BUTTERWORTH HEINEMANN

Detection and geometric modeling of molecular surfaces and cavities using digital mathematical morphological operations

Masato Masuya and Junta Doi

Department of Biotechnology, The University of Tokyo, Tokyo 113, Japan

We developed a digital method based on mathematical morphological operations to obtain three types of surfaces: van der Waals surface, solvent-accessible surface, and molecular surface, to extract the cavities on the surface and interior part of the molecule and to extract the ligand portions in contact with the cavities. The molecular surface, the cavities and the portions, and the heme region are visualized using solid modeling.

The method enables us to obtain the volumes of the cavities and inhibitor portions and the areas of the surfaces. Solid modeling enables us to obtain cross-sections at arbitrary positions. This will have considerable utility in docking studies.

Keywords: Docking studies, mathematical morphology, molecular cavities, molecular surfaces, solid modeling

INTRODUCTION

Geometric features play an important role in molecular binding. The ordinary approach to examining molecular binding deals with the molecular surfaces for the receptor, cavities of the receptor in contact with the ligand, and potential ligands to the cavities of the protein.

Three types of surfaces are used for the molecular surface representation. They are the van der Waals surface, the solvent-accessible surface,[1] and the solvent contact/reentrant surface.[2] They have been implemented using several computer algorithms.[3–6] Connolly[7] has developed a method of placing dots over the surface.

Several algorithms have been developed to extract the cavities in contact with the ligand. Kuntz et al.[8] approximated a cavity to a set of spheres. Ho and Marshall[9] produced a program that solid-fills a defined protein cavity. Their program is based on a two-dimensional filling routine and requires the definition of a seed point to set a surface boundary between the cavity and exterior free space. Delaney[10] has produced a cavity-finding algorithm using a three-dimensional cellular logic operation. The program of Kisljuk et al.[11] is specified to detect the channels inside protein. No detecting method for the portions of the ligand in contact with the cavities has been proposed.

We focused on the detection of the cavities and portions of the ligand in contact with the cavities and developed a robust method for geometric modeling and visualization. In our method, all molecules are treated as a set of three-dimensional discrete grid points. The method is based on mathematical morphology,[12,13] which is a set theoretical approach to image processing. We extended the method, which is applicable to digital processing, for the molecular docking study.

We calculated three types of surface representation and detected the cavities and portions in contact with the cavities. All of them are reconstructed into a solid model format using a modified marching cube method.[14–16] The model enables us to visualize the geometric properties of the shape in the binding feature.

METHOD

Mathematical morphological operations

Mathematical morphology has been developed as a set theoretic approach to image analysis. It was formalized by Matheron[12] and extended by Serra.[13] It has been applied mainly to two-dimensional image analysis and image processing. However, no practical three-dimensional studies have been reported so far. The method is defined in Euclidean R^N space; therefore it is applicable to the three-dimensional problem.

Color Plates for this article are on pages 349–350.

Address reprint requests to Junta Doi, Department of Biotechnology, The University of Tokyo, Yayoi 1-1-1, Bunkyo-ku, Tokyo 113, Japan.
Received 13 June 1995; revised 7 August 1995; accepted 8 August 1995

Journal of Molecular Graphics 13:331–336, 1995

655 Avenue of the Americas, New York, NY 10010

0263-7855/95/$15.00
SSDI 0263-7855(95)00071-2

Mathematical morphology extracts information about the geometric structure of the object by transforming it through its interaction with another object, which we call a *probe*. In this article, all objects are aligned to a three-dimensional discrete grid and we use the probe which is symmetrical in the three-dimensional Cartesian coordinate systems, such as a sphere.

Erosion and dilation are two fundamental operation of mathematical morphology. They are defined as a set transformation that shrinks and expands a set. Let an object that consists of points aligned to a three-dimensional grid be X, the probe be P, and the erosion of X by P, which is denoted as $X \ominus P$, be defined as a set of points, a, such that the translated set, P_a, is contained in X. It is expressed as

$$X \ominus P = \{a: P_a \subseteq X\} = \bigcap_{p \in P} X_p \tag{1}$$

where

$$P_a = \{p + a: p \in P\} \tag{2}$$

An explanation of erosion is illustrated in Figure 1.

