Tetrahedron Vol. 43, No. 22, pp. 5229 to 5236, 1987
Printed in Great Britain.

0040-4020/87 $3.00+.00

A NOVEL METHOD FOR SUPERIMPOSING MOLECULES AND RECEPTOR MAPPING

Yuichi Kato, Akiko Itai* and Yoichi Iitaka

Faculty of Pharmaceutical Sciences,
University of Tokyo, Hongo, Bunkyo-ku, Tokyo, Japan

(*Received in Japan* 15 *July* 1987)

ABSTRACT

Superposition of molecules on a three-dimensional computer graphic display is an efficient means to compare three-dimensional molecular structures. Biologically active molecules, which are presumed to bind to the same receptor site, are thought to have common structural features. But, it is the physical and chemical properties arranged spatially through chemical structures that are important for specific binding to a receptor. Therefore, for the purpose of studies on biological activities, molecules should be superposed to those properties, not to the atomic positions as in the traditional methods. We have developed a program system for realizing this new concept. The concept stands on the general perceptions of organic chemists about hydrogen bondings and chemical isosterisms. The 'goodness of fit' values, which are estimated in realtime on the basis of spatial similarity of those properties between molecules, are displayed and updated throughout the superposing process. This program can construct a receptor cavity model and provide with the cavity size and shape, surface electrostatic potentials, hydrogen bonding sites and so on, by using all information supplied by the superposed molecules. This model can be modified by further superposing of another molecule. These constructed models would be of help for rational drug design, when the receptor structures are not yet known.

It is well established that the three-dimensional structure is important for biological activity of molecules. And consequently the superposition of molecules on a three-dimensional computer graphic display (abbreviated as 3D-CG hereafter) is an efficient means to compare molecular structures three-dimensionally.[1] Biologically active compounds, which are supposed to bind specifically to the same receptor macromolecule, should have common structural and physical features required for the specific molecular recognition. Molecules with apparently different chemical structures often exhibit the same kind of biological activities and show similar biochemical and pharmacological behaviors. This fact strongly suggests that it is the physical and chemical properties, not the chemical structure, that is responsible for the biological activities. Therefore, molecules should be superposed to those properties not to the atomic positions. The superpositions of molecules in terms of hetero atoms, as were attempted quite often, are sometimes meaningless between fairly different structures, because the coincidence between them in atomic species and atomic positions are unnecessary for the binding to the same receptor. Methods for superposition conventionally used so far are: 1) by least-squares calculation specifying the atom-pairs between molecules, and 2) by local manipulation of individual molecules on 3D-CG with visual judgment of the fit. The former method can not be applied to molecules in which the atom-pair specifications are difficult because of large discrepancies between the chemical structures, while the latter can not yield any numerical index of the goodness of fit in spite of its wide applicability to different structures.

By using the structural information of superposed plural molecules, images of a receptor cavity can be described more accurately than by using the information

template, the dihydrofolate molecule was superposed as a trial molecule. As for the conformation of dihydrofolate molecule trapped in the enzyme active site, two representative conformers have been considered so far. The conformer A corresponds to that of methotrexate molecule found in the crystal of the ternary complex, and the conformer B is presumed on the basis of reported one,which is favorable for hydrogen bondings to the known receptor molecule.[8] The difference between the conformer A and B is that the pteridine rings are reversed about C6-C9 bond for each other. Both conformers of dihydrofolate molecule were independently superposed to the methotrexate template molecule and the similarity of the physical and chemical properties between them were monitored by the 'goodness of fit' values displayed on the upper part of the screen (see Figures 2(a) and 2(b)). Figures 3(a) and 3(b) illustrate the superposed wire skeletal models for both conformers on the methotrexate template. Table 1 shows a comparison between conformer A and B employing the various 'goodness of fit' criteria(final values). The reference values, which are expected for the perfect superposition of the same molecules, are shown in the first line of the table. The values for shape and charge distribution favored conformer A, whereas those for electrostatic potential and hydrogen bonding favored conformer B. Now, the conformer B is firmly supposed to be the crucial conformer for the receptor binding of dihydrofolate molecule on the basis of the stereochemistry of the enzyme reaction product, tetrahydrofolate molecule. The conformer A should be converted into a hydrogenated product with the different stereochemistry about the position C-6, because the pteridine ring in the conformer A is rotated by 180° compared to that of the conformer B.

