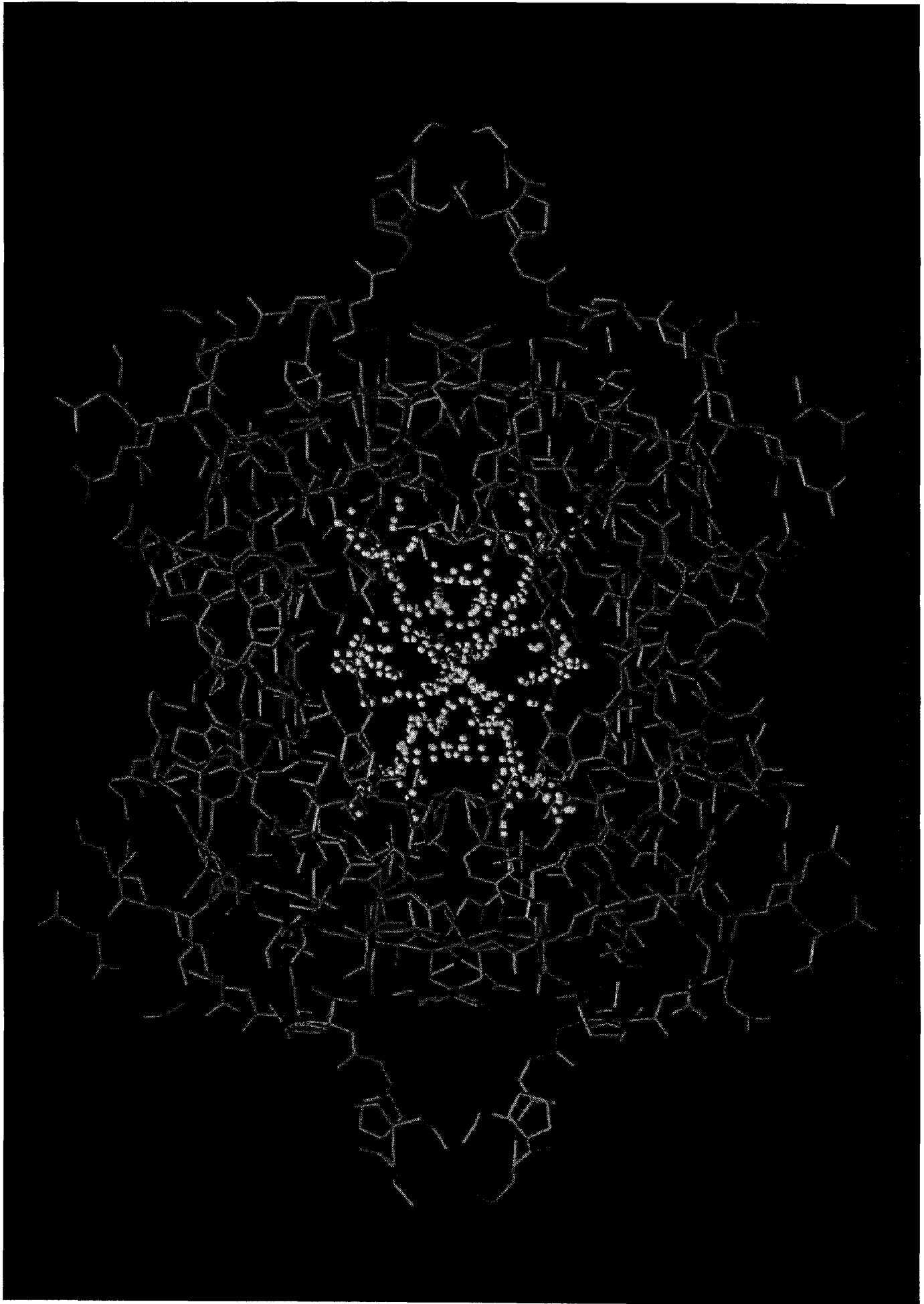


Figure 2

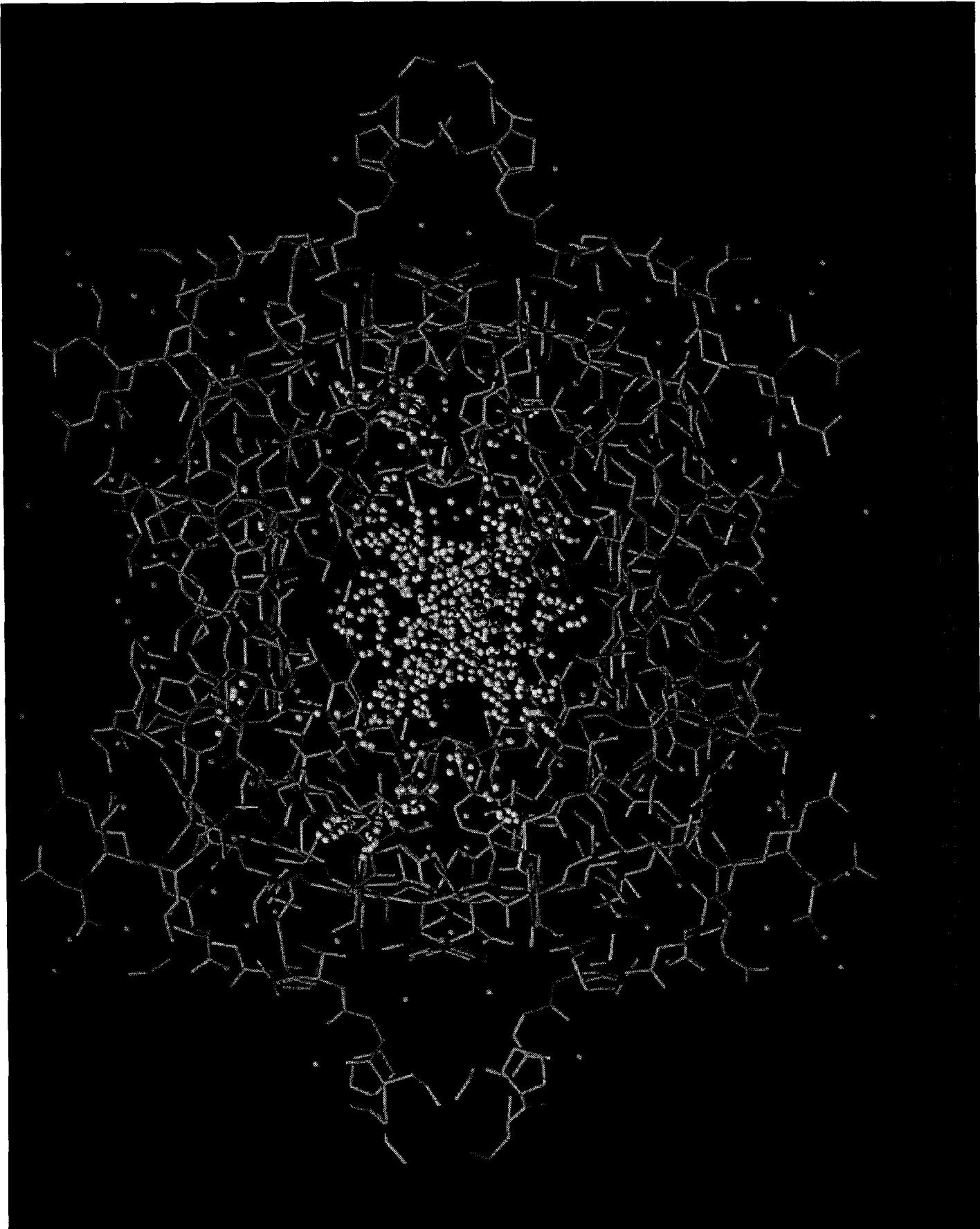