The dilation of X by P, which is denoted as $X \oplus P$, is defined as a set of all points, a, such that P_a intersects X. It is expressed as

$$X \oplus P = \{a: P_a \cup X = \emptyset\} = \bigcup_{p \in P} X_p \tag{3}$$

An explanation of dilation is illustrated in Figure 2.

On a three-dimensional discrete grid, the object and probe are sets of grid points. We define the set of n points, which are included in the object, as $\{x_1,x_2, \ldots, x_n\}$ and the set of m points, which are included in the probe, as $\{p_1,p_2, \ldots, p_m\}$, where x_i and p_j are grid points that have discrete coordinates such as (0,0,1) and (0,1,1). The digital erosion is expressed as

$$\{x_1,x_2, \ldots, x_n\} \ominus \{p_1,p_2, \ldots, p_m\} = \bigcap_{i=1}^{n} \bigcup_{j=1}^{m} \{x_i + p_j\} \tag{4}$$

and the digital dilation is expressed as

$$\{x_1,x_2, \ldots, x_n\} \oplus \{p_1,p_2, \ldots, p_m\} = \bigcup_{i=1}^{n} \bigcup_{j=1}^{m} \{x_i + p_j\} \tag{5}$$

Opening and closing can be thought of as nonlinear filters, which smooth the contours of the object. If we erode X by P and then dilate the eroded set, $X \ominus P$, by P, we can obtain a simplified and less detailed object of X. This operation is called the opening of X by P and is denoted as $X \circ P$. The closing of X by P, which is denoted as $X \bullet P$, results from first dilating and then eroding. These operations are expressed as follows:

$$X \circ P = (X \ominus P) \oplus P \tag{6}$$

$$X \bullet P = (X \oplus P) \ominus P \tag{7}$$

Opening can suppress the sharp edges and cut the narrow necks of the object, whereas closing can fill up thin gaps and small holes. Opening and closing are explained in Figures 3 and 4, respectively.

The digital expressions of opening and closing are derived from combining the digital erosion and digital dilation. They are as follows:

$$\{x_1,x_2, \ldots, x_n\} \circ \{p_1,p_2, \ldots, p_m\} = (\{x_1,x_2, \ldots, x_n\} \ominus \{p_1,p_2, \ldots, p_m\}) \oplus \{p_1,p_2, \ldots, p_m\} \tag{8}$$

$$\{x_1,x_2, \ldots, x_n\} \bullet \{p_1,p_2, \ldots, p_m\} = (\{x_1,x_2, \ldots, x_n\} \oplus \{p_1,p_2, \ldots, p_m\}) \ominus \{p_1,p_2, \ldots, p_m\} \tag{9}$$

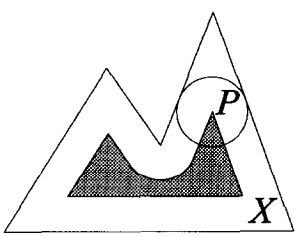

Figure 1. Erosion: The object X *is shrunk by the probe* P.

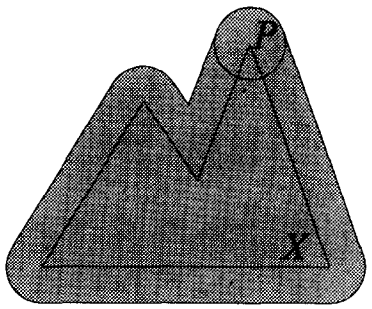

Figure 2. Dilation: The object X *is expanded by the probe* P.