Accordingly, the 'goodness of fit' for electrostatic potential and hydrogen bonding proved to be effective for discriminating the correct superposition in this case. As the two molecules with a similar shape were superposed in case of superposition of the conformer A, it seems quite natural that the 'goodness of fit' value for molecular shape, and consequently the value for charge distribution in the molecules are favorable for the conformer A. This example strongly suggests that the coincidence of molecular shapes is not always essential for the binding to the same receptor and the superposing molecules in terms of atomic positions is useless in some cases. It is widely accepted that among various intermolecular forces between drugs and receptors, hydrogen bondings, electrostatic interactions and hydrophobic interactions are very important for the binding specificity to the receptor. And the correlation of electrostatic potentials can indicate directly the similarity of electrostatic properties required for receptor binding between molecules better than the charge distribution in the molecules. Although we have not treated the fitness of hydrophobic character in this paper, our programs provide a function to assess it.

In Table 2, the indices for the correlation of electrostatic potentials calculated for the superposed molecules by two different methods, which were explained in 'goodness of fit' section, are compared to the ideal one obtained by using the van der Waals surface points of the actual receptor protein, dihydrofolate reductase. Although three kinds of indices show the same trends for the conformer A and B, the values from the current molecular surface method are in better agreement with the ideal ones than the values from the spherical approximation method.

The constructed receptor model based on conformer B is shown in Figures 4(a) (clip off) and 4(b) (clip on). The size and shape of the cavity are represented by cage-expression which is color-coded according to the computed surface electrostatic potential with the reversed sign. The colored dots express the locations of hydrogen bonding functional hetero atoms in the receptor. The receptor model comprising the two molecules based on is shown in Figure 4(c).

from a single molecule. Receptor cavity models would be useful not only to design new drugs, but also to elucidate structure-activity relationships, when the structures of the receptor macromolecules are unknown. More efficient and reasonable methods of superposing molecules has been expected to be developed for a long time. We have developed a new rational method for superposing molecules on the prerequisite of specific binding to a common receptor, and for three-dimensional receptor mapping to describe the environment of receptor cavity.[2]

Method

Molecules are superposed to the physical and chemical properties through a three-dimensional grid in our method, whereas they are superposed to the atomic positions of the molecules in the conventional methods. First, a template molecule must be chosen, whose structure should be rigid or conformationally defined. On a 3D-CG, a rectangular box is set up to extract the essential region for specific binding to the receptor and to define the ranges for grid point calculation. The lengths of three edgelines and the position of the box are determined interactively so as not only to cover the required region of the template molecule but also to have a sufficient reserve of space for the subsequent superposition of other molecules. Then three-dimensional grids with a regular interval of 0.4-1.0 Å are generated inside the box. For each grid point, the following physical and chemical properties are calculated and stored: electrostatic potential, charge distribution, hydrogen bonding character, flag on occupancy by each molecule, and flag for molecular surface. New molecules (named trial molecules) are superposed to the graphic expression of these three-dimensionally tabulated data. The 'goodness of fit' values are calculated on the basis of spatial similarity of the physical and chemical properties between molecules using the tabulated data. The values are displayed on 3D-CG and updated during interactive manipulation of the trial molecule. Trial molecules are superposed one after another, and the new coordinates are stored into a file successively. After that, the grid point data are calculated for the new atomic coordinates of all the molecules, from which the united grid point data are obtained by using weights of biological activities. A receptor cavity model, providing information on cavity size and shape, surface electrostatic potential, locations of hydrogen bonding hetero atoms and so on, can be described using the united grid point data. The obtained receptor cavity model can be represented on a 3D-CG in various ways and can be further modified (including enlargement of size) by superposing additional molecules.

