ELSEVIER

Journal of Molecular Graphics and Modelling 20 (2002) 313–328

Journal of Molecular Graphics and Modelling

www.elsevier.com/locate/JMGM

Protein–ligand recognition using spherical harmonic molecular surfaces: towards a fast and efficient filter for large virtual throughput screening

Wensheng Cai [a], Xueguang Shao [b], Bernard Maigret [c,*]

[a] *Department of Applied Chemistry, University of Science and Technology of China, Hefei, Anhui 230026, PR China*
[b] *Department of Chemistry, University of Science and Technology of China, Hefei, Anhui 230026, PR China*
[c] *UMR 7565 CNRS/UHP, H. Poincare University, B.P. 239, 54506 Vandoeuvre-Les-Nancy Cedex, France*

Received 22 March 2001; received in revised form 20 September 2001; accepted 24 September 2001

Abstract

Molecular surfaces are important because surface-shape complementarity is often a necessary condition in protein–ligand interactions and docking studies. We have previously described a fast and efficient method to obtain triangulated surface-meshes by topologically mapping ellipsoids on molecular surfaces. In this paper, we present an extension of our work to spherical harmonic surfaces in order to approximate molecular surfaces of both ligands and receptor-cavities and to easily check the surface-shape complementarity. The method consists of (1) finding lobes and holes on both ligand and cavity surfaces using contour maps of radius functions with spherical harmonic expansions, (2) superposing the surfaces around a given binding site by minimizing the distance between their respective expansion coefficients. This docking procedure capabilities was demonstrated by application to 35 protein–ligand complexes of known crystal structures. The method can also be easily and efficiently used as a filter to detect in a large conformational sampling the possible conformations presenting good complementarity with the receptor site, and being, therefore, good candidates for further more elaborate docking studies. This "virtual screening" was demonstrated on the platelet thrombin receptor.

Keywords: Spherical harmonics; Molecular surface; Shape complementarity; Molecular docking; Molecular recognition

1. Introduction

The interaction models of protein–protein, protein–ligand or substrate–enzyme are important in understanding these interactions, especially in the field of computer-aided drug design. A lot of protein docking algorithms have been reviewed and discussed for their application [1–3], such as simplified representation of protein geometry, Fourier correlation algorithm, sparse critical points method, buried surface area, and empirical potentials and functions which are being increasingly used [3,4]. In predicting protein–protein and protein–ligand interactions, the challenge is to start with the coordinates of the unbound component molecules and to obtain computationally a model for the bound complex, including the conformational changes on docking [2]. Although several methods have been proposed for that [2], it is still a difficult problem.

The accurate simulation of molecular interactions and recognition processes requires appropriate molecular models. Molecular surfaces are often used because they represent the interacting part of molecules and they have also been shown to exhibit very high level of complementarity both for a geometric point of view and for chemical properties [5–14]. Several methods have been developed for that purpose [15–18]. Spherical harmonic models based on expansions of spherical harmonic functions have also been used for quantitative description of molecular shapes [19–25]. The models offer several advantages which include being analytical models, controlling over level of approximation, allowing representation of not only geometry, but also chemical properties and quantitative measure of surface similarity.

But the models also present challenges. Since spherical harmonic functions are defined on a unit sphere, a genus 0 surface is needed. Therefore, before computing harmonics, the input surface must be topologically mapped onto a unit sphere. After the topological mapping procedure, a star-like surface, i.e. the surface must be single-valued with respect to rays from the chosen origin inside the surface, was obtained. The spherical harmonic expansion coefficients can be determined from this surface.

Previous topological mapping method [21] maps the molecular surface computed by Connolly's programs (AMS, CT, and TS) onto a unit sphere by representing it as an

* Corresponding author. Tel.: +33-383-91-2528; fax: +33-383-91-2530.
E-mail address: bernard.maigret@lctn.uhp-nancy.fr (B. Maigret).

1093-3263/02/$ – see front matter
PII: S1093-3263(01)00134-6

elastic surface and modeling its continuous deformation into a spherical shape. We have presented another approach [26] that starts with the triangular mesh distributed on an ellipsoid or a sphere embracing the molecule. This ellipsoid is deflated to approximate the molecular surface. During this stepwise deflation procedure, it is ensured that each ray from the origin intersects the triangle on the surface at only one point. Therefore, an expansion of spherical harmonic functions can be computed. But the deflation method cannot provide good model for deep and narrow cavities where most substrate–enzyme or receptor–ligand interactions occur. In order to investigate the molecular complementarity in binding site regions, it is necessary to model the surfaces of the cavities [27–29].

In this paper, we present a variation of that technique allowing to represent with more accuracy molecular cavities. We do so by inverting the deflating process and inflate a sphere placed initially inside the cavity until it approximates the cavity's shape.

It has also been shown that protein–ligand interactions often involve lobe–hole type interactions. The method we present here is capable of identifying these regions of interactions from contour maps computed using spherical harmonic surfaces. The complementarity of lobes and holes by visual inspection is very clear. This was demonstrated by application to the protein–ligand complexes with known structures.

The spherical harmonic technique can be used to quantify the complementarity of the shape and properties of molecules [30] and is, therefore, useful to search for relative orientation of molecules that are favorable for their interaction. Since rotations of spherical harmonic surfaces can be done by rotating the expansion coefficients, the complementarity of interactive surfaces around a given binding site can be easily searched by minimizing the distance between their respective expansion coefficients [31]. But the global docking search with six degrees of freedom(translation and rotation) is still difficult because the surface integration for calculating the expansion coefficients is computationally expensive. Thirty-five protein–ligand complexes selected from Protein Data Bank (PDB) are tested as examples. The ligands are small molecules and are docked in the corresponding cavities of their receptors. The spherical harmonic surfaces for the separated ligands and cavities, respectively are calculated. A docking procedure combined with genetic algorithm (GA) was developed in this study, and used to find the rotation of a ligand relative to its receptor that minimizes the shape difference between surfaces of the ligand and the cavity. Staring with their spherical harmonic representations, it rotates the ligand surface by rotations of its spherical harmonic expansion coefficients to fit the cavity surface. The GA is adopted to minimize the distance between their expansion coefficients to obtain the corresponding Euler rotation angles.

Computational screening or virtual screening is used to select candidates in very large combinatorial libraries. The method described above was also applied to virtual screening as a fingerprint map procedure.

2. Methods

2.1. Spherical harmonic representations of molecular surfaces

Spherical harmonics are single-valued, continuous bounded, complex functions of the spherical coordinates (θ, ϕ), which can be considered as "standing waves on a sphere". They are characterized by two quantum numbers L and M, which together determine the number and spatial arrangement of nodes in each function. Any single-valued three-dimensional surfaces can be approximated by encoding the radial distance of surface points from the origin as a sum of spherical harmonic functions. Spherical harmonic functions are defined as follows [19–21]:

$$Y_l^m(\theta,\phi) = \sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}}P_l^m\cos\theta\,\mathrm{e}^{\mathrm{i}m\phi} \tag{1}$$

where l and m are integers, $m = -1, -l+1, \ldots, 0, \ldots, l$ and $P_l^m\cos\theta$ are the associated Legendre functions. They form a complete orthonormal basis set.

The use of complex quantities in computer programs increases storage and CPU time requirements. In our study, the real spherical harmonics $S_l^m(\theta,\phi)$ are used which can be represented by the linear combination of the complex functions as follows:

$$\begin{aligned}
S_l^m(\theta,\phi) &= \frac{1}{\sqrt{2}}(Y_l^m + Y_l^{-m}) \\
&= \sqrt{\frac{2l+1}{2\pi}\frac{(l-m)!}{(l+m)!}}P_l^m\cos\theta\cos m\phi \\
S_l^0(\theta,\phi) &= Y_l^0 = \sqrt{\frac{2l+1}{4\pi}}P_l^0\cos\theta \\
S_l^{-m}(\theta,\phi) &= \frac{1}{i\sqrt{2}}(Y_l^m - Y_l^{-m}) \\
&= \sqrt{\frac{2l+1}{2\pi}\frac{(l-m)!}{(l+m)!}}P_l^m\cos\theta\sin m\phi
\end{aligned} \tag{2}$$

where $m > 0$.

A single-valued function $f(\theta, \phi)$ can be expressed as an expansion of real spherical harmonic functions:

$$f(\theta,\phi) = \sum_{l=0}^{L}\sum_{m=-l}^{l} C_{lm}S_l^m(\theta,\phi) \tag{3}$$

The coefficient C_{lm} in the expansion of the function $f(\theta, \phi)$ can be found as an inner product:

$$C_{lm} = \langle f, S_l^m\rangle = \int_0^{\pi}\int_0^{2\pi} f(\theta,\phi)S_l^m(\theta,\phi)\sin\theta\,\mathrm{d}\theta\,\mathrm{d}\phi \tag{4}$$

A star-like molecular surface can, thus, be represented as a sum of real spherical harmonic functions up to a limiting value of L:

$$r(\theta, \phi) = \sum_{l=0}^{L} \sum_{m=-l}^{l} C_{lm} S_l^m(\theta, \phi) \quad (5)$$

where L is the order which determines the accuracy of the representation, $r(\theta, \phi)$ are the distance functions of surface points from the origin inside.

The expansion coefficients are determined by the surface integral similar to Eq. (4). Once the coefficients are known, the surface can be rendered using Eq. (5).