Figure 3

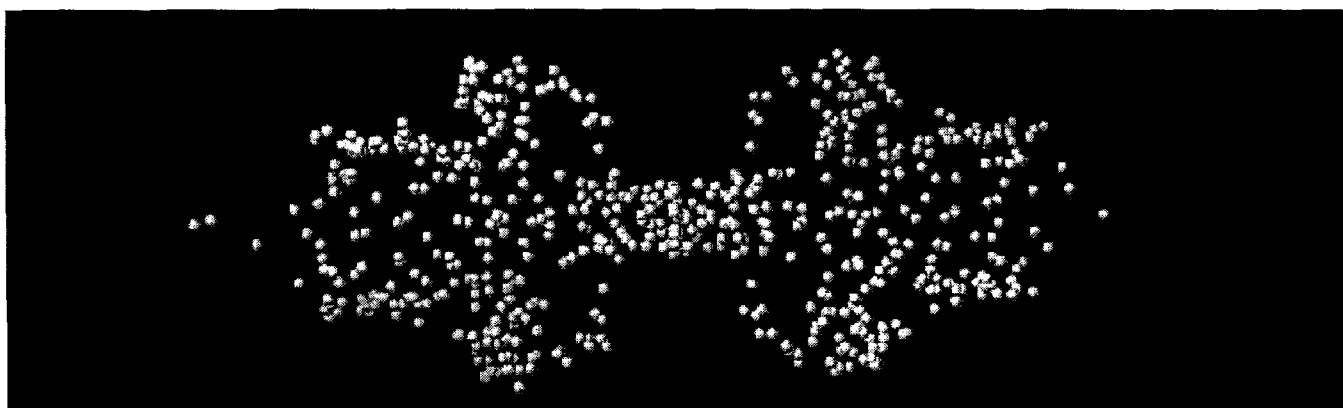