Surface representations

There are three types of surface representation[17]: van der Waals surface, solvent-accessible surface,[1] and solvent contact and reentrant surface.[2] They are shown in Figure 5. The van der Waals surface is the boundary of overlapping regions of the spheres of the appropriate van der Waals radius. The solvent-accessible surface is defined by the center of the probe as it moves over the van der Waals surface of the object molecule. The probe is usually taken to be a water molecule, which is approximated as a sphere with a radius of 1.4 Å. The solvent contact surface is the part of

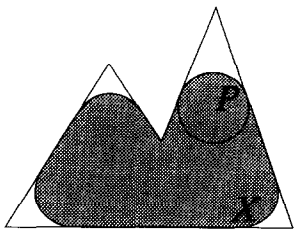

Figure 3. Opening: The object X *is smoothed by the probe* P.

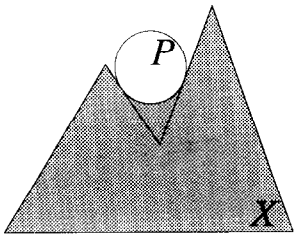

Figure 4. Closing: The object X *is filled by the probe* P.

the van der Waals surface that makes contact with the surface of the probe, and the reentrant surface is defined as the interior surface of the probe, which is simultaneously in contact with more than one molecular atom. Together, the solvent contact and reentrant surfaces define a continuous surface, which is known as the molecular surface.

Each surface and its interior can be represented with a mathematical morphological operation. Let a sphere of the probe be R_w, the center of the ith atom of an object protein be X_i, a sphere with the van der Waals radius of the ith atom be R_i, and number of atom included in the object protein be N. Each surface and its interior, which are denoted as SI_v, SI_a, and SI_m, is expressed as follows:

The van der Waals surface, S_v, and its interior, I_v:

$$SI_v = S_v \cup I_v = \bigcup_{i=1}^{N} \{X_i \oplus R_i\} \tag{10}$$

The solvent-accessible surface, S_a, and its interior, I_a:

$$SI_a = S_a \cup I_a = (S_v \cup I_v) \oplus R_w \tag{11}$$

The molecular surface, S_m, and its interior, I_m:

$$SI_m = S_m \cup I_m = (S_v \cup I_v) \bullet R_w \tag{12}$$

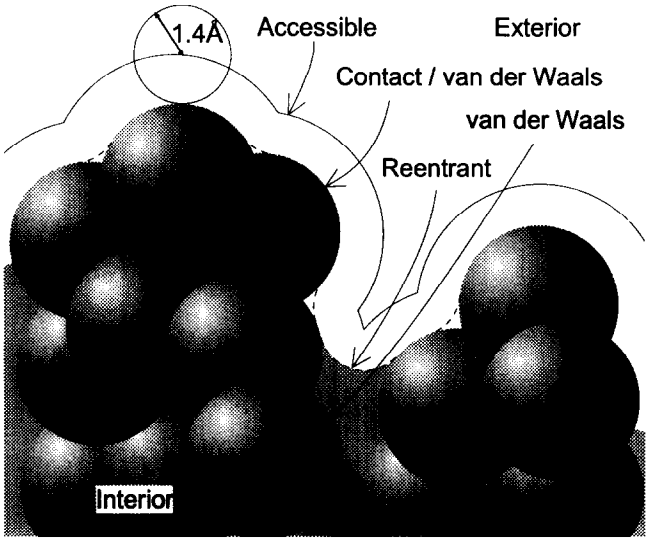

Figure 5. Three types of surface representation: van der Waals surface, S_v*; solvent-accessible surface,* S_a*; and solvent contact and reentrant surface,* S_m *(enveloped by a dotted line).*

The molecular surface and its interior, SI_m, are used to detect the cavities on the molecular surface, because SI_m is closely related to the molecular interactions.

