Molecular Surface Walk

Janez Konc, Milan Hodošček, and Dušanka Janežič*

National Institute of Chemistry, Hajdrihova 19, SI-1000 Ljubljana, Slovenia

Keywords
solvent accessible surface
molecular graphics
algorithm

A new algorithm to compute a solvent accessible molecular surface (MS) and a graphical computer program for a molecular surface walk have been designed. The surface is generated by rolling a spherical probe representing a solvent molecule over the atoms of the investigated molecule. This surface representation is used in the graphical computer program MS walk to achieve a new way of viewing the molecules. The advantage of the MS walk approach is that the molecular cavities, only partially visible by other methods, can be inspected. The new algorithm's performance is comparable to that of similar existing algorithms and its time complexity is linear with respect to the number of atoms. The MS algorithm and the MS walk program can be accessed on the Web (http://www.cmm.ki.si/konc/ms_walk).

INTRODUCTION

The solvent accessible surface of a molecule is defined as the surface that a spherical probe of radius r_p, representing a solvent molecule, touches as it is rolled over the spherical atoms of this molecule.[1,2] This surface consists of toroidal patches and concave and convex spherical triangular regions. It is useful for studying the structure and interactions of proteins,[3] calculating the volume and surface area,[4] as well as for molecular docking,[5,6] where the solvent accessible surface can serve as a simplified representation of the binding cavity and can be used as a measure of the loss of solvation energy upon binding of a small molecule.[4,7,8] Also, it has some relevance for defining a dielectric boundary of a molecule in numerical solutions of the Poisson-Boltzmann equation.[7,8] There are several computer programs for computing the smooth molecular surface, amongst others MSMS[9] and SURF,[10] both of which are a part of the molecular graphics program VMD.[11]

In this paper we present: first, the molecular surface (MS) algorithm for the calculation of the smooth molecular surface, which differs from the existing algorithms[9,10] in the method used to calculate the surface, the efficiency of the triangulation of this surface, and in the calculation speed; second, a visualization program, which implements a new perspective of the molecule that gives an experience of walking on the molecular surface. The molecular surface walk could be used in combination with molecular dynamics simulations[12–14] where a drug molecule is guided into the active site, which is often concealed inside a protein.[15,16] Using this approach, interactions of the drug molecule with the active site could be more reliably observed.

METHODS

Molecular Surface Algorithm

One of the standard graphical representations of a molecule is the CPK model, which represents atoms in the molecule as a set of partially overlapping spheres of different radii. The input to our MS algorithm is the CPK model of the molecule, for which the smooth molecular surface is calculated (*i.e.*, coordinates of the atom centers and van der Waals radii). The first step in the algorithm is to define which atoms in the molecule are close together. For each atom we need to know the coordinates of all the atoms that are less than some distance away from that particular atom. We achieve this by dividing the space occupied by the molecule into smaller box-shaped segments and mapping the atoms into these segments. A brute force algorithm that calculates the distance for every pair of atoms in a molecule takes $O(n^2)$ time. The MS algorithm compares only atoms that are mapped to neighboring segments, which reduces the time complexity of determining the closeness of atoms to linear time $O(n)$.

The above preparation of the input is necessary for an efficient implementation of the second part of the MS algorithm. Here, a spherical probe with radius r_p is rolled across the surface of the molecule avoiding steric clashes (overlaps) with the molecule's atoms. This spherical probe is used instead of a solvent molecule to determine which parts of the investigated molecule are accessible to solvent. If the probe sphere touches the molecule and has no steric clashes with neighboring atoms, the MS algorithm marks this part of the molecule as solvent accessible.

The procedure by which the MS algorithm computes the solvent accessible surface is the following. The algorithm first assigns the atoms of the investigated molecule to groups of three atoms – triples, where the interatomic distance between any two of the three atoms in a triple is less than $2r_p$. An atom can be assigned to more than one group at the same time. The algorithm then selects triples of atoms of the investigated molecule until all possible triples have been considered. For each triple, the center of the probe sphere is calculated so that the surface of the probe is tangential to all three atoms in the triple, which is depicted in Figure 1.

If the probe overlaps with a nearby atom, then the algorithm rejects such a position, since it is not on the molecule's surface (grey region in Figure 1). If, on the other hand, the probe does not overlap with any of the nearby atoms of the molecule, then the vectors r_1, r_2, r_3, which define the points of tangency of the probe with the atoms at$_1$, at$_2$, and at$_3$, are calculated. The triangle with sides $r_1 - r_2$, $r_2 - r_3$, and $r_3 - r_1$ is then projected onto the curved surface of the probe atom, forming a concave triangular spherical region of the solvent accessible surface (blue region in Figure 1).