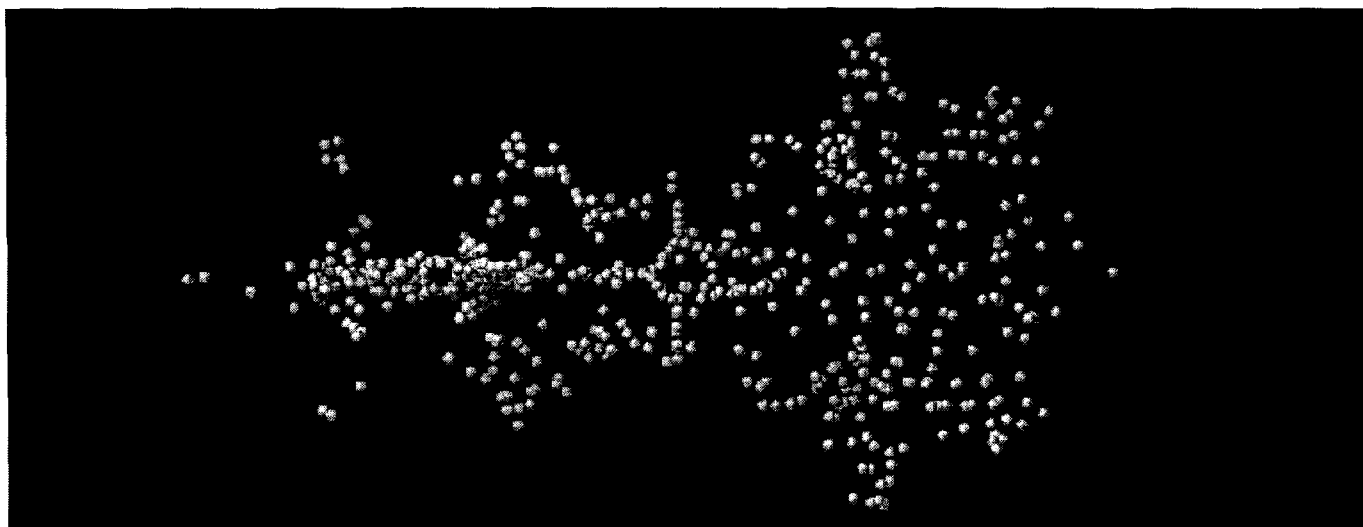
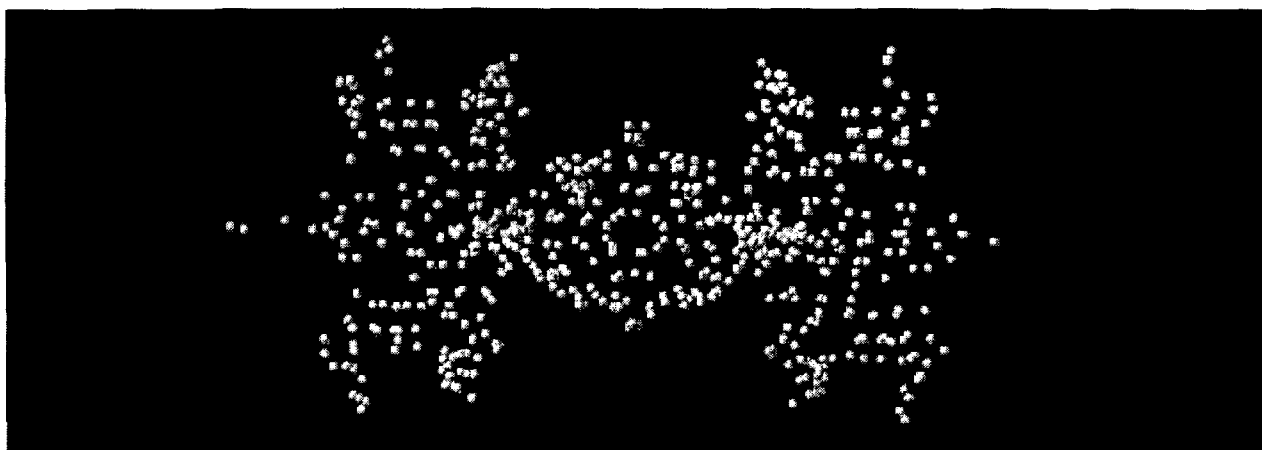