Suppose the truncated grid point of the center of the ith atom is x_i and the grid points included in R_i are $\{r_{i1}, r_{i2}, \ldots, r_{im}\}$; the digital processing to obtain the van der Waals surface and its interior is then

$$SI_v = \bigcup_{i=1}^{N} \bigcup_{j=1}^{m} \{x_i + r_{ij}\} \tag{13}$$

Let the grid points included in SI_v be $\{v_1, v_2, \ldots, v_n\}$ and the grid points included in R_w be $\{r_{w1}, r_{w2}, \ldots, r_{wm}\}$. The solvent-accessible surface and its interior is digitally expressed as follows:

$$\begin{aligned} SI_a &= \{v_1, v_2, \ldots, v_n\} \oplus \{r_{w1}, r_{w2}, \ldots, r_{wm}\} \\ &= \bigcup_{i=1}^{n} \bigcup_{j=1}^{m} \{v_i + r_{wj}\} \end{aligned} \tag{14}$$

The molecular surface in digital form is expressed as

$$\begin{aligned} SI_m = (\{v_1, v_2, \ldots, v_n\} \oplus \{r_{w1}, r_{w2}, \ldots, r_{wm}\}) \\ \ominus \{r_{w1}, r_{w2}, \ldots, r_{wm}\} \end{aligned} \tag{15}$$

The van der Waals interior, I_v, is a result of the erosion of SI_v by the probe E. The probe consists of its center and its six neighbors, which are expressed as:

$$\begin{aligned} E = \{&(0,0,0), (1,0,0), (0,1,0), (0,0,1), (-1,0,0), \\ &(0,-1,0), (0,0,-1)\} \end{aligned} \tag{16}$$

Erosion of SI_v by the probe E is the operation that removes the 26-connected surface from the van der Waals surface and its interior, SI_v. The result of this operation indicates the van der Waals interior, I_v. It is denoted as:

$$I_v = (SI_v) \ominus E \tag{17}$$

$$\begin{aligned} &= \{v_1, v_2, \ldots, v_n\} \ominus \{(0,0,0), (1,0,0), (0,1,0), \\ &\quad (0,0,1), (-1,0,0), (0,-1,0), (0,0,-1)\} \end{aligned} \tag{18}$$

The van der Waals surface is obtained by subtracting the interior, I_v, from SI_v. It is expressed as follows:

$$S_v = SI_v \cap I_v^c \tag{19}$$

where I_v^c denotes the complement set of I_v.

The solvent-accesible interior I_a, molecular interior, I_m, solvent-accessible surface S_a, and molecular surface S_m can be derived using the above operations. They are

$$I_a = SI_a \ominus E \tag{20}$$

$$I_m = SI_m \ominus E \tag{21}$$

$$S_a = SI_a \cap I_a^c \tag{22}$$

$$S_m = SI_m \cap I_m^c \tag{23}$$

Detection of cavities

In this article we regard protein molecules as rigid bodies having three-dimensional geometric shapes, since associating molecules are generally observed to behave as rigid bodies.

The active sites and other molecule-binding regions tend to be concave. The concavities and interior gaps of a mol-

ecule are defined as cavities. The detection of cavities is important in understanding the binding mechanism. However, the boundary between the cavity and exterior free space is difficult to discern.

We proposed a method to detect these boundaries using closing. The closing of a molecule by a probe sphere is an operation that fills the cavities. The boundaries of the filled molecule are divided into two parts: exposed molecular surfaces and cavity boundaries. These cavity boundaries are spherical surfaces that are part of the probe sphere. The cavities are detected by subtracting the molecule from the filled molecule (Figure 6).

Let the molecular surface of the object protein and its interior be SI_m^o and the probe sphere be R; the filled molecule, C^f, is then obtained by:

$$C^f = SI_m^o \bullet R \tag{24}$$

The cavities, C, which are distributed on the molecular surface and embedded in the molecular interior, are detected by subtracting the molecule, SI_m^o, from the filled molecule, C^f. It is expressed as:

$$C = C^f \cap (SI_m^o)^c \tag{25}$$

Some of the cavities are smaller than a water molecule. They should be removed because they do not contribute to the molecular binding. This removing operation is performed with the opening of C by the probe sphere, R_w. The cavities, C', which exclude the smaller cavities, are as follows:

$$C' = C \circ R_w \tag{26}$$

If we are concerned with the phenomenon of proton transfer, the radius of the probe sphere should be set smaller than that of a water molecule.

