NEW PROGRAMS

An algorithm to find channels and cavities within protein crystals

Oleg S. Kisljuk, Galina S. Kachalova,* and Nadejda Ph. Lanina*

Institute of Mathematical Problems in Biology, Russian Academy of Sciences, Pushchino, Russia

**Institute of Theoretical and Experimental Biophysics, Russian Academy of Sciences, Pushchino, Russia*

We have written a package called CHANNEL for determining and analyzing the channels and cavities within protein crystals. By using CHANNEL, the intermolecular space within a crystal lattice can be divided into closed cavities and channels. This package is certainly useful in determining the channel topological structure and quantitative characteristics. The package allows also the volume, and maximal and minimal areas of channels along a required direction to be calculated.

Keywords: channel, cavities, protein crystal

INTRODUCTION

The knowledge of the channels and cavity structures within crystals of globular proteins may be useful[1-9] in detecting the binding sites of specific ligands and ions on accessible surfaces of protein molecules within channels; in studying the structure and physical properties of the water near protein surfaces within the channels; for predicting the narrowest sites of the channels for diffusion of molecules of arbitrary sizes, etc. In many such studies, not only graphic representation of channels, but also their topological structure and quantitative characteristics will be needed.

In view of the increasing interest in these problems, we have written a package, called CHANNEL, which offers the following capabilities:

- To build up a space net of nodes which fill the unit cell of the crystal lattice—if a sphere of radius R with the center in a given node doesn't intersect with any molecule's atom, then this node is marked. (Each atom is represented as a sphere with its corresponding van der Waals radius);
- To find connected sets of marked nodes and classify them into two groups: closed cavities and the channels. A sphere of radius R can move along these determined channels.
- To estimate the volume, the maximum and minimum areas of channel cross section.
- To build up a set of spheres that approximate channels for further visualization.

HARDWARE REQUIREMENTS

All programs are implemented in FORTRAN. The overall length of programs is about 8000 lines.

The package CHANNEL runs on all version of PC/AT that have sufficient memory. The structure of a channel may be displayed as a set of spheres of arbitrary radii with any available graphic package for protein structures.

PROGRAM DESCRIPTION

The basic information required by package CHANNEL is a list of atomic coordinates in PDB format, coordinates of three vectors along the axis of a unit cell, a list of definitions of atom types and the van der Waals radius. The package consists of five essential, connected programs YYMAS, YYCNL, YYDATA, YYPIC, and YYCUB.

Program YYMAS forms the space net of nodes marked as O/1 from the list of original atomic coordinates. Dimensions of the space net are NX, NY, and NZ along the respective axis of the unit cell. The nodes that are within van der Waals spheres centered at atoms are marked 0; 1 is for nodes beyond these van der Waals spheres.

Program YYCNL classifies the space set of nodes marked with 0 by program YYMAS into channels and closed cavities.

Program YYDATA calculates the mean, maximal, and minimal areas of channel cross sections along a required direction. The volume of a channel within a unit cell also can be calculated.

Program YYPIC forms a list of coordinates of the spheres that approximate the structure of a channel or cavities.

Program YYCUB generates coordi-

Color Plates for this article are on page 296.

Address reprint requests to Dr. Kisljuk at the Institute of Mathematical Problems in Biology, Russian Academy of Sciences, Pushchino, Moscow region, 142292, Russia.

Received 8 March 1993; revised 1 February 1994; accepted 31 March 1994

nates of the vectors along the axis of the unit cell from the parameters of the latter.

An algorithm to find channels in a crystal lattice

Consider a crystal lattice filled with atoms. The atoms are approximated by van der Waals spheres.

The task is: for the spheres of radius R to find all cavities in which one or more spheres can be included. That is, we want to divide all the cavities into two groups: channels in which a sphere of radius R can move through the cell, and closed cavities.

Consider a space net that fills the unit cell of the crystal lattice. The axes of the net are parallel to the axes of the unit cell. The distance between the nodes of the net depends upon the accuracy of calculations. Let the unit cell have a periodic boundary condition so that each node has six neighboring nodes.

Let some nodes have marks. Two marked nodes P_1 and P_n are connected if we have a sequence of marked nodes $P_1, P_2, \ldots, P_n$, and P_i and P_{i+1} are neighbors.

A set of marked nodes is a maximally connected set if

- Each two nodes of the set are connected.
- If node P belongs to the set and nodes P and Q are connected, then Q belongs to the set.

We approximate atoms of the cell by spheres of $R + R_{VW}$ radius, where R_{VW} is van der Waals radius, and R is a radius of sphere that can be included into a cavity. We mark the nodes of a net that belong merely to one sphere with 1, and all other nodes with 0. Thus if we put a sphere of radius R at the node marked 0, it does not intersect with the spheres of van der Waals radius that approximate atoms. The maximally connected sets of nodes marked with 0 are models of channels and cavities for spheres of radius R, and the spheres cannot penetrate from one maximally connected set to another.

The problem is to classify the maximally connected sets into two groups: the channels and closed cavities. First, we have to define more precisely what the word *channel* means.

Definition 1

Consider an infinite three-dimensional net which fills a crystal lattice. Then the channel is the maximally connected set of nodes in which for an arbitrary real L we can find two nodes with a distance between them more than L.

Proposition 1

Consider the maximally connected set of nodes of the net that fills the unit cell with a periodic boundary condition. Consider the closed chain of nodes $P_1, P_2, \ldots, P_n$. Let us denote the number of crossings of this chain with the unit cell boundaries along a given axis as N_{+x}, N_{+y}, N_{+z} for the straight direction and as N_{-x}, N_{-y}, N_{-z} for the opposite one. Let us define this consideration as a movement along the closed chain of nodes.