Figure 4: Sphere Clusters Viewed Along the Length of the Molecule



Figure 5: 3',5'-Di-O-acetyl thymidine docked with R67 DHFR

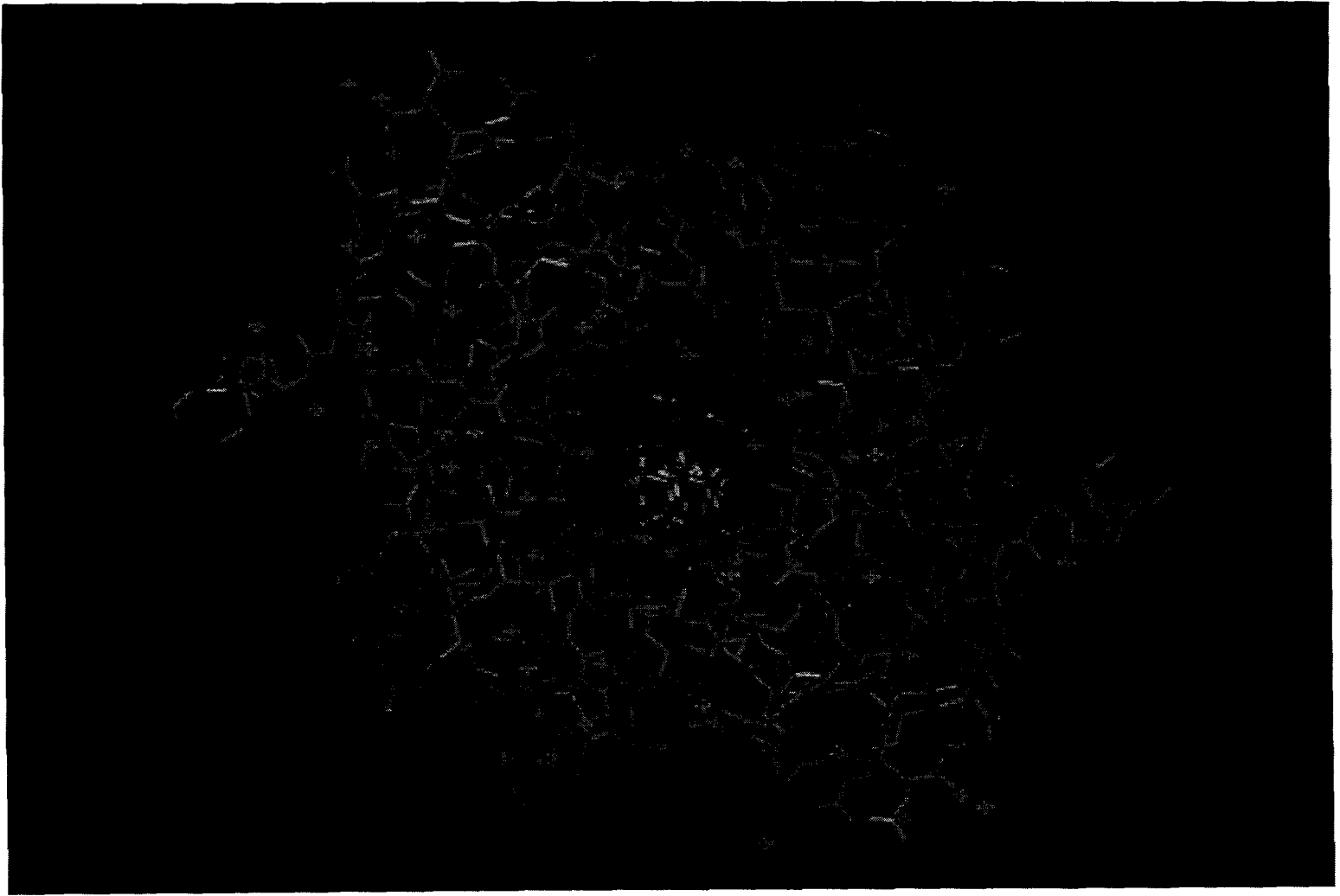


Figure 6: Deoxycholic Acid (co-crystallized with salicylic acid)  
docked with R67 DHFR



Figure 7: 15-crown-5-clathrate docked with R67 DHFR

hydrophobic to interact with R67 in solution, or were toxic and thus had to be eliminated as potential candidates for inhibition. Mr. J. Helton screened the kinetic effect of 15-crown-5 clathrate, 3'-azido-3'-deoxythymidine, deoxycholic acid, 1,4,5,8-naphthalene tetracarboxylic anhydride, and 3',5'-di-O-acetyl thymidine on the activity of DHFR (16).

#### DelPhi Potential Maps

The potential map generated by DelPhi showed a concentration of positive charge in the active site pore. Since there are no positive residues in the pore itself, the influence of the lysine residues near the pore was analyzed. When the charge on all eight symmetry related lysines (residues 32 and 33) is removed, the active site pore becomes red colored (negatively charged). Thus, the presence of the Lysines is essential to the observed charge distribution. To analyze the individual contribution of each Lysine, they were then treated separately. When lysine 33 is removed, the charge becomes more concentrated in the pore. This is logical since lysine 32 is physically closer to the pore. However, when the charge on Lysine 32 is removed, a blue area is still observed in the pore. This is due to the presence of Histidine 62. When the charge on Histidine is removed in addition to that on Lysine 32, the blue in the pore disappears (Figure 8). Interestingly, the Arginines, though far removed from the pore, influence the charge distribution as well. When their charge is removed, the active site becomes neutral (white). Since they are on the outside of the molecule, they have a greater influence on the induced dipoles in the solvent. Additionally, arginines tend to be more exposed to solvent.