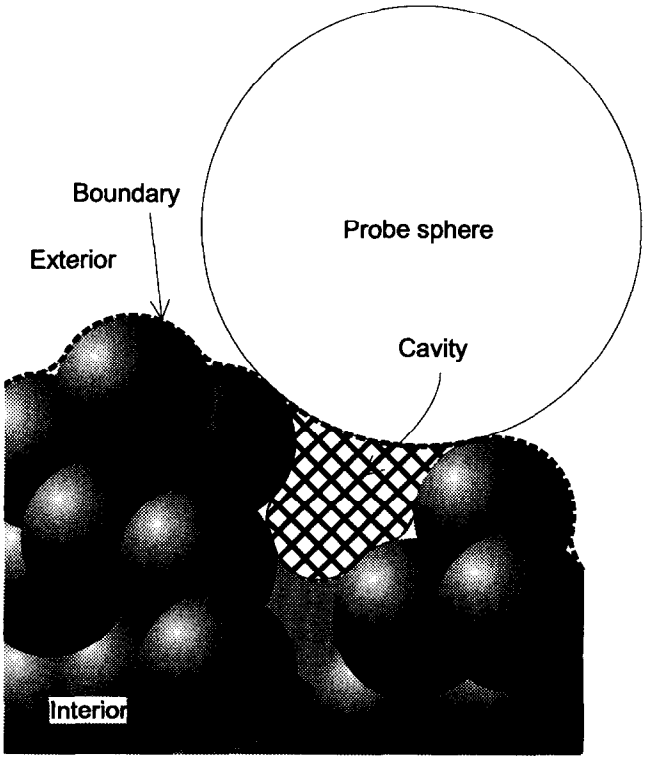

Figure 6. A cavity boundary (thick dotted line in contact with the probe sphere) and exposed molecular surface (thick dotted line without contacting the probe sphere) and their interior (region under the thick dotted line) using closing by the probe sphere. Cavities are extracted by subtracting the molecular surface and its interior from that.

Detection of portions in contact with cavities

It is necessary to determine the convex portions of the ligand in contact with the cavities in the case of the docking study. However, it is difficult to define the boundary between each portion in contact with each cavity and its exterior portion, which has no contact with the binding region.

We proposed a method to detect these boundaries using opening. The opening of the ligand by a probe sphere is an operation that removes convex portions of the ligand. The surfaces of removed portions are divided into two parts: exposed molecular surfaces and boundaries between removed portions and the rest of the ligand, which are parts of the probe sphere. We can detect the removed portions by subtracting the ligand from the result of the opening (Figure 7).

Let the molecular surface of the ligand protein and its interior be SI_m^l and the probe sphere be R; the exterior portion, which is the rest of the ligand, P^r is as follows:

$$P^r = SI_m^l \circ R \tag{27}$$

The portions of ligand in contact with cavities, P, are detected by subtracting P^r from SI_m^l. It is denoted as:

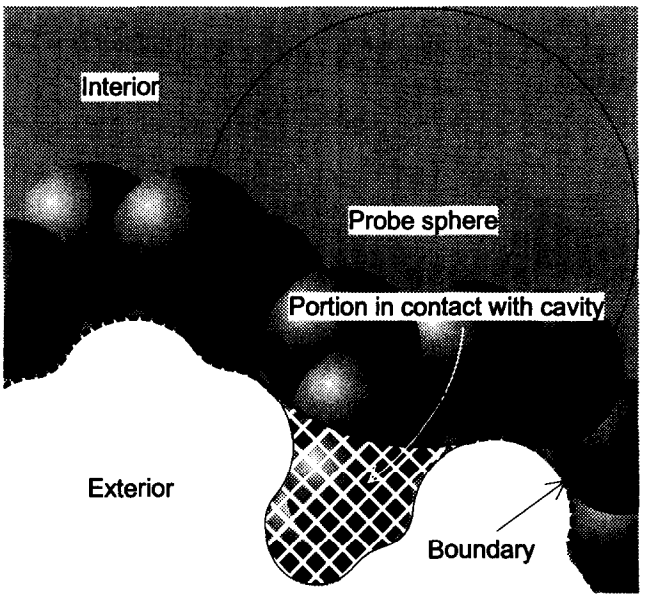

Figure 7. A boundary of portions of the ligands in contact with the cavities (thick dotted line in contact with the probe sphere) and exposed molecular surface (thick dotted line not in contact with the probe sphere) and their interior (region upward of the thick dotted line) using opening by the probe sphere. Portions in contact with the cavities are extracted by subtracting that from the molecular surface and its interior.

$$P = SI_m^1 \cap (P^r)^c \quad (28)$$

If the small portions of the ligand are unnecessary, they should be removed using the following opening, for instance, by the probe sphere, R_w. The portions that do not include the small portions, P', are as follows:

$$P' = P \circ R_w \quad (29)$$

If the cavities, C, and portions, P, of the ligand are detected using the same probe sphere, the boundary between a cavity and its exterior free space is identical to the removed boundary of the ligand, which is bound to the cavity.

When a probe with a large radius is used, the cavities can be detected but the portions of the ligand in contact with the cavities vanish. This is explained by the fact that the result of the opening becomes empty when the probe is large. When a probe with a small radius is used, many small cavities of the molecule and many small convex portions of the ligand in contact with the cavities are detected and it is difficult to determine which of them are bound. We defined the optimal radius as that of the largest sphere that is included in the ligand.

RESULTS

We applied the method to a trypsin–pancreatic trypsin inhibitor (PTI) complex (2PTC).[18] The complex is shown in Color Plate 1. Trypsin, which consists of 223 residues, is shown in magenta and the inhibitor of 58 residues is shown in yellow. The molecular surface of the inhibitor, which is calculated using Eq. (12), is shown in Color Plate 2 using a solid model. The inhibitor is depicted with a cyan wireframe representation and trypsin is shown with a magenta wire model. In the representation, the hidden wires are suppressed. The solid model is generated using a modified marching cube algorithm.[14–16] In the plate, the grid spacing of the wireframe is 0.77 Å. The resolution of the X-ray crystallography is 1.9 Å; therefore the grid spacing is less than one-half of the resolution. The surfaces of all of the detected cavities of trypsin are shown in Color Plate 3 using the same solid modeling. They are detected using Eq. (25). In the equation, we used a probe with a radius of 13 Å, which is the largest sphere included in the inhibitor. In this case, the cavities smaller than a water molecule have been removed using Eq. (26). Some 20 cavities are observed. The detected portions of the inhibitor that are in contact with the cavities are extracted using Eq. (28) with a probe with a radius of 13 Å. They are shown in Color Plate 4 using the same solid modeling. The binding portions are observed to be divided into two regions. In Color Plate 5, the atoms that belong to the portion of the inhibitor in contact with the cavities are expressed as space-filling models (yellow balls). The cavities of trypsin that are in contact with the inhibitor are extracted using Eq. (25), which is a region that belongs to the cavities shown in Color Plate 3, and are expressed as the solid model of the cyan grid. They agreed with each other.

We attempted to extract a heme region, which locates the interior of the myoglobin (4MBA)[19] molecule of 147 residues, using closing. Closing by a sphere with a radius of 6.0 Å was applied. The size of the sphere is determined to be similar to that of heme. The heme region, a region bounded by the space-filling model of the heme, is depicted using the same solid modeling in Color Plate 6. In the plate, the grid spacing of the wireframe is 0.60 Å. Myoglobin is depicted using a ball-and-stick model. The radius of the ball is one-tenth of the space-fill model. The heme is magenta and the others are green. The method can extract such an interior part. The extracted solid model agrees with the space-filled region (illustration not shown).

This method is able to calculate the volume of the object molecule by counting its consisting grid points. The volumes of trypsin and myoglobin are 28,065 and 18,291 Å^3, respectively.