'Goodness of fit'

For the time being, address of each grid point, flag on occupancy by molecules, charge distribution, electrostatic potential and hydrogen bonding character are tabulated three-dimensionally as grid point data. They are used to represent the spatial arrangement of properties of molecules and are used to calculate the 'goodness of fit' in realtime. 'Goodness of fit' values are calculated by using tabulated data for the template molecule and atomic data for the trial molecule, which are varied by the interactive manipulation. Goodness of fit terms tested in this study are for shape, charge distribution, electrostatic potentials and hydrogen bonding. The goodness of fit for shape, F_{shape}, is defined as the ratio of number of commonly occupied grid points by both molecules to the number of occupied ones by the template molecule. The charge distributions, which we have tentatively defined from the atomic charges so as to be distributed on the grid points around the atoms by a Gaussian distribution, are calculated inside the van

der Waals volume of each molecule, whereas the electrostatic potentials are calculated outside it. The goodness of fit for charge distribution, F_{charge}, and for electrostatic potential, F_{elpo}, are defined as shown in chart 1. The goodness of fit for hydrogen bonding, $F_{H\text{-}bond}$, is defined as the ratio the number of commonly formed hydrogen bondings to the number of possible hydrogen bondings in the template molecule. Suitable terms should be selected on the basis of their effectiveness and the equations should be improved by applying them to many cases, in order to discriminate effectively the correct superposition from incorrect ones.

$$F_{charge} = \frac{\sum |c_j - q_i|^n}{\sum |c_j|^n}$$

j : the nearest grid point to i-th atom

c_j: charge distribution on j-th grid point

$$F_{elpo} = \frac{\sum_k (V_{k(temp)} \times V_{k(trial)})^2}{\sqrt{\sum_k |V_{k(temp)}|^2} \times \sqrt{\sum_k |V_{k(trial)}|^2}}$$

where $V_k = K\sum q_i/\varepsilon \cdot r_{ik}$

Chart 1

As regards hydrogen bonds, probable locations of hydrogen bonding partners in the receptor are predicted by considering the distances and directions from all hydrogen bonding functional groups in the template molecule, and are described with a flag on each grid point. Flags are used to indicate hydrogen donor or acceptor or amphiphilic character. Allowable range for the distances between two hydrogen bonding hetero atoms were set up as from 2.5 Å to 3.1 Å, by taking into account of the grid interval. As for the direction of the presumed location of the hydrogen bonding partner, allowable deviation from the orientation vector of X-H or Y-lone-pair electrons (X, Y = N, O, S...) were assumed to be 30°. This criterion was chosen by referring to the studies on statistical analysis of crystallographic database[3]. 'Goodness of fit' for hydrogen bonds between two molecules is estimated from the coincidence of both the character and region. For all molecules, we take into account the bond-rotational ambiguity of the C-X bond in C-X-H and C-Y bond in C-Y-lone pair electrons (X, Y = N, O, S...), as well as the distinct distances and angles between the lone-pairs, hydrogens and hetero atoms.

For the assessment of the 'goodness of fit' for electrostatic potentials, we have attempted another method of calculation, in addition to the correlation of those on the surface of superposed plural molecules, current molecular surface method. That is, we have assumed a series of concentric spheres, whose center and radii are determined to fully include both the template and trial molecules, as a temporary receptor cavity. On the randomly generated points on each sphere, the correlation is calculated between the electrostatic potentials from the template molecule and those from trial molecules. For the realtime use, this fixed spherical approximation would be useful, because the assessment on the real surface of the superposed plural molecules is not so fast due to the change of the surface points at every stage of manipulation.