2.2. Molecular embracing surface

We have described an algorithm to define a uniform mesh on an ellipsoid embracing a molecule [26]. Such a mesh could be "deflated" step by step to obtain an approximation of a molecular surface as defined by Connolly [15–17]. The deflated ellipsoid can then be used to obtain models of the molecular surface based on spherical harmonic expansions. But the deflation technique does not provide good models for deep and narrow cavities where typical protein–ligand interactions take place because the triangles cannot reach the molecular surfaces of the cavities.

2.3. Surface of cavity

To obtain an usable models describing a cavity we adapted our algorithm to inflate a sphere inside the cavity(as in Fig. 1). The inverted mapping procedure from the sphere to the molecular surface of the cavity is achieved by inflating the triangular mesh on the sphere step by step with the similar technique. During inflation procedure the only difference from deflation is the direction along which the triangular vertices will move.

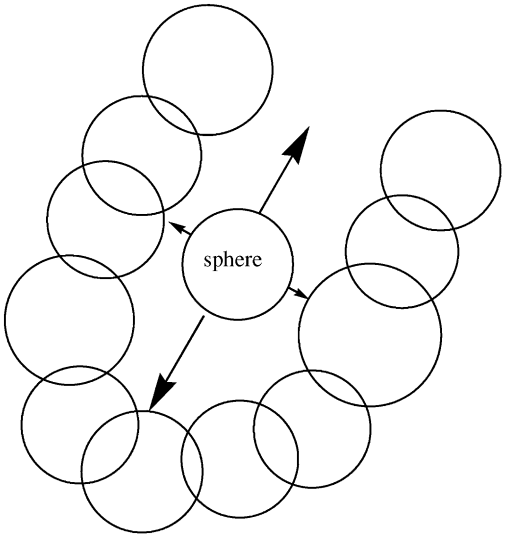


Fig. 1. Inflation of a small sphere inside the cavity to approximate the surface of it.

In the algorithm, two types of inflation directions are defined, a normal direction and a centrifugal direction. For each sphere's vertex, the normal direction is defined as the average normal vector of its adjacent triangles backing toward the center of the sphere. In each inflation step, this direction will be used first. If this inflation is not modified after a number of steps some triangles will not remain inward facing. The centrifugal direction is then used which is defined as the direction from the center of the sphere to surface points. This guaranties that each intermediate surface is star-like. Along the inflation direction, the step size for moving points is changeable within a given range. Similar types of deflation direction were also implemented in our modified deflation algorithm.

In our study, the center of the ellipsoid embracing the binding ligand is used as the center of the sphere in the cavity. The radius of the sphere is 1 Å. The whole algorithm to approximate molecular surfaces of cavities can be described as follows:

1. Build a triangular mesh on an unit sphere by subdivision of an icosahedron on it. Put the sphere at a given position inside the cavity.
2. Calculate the elements of the molecular surface of the cavity from its atomic coordinates, including convex pieces, spherical triangles and tori.
3. Calculate the inflation direction of each surface point. This is the direction along which each point will be moved in the next step. In each inflation step, the normal direction of each surface point is calculated first. When a point cannot moved in this direction, the centrifugal direction of the point will be calculated.
4. Move each surface point along its inflation direction. Calculate its new position. During this iterative procedure, all possible intersections between segments(from points to their next positions) and surface elements are computed.
5. Check all adjacent triangles of each surface point. If some of them will not remain outward facing, reduce the movement step size and repeat 4 and 5 until the smallest step size. If two types of inflation directions have been tested but the point cannot be moved in the given range of step size, the point is fixed and will not be moved again.
6. Repeat steps 3–5 until no surface points can be moved again. The triangular mesh is stuck to the molecular surface of the cavity as close as possible.

During the inflation procedure, all triangles on the surface are kept outward facing. It is ensured that the resulting surface is a star-like surface, i.e. each ray from the origin intersects the surface at only one point. Another restriction is the side length of each triangle on the surface which is necessary for avoiding the divergence of the inflation in the regions exposed to the solution. Therefore, the approximation of the molecular surface of the cavity obtained by our

inflation method can be represented by spherical harmonic surfaces.

This inflation method to describe the surface of the cavity has an advantage that the surface is only a closed surface to approximate the shape of the binding cavity, not the whole surface of the receptor.

2.4. Computation of expansion coefficients

As the molecular surfaces obtained are topologically equivalent to a sphere, each point on the surface has a unique spherical coordinate (θ, ϕ) on a unit sphere. Therefore, the radial distance functions $r(\theta, \phi)$ can be expressed as an expansion of spherical harmonic functions using Eq. (5). The spherical harmonic expansion coefficients C_{lm} can be computed by evaluating the double level integrals as in Eq. (4). But to obtain a regular mesh θ and ϕ is computationally expensive.

Our integration procedure uses the solid angle approach described by Max and Getzoff [20]:

$$C_{lm} = \int_F r(\theta, \phi) S_l^m(\theta, \phi)\, \mathrm{d}\omega \tag{6}$$

where $\mathrm{d}\omega = \sin\theta\, \mathrm{d}\theta\, \mathrm{d}\phi$ is the element of surface area on F. In our study, it can be calculated using each triangle on the surface obtained by our deflation or inflation methods through projection. Therefore, the spherical harmonic expansion coefficients can be expressed as [20]:

$$C_{lm} = \int_F \frac{S_l^m(\theta, \phi)}{r(\theta, \phi)} U(\theta, \phi) N(\theta, \phi)\, \mathrm{d}A \tag{7}$$

where $\mathrm{d}A$ is the area of each triangle, $N(\theta, \phi)$ the normal direction to $\mathrm{d}A$, $U(\theta, \phi)$ is a unit vector from the origin to the center of $\mathrm{d}A$. Integration procedure is performed over all triangles on surface F.

Calculation of the associated Legendre functions at specific θ values is another important step in the integration procedure. For each l from 0 to L, the value of $P_l^l \cos\theta$ can be evaluated as [21]:

$$P_l^l = \frac{(2l)!}{(l)!} \left(\tfrac{1}{2} \sin\theta\right)^l \tag{8}$$

The value of $P_l^m \cos\theta$, where $0 < m < l$, can be evaluated using the backwards recursion relation as follows [21]:

$$(l + m + 1)(l - m) P_l^m \cos\theta = 2(m + 1) \cot\theta P_l^{m+1} \cos\theta - P_l^{m+2} \cos\theta \tag{9}$$

where $P_l^m \cos\theta = 0$, if $m > l$. With Eqs. (8) and (9), all values of $P_l^m \cos\theta$ at specific θ, for any $0 \le l \le L$, $0 \le m \le l$ can be computed.

The real spherical harmonic function values at the center of each triangle with spherical coordinate (θ, ϕ) can be evaluated by Eq. (2), and the expansion coefficients are then computed by evaluating the integrals in Eq. (7).

2.5. Recognition of lobes and holes

Previous studies showed that lobes and holes on two interactive surfaces are of obvious complementarity and these lobes and holes play very important role in molecular recognition. In order to alleviate the computationally expense of a global six-dimensional docking search, we first consider the lobes and the holes distributed on the surfaces of the ligand and the binding site cavity instead of the whole surfaces. To recognize lobes and holes on surfaces of ligands and cavities, respectively contouring techniques are used.

Based on spherical harmonic expansions contour lines of radius functions can be calculated. To recognize lobes or holes, useful contour lines are selected and grouped into a number of sets. Each set of contour lines represents a lobe or a hole. A function of two variables, obtained by translating and scaling Gaussian distributions is used as an example as in Fig. 2. In Fig. 2(b), five sets of contour lines are recognized corresponding to five peaks in Fig. 2(a).

For pair of complementary surfaces with lobes and holes, respectively, distributions of the lobes and the holes in

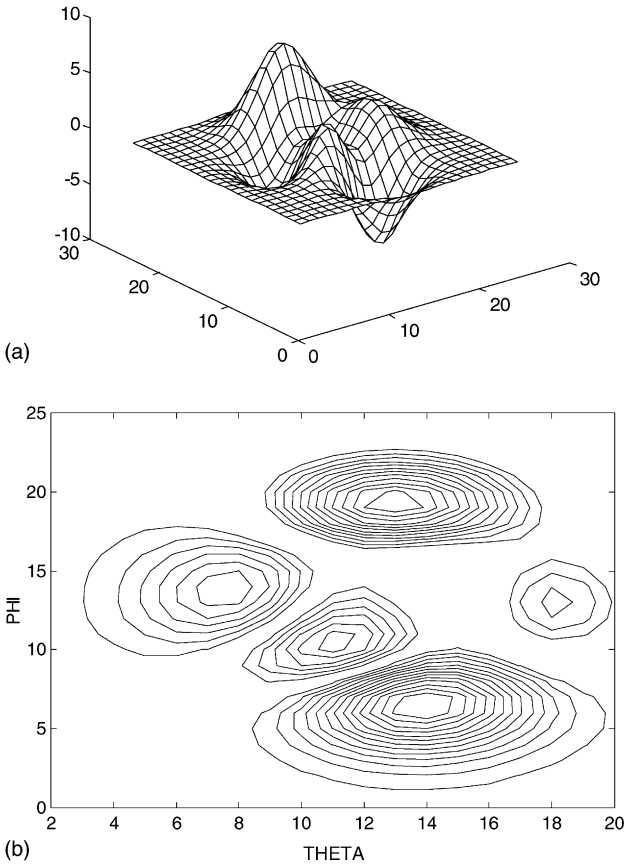


Fig. 2. Contour maps of the function of two variables, obtained by translating and scaling Gaussian distributions. (a) The mesh display of the function. (b) The contour map of the function, on which five peaks are recognized by five sets of contour lines.

complementary regions are the same. This can be seen clearly from their two-dimensional contour maps. Therefore, the complementarity between surfaces of a ligand and its cavity can be inspected by comparing their respective contour maps.