WT R67 DHER

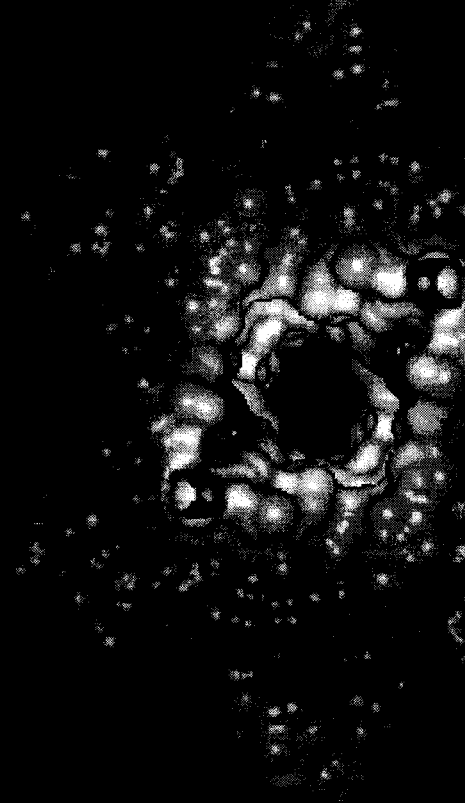
K32M+K33M

K32 present,  
K33M

K33 present,  
K32M

Figure 8

R67 DHER with no arginines



R67 DHER with K32M, no histidines

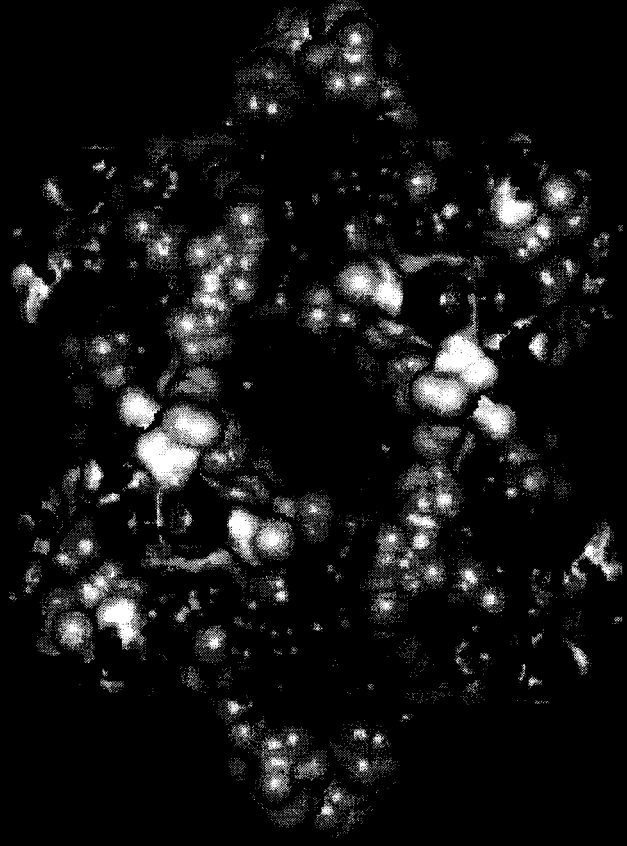


Figure 9

## **Future Studies**

The Heavy Atom limit can be circumvented by dividing the database into small portions. This is currently being done in the lab, and should result in more accurate predictions. Some predicted compounds were also tested for inhibition (16). Lysine and Histidine mutants were created in the lab. Kinetic studies should be done on these mutants to determine the accuracy of the DelPhi predictions. Electrostatic considerations should be incorporated into DOCK, as well as energy scoring and chemical matching. Also, it is unrealistic to assume that the ligands are rigid. Use of the flexible ligand model within DOCK should yield a more realistic representation of ligand binding.

## **Acknowledgements**

I would like to thank my mentor, Dr. Elizabeth Howell for her support and guidance, Mr. Rod Bunn for his invaluable computational expertise, and Jimmy Helton for his kinetic studies. Also, I would like to thank the Threshold Program and Howard Hughes Medical Institute for giving me this opportunity. Lastly, I'd like to thank all the members of Dr. Howell's lab for their support.

## References

- 1) Madigan, M.T., Martinko, J.M., Parker, J., eds. Brock: Biology of Microorganisms. Prentice Hall: NJ, 1997. p 423-427.
- 2) Gschwend, D.A., Sirawaraporn, W., Santi, D.V., Kuntz, I.D. Specificity in structure-based drug design: Identification of a novel, selective inhibitor of *Pneumocystis carinii* Dihydrofolate Reductase. PROTEINS: Structure, Function, and Genetics. 29: 59-67 (1997).
- 3) Narayana, N., Matthews, D.A., Howell, E.E., Xuong, N. A plasmid-encoded dihydrofolate reductase from trimethoprim-resistant bacteria has a novel D2-symmetric active site. Nature Structural Biology, 2:11: 1018-1025 (1995).
- 4) Bradrick, T.D., Beechem, J. M., Howell, E.E. Unusual binding stoichiometries are observed during binary and ternary complex formation in the single active pore of R67 Dihydrofolate Reductase, a D2 symmetric protein. Biochemistry 35:11414-11424 (1996).
- 5) Kuntz, I.D. Structure-based strategies for drug design and discovery. Science 257:1078-1082 (1992).
- 6) Kuntz, I.D., Meng, E.C., Shoichet, B.K. Structure-based molecular design. Accounts of Chemical Research 27:5, (May 1994).
- 7) Connolly, M.L. Solvent accessible surfaces of proteins and nucleic acids. Science 221: 709-713 (1983).
- 8) Ewing, Todd, ed. DOCK version 4.0 (manual). San Francisco, CA: Regents of the University of California, 1998.