Program

The program system RECEPS was developed to realize the ideas described above. The RECEPS system consists of three programs PRECEP, GRDFIT and CONVRG. These programs are linked through each other through atomic coordinates files and grid point data files. The role of each program is as follows.

Program PRECEP: Inside the rectangular box, in which the template molecule is properly arranged on 3D-CG, three-dimensional grids with a regular interval are generated. Then, various physical and chemical properties are calculated on every grid point, which are stored and used for superposing other molecules. In addi-

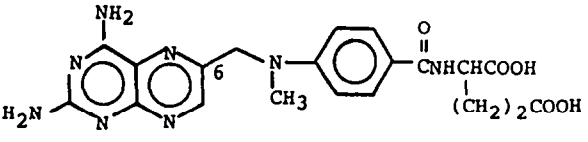
tion, this program is used to calculate the same kinds of grid point data for molecules already superposed by program GRDFIT.

Program GRDFIT: In this program, trial molecules are superposed on 3D-CG, displaying the grid point data expression as well as a wire skeletal model of the template molecule. The trial molecule is interactively manipulated (with local rotation, translation and bond rotation). The 'goodness of fit' values are calculated and displayed throughout the superposing process in order to lead the trial molecule to an appropriate position. The new atomic coordinates for superposed molecules are stored into an atomic coordinates file one after another and are transferred to program CONVRG, together with the grid point data file resulted from PRECEP.

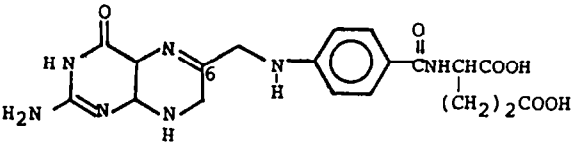
Program CONVRG: The grid point data for plural molecules calculated in program PRECEP, using the atomic coordinates obtained from program GRDFIT, are united in this program. The united grid point data can be used to represent a receptor cavity model showing the size, shape, electrostatic potential, charge distribution, hydrogen bonding character and so on. Moreover, this program can perform the various expression about superposed molecules, such as an enzyme excluded volume, essential volume and differential volume[4], in addition to the color-coded dots expression of Connolly's molecular surface for each molecule.[5]

Atomic coordinates data might be taken from a crystallographic database or prepared by molecular modeling on 3D-CG. If necessary, molecular mechanics calculations can be performed. Atomic charges should be calculated in advance by molecular orbital calculations. Category number should be given in advance to all hetero atoms in hydrogen bonding functional groups in order to generate the lone-pair positions automatically. Category number corresponds to each hydrogen bonding functional group, such as hydroxyl O, carbonyl O, ether O, carboxyl O, amino N, amide N, aromatic N, sulfhydryl S and so on, whose geometry about hydrogens or lone-pairs and rotational ambiguity as well as the hydrogen bonding character are prepared in program PRECEP. All programs are written in FORTRAN77 and work on a Daikin DS301B (or DS361B) raster three-dimensional color graphic display with a HITAC M-280H(or 680H) computer. The programs are available for the users who have the same type of graphic devices on request. And moreover, we are preparing for the new versions for other graphic displays. Details of this program system will be presented elsewhere.

As an example to illustrate the usefulness of this program system, we attempted to superpose methotrexate (I) and dihydrofolic acid (II) molecules. Both molecules have a pteridine ring system. Because detailed structural information of the ternary complex of dihydrofolate reductase-methotrexate-NADPH is available from the X-ray crystal analysis of Bolin et al[6], this system seemed to be a good one to test our method. We examined whether or not the receptor cavity model obtained by superposing metho-trexate and dihydrofolate molecules resembles the cavity of the actual receptor, dihydro-folate reductase. The conformation of methotrexate molecule was fixed to that found in the crystal and was used as the template molecule. The structural part which is supposed to be essential for specific binding to the receptor was extracted with a rectangular box as shown in Figure 1, because the whole structure is not necessarily required for specific binding. To the representation of grid point data for model of the



methotrexate (I)



dihydrofolic acid (II)