2.6. Molecular docking using expansion coefficients

For a known complex, the molecular surfaces of a ligand and the cavity in its binding site region can be modeled by our deflation and inflation technique, respectively. Then their spherical harmonic representations can be calculated. Rotations of molecular surfaces can be calculated by rotations of their spherical harmonic expansion coefficients. Therefore, shape comparison of two surfaces with the same origin can be implemented by fixing the surface of the cavity and rotating the expansion coefficients of the ligand and then minimizing the distance between their expansion coefficients.

In this study, the methods for the rotation and comparison of spherical harmonic surfaces described by Ritchie and Kemp [31] were used to superpose two interactive surfaces. They introduced the calculation of rotation matrix and shape difference function in detail in the paper. The only modification here is the shape difference function. Since the surface of the cavity is generally not a closed surface, the surface points in the open region, which are identified during the inflation procedure, are not taken into account. The root

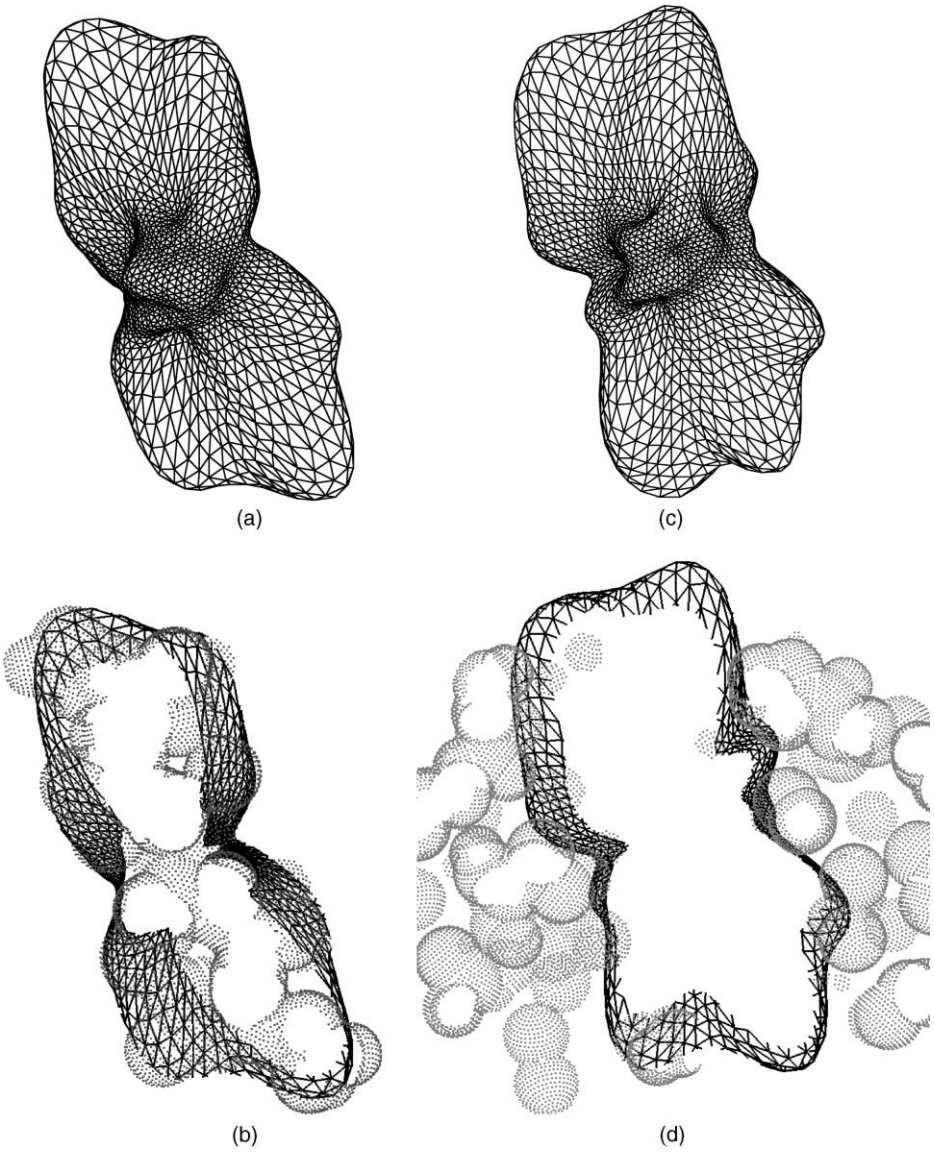


Fig. 3. (a) An order 10 spherical harmonic surface of the ligand in the complex 5hvp. (b) Section view of the spherical surface of the ligand and it is space filling atomic figure. (c) An order 10 spherical harmonic surface of the binding cavity of the ligand in 5hvp. (d) Section view of the spherical surface of the cavity and it is space filling atomic figure.

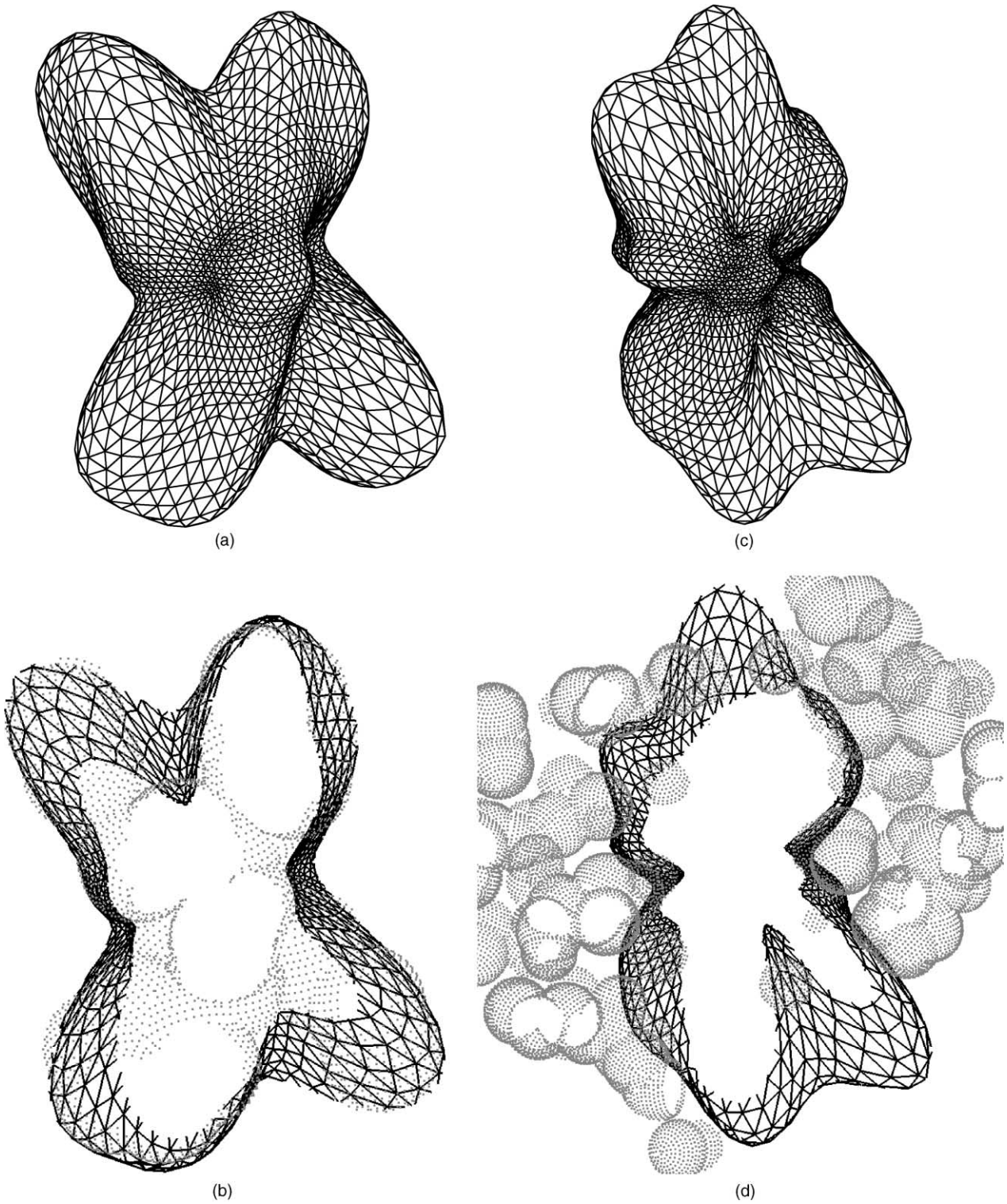


Fig. 4. (a) An order 10 spherical harmonic surface of the ligand in the complex 1ajx. (b) Section view of the spherical surface of the ligand and it is space filling atomic figure. (c) An order 10 spherical harmonic surface of the binding cavity of the ligand in 1ajx. (d) Section view of the spherical surface of the cavity and it is space filling atomic figure.

mean squared (RMS) distance difference between corresponding surface sample points is used as the shape difference function as follows:

$$D_{\mathrm{RMS}} = \left[\frac{1}{N} \sum_{n=1}^{N} (r_{\mathrm{cavity}}(\theta_n, \phi_n) - r'_{\mathrm{ligand}}(\theta_n, \phi_n))^2 \right]^{1/2} \quad (10)$$

where N is the number of surface points on cavity which are taken into account.

