

# Molecular Simulations using Spherical Harmonics

CAI, Wen-Sheng\* (蔡文生) XU, Jia-Wei(徐佳维) SHAO, Xue-Guang(邵学广)  
MAIGRET, Bernard<sup>b</sup>

<sup>a</sup> Department of Chemistry, University of Science and Technology of China, Hefei, Anhui 230026, China

<sup>b</sup> UMR 7565 CNRS/UHP, Henri Poincaré University, B. P. 239 54506 Vandoeuvre-Les-Nancy Cedex, France

Computer-aided drug design is to develop a chemical that binds to a target macromolecule known to play a key role in a disease state. In recognition of ligands by their protein receptors, molecular surfaces are often used because they represent the interacting part of molecules and they should reflex the complementarity between ligand and receptor. However, assessing the surface complementarity by searching all relative position of two surfaces is often computationally expensive. The complementarity of lobe-hole is very important in protein-ligand interactions. Spherical harmonic models based on expansions of spherical harmonic functions were used as a fingerprint to approximate the binding cavity and the ligand, respectively. This defines 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. This method can be used as a filter to eliminate candidates among a large number of conformations, and it will speed up the docking procedure. Therefore, it is possible to select complementary ligands or complementary conformations of a ligand and the macromolecules, by comparing their fingerprints previously stored in a database.

**Keywords** molecular surface, shape complementarity, molecular docking, molecular recognition, spherical harmonics

## Introduction

Recognition of protein-protein, protein-ligand or substrate-enzyme is important in understanding these interactions, especially in the field of computer-aided drug design. The recognition of small molecules by proteins is largely mediated by molecular surface complementarity. Therefore, molecular surfaces are often used because they represent the interacting part of molecules and have also been shown to exhibit very high level of complementarity both for a geometric point of view and for chemical properties.<sup>1-10</sup>

Several methods have been developed for that purpose.<sup>11-14</sup> Spherical harmonic models based on expansions of spherical harmonic functions have also been used for quantitative description of molecular shapes.<sup>15-21</sup> The models offer several advantages which include being analytical models, controlling over level of approximation, allowing representation of not only geometry but also chemical prop-

erties and quantitative measure of surface similarity.

We have presented a deflating approach that starts with the triangular mesh distributed on an ellipsoid or a sphere embracing the molecule.<sup>22</sup> This ellipsoid is deflated step by step to approximate the molecular surface. Based on the resulting surface its spherical harmonic representation can be calculated directly. But the deflation method can not provide a good model for deep and narrow cavities where most enzyme-substrate 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.<sup>23-25</sup> A variation of the deflation technique allowing to characterize molecular cavities with more accuracy is also presented. This is achieved by inverting the deflating process and we inflate a sphere placed initially inside the cavity until it approximates the cavity's shape.

By using the analytical model of spherical harmonics, the critical points on the surface which indicate the holes and the lobes can be determined mathematically. The holes and lobes distributed on the surfaces are often involved in the protein-ligand interactions. The method we present here is also capable of identifying these regions of interactions from contour maps computed by using spherical harmonic surfaces. In the next step, the critical points and contours are used as the fingerprint for each 3D structure, and the complementarity of the ligand and the protein can be quantified by comparing fingerprints. This technique is useful in computational screening or virtual screening of very large combinatorial libraries as a fingerprint recognition procedure.

## Methods

### *Creation of the molecular surface of a ligand*

We have described a deflation algorithm to approximate molecular surfaces, which can be described as follows:<sup>22</sup>

(1) create a uniform mesh on an ellipsoid embracing the molecule;

\* E-mail: wscail@ustc.edu.cn; Tel.: 0551-3606160; Fax: 0551-3631760

Received April 4, 2003; revised and accepted July 9, 2003.

(2) calculate the elements of the molecular surface from its atomic coordinates, including convex pieces, spherical triangles and tori;

(3) deflate the mesh step by step, and calculate all possible intersections between segments (from points to their next positions) and surface elements, keeping the embracing surface in each step is starlike;

(4) repeat the deflation procedure until the triangular mesh is stuck to the molecular surface as close as possible.