Figure 5(a) shows a comparison between the constructed model obtained as described above (with the hydrogen bonding functional heteroatoms in the receptor expressed in white and the actual receptor enzyme dihydrofolate reductase whose surface is expressed by red dots). Both boundaries well agree with each other. By contrast, the preliminary receptor model (shown as green cross-hatching) constructed from the methotrexate template only shows large vacant space when compared with the actual enzyme, as shown in Figure 5(b). Thus the more correct model for the receptor cavity can be deduced by our new method.

Conclusion

A new reasonable method for superposing molecules and constructing a receptor model on 3D-CG have been developed. The validity and the usefulness of our system are supported by the results of its application to methotrexate and dihydrofolic acid. A distinct merit of superposing plural molecules is exemplified.

This method offers significant advantages: (1) Molecules whose chemical structures are quite different may be compared with each other. (2) The method may enable to consider the cases, where similar stabilization can be attained even from different directions of interaction with receptor functional groups, through hydrogen bonding or ionic bondings. (3) The method may be used in those cases in which delocalized electron distributions are significant in biological activity.

The characteristic of our method is that the 'goodness of fit' value can be estimated in real time throughout the interactive superposing process on a 3D-CG. Although further improvement and refinement of 'goodness of fit' criteria are desirable, our method should already be of great use in many cases. The more reasonable the superposition the more accurate the receptor image that would be constructed. This enables not only the extraction of structural requirements for biological activity but also the determination of the allowable spatial arrangement among them. The receptor cavity description estimated by this method would be of help for rational drug design, in the following ways: (1) for explaining the relationships between three-dimensional structures and biological activities. (2) for predicting whether a certain drug candidate molecule can bind strongly to the receptor or not. (3) for constructing chemical structures which are expected to bind more strongly to the receptor, by fitting molecules or fragments to the visualized receptor cavity on 3D-CG.

References

1) Marshall, G. R. "Drug Design: Fact or Fantasy", Academic Press: New York, 1984;p.35.
2) Itai, A., Kato, Y., Iitaka Y., Abstract of Symposium on Three-Dimensional Structures and Drug Action, 1986, p.54. Tokyo.
3) Taylor, R., Kennard, O., Versichel, W., Acta.Cryst. 1984, B40, 280.
Taylor, K., Kennard, O., Versichel, W., J.Am.Chem.Soc. 1983, 105, 5761.
Vedani, A., Dunitz, J. D., J.Am.Chem.Soc, 1985, 107,7653.
Murray-Rust, P., Glusker, J. P., J.Am.Chem.Soc. 1984, 106, 1018.
4) Marshal, G. R., "Computer-Assisted Drug Design"; American Chemical Society: Washington, DC, 1979.
5) Connolly, M., QCPE Bull. 1981, 1, 75.
6) Bolin, J. T., Filman, D. J., Matthews, D. A., Hamlin, R. C., Kraut, J. J. Biol. Chem. 1982,257, 13650.
7) Camerman, A., Mastropaolo, D., Camerman, N., "X-ray Crystallography and Drug Action"; Oxford University press: Oxford, 1984.

Table 1. Tested 'goodness of fit' values for comparison between dihydrofolate conformer A and B superposing on methotrexate molecule. Reference means the perfect superposing of a molecule (MTX) on itself.
F_{shape}: similarity of shape, F_{charge}: similarity of charge distribution, F_{elpo}: correlation index of electrostatic potential by the current molecular surface method, $F_{H\text{-}bond}$: ratio of number of formed hydrogen bond to number of possible hydrogen bonding sites.