A GA is applied to minimize the function to give the rotation of a ligand relative to its receptor. In fact, the surface matching procedure is a docking procedure with three degrees of freedom. But accurate calculations of the complementarity between two spherical harmonic surfaces with different origins are still difficult, because it is a geometrical searching procedure which has six degrees of freedom (translation and rotation) and one surface must be re-modeled for each new origin.

3. Results

3.1. Test of the method

The capabilities of our methods are demonstrated by application to 35 protein–ligand complexes selected from PDB. The selection criteria was in a first stage, the diversity of ligands for a same or mutated protein. Because of our interest in this field, we have chosen molecules complexed with HIV protease (PDB codes 1A30, 1A9M, 1AJV, 1AJX, 1DIF, 1HBV, 1HIH, 1HIV, 1HPO, 1HPV, 1HTG, 1HXB, 1ODW, 1ODX, 2UPJ, 3UPJ, 5UPJ, 6UPJ, 7UPJ, 4HVP, 5HVP, 9HVP). Next, other complexes, outside the HIV problem, were retained to check the method on other systems. The selection criteria here was the quality of the structures and their diversity, in order to obtain a large panel of examples. The water molecules of these complexes have been removed. No water molecule was taken into account as we considered that the interactions, when found, with water, are

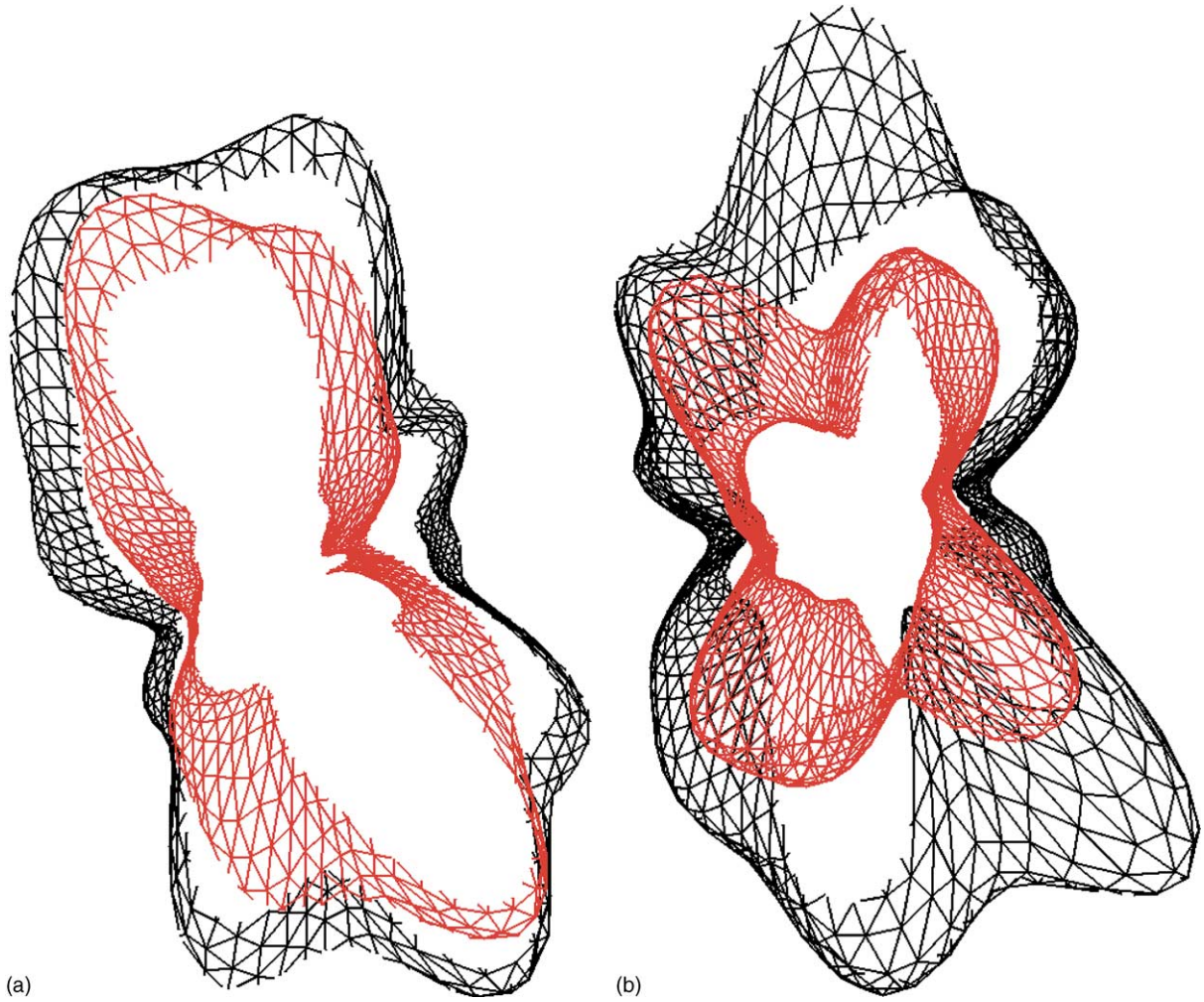


Fig. 5. Section view the complementarity between the molecular surface of the ligand and its binding cavity by using clipping techniques: (a) 5hvp; (b) 1ajx.

the result of a dynamic process which can be reproduced properly only by free energy calculations. Moreover, examination of several examples in which water molecules were found bridged between the ligand and the protein, showed that they are mostly outside the cavity in which the ligands are embedded. In the present paper, we were considering only the steric possibilities of matching between the partners surfaces. The method we propose is being extended to take into account other properties.

Our deflation and inflation techniques are used to approximate the molecular surfaces of ligands and cavities from their atomic coordinates, respectively, with which the corresponding spherical harmonic expansions are calculated. Figs. 3 and 4 show the spherical harmonic surfaces for ligands and their binding cavities of two complexes (5hvp, 1ajx), respectively. The surface shape complementarity between the ligands and their receptors are also shown in Fig. 5. Each one in Fig. 5 is the section view of the cavity

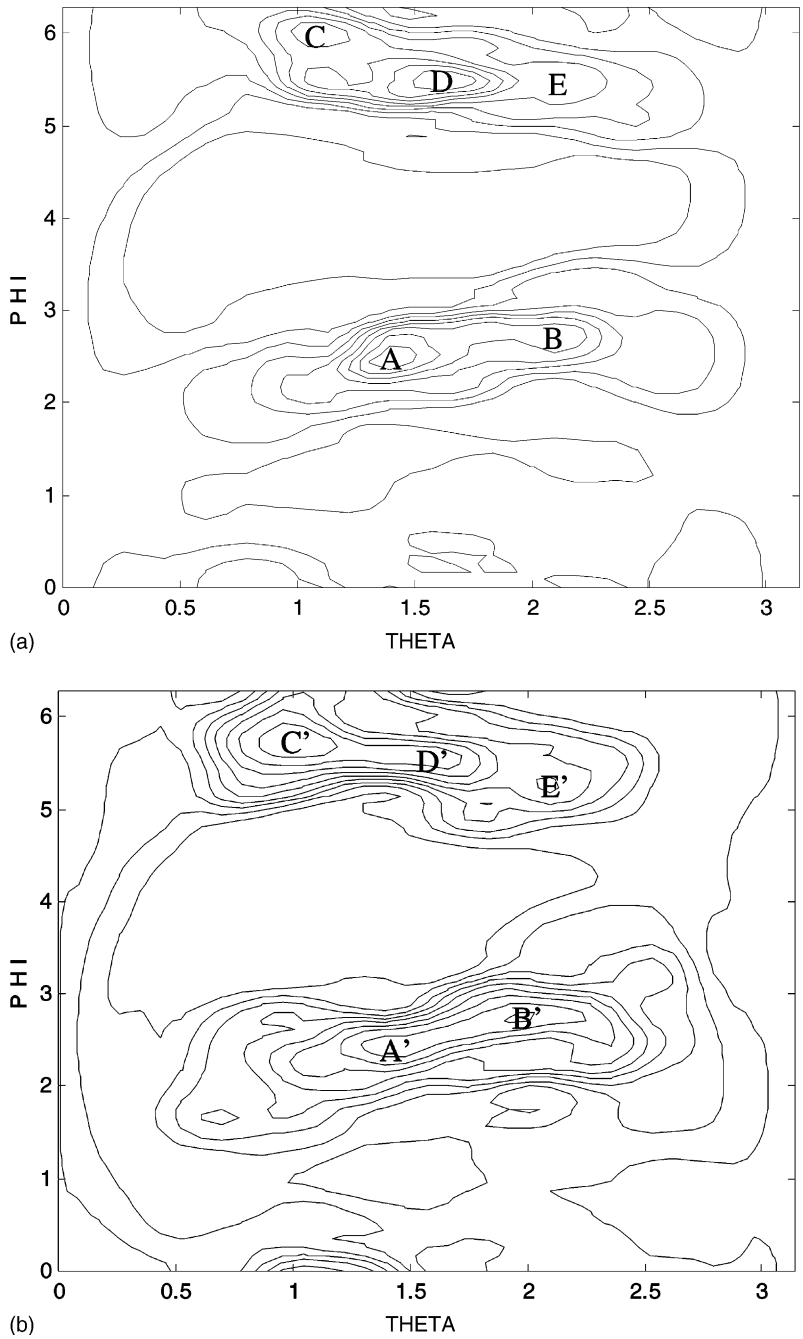


Fig. 6. Contour maps of the surfaces of the ligand and its binding cavity in complex 5hvp. The main complementary area locates in the lobes or holes. They are A–E in (a) corresponding to A′–E′ in (b), respectively. (a) The contour map of the ligand. (b) The contour map of the binding cavity.

including the surface of the ligand, which is offered by using clipping techniques. Steric relationship can be seen clearly and understood.