The probe atom then rolls around the axis of atoms at$_1$ and at$_2$ until it hits atom at$_4$ in position p', as shown in Figure 2. During this motion, the probe is in contact only with atoms at$_1$ and at$_2$, as shown in Figure 3. The points through
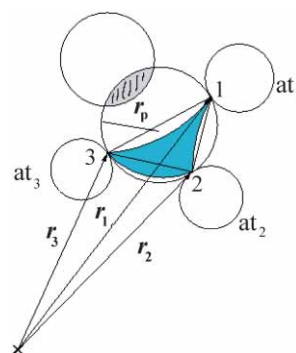

Figure 1. The probe atom with radius r_p sits between the three molecule atoms at$_1$, at$_2$, at$_3$. If there are no steric clashes (overlaps) of the probe with nearby atoms, then a triangle is drawn between the vectors **r_1**, **r_2**, **r_3**, (*i.e.*, between the points of tangency numbered 1, 2 and 3). Its projection onto the curved probe atom surface, which forms a concave patch, is colored blue. If a steric overlap (colored grey) of the probe with a nearby atom exists, then this position of the probe is rejected.
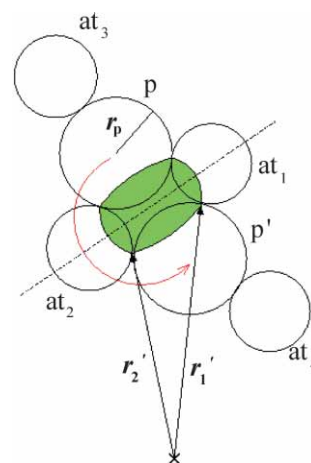

Figure 2. The toroidal saddle-shaped region (green) is defined by the points on the surface of the probe atom, which moves from position p to position p', in each of which the probe is tangential to a triple of atoms. This motion is illustrated with a red arrow. Vectors **r_1'** and **r_2'** define the points of tangency of the probe with atoms at$_1$ and at$_2$ in position p'.
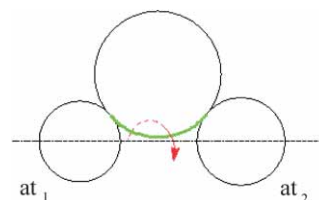

Figure 3. The probe atom rolls around the at$_1$-at$_2$ axis touching only these two atoms. The toroidal surface is colored green. The red arrow delineates the motion.

which the probe atom surface moves from position p to position p' define a toroidal saddle-shaped region of the molecular surface. The MS algorithm calculates these toroidal regions after the concave spherical regions have been triangulated.

Calculation of the concave regions gives vectors r_1, r_2 and r_1', r_2' for each pair of the neighboring surface atoms (Figures 1 and 2). Vector points r_1, r_2 and r_1', r_2' define the borders of the saddle regions and are used for calculating these regions. The MS algorithm takes two neighboring surface atoms at a time and selects the four vector points (*e.g.*, r_1, r_2 and r_1', r_2' for atoms at_1, at_2). The flat rectangular surface between these four vector points is then triangulated by computing a few additional surface points to reflect the curved motion of the probe atom. If the probe touches a single atom, then it is on a convex spherical patch, which is further triangulated. The whole algorithm scales linearly with respect to the number of atoms. The MS algorithm for computing the solvent accessible surface is shown in Figure 4.

while there are triples of neighboring atoms *do*

    select three atoms so that the probe can sit in between them

    calculate the center of the probe

    detect steric clashes (overlaps) between the probe and other atoms

    calculate points of tangency of the probe with the three atoms

    triangulate the concave triangular spherical region

    forall pairs of atoms in this triple

        if two atoms are shared by a second triple *then*

            connect the two concave regions by a toroidal surface

forall surface atoms

    triangulate the remaining convex spherical regions on each atom

Figure 4. Overview of the MS algorithm.

Molecular Surface Walk

The output of the MS algorithm are *x, y, z* coordinates of the vertices of the triangles describing the molecular surface. To model the graphical shading of the molecular surface, a normal vector is calculated for each triangle. We take this representation of the surface as an input for our MS walk visualization program.

In the current molecular graphics programs[11] three functions, rotation, translation, and scaling, are applied to the molecular coordinates. These rotate, translate, and scale the molecule about its center. The user has to apply these operations to examine the molecule from the position he wants.