- 9) Shoichet, B.K., Kuntz, I.D. Matching chemistry and shape in molecular docking. Protein Engineering 6: 723-732 (1993).
- 10) Bajorath, J., Kitson, D.H., Kraut, J., Hagler, A.T. The Electrostatic Potential of *Escherichia Coli* Dihydrofolate Reductase. PROTEINS: Structure, Function, and Genetics 11:1-12 (1991).
- 11) Honig, B. Nicholls, A. Classical Electrostatics in Biology and Chemistry. Science 268:1144-1149 (1995).
- 12) Gilson, M.K., Honig, B. Calculation of the Total Electrostatic Energy of a Macromolecular System: Solvation Energies, Binding Energies, and Conformational Analysis. PROTEINS: Structure, Function, and Genetics 4:7 (1988).
- 13) Cambridge Structural Database. Cambridge Crystallographic Centre, Cambridge, UK.
- 14) Yeh, Tina. (at Biosym) Personal Communication.
- 15) Bunn, Rod. Personal Communication.
- 16) Helton, James D. Kinetic Studies on the Inhibition of R67 Dihydrofolate Reductase: Threshold Senior Thesis. May 1999.

Appendix 1  
Protocol for DOCK

## Protocol for DOCK

A step by step guide for searching for potential ligands

Ingredients: A Silicon Graphics Machine

Insight

SYBYL

DOCK

MS

The Cambridge Structural Databank

Delphi

A researcher,

A computer guy,

and...

A student ( to do things like type up this protocol) ☺

To run dock, we must start with a receptor file from the protein databank. The first step is to remove any ligands that were co-crystallized and to remove crystallographic waters. (do this in INSIGHT)

```
r67 without ligands and waters /usr/people/ushma/narayanaternary1.pdb
r67 with ligands                /usr/people/ushma/narayanaternary2.pdb
r67 with waters                 /usr/people/ushma/narayanaternary3.pdb
r67 with ligands and waters     /usr/people/ushma/narayanaternary4.pdb
```

Next we must create a Connolly surface using the program MS (molecular surface) from the QPCE (Quantum Program Chemistry Exchange).  
go to /usr/people/ushma/connollysurf

Dock will automatically run MS and create a dock input file using the program AutoMS. However AutoMS has a fatal program error. A corrected copy of the program is included with MS. The file is called DOCK\_interface(CHECK PATH). We can find the differences between the two programs with the unix difference command:

```
$diff DOCK_interface /usr/data/ushma/dock4/bin/autoMS
```

Using jot, we then update the paths in the DOCK\_interface file and take off the header.  
save as autoMS.

```
dock_root= /usr/data/ushma/dock4/bin
MS_executable= /usr/people/ushma/connollysurf/429_sgrw/connolly
```

(AutoMS requires an extract.pdb file to run.)

First we must exclude all portions of the molecule that are irrelevant to our search. We do this by creating an exclude.pdb file using the molecular modeling program SYBYL. We use SYBYL remotely on Dr. Baker's machine in the Chemistry Department. To go to Dr. Baker's machine:

```
$xhost +  
$telnet sugar.chem.utk.edu  
login: howell  
password: folate1  
$sybyl
```

This will bring up the SYBYL program.  
Under the file tab select read  
Select the desired pdb file (see sybyl files list)

When asked if the molecule should be centered, select yes.  
In order to create a shell that encompasses a spherical or symmetrical active site, we must add a raw atom to the center of the active site. Select C3.  
For the coordinates enter (0,0,0) since we want the molecule in the center of the active site.

Go the extract command  
When prompted to select atoms select the center atom we added by clicking on it.  
Then hit the sets tab on the extract command window. Select sphere.  
Enter a radius of 15 angstroms. (Adjust this to include whatever parts of the molecule are of interest).  
Then hit the invert tab. It will ask where you wish to put the extracted atoms. Put them in M2 (molecular area two). Save M2 as a Brookhaven file called exclude.pdb (Be sure to rename any old exclude.pdb files you may want to keep).

Use xftp to move the files from sugar to how2.bio.utk.edu. Put in  
/usr/data/ushma/dock4/bin  
Log off of Dr. Baker's machine. Look at the created file in INSIGHT.

```
$cd /usr/data/ushma/dock4/bin  
$autoMS /usr/people/ushma/narayanaternary1.pdb  
• location of receptor pdb file
```

The output file should have the extension \*.ms.

The screen should read: Ready to run SPHGEN

Now we must create an INSPH File! Using jot, type the following (excluding the parts in parentheses). Modify as needed. For an explanation of the values see the DOCK manual p.84.

/usr/people/ushma/narayanaternary1.ms ( ← MS output file)

R

X

0.0

4.0

/usr/people/ushma/narayanaternary1.sph ( ← Sphgen output file name)

\$sphgen

It takes a while to run, so this is a good time for a snack. ☺

The OUTSPH file contains useful information including the number of sphere clusters generated.

density type = X

reading /usr/people/ushma/narayanaternary3.ms

# of atoms = 2000 # of surf pts = 21996

finding spheres for /usr/people/ushma /narayanaternary1.ms

dotlim = 0.000

radmax = 4.000

Minimum radius of acceptable spheres?