The contour maps of the ligands and the receptors are calculated from their radius spherical harmonic expansion functions. As examples, the results obtained from the 5hvp and 1ajx PDB files are given in Figs. 6 and 7, respectively from which the lobes or holes are recognized easily, and also the high level of complementarity by visual inspection between the surfaces of the ligands and their binding cavities is clear. It is the argument to dock a ligand in a cavity by matching lobes and holes on their respective surfaces.

The docking method to obtain the best superposition of surfaces of each ligand and its cavity is also tested using the 35 complexes from PDB which are listed in Table 1. The spherical harmonic expansion coefficients (of order 10) of each ligand were randomly rotated around its origin, the

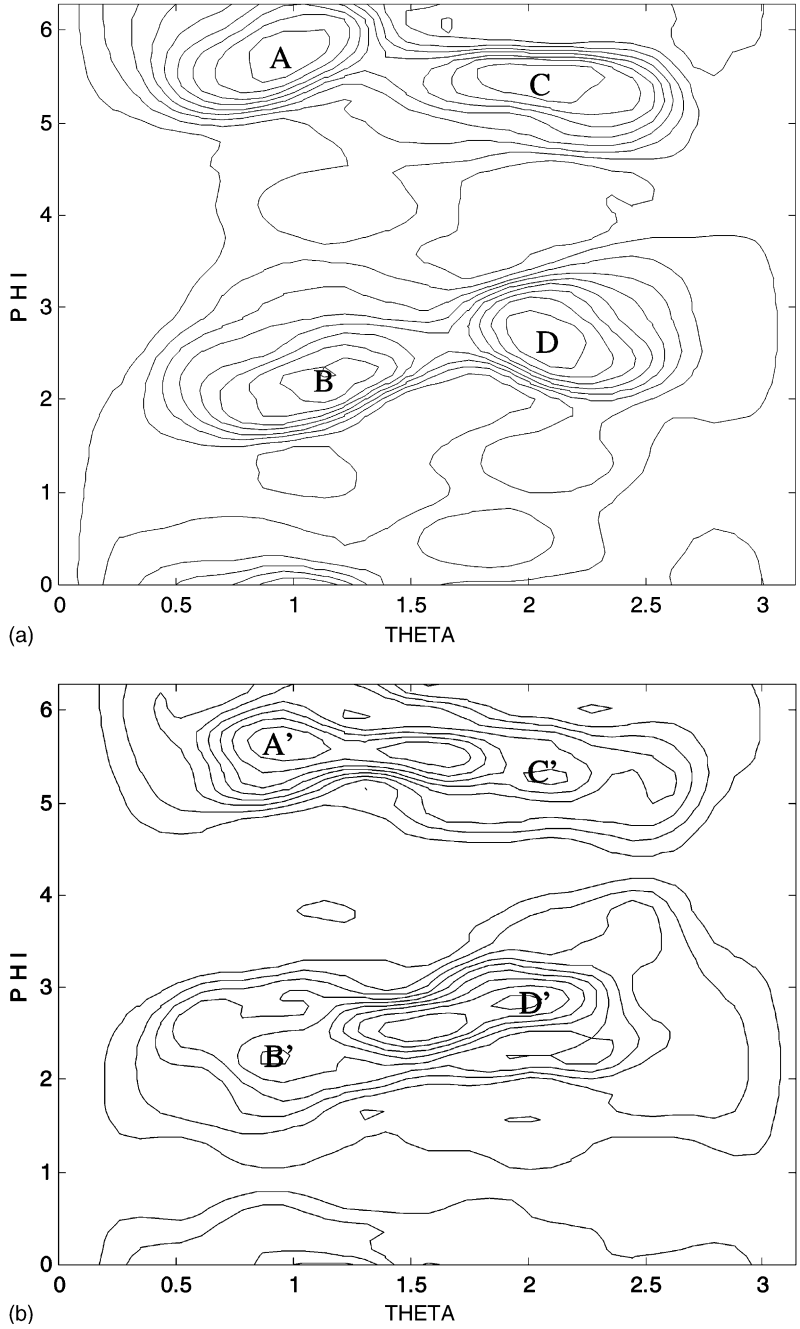


Fig. 7. Contour maps of the surfaces of the ligand and its binding cavity in complex 1ajx. The main complementary area locates in the lobes or holes. They are A–D in (a) corresponding to A′–D′ in (b), respectively. (a) The contour map of the ligand. (b) The contour map of the binding cavity.

Table 1
Comparison of docking results on ligand into protein with crystallized structure in complex

Number	Complex PDB	Atoms		RMS (Å)
		Receptor	Ligand	
1	1a30	1510	26	0.6390
2	1a9m	1528	42	0.6428
3	1ajv	1516	41	0.7245
4	1ajx	1516	40	0.5095
5	1dif	3144	116	0.2647
6	1erb	1401	24	0.8517
7	1fmp	2114	23	0.6333
8	1hbv	1520	42	0.8541
9	1hih	1520	41	0.2675
10	1hiv	1515	59	0.3105
11	1hpo	1491	36	0.4091
12	1hpv	1516	35	0.5703
13	1htg	1516	110	1.1111
14	1hxb	1514	98	0.6481
15	1lif	1017	20	0.5801
16	1odw	1514	39	0.4769
17	1odx	1516	39	0.9610
18	1ppm	2366	42	0.2531
19	1sgc	1259	45	1.0033
20	1stp	901	16	0.1716
21	1tlp	2432	37	0.9295
22	2upj	1516	41	0.5135
23	3apr	2403	57	1.1465
24	3upj	1492	23	0.2877
25	4dfr	2547	33	0.4775
26	4dfr	2547	33	0.0480
27	4hvp	1516	54	1.5927
28	4mbn	1217	44	0.7644
29	5hvp	1540	90	1.3136
30	5sga	1259	35	0.0811
31	5upj	1500	23	0.6045
32	6upj	1508	22	0.6817
33	7can	5815	13	0.5014
34	7upj	1516	34	0.6109
35	9hvp	1520	54	1.5842

distance functions for the grid points of 2562 were calculated using Eq. (10), and then the GA procedure was employed to optimize the Euler rotation angles by minimizing the functions. The RMS errors between the crystal coordinates of ligands and their rotated ones which giving the best complementarity at the binding site are listed in Table 1. The fact that the RMS <1.60 Å in Table 1 shows that the molecular surface shape complementarity is crucially important in protein–ligand interactions. The docked results for the complexes 5hvp and 1ajx obtained by the GA are also shown in Fig. 8(a) and (b), respectively. The complementarity in the complexes docked is in good agreement with the X-ray results.

3.2. Application to "virtual screening"

Most docking procedures using heuristic searching methods produce accurate predictions of binding modes but are too computationally expensive to be used for selecting the best candidates in a very large sample of molecular structures and/or conformations. As we have now large databases of compounds with a wide diversity of conformations, we aim to identify reliable and computationally efficient docking strategies for virtual high throughput screening of large combinatorial chemical libraries [32,33].

The method described above may be used for that purpose as a fingerprint searching procedure, giving a known fingerprint and looking in a database for possible similarities with a collection of other fingerprints. The surface map of the receptor-cavity can be used as the target fingerprint and the "fingerprint" map produced by each structure or conformation can be checked for similarities with the target.

We have applied this procedure to the search of the best candidate conformation of the SFLLRN peptide which is the minimum peptide sequence activating the thrombin platelet receptor. Thrombin receptor has been identified, cloned and