In our MS walk program, the user's viewpoint moves and the user's viewing direction rotates, the molecule being fixed, achieving an experience of walking on the molecular surface. To our knowledge, this concept has not been implemented in full extent in other programs dealing with molecular graphics. The movement is controlled with a computer mouse and the keyboard rather than with different six degrees of freedom pointers, which enables one to quickly inspect the inner cavities of molecules from different perspectives. This allows faster identification and better perception of important atoms, groups, or residues for drug-receptor interactions.

RESULTS

The MS algorithm and the MS walk program have been implemented in the C++ programming language. Table I shows our timings for computation of the solvent accessible surface for various proteins for a probe radius of 1.4 Å. These results were obtained on a 1.5 GHz Intel Centrino configuration. The proteins for which we performed the molecular surface calculation are 17β-hydroxysteroid dehydrogenase,[17] deoxy human hemoglobin A,[18] src family kinase,[19] and human erythrocyte catalase.[20] The RCSB Protein Data Bank[21] files for these molecules, with the exception of 17β-hydroxysteroid dehydrogenase, which is a homology built model, are 1BIJ.pdb, 1AD5.pdb and 1DGB.pdb, respectively. We have removed all extra water molecules and complexed ligands that were at the end of these pdb files, since they are not parts of the molecules.

TABLE I. Times and numbers of triangles for generation of a solvent accessible surface with the MS algorithm for the probe radius 1.4 Å

Protein	Atoms	Time / s	Triangles / $\times 10^3$
17β-HSD	2024	0.210	68
1BIJ	4513	0.446	144
1AD5	8474	0.834	264
1DGB	16066	1.390	411

At the present stage, our program calculates molecular surfaces that are colored by the atom type of surface atoms. Color codes are: oxygens are red, nitrogens are blue, carbons are grey, hydrogens are white, and sulfurs
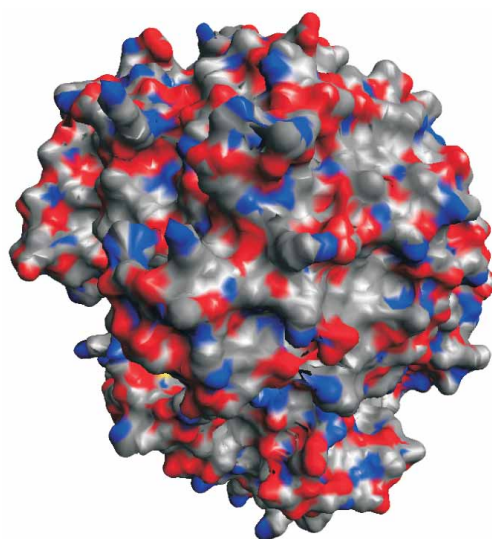

Figure 5. The smooth molecular surface of a 17β-hydroxysteroid dehydrogenase with 2024 atoms generated by the MS algorithm visualized in the MS walk computer program. The probe radius is 1.4 Å.
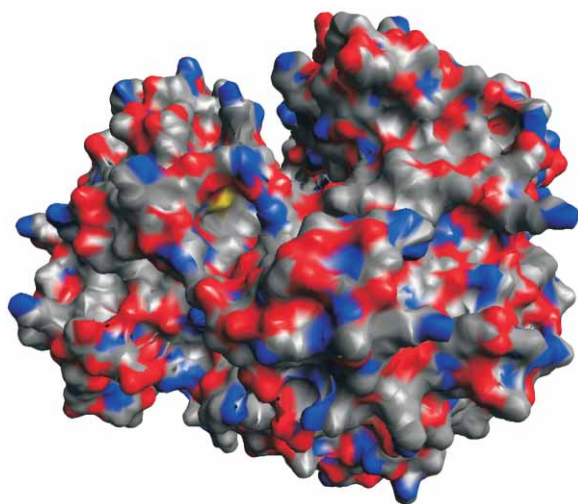
Figure 6. The smooth molecular surface of a human deoxy hemoglobin A consisting of 4513 atoms for a probe radius of 1.4 Å.
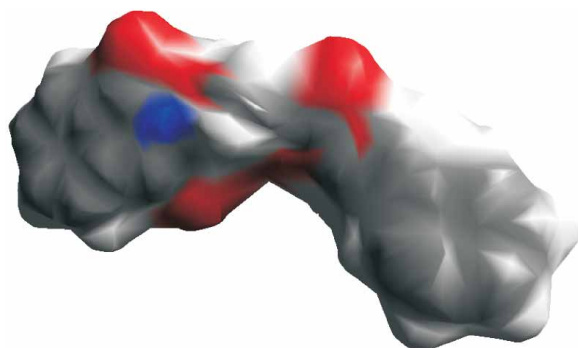

Figure 7. Molecular surface of a molecule of cinnamate (39 atoms) for a probe radius of 1.4 Å.