	F_{shape}	F_{charge}	F_{elpo}	$F_{H\text{-}bond}$
reference (MTX)	1.00	0.00	1.00	1.00
conformer A	0.85	0.94	0.07	0.29
conformer B	0.67	1.62	0.29	0.71

Table 2. Comparison of the two methods for assessing the correlation of electrostatic potentials on the superposed molecules with the ideal values calculated by using the actual receptor protein structure.

methos	current molecular surface method	supherical approximation method	ideal values obtained from the actual protein structure
conformer A	0.07	0.16	0.10
conformer B	0.29	0.27	0.30

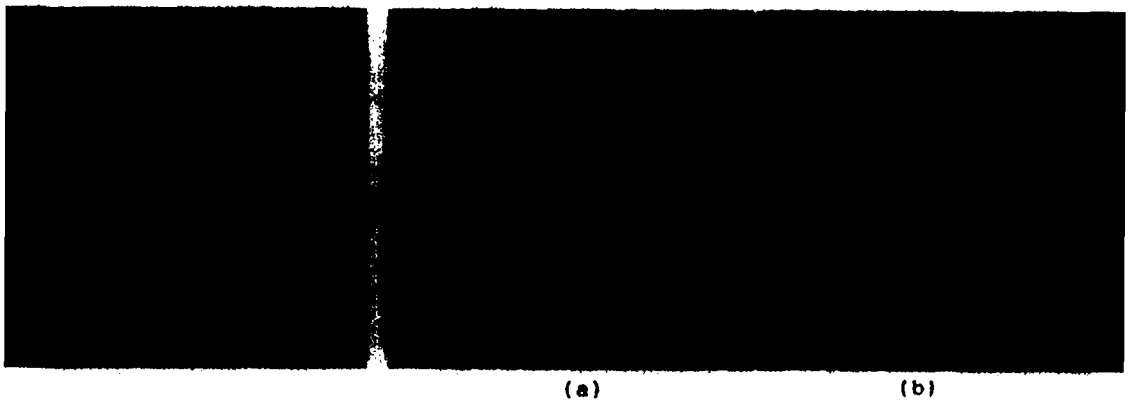


Figure 1. A rectangular box, by which the ranges for grid points calculations are determined, and the structural parts required for specific binding to receptor are extracted from the template molecule (N-1 protonated state of methotrexate). Three-dimensional grid points with a regular interval are generated inside this box.

Figure 2. Dihydrofolate molecule is superposed on the grid point data. The locations of hydrogen bonding functionalities in the receptor expected from the template molecule are represented (pink, donor; light blue, acceptor). Other properties are not displayed in this plate for clarity. The various 'goodness of fit' values for this position of the trial molecule are shown at the upper part of the display.
(a) conformer A
(b) conformer B

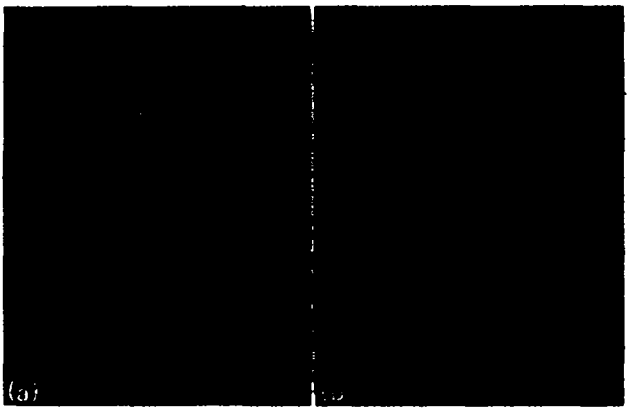


Figure 3. Dihydrofolate (sky blue) superposed on the methotrexate molecule (yellow). Color codes for atoms are: cobalt blue, nitrogen atoms; red, oxygen atoms. (a) conformer A (b) conformer B