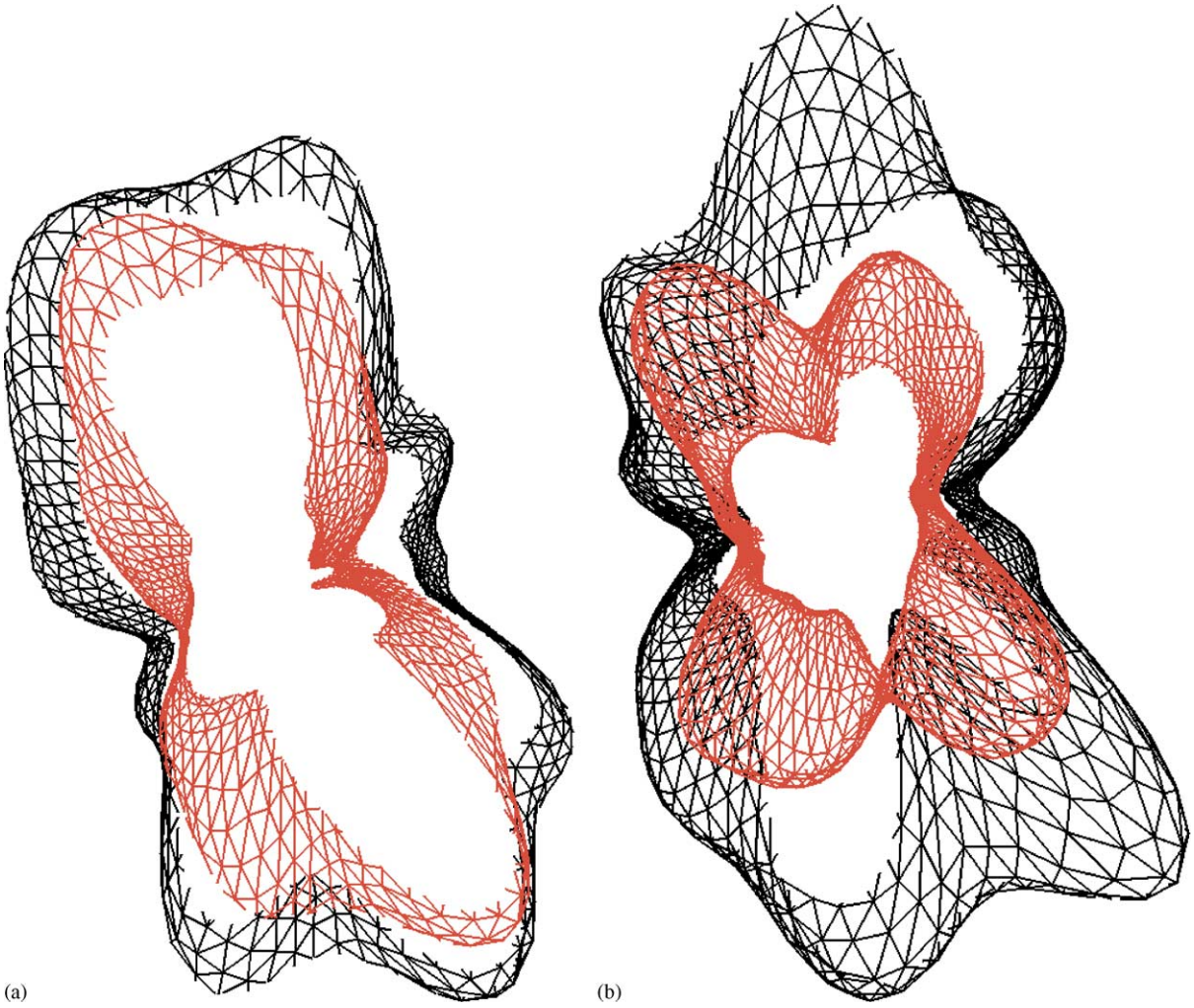


Fig. 8. Section view the complementarity between the molecular surface of the ligand docked and its binding cavity by using clipping techniques: (a) 5hvp; (b) 1ajx.

shown to be present on platelet, endothelial cell, fibroblasts and vascular smooth muscle cells. In contrast to the other GPCRs, thrombin receptor activation occurs through a special mechanism in which thrombin cleaves its receptor at the R41/S42 peptide bond within the LDPR/SFLLRN sequence. This event unmask the peptide ligand SFLLRN, which is silent in the un-cleaved receptor. The SFLLRN sequence then binds to the receptor functioning as an intramolecular peptide ligand to effect receptor activation. Free SFLLRN mimics the tethered ligand to activation the receptor [34].

We have modeled the thrombin receptor and investigated the conformational possibilities of the SFLLRN peptide by intensive MD simulated annealing, producing, therefore, about 1000 minimum energy conformations suitable for docking inside the receptor groove.

The following procedure was used to sample the peptide conformational space using the InsightII Accelrys softwares [35].

A starting conformation of SFLLRN peptide was first built using the InsightII fragment library in an extended conformation and refined by 10,000 steps of energy minimization using a conjugate gradient algorithm. The CVFF force field was used in the Discover molecular mechanics and dynamics program. This starting conformation was then used for a pseudo-simulated annealing conformational sampling. Simulated annealing involves a temperature increase of the system, followed by a slow cooling to avoid local minima, thereby trying to locate the global minimum region of the energy function. This conformational sampling was performed using molecular dynamics (MD) in a vacuum (over 40,000 steps with a time step of 1 fs). The equations of motion were integrated using the Leapfrog version of the Verlet algorithm 21. The dielectric constant used was distance-dependent ($\varepsilon = 1r$) to simulate roughly the electrostatic shielding due to the solvent. The (N, V, T) ensemble was used at a fixed value, with the Berendsen algorithm. This method allows to maintain temperature, at a fixed value

by means of a coupling to an external bath. The simulated annealing-like method used here consists of one thousand loops of slow cooling, each one leading to a low energy conformation. Each loop begins by fixing the temperature to 1000 K, followed by 5000 steps of MD. The temperature was then decreased by steps of 100 K. Decreasing the temperature by 100 K every 5000 steps, so that after 40,000 steps, the temperature of the system corresponds approximately to 300 K. The final conformation obtained at the end of this process was energy-refined using again conjugate gradient algorithm, and, after storage, was used to start a new simulation at high temperature with a slow cooling stage, as described above. This procedure produced 1000 minimized conformations for SFLLRN.

The thrombin receptor was modeled according to a procedure already described [36].

The filtering of the SFLLRN candidate conformations by usual screening methods is tedious while the surface map fingerprint-like comparison produced fast a few "good" candidates to be used next for more elaborate docking procedures. We illustrated on Fig. 9 the obtained results between the thrombin GPCRs cavity surface and the best SFLLRN molecular surface matching. Fig. 9(a) shows the spherical harmonic surfaces of two different conformations of the SFLLRN peptide produced by MD simulated annealing, and the associated fingerprints. Fig. 9(b) is for the cavity in thrombin receptor. By comparing the contour maps of two candidate conformations with the cavity's, the complementarity between the lower in Fig. 9(a) and the cavity can be found easily, which is shown in Fig. 9(c), and the upper in Fig. 9(a) is filtered out.

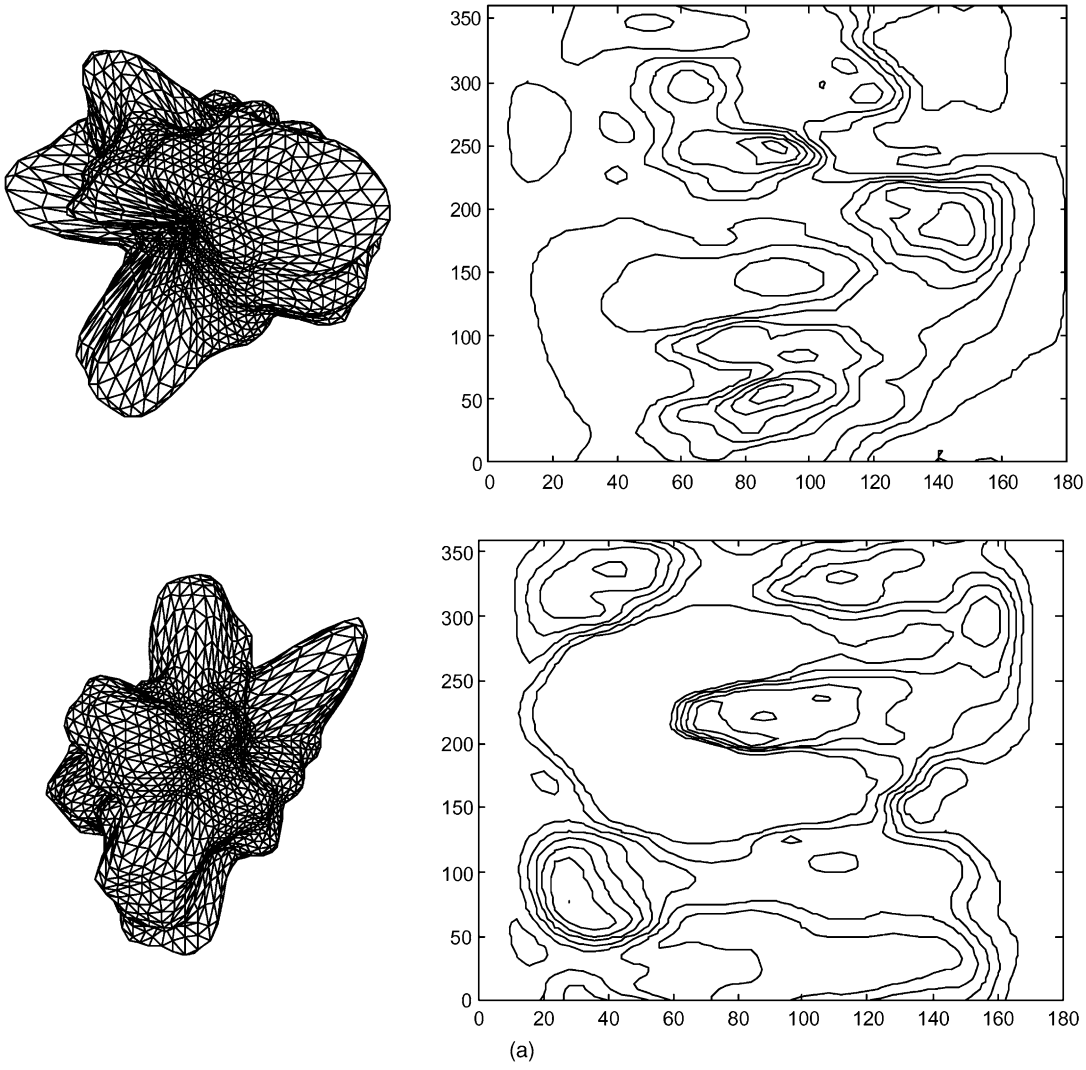
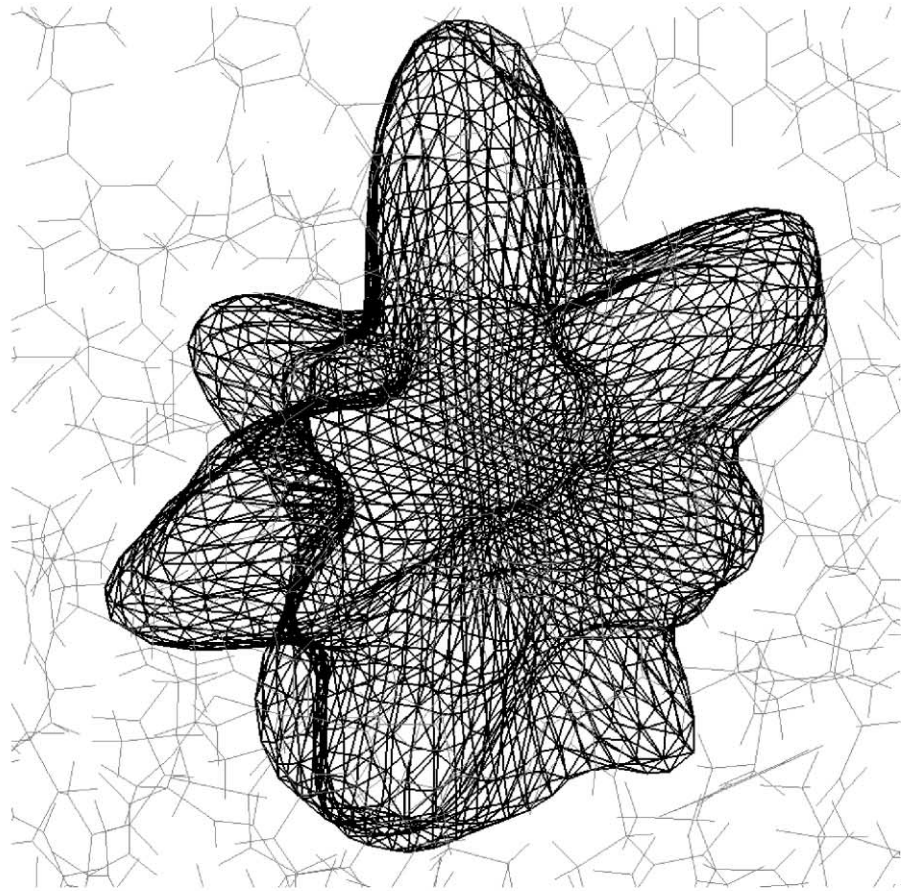
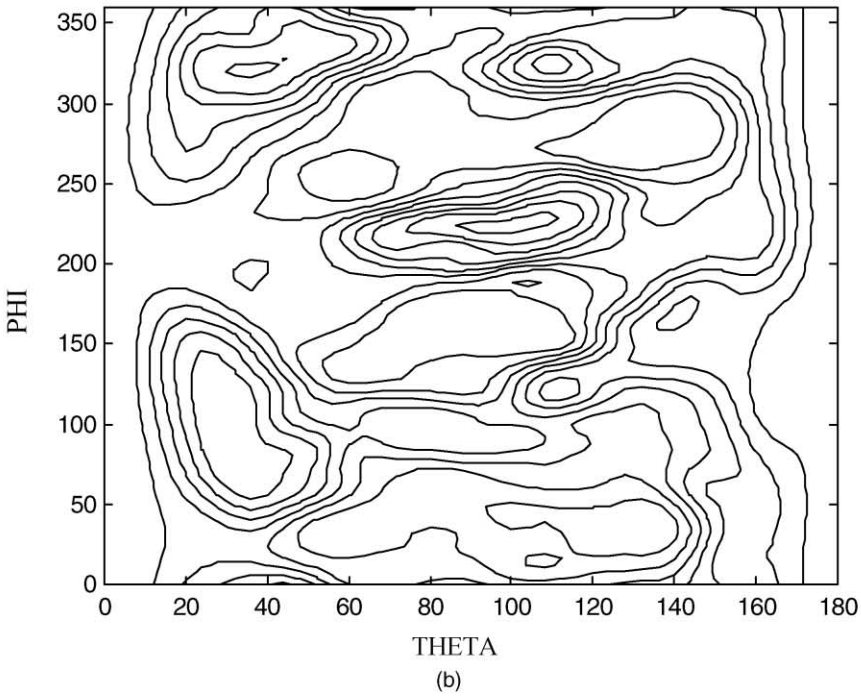