1.400000

output to /usr/people/ushma/narayanaternary1.sph

clustering is complete 27 clusters

There are probably more sphere clusters than you need. We can visualize the sphere clusters using the program showsphere. It is interactive.

\$showsphere

the input is as follows:

/usr/people/ushma/narayanaternary1.sph

0

n

/usr/people/ushma/showsphere.pdb

Transfer the created showsphere file to Dr. Baker's machine so that it can be viewed with SYBYL.

Unwanted Spheres must be removed by hand. Choose the EXTRACT command, click on atoms, then sets, choose sphere with a radius of 1-3 angstroms. Invert. Save as another molecule (M1 or M2 whichever is unoccupied). Now zap the first Molecular area. The second area now contains the molecule minus the selected sphere clusters. Repeat this process until all unwanted clusters have been removed. Rinse (just kidding). Save the number cluster file with an appropriate name (eg. reducedsph.pdb).

FTP the new file back to how2.bio.utk.edu. Place in /usr/data/ushma/dock4/bin/(filename).

Run pdb to sph to change the file into the sphere format required by dock

```
$pdbtosph
input file:   reducedsph.pdb
output file:  reducedsph.sph
```

Next we must create a scoring grid, using GRID.  
See p.80 of the Dock reference manual.

Read the receptor file into INSIGHT, Set the potentials, add H's, and charges. Modify the hydrogens and fix forcefield potentials. Using Molecule/put save the receptor as a mol2 file. (this takes a while to save)

Run using created box file box.pdb . The following is sample input for GRID.

```
compute_grids           yes
output_molecule       no
grid_spacing           0.3
contact_score          yes
contact_cutoff_distance 4.5
chemical_score         no
energy_score           no
bump_filter            yes
bump_overlap           0.75
receptor_file          /usr/people/ushma/narayanaternary1.mol2
box_file               /usr/data/ushma/doc4/bin/box.pdb
vdw_definition_file    /usr/data/ushma/dock4/parameter/vdw.defn
score_grid_prefix      grid2
```

The Grid Command line is \$grid -i grid.in

GRID takes a really long time to run and the output file size is quite large.

Look at the GRID output using showbox. This is also an interactive program.

The output when viewed in Insight, should encompass all areas of the molecule that are of interest. Reference the grid file to the receptor file for comparison.

(Yes, it's supposed to be a boring yellow box)

#### PREPARE LIGAND FILE in Cambridge.

```
$cd /usr/data/CAMBRIDGE
$setenv CSDHOME `pwd`
$set path = ($CSDHOME/bin $path)
$rehash
$cd trials
$ls
```

\$quest trialx (where x is the next number after the last trial in this directory)

\$term x

\$menu

The x-windows version of Cambridge should appear. Click in the window to begin. Go to the TO SEARCH tab in the upper right hand corner. Choose NUMERICAL. Go to CLASS enter RANGE and enter 1-64. Then return to the main menu. Save as mol2 and as REFCODE LIST. Go to the QUEST... command and selected the 1-64 search and hit ONLY ORGANICS. Then hit startsearch. When the first result comes up, click DIAG with the right mouse button (this speeds up anything done with the left mouse button). The compounds should flash on the screen very rapidly. ( This is neat to watch, but takes a couple of hours). At the end, the message End of Database Encountered! comes up and you are asked if you want to quit. Say yes. Once you are out of Cambridge check to see that a trialx.mol2 and a trialx.gcd file have been created, by using the unix list command: \$ls This mol2 file will be the ligands file for DOCK.

Now we are (finally) ready to run DOCK!!! ( a dance of joy may be appropriate)

First we must create an dock.in file.

here is a sample file:

flexible_ligand	no
orient_ligand	yes
score_ligand	yes
minimize_ligand	no
multiple_ligands	no
parallel_jobs	no
random_seed	0
match_receptor_sites	yes
random_search	no
automated_matching	yes
maximum_orientations	50
write_orientations	yes
rank_orientations	yes
rank_orientations_total	10
intermolecular_score	yes
gridded_score	yes
grid_version	4
bump_filter	yes
bump_maximum	3
contact_score	yes
chemical_score	no
energy_score	no

atom_model	u
vdw_scale	1
electrostatic_scale	1
ligands_maximum	20000 (note: this number must be larger than the total number of molecules in the ligands file)
initial_skip	0
interval_skip	0
heavy_atoms_minimum	4
heavy_atoms_maximum	40 (this limits the time that the run takes)
rank_ligands	yes
rank_ligands_total	1000
restart_interval	100
ligand_atom_file	/usr/data/CAMBRIDGE/trials/trial2.mol2 (change this to subset)
receptor_site_file	/usr/people/ushma/narayanaternary3.sph
score_grid_prefix	grid3
vdw_definition_file	/usr/data/ushma/dock4/parameter/vdw.defn (this file is provided with dock. The path MUST be correct)
quit_file	dock15.quit
dump_file	dock15.dump
info_file	dock15.info
restart_file	dock15.rst (restart commands are in the DOCK manual)
contact_clash_penalty	50
ligand_contact_file	dock_cnt15.mol2 (this is the results file)
ligand_centers	no

The DOCK command line is:

```
$dock -i dock.in
```

This gives the top 1000 scoring ligands. We then reran these thousand with more orientations.