If all of the following equations are true at once for this closed chain:

$$N_{+x} = N_{-x}$$
$$N_{+y} = N_{-y}$$
$$N_{+z} = N_{-z}$$

then this maximally connected set of nodes is the model of a closed cavity; in any other case, of a channel.

Proof

The movement along a closed chain of nodes $P_1, P_2, \ldots, P_n, P_1$ of the net that fills the unit cell of the crystal lattice with a periodic boundary condition corresponds to the movement along a chain of nodes of an infinite three-dimensional net that fills the entire crystal lattice.

Along this chain, if the numbers in any of the pairs N_{+x} and N_{-x}, N_{+y} and N_{-y}, or N_{+z} and N_{-z} in the straight and opposite direction are not equal to each other, this means that we move from one cell to another during each cycle $P_1, P_2, \ldots, P_n, P_1$ into an infinite three-dimensional net.

Since the size of the unit cell is not equal to zero, we can find two nodes of an infinite net with an arbitrary distance between them.

Consider a maximally connected set of nodes of this infinite net that fills the crystal lattice. If the set is a channel we can find a chain of nodes $Q_1, Q_2, \ldots, Q_i$ of an arbitrary length. In this chain we can find a subchain $Q_m, Q_{m+1}, \ldots, Q_n$ where Q_m and Q_n are located in different cells but have the same coordinates inside the cell. Thus, for the corresponding subchain of the net that fills the unit cell with a periodic boundary condition, the condition of this proposition is valid.

Using this proposition as a base, we can write a recursive algorithm that classifies the maximally connected sets into two groups: channels and closed cavities.

Let a node of the net have three marks Mx,My,Mz with interger values.

The procedure CHANNEL writes a message if the set of nodes is a channel.

```
CHANNEL(node P,integer Nx,integer Ny,integer Nz)
BEGIN
  IF P does not belong to the set
      THEN return
  ELSE IF P has marks Mx,My,Mz and Mx=Nx and My=Ny and Mz=Nz
      THEN return
  ELSE IF P has marks Mx,My,Mz and (Mx<>Nx or My<>Ny or Mz<>Nz)
      THEN write ("The set is a channel")
  ELSE IF P has not marked
      THEN
        BEGIN
          Mark the node P by Nx,Ny,Nz;
          FOR each P' neighboring to P
            increase (decrease) the values of Nx,Ny,Nz
            if we cross the corresponding boundary and
            call CHANNEL(P',Nx,Ny,Nz);
        END
END
```

To start the procedure we must call

CHANNEL(P,0,0,0)

where P is an arbitrary node belonging to the set under investigation.

APPLICATIONS

During the studies of the transparency of lysozyme tetragonal crystals for water[10] we have used the package CHANNEL. The possibility to calculate the space net of nodes that have been marked as channels and cavities enabled us to determine the ratio of effective diffusion coefficients along [001] and [100] directions.

To compare the calculated values of this ratio with the experimental one, we have determined the structure of channel with the atom's radii $R = R + R_{vw}$, where R_{vw} is the van der Waals radius of an atom and R is an additional radius, which accounts for the finite size of a probing diffusing body.

Thus we have calculated a number of the space nets of the channel for a moving sphere of radius 0.0, 0.5, 1.0, 1.5, 2.0, 2.5 Å. Color Plates 1–3 present the structures of channels within a protein crystal with R values of 0.0, 1.0, 2.0 Å, respectively. As seen from these figures: the channel presents a very irregular set of bottleneck-shaped widenings; and the dimensions of channels within a protein unit cell decrease with increasing R. As has been found, the best correlation between the experimental and calculating values of the ratio between diffusion coefficients was where R was equal to 1.9 Å. The remarkable closeness of this value to the average radius of water molecule gives us an additional argument in favor of the validity of our calculation.

ACKNOWLEDGMENTS

Dr. V.N. Morozov is gratefully acknowledged for useful discussion and for providing us with the experimental data of the transparency of lysozyme crystal for water.

REFERENCES

1 Pesek, J.J. and Schneider, J.F. *J. Inorganic Biochem.* 1988, **32,** 233–238
2 Eisenman, G. and Oberhauser, A. *Biophys. J.* 1988, **53,** A631
3 Hannon, R.A. and Ford, G.C. *Biochem. Soc. Trans.* 1988, **16,** 961–962
4 Olah, Glenn A., Huang, Huey W., Liu, Wenhan, and Wu, Yili. *J. Mol. Biol.* 1991, **218,** 847–858
5 Leach, A.R. and Kuntz, I.D. *J. Comp. Chem.* 1992, **13,** pp. 730–748
6 Meng, E.C., Shoichet, B.K., and Kuntz, I.D. *J. Comput. Chem.* 1992, **13,** 4, 505–524
7 Shoichet, B.K., Bodian, D.L., and Kuntz, I.D. *J. Comput. Chem.* 1992, **13,** 3, 380–397
8 Lewis, R.A., Roe, D.C., Huang, C., Ferrin, T.E., Langridge, R., and Kuntz, I.D. *J. Mol. Graphics.* 1992, **10,** 66–78
9 Kuntz, I.D. *Science,* 1992, **257,** 5073, 1078–1082
10 Morozov, V.N. Evtodienko, Yu.V., Lanina, N.Ph., and Kachalova, G.S. *Biophys. J.* 1992, in press