Fig. 9. (a) Order 20 spherical harmonic surfaces of two different conformations of the SFLLRN peptide and the associated fingerprints. (b) An order 20 spherical harmonic surface of binding site cavity in platelet thrombin receptor and the associated fingerprint. (c) Section view of the complementarity between the surface of the binding site cavity in thrombin receptor and the surface of the ligand (red is for the ligand, and black is for cavity).





(b)

Fig. 9. (*Continued*).

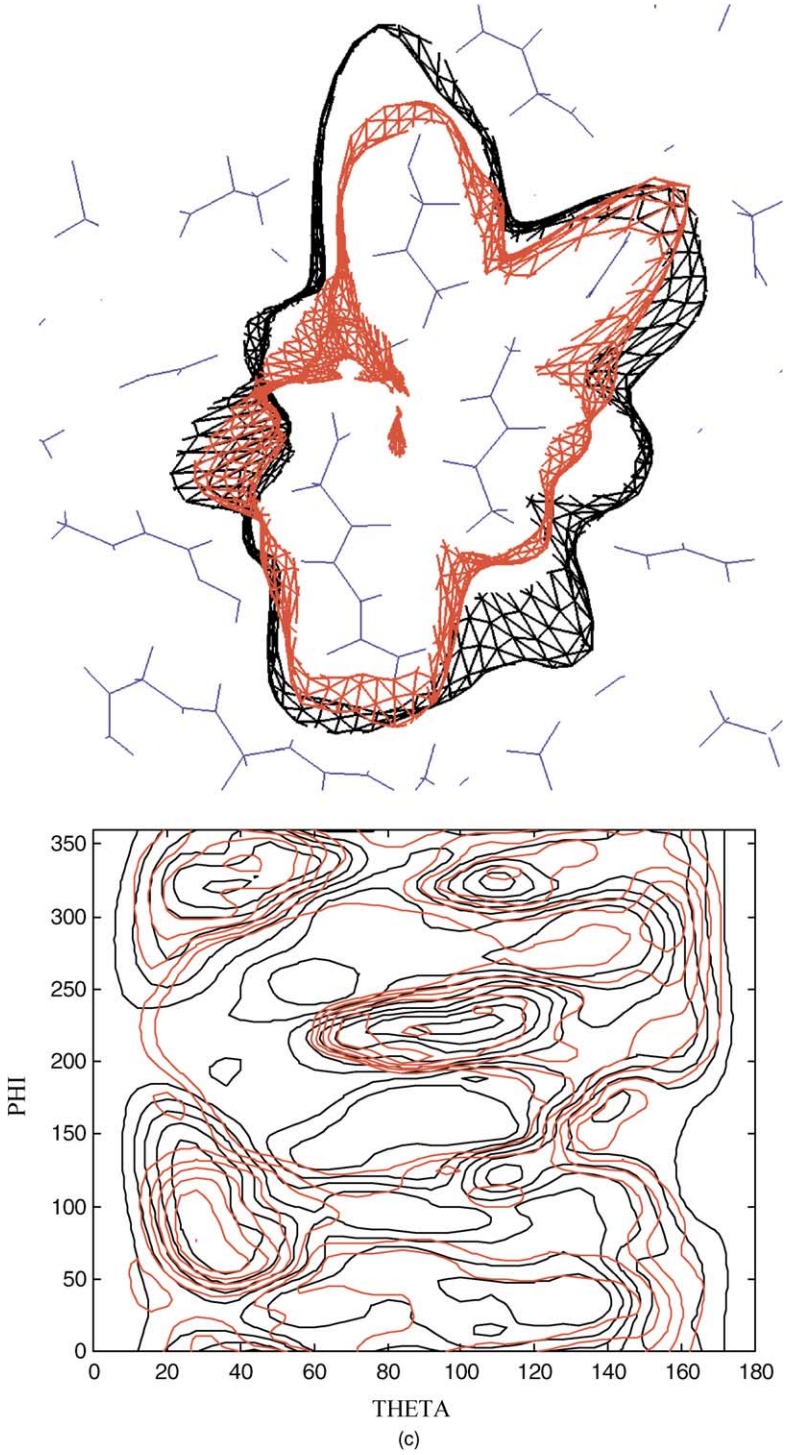


Fig. 9. (*Continued*).

4. Discussion

The previous methods started with molecular surfaces computed by well-known Connolly's program, and mapped the surface even with overlapping regions to a sphere using expansion techniques. In a different way, our method to calculate the spherical harmonic molecular surfaces starts from the atomic coordinates of molecules, calculates the embracing surfaces of molecules using the deflation technique in our work. Since the aim of deflation is to obtain a star-like surface, therefore, it cannot be used to describe the surface of deep and narrow cavity. The inflation technique was proposed to solve this problem. The advantage of the technique is that it can focus on the cavity which is interesting and the

surface of the cavity approximated is a closed star-like surface (in fact the real cavity is probably open in the entrance) which allows us to manipulate it as the embracing surface of the molecule. The limitation of the inflation technique is that the position where the unit sphere is placed should be given. Our program cannot estimate this position automatically.

The fineness of the triangulation that used to generate the surface is that the numerical integration in computation of the spherical harmonic coefficients can directly use the triangle elements and their surface normals for each surface point. The degree of the spherical harmonic approximation of the surface is determined by the size of the grid as well as the order of the expansion L. Small grid results high accurate, but much computation time is needed. Initial triangle mesh is a regular icosahedron with 20 triangles, and higher resolution meshes are obtained by dividing each triangle into four smaller triangles. But very small grid will probably cause the deflation or inflation procedure to be unstable. All the spherical harmonic surfaces involved in this study are calculated based on 5120 triangles, and represented using the same resolution. The resulted surfaces are considerably accurate.