Thus such a mesh could be "deflated" step by step to obtain an approximation of a molecular surface as defined by Connolly.<sup>11-13</sup> During the deflation procedure, all triangles on the surface are kept outward facing. It is ensured that the resulting surface is a starlike surface, *i. e.* each ray from the origin intersects the surface only at one point. Therefore, the approximation of the molecular surface can be represented by 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 can not reach the molecular surfaces of the cavities.

#### Creation of the surface of a binding cavity

In order to model a cavity we adapted our deflation algorithm to inflate a sphere inside the cavity. 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. In our study, the center of the ellipsoid embracing the binding ligand is used as the center of the sphere in the cavity.

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.

#### Spherical harmonic representation of molecular surface

Spherical harmonics are single-valued, continuous bounded, complex functions of the spherical coordinates  $(\theta, \phi)$ , 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<sup>19-21</sup>:

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

where,  $l$  and  $m$  are integers,  $m = -l, -l+1, \dots, 0, \dots, l$  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:

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

where  $m > 0$ .

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

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

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

$$C_{lm} = [f, S_l^m] = \int_0^\pi \int_0^{2\pi} f(\theta, \phi) S_l^m \sin \theta d\theta d\phi \quad (4)$$

A starlike 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).

#### 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  $(\theta, \phi)$  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 coeffi-

icients  $C_{lm}$  can be computed by evaluating the double level integrals as in Eq. (4). But to obtain a regular mesh  $\theta$  and  $\phi$  is computationally expensive.

Our integration procedure uses the solid angle approach described by Max and Getzoff,<sup>20</sup>

$$C_{lm} = \int r(\theta, \phi) S_l^m(\theta, \phi) d\omega \quad (6)$$

in which,  $d\omega = \sin \theta d\theta d\phi$  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:<sup>20</sup>

$$C_{lm} = \int_F \frac{S_l^m(\theta, \phi)}{r(\theta, \phi)} U(\theta, \phi) \cdot N(\theta, \phi) dA \quad (7)$$

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

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

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

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

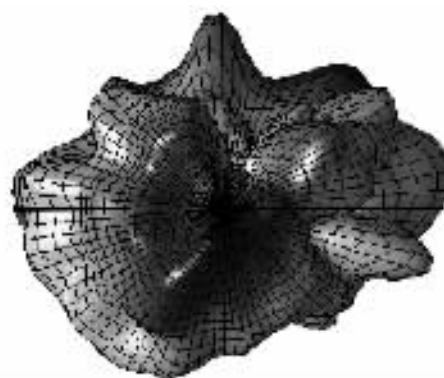
$$(l+m+1) \{ l-m \} P_l^m \cos \theta = \{ m+1 \} \cot \theta P_l^{m+1} \cos \theta - P_l^{m+2} \cos \theta \quad (9)$$

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  $\theta$ , for any  $0 \leq l \leq L$ ,  $0 \leq m \leq l$  can be computed.

The real spherical harmonic function values at the center of each triangle with spherical coordinate  $(\theta, \phi)$  can be evaluated by Eq. (2), and the expansion coefficients are then computed by evaluating the integrals in Eq. (7). As an example, Fig. 1 shows a spherical harmonic surface (order = 20) of trypsin in 3TPI complex calculated using the above method.

#### Identification of critical points and 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 roles in molecular recognition. To quickly determine the complementary region we first consider the critical points on the surfaces of the ligand and the binding site cavity, each critical



**Fig. 1** A spherical harmonic surface (order = 20) of trypsin in 3TPI complex calculated using the presented method.

point indicates a hole or a lobe distributed on the surface. These critical points together with the lobes and the holes can be used as criteria in molecular docking and recognition, instead of the whole surfaces. The critical points have the local maximal or minimal  $r$  values, and can be identified using the analytical model of spherical harmonic functions in Eq. (5). To recognize lobes and holes on surfaces of ligands and cavities respectively, contouring techniques are used.<sup>26</sup>

## Results and discussion

### Molecular docking using critical points

The recognition of small molecules by their protein receptors has been proved to be driven by surface complementarity. However, assessing the surface complementarity by searching all relative position of two surfaces is often computationally expensive. The complementarity of lobe-hole is very important in protein-ligand interactions. In order to develop a fast and also a six degrees of freedom (translation and rotation) docking algorithm, only the critical points are used. Therefore, shape comparison of two surfaces can be implemented by fixing the critical points on the surface of the binding cavity and rotating and translating the critical points on the surface of the free ligand and then minimizing the distance between the corresponding points.