These parameters have to be changed:

maximum_orientations	400
ligands_maximum	1000

To view the results:

First, convert the dock.info (save to disk and move to the Gateway Computer) file to a REFCODE list for Cambridge, using Rod's program.

Click on DOCK\_converter

Go to File/ Alpha

Put source file name in source Box.

Then Select files and save conversions

It should read finished when its done.

save file as trialx.gcd

```
$cd /usr/data/CAMBRIDGE
$setenv CSDHOME `pwd`
$set path = ($CSDHOME/bin $path)
$rehash
$cd trials
$setenv CSDVIRTDB trialx.gcd
$quest trial(x+1)
```

At this point you should see information about the Cambridge Structural Databank flash across the screen.

Near the top it should say: search being run on VIRTDB trialx  
(this greatly speeds up the viewing process)

```
$term x
$menu
```

[to remove the virtual database option, type unsetenv CSDVIRTDB]

Now the x-windows version of CSD appears. Click in the window to begin.

In the upper right hand corner, click on TO SEARCH

From the search options choose TEXT

From the text options choose REFCODE

Enter the reference codes (eg. BZANTC 10) one by one for the compounds that you wish to view.

When you are done return to the main menu.

Go to the QUEST... option. You should see a REFCODE list. Choose only organics on the right side's menu. Select 1 of the refcodes listed. Then hit SEARCH. The selected compounds will appear on the screen. Go to view 1D to see the compounds chemical name.

Hit reject from the 2D/3D option. It will then tell you: End of Database Encountered! Do you wish to exit? If you do not exit at this point you will lose all hits from this search!

Enter no. (This is a flaw in the program. There must be a way to save the compounds that you like from the reference list, but I haven't been able to do it without getting this message.)

Then repeat with each of the compounds of interest.

When you are ready to exit Cambridge, go to the exit option on the main menu.

To view the results oriented to the receptor:

```
$cd /usr/data/ushma/dock4/bin
```

The dock\_cnt.mol2 files are too big to open !!! Split them up using the program from Tina Yeh at Biosym.

See Media Mail for instructions

The program is in /usr/data/ushma/dockrun1 called renamefiles.

```
$mkdir /usr/data/ushma/dockrunx
```

```
$ cp /usr/data/ushma/dockrun1/renamefiles /usr/data/ushma/dockrunx/renamefiles
```

```
$cp /usr/data/ushma/dock4/bin/dock_cntx.mol2
```

```
/usr/data/ushma/dock4/dockrunx/dock_cntx.mol2
```

Now we have to change the access mode to the new folder.

```
$chmod -R u+rwx /usr/data/ushma/dockrunx  
$csplit -f dock_cnt_new dock_cntx.mol2 '/Number :/' '{49}'  
• (six spaces)
```

```
$renamefiles dock_cnt_new .mol2
```

jot to open and check files for refcodes of interest.

Go into Insight. Molecule → get → narayanaternary1.pdb

Get the compounds of interest from /usr/data/ushma/dockrunx

Pull them up while referencing the receptor molecule.

Save images if desired.

### Delphi Electrostatics

To create the delphi potential map:

```
$ cd /usr/data/delphi/export
```

The charge file and parameter file must be modified.

```
$cd /usr/data/delphi/export/examples/SOD  
$ls  
$jot r67.prm  
An example parameter file:  
scale=0.8  
perfil=80  
in (pdb, file= "/usr/people/ushma/narayanaternary1.pdb")  
in (crg, file="examples/SOD/r67.crg")  
in (siz, file="examples/SOD/sod.siz")  
out (modpdb, file="r67_delphi.pdb")  
out (phi, file="r67.grd", form="BIOSYM")  
energy (c,s)
```

Change the parameter file. Make sure that the output file extension is \*.grd, so that it can be viewed with Insight. The size (\*.siz) file is based on Van der Waals Radii and does not need to be modified.

The charge file should be adapted for the molecule of interest.

Go back to (\*crg) /usr/data/delphi/export

The command line of Delphi is as follows:

```
$delphi examples/SOD/parameterfile.prm > logfile.log
```

[look at the logfile to make sure it ran]

Viewing the results:

Enter Insight. Pull up receptor molecule with Molecule/ get. Use Molecule/color to change backbone color to gray or white. Then go to the third icon on the left side of the screen and click. Pull in \*.grd file from /usr/data/delphi/export . A yellow box should appear. Then graphically create a Connolly surface by going to Molecule/Surface. Select low resolution or it will take ½ hour to load. To Color the surface go to Molecule/Color. Under attribute choose surface. Under Color Method choose grid. Under Spectrum Name choose Delphi\_Spectrum. This will display the potential map.