Comparing pairs of surfaces by rotating spherical harmonic expansion coefficients is useful and efficient. In our test examples, the surface does not change much for L values greater than 10. Therefore, order 10 spherical harmonic surfaces are used in our docking procedure. For only a few number of coefficients (121) are involved, rotating coefficients is fast. In fact, considering the cost of computation, 2562 grid points (with 5120 triangles) sampling on the surface are not necessary, and 642 vertices (with 1280 triangles) are sufficient. Note that the above method is a three-dimensional docking, i.e. the binding site hypothesized must be given. It is computationally expensive to do the global six-dimensional search of the complementarity using this method.

The recognition of small molecules by their protein receptors has been proved to be driven by surface complementarity [37]. But often, assessing the surface complementarity by searching all relative position of two surfaces is computational expansive. The complementarity of lobe–hole is very important in protein–ligand interactions. The contouring method is a new way to identify the complementarity between lobes and holes. The advantage of this method is that two spherical harmonic surfaces to be compared can be defined separately. The important advantage is that it can be used as a filter to eliminate candidates among a large number of conformations which will speed up the docking procedure. Our aim is to develop a virtual screening process using the above method, which is carried out in four steps. First, the surfaces of the binding cavity and candidate ligand are represented by expansions of spherical harmonic functions. Second, the lobes and holes on the surfaces are recognized based on the radius spherical harmonic expansion functions. The complementarity between lobes and holes is then searched. Finally, a score is calculated for the ligand in that orientation.

Completely comparing two contour maps is a complicated procedure like fingerprint identification. An automatic docking algorithm using the contour map is under development.

5. Conclusion

In this paper, we present a method to approximate the molecular surfaces of a cavity by inflating a triangular mesh on an unit sphere placed in the cavity which can be used to calculate model surfaces based on spherical harmonic expansion. Such spherical harmonic expansion coefficients can be used to compare two surfaces, especially between a ligand and its receptor. Lobes and holes on molecular surfaces can be recognized from the contour maps calculated by their spherical harmonic representations. This fingerprint recognition method based only on molecular surface complementarity can be easily extended for efficient and fast discrimination in large data sets of chemical structures or molecular conformations. Work is in progress to extend the fingerprint map concept, including additional information to evaluate the shape/electrostatic complementarities in our docking program.

Acknowledgements

This work is supported by Franco-Chinese PRA project (Biology B9606).

References

[1] J. Janin, Protein–protein recognition, Prog. Biophys. Mol. Biol. 64 (1995) 145–166.
[2] M.J.E. Sternberg, H.A. Gabb, R.M. Jackson, Predictive docking of protein–protein and protein–DNA complexes, Curr. Opin. Struct. Biol. 8 (1998) 250–256.
[3] S. Vajda, M. Sippl, J. Novotny, Empirical potentials and functions for protein folding and binding, Curr. Opin. Struct. Biol. 7 (1997) 222–228.
[4] I. Muegge, Y.C. Martin, A general and fast scoring function for protein–ligand interactions: a simplified potential approach, J. Med. Chem. 42 (1999) 791–804.
[5] L. Le Conte, C. Chothia, J. Janin, The atomic structure of protein–protein recognition sites, J. Mol. Biol. 285 (1999) 2177–2198.
[6] M.C. Lawrence, P.M. Colman, Shape complementarity at protein–protein interfaces, J. Mol. Biol. 234 (1993) 946–950.
[7] F. Jiang, S.-H. Kim, Soft docking: matching of molecular surface cubes, J. Mol. Biol. 219 (1991) 79–102.
[8] M. Meyer, P. Wilson, D. Schomburg, Hydrogen bonding and molecular surface shape complementarity as a basis for protein docking, J. Mol. Biol. 264 (1996) 199–210.
[9] V. Sobolev, R.C. Wade, G. Vriend, M. Edelman, Molecular docking using surface complementarity, Proteins Struct. Func. Genet. 25 (1996) 120–129.
[10] I.D. Kuntz, J.M. Blaney, S.J. Oatley, R. Langridge, T.E. Ferrin, A geometric approach to macromolecule–ligand interactions, J. Mol. Biol. 161 (1982) 269–288.

[11] E. Katchalski-Katzir, I. Shariv, M. Eisenstein, A.A. Friesem, C. Aflalo, I.A. Vakser, Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques, Proc. Natl. Acad. Sci. U.S.A. 89 (1992) 2195–2199.
[12] M. Helmer-Citterich, A. Tramontano, Puzzle: a new method for automated protein docking based on surface shape complementarity, J. Mol. Biol. 235 (1994) 1021–1031.
[13] R. Norel, S.L. Lin, H.J. Wolfson, R. Nussinov, Molecular surface complementarity at protein–protein interfaces: the critical role played by surface normals at well placed, sparse, points in docking, J. Mol. Biol. 252 (1995) 263–273.
[14] M. Eisenstein, S.G. Koren, A.A. Friesem, E. Katchalski-katzir, Modeling supra-molecular helices: extension of the molecular surface recognition algorithm and application to the protein coat of the tobacco mosaic virus, J. Mol. Biol. 266 (1997) 135–143.
[15] M.L. Connolly, Solvent-accessible surfaces of proteins and nucleic acids, Science 221 (1983) 709–713.
[16] M.L. Connolly, Analytical molecular surface calculation, J. Appl. Crystals 16 (1983) 548–558.
[17] M.L. Connolly, Molecular surface triangulation, J. Appl. Crystals 18 (1985) 499–505.
[18] M.F. Sanner, A.J. Olson, J.-C. Spehner, Reduced surfaces: an efficient way to compute molecular surfaces, Biopolymers 38 (1996) 305–320.
[19] S. Leicester, J.L. Finney, R.P. Bywater, Description of molecular surface shape using Fourier descriptors, J. Mol. Graph. 6 (1988) 104–108.
[20] N.L. Max, E.D. Getzoff, Spherical harmonic molecular surfaces, IEEE Comput. Graph. Appl. 8 (1988) 42–50.
[21] B.S. Duncan, A.J. Olson, Approximation and characterization of molecular surfaces, Biopolymers 33 (1993) 219–229.
[22] S. Leicester, J.L. Finney, R.P. Bywater, A quantitative representation of molecular surface shape: theory and development of the method, J. Math. Chem. 16 (1994) 315–341.
[23] S. Leicester, J.L. Finney, R.P. Bywater, A quantitative representation of molecular surface shape. II. Classification of protein shapes, J. Math. Chem. 16 (1994) 342–365.
[24] B.S. Duncan, A.J. Olson, Approximation and visualization of large-scale motion of protein surfaces, J. Mol. Graph. 13 (1995) 250–257.
[25] D.W. Ritchie, Spherical Harmonics, http://www.csd.abdn.ac.uk/~dritchie/graphics.html.
[26] W. Cai, M. Zhang, B. Maigret, New approach for representation of molecular surface, J. Comput. Chem. 19 (1998) 1805–1815.
[27] C.A.D. Carpio, Y. Takahashi, S.-I. Sasaki, Anew approach to the automatic identification of candidates for ligand receptor sites in proteins. I. Search for pocket regions, J. Mol. Graph. 11 (1993) 23–29.
[28] C.M.W. Ho, R. Marshall, Cavity search: an algorithm for the isolation and display of cavity-like binding regions, J. Comput. Aided Mol. Des. 4 (1990) 337–354.
[29] J. Ruppert, W. Welch, A.N. Jain, Automatic identification and representation of protein binding sites for molecular docking, Protein Sci. 6 (1997) 524–533.
[30] B.S. Duncan, A.J. Olson, Shape analysis of molecular surfaces, Biopolymers 33 (1993) 231–238.
[31] D.W. Ritchie, G.J.L. Kemp, Fast computation, rotation, and comparison of low resolution spherical harmonic molecular surfaces, J. Comput. Chem. 20 (1999) 383–395.
[32] J.W. Godden, F. Stahura, J. Bajorath, Evaluation of docking strategies for virtual screening of coumpound databases: cAMP-dependent serine/threonine kinase as an example, J. Mol. Graph. Model. 16 (1999) 139–143.
[33] C.A. Baxter, C.W. Murray, B. Waszkowycz, J. Li, R.A. Sykes, R.G.A. Bone, T.D.J. Perkins, W. Wylie, New approach to molecular docking and its application to virtual screening of chemical databases, J. Chem. Inf. Comput. Sci. 40 (2000) 254–262.
[34] S.R. Coughlin, Thrombin signaling and protease-activated receptors, Nature 407 (2000) 258–264.
[35] © 2001 Accelrys Inc., Accelrys is a wholly owned subsidiary of Pharmacopeia Inc., http://www.accelrys.com.
[36] J. Marie, E. Richard, D. Pruneau, J.L. Paquet, C. Siatka, R. Larguier, C. Ponce, P. Vassault, T. Groblewski, B. Maigret, J.C. Bonnafous, Control of conformational equilibria in the human B2 bradykinin receptor: modeling of non-peptidic ligand action and comparison to the rhodopsin structure, J. Biol. Chem. 2001, in press.
[37] R. Norel, H.J. Wolfson, R. Nussinov, Small molecule recognition: solid angles surface representation and molecular shape complementarity, Comb. Chem. High Throughput Screen 2 (1999) 223–237.