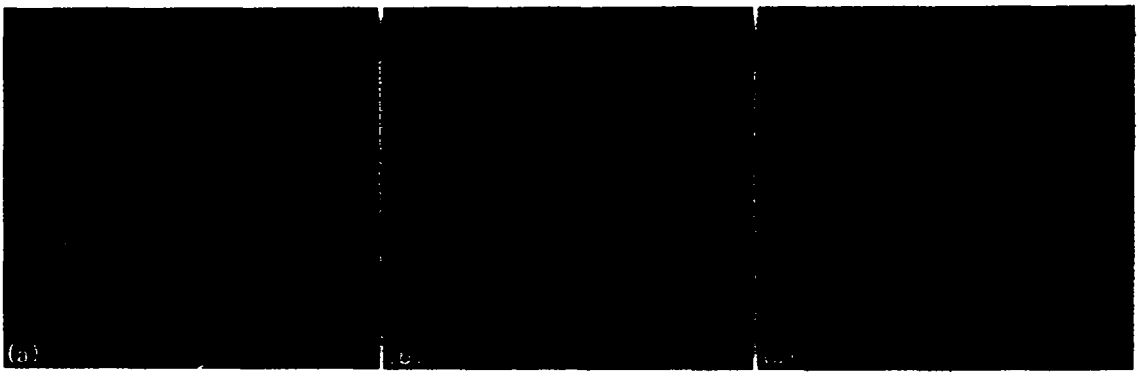


Figure 4. Constructed receptor cavity model obtained from the superposition of methotrexate and dihydrofolate molecules. The size and shape of the receptor cavity are shown by the cage. Color codes express the reversed sign of the surface electrostatic potential for superposed molecules (yellow, negative; blue, positive). Colored dots around the cage express the locations of hydrogen bonding functionalities in the receptor (color codes are the same as in Figure 2). (a) not clipped (b) clipped (c) Clipped repreaentation of the receptor cavity containing the two molecules concerned.

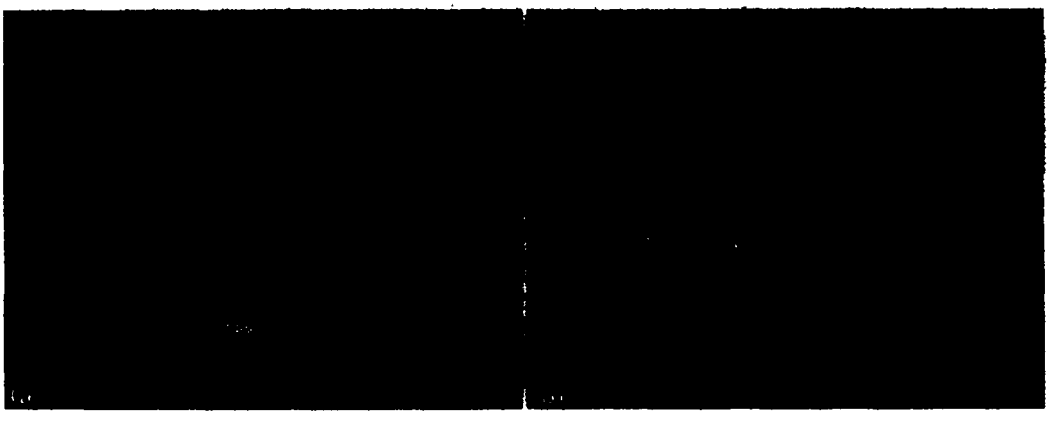


Figure 5. A comparison of the cavity shape between the receptor model and the actual enzyme. The white cage indicates the cavity of the receptor. Red dots express the molecular surface of the actual receptor enzyme. Both expressions are clipped with the same depth level. A good agreement between the two boundaries is seen in Figure 5(a). Comparing 5(a) with 5(b), a large unoccupied space is left at the rifht side of the cavity in the latter. (a) receptor model based on the superpostion of methotrexate and dihydrofolate molecules. (b) based on the single molecule of methotrexate.