### 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 now we have large databases of compounds with a wide diversity of conformations, our aim is to identify reliable and computationally efficient docking strategies for virtual high-throughput screening of large combinatorial chemical libraries.

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.

## Conclusion

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, with which the critical points and the contour lines on the surface can be obtained. These can be used to identify the complementarity between lobes and holes as fingerprints. This fingerprint recognition method based only on molecular surface complementarity can be easily extended to efficient and fast discrimination for large data sets of chemical structures or molecular conformations. An automatic docking algorithm using the fingerprint is under development.

One of the advantages of this method is that two spherical harmonic surfaces to be compared can be defined separately. Another attractive advantage is that it can be used as a filter to efficiently screen huge amount of molecular structures against large protein targets collections, which will speed up the docking procedure. Our aim is to develop a virtual screening process using the above method to predict which compounds will have the highest affinity for the target. It is important for lead generation and subsequent testing in drug discovery.

## References

- 1 Le Conte, L.; Chothia, C.; Janin, J. *J. Mol. Biol.* **1999**, *285*, 2177.
- 2 Lawrence, M. C.; Colman, P. M. *J. Mol. Biol.* **1993**, *234*, 946.
- 3 Jiang, F.; Kim, S. H. *J. Mol. Biol.* **1991**, *219*, 79.
- 4 Meyer, M.; Wilson, P.; Schomburg, D. *J. Mol. Biol.* **1996**, *264*, 199.
- 5 Sobolev, V.; Wade, R. C.; Vriend, G.; Edelman, M. *Proteins: Struct., Funct., Genet.* **1996**, *25*, 120.
- 6 Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. *J. Mol. Biol.* **1982**, *161*, 269.
- 7 Katchalski-Katzir, E.; Shariv, I.; Eisenstein, M.; Friesem, A. A.; Aflalo, C.; Vakser, I. A. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89*, 2195.
- 8 Helmer-Citterich, M.; Tramontano, A. *J. Mol. Biol.* **1994**, *235*, 1021.
- 9 Norel, R.; Lin, S. L.; Wolfson, H. J.; Nussinov, R. *J. Mol. Biol.* **1995**, *252*, 263.
- 10 Eisenstein, M.; Koren, S. G.; Friesem, A. A.; Katchalski-katzir, E. *J. Mol. Biol.* **1997**, *266*, 135.
- 11 Connolly, M. L. *Science* **1983**, *221*, 709.
- 12 Connolly, M. L. *J. Appl. Crystallog.* **1983**, *16*, 548.
- 13 Connolly, M. L. *J. Appl. Crystallog.* **1985**, *18*, 499.
- 14 Sanner, M. F.; Olson, A. J.; Spehner, J. C. *Biopolymers* **1996**, *38*, 305.
- 15 Leicester, S.; Finney, J. L.; Bywater, R. P. *J. Mol. Graphics.* **1988**, *6*, 104.
- 16 Max, N. L.; Getzoff, E. D. *IEEE Comput. Graphics Appl.* **1988**, *8*, 42.
- 17 Duncan, B. S.; Olson, A. J. *Biopolymers* **1993**, *33*, 219.
- 18 Leicester, S.; Finney, J. L.; Bywater, R. P. *J. Math. Chem.* **1994**, *16*, 315.
- 19 Leicester, S.; Finney, J. L.; Bywater, R. P. *J. Math. Chem.* **1994**, *16*, 342.
- 20 Duncan, B. S.; Olson, A. J. *J. Mol. Graphics* **1995**, *13*, 250.
- 21 Ritchie, D. W. <http://www.csd.abdn.ac.uk/dritchie/graphics.html>.
- 22 Cai, W.; Zhang, M.; Maigret, B. *J. Comput. Chem.* **1998**, *19*, 1805.
- 23 Carpio, C. A. D.; Takahashi, Y.; Sasaki, S. I. *J. Mol. Graphics* **1993**, *11*, 23.
- 24 Ho, C. M. W.; Marshall, R. J. *Comput.-Aided Mol. Des.* **1990**, *4*, 337.
- 25 Ruppert, J.; Welch, W.; Jain, A. N. *Protein Science* **1997**, *6*, 524.
- 26 Cai, W.; Shao, X.; Maigret, B. *J. Mol. Graphics & Modell.* **2002**, *20*, 313.