CONCLUSIONS

We have developed a digital method based on mathematical morphological operations to obtain three types of surfaces—a van der Waals surface, a solvent-accessible surface, and a molecular surface—to extract the cavities on the surface and interior part of the molecule and to extract the ligand portions in contact with the cavities. The molecular surface, the cavities and the portions, and the heme region are visualized using solid modeling.

The method enables us to obtain the volumes of the cavities and inhibitor portions and the areas of the surfaces. The solid modeling enables us to obtain cross-sections at arbitrary positions. This is a useful approach to define the structural features of molecules. It can be widely used for modling arbitrary molecular structures.

REFERENCES

1 Lee, B., and Richards, F.M. The interpretation of protein structures: Estimation of static accessibility. *J. Mol. Biol.* 1971, **55,** 379–400.

2 Richards, F.M. Areas, volumes, packing, and protein structure. *Annu. Rev. Biophys. Bioeng.* 1977, **6,** 151–176

3 Richmond, T.J., and Richards, F.M. Packing of α-helices: Geometrical constrains and contact areas. *J. Mol. Biol.* 1978, **119,** 537–555

4 Alden, C.J., and Kim, S.-H., Solvent-accessible surfaces of nucleic acids. *J. Mol. Biol.* 1979, **132,** 411–434

5 Finney, J.L. Volume occupation, environment, and accessibility in proteins. Environment and molecular area of RNase-S. *J. Mol. Biol.* 1978, **119,** 415–441

6 Greer, J., and Bush, B.L. Macromolecular shape and surface maps by solvent exclusion. *Proc. Natl. Acad. Sci. U.S.A.* 1987, **75,** 303–307

7 Connolly, M.L. Analytical molecular sufrace calculation. *J. Appl. Crystallogr.* 1983, **16,** 548–558

8 Kuntz, I.D., Blaney, J.M., Oatley, S.J., Langridge, R., and Ferrin, T.E. A geometric approach to macromolecule–ligand interactions. *J. Mol. Biol.* 1982, **161,** 269–288.

9 Ho, M.W., and Marshall, G.R. Cavity search: An algorithm for the isolation and display of cavity-like binding regions. *J. Comput.-Aided Mol. Design* 1990, **4,** 337–354

10 Delaney, J.S. Finding and filling protein cavities using cellular logic operations. *J. Mol. Graphics* 1992, **10,** 174–177
11 Kisljuk, O.S., Kachalova, G.S., and Lania, N.P. An algorithm to find channels and cavities within protein crystals. *J. Mol. Graphics* 1994, **12,** 305–307
12 Matheron, G. *Random Sets and Integral Geometry,* John Wiley & Sons, New York, 1975
13 Serra, J. *Image Analysis and Mathematical Morphology.* Academic Press, New York/London, 1982
14 Ikeguchi, M., and Doi, J. Solid modeling for iso-value potential surface of bio-molecules. *J. Visual. Soc. Jpn.* 1991, **11**(Suppl. 2), 105–108 [in Japanese]
15 Ikeguchi, M., Ishii, T., and Doi, J. Visualization of bio-molecular structures and activities. *Trans. Soc. Instrument Control Eng.* 1992, **28,** 154–155 [in Japanese]
16 Ikeguchi, M., and Doi, J. Electrostatic field visualization on HIV-1 protease surface. *Album Visualization,* 1994, **11,** 19–20
17 Creigton, T.E. *Proteins: Structures and Molecular Properties,* 2nd Ed. W.H. Freeman and Company, New York, 1993
18 Marquart, M., Walter, J., Deisenhofer, J., Bode, W., and Huber, R. The geometry of the reactive site and of the peptide groups in trypsin, trypsinogen and its complexes with inhibitors. *Acta Crystallogr. Sect. B* 1983, **39,** 480–490
19 Bolognesi, M., Onesti, S., Gatti, G., Coda, A., Ascenzi, P., and Brunori, M. *Aplysia limacina* myoglobin. Crystallographic analysis at 1.6 Å resolution. *J. Mol. Biol.* 1989, **205,** 529–544.