are yellow. Molecular surfaces of two proteins, 17β-hydroxysteroid dehydrogenase and human deoxy hemoglobin, colored by this atom type color map are represented in Figures 5 and 6. The surface of a small molecule is illustrated for a molecule of cinnamate in Figure 7. These
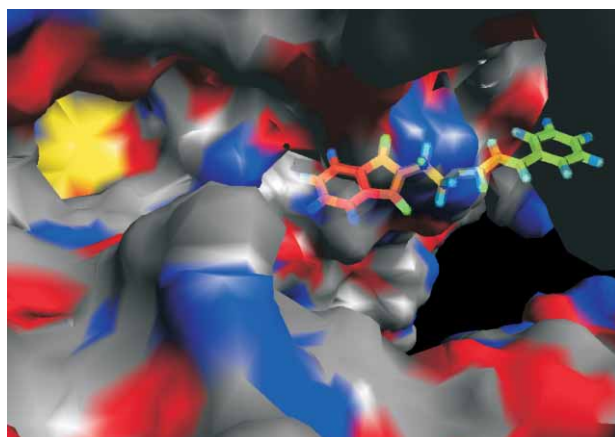

Figure 8. A small molecule in a protein cavity.

surfaces can be viewed from any viewpoint or angle in the MS walk molecular graphics program. A molecular surface walk approach could also be used in manual docking to guide small molecules into proteins, to get an idea of how interactions are involved in the molecular recognition process. This approach, not yet implemented, is shown in Figure 8.

CONCLUSIONS

The paper presents a newly developed MS algorithm for computation of a solvent accessible molecular surface. The algorithm efficiently triangulates molecular surfaces. Also, a new MS walk graphical computer program for visualizing molecules is described. The program implements a new way of viewing molecules. The position of the viewer rather than that of the molecule changes. However, much work remains to be done. For example, the problem of self-intersecting parts of the surface has not been addressed. In the future, more complex functions concerning triangulation will be added and parallelization of the MS algorithm will be performed.[22,23]

Availability

The MS algorithm and the MS walk, both implemented in the C++ programming language, are available for Linux platforms. The programs can be accessed on the Web at http://www.cmm.ki.si/konc/ms_walk.

REFERENCES

1. M. L. Connolly, *J. Appl. Cryst.* **16** (1983) 548–558
2. B. Lee and F. M. Richards, *J. Mol. Biol.* **55** (1971) 379–400.
3. F. Merzel and J. C. Smith, *Acta Crys.* **58** (2002) 242–249.
4. L. Cavallo, J. Kleinjung, and F. Fraternali, *Nucleic Acid Res.* **31** (2003) 3364–3366.
5. T. J. Richmond, *J. Mol. Biol.* **178** (1984) 63–89
6. J. Wang, W. Wang, S. Huo, M. Lee, and P. A. Kollman, *J. Phys. Chem.* **105** (2001) 5055–5067.
7. A. Nicholls and B. Honig, *J. Comp. Chem.* **12** (1991) 435–445.
8. J. Tomasi and M. Persico, *Chem. Rev.* **94** (1994) 2027–2094.
9. M. F. Sanner, A. J. Olson, and J.-C. Spehner, *Proceedings of the Eleventh Annual Symposium on Computational Geometry*, Vancouver, British Columbia, Canada, 1995, pp. 406–407.
10. A. Varshney, F. P. Brooks, and W. V. Wright, *IEEE Computer Graphics and Applications*, 1994, pp. 19–25.
11. W. Humphrey, A. Dalke, and K. Schulten, *J. Mol. Graphics* **14** (1996) 33–38.

12. D. Janežič, M. Praprotnik, and F. Merzel, *J. Chem. Phys.* **122** (2005) 174101.

13. M. Praprotnik and D. Janežič, *J. Chem. Phys.* **122** (2005) 174102.

14. M. Praprotnik and D. Janežič, *J. Chem. Phys.* **122** (2005) 174103.

15. R. A. Laskowski, N. M. Luscombe, M. B. Swindells, and J. M. Thornton, *Protein Science* **5** (1996) 2438–2452.

16. R. A. Laskowski, *J. Mol. Graphics* **13** (1995) 323–330.

17. K. Kristan, K. Krajnc, J. Konc, S. Gobec, J. Stojan, and T. Lanišnik-Rižner, *Steroids* **70** (2005) 626–635.

18. E. J. Fernandez, C. Abad-Zapatero, and K. W. Olsen, *J. Mol. Biol.* **296** (2000) 1245–256.

19. F. Sicheri, I. Moarefi, and J. Kuriyan, *Nature* **385** (1997) 602.

20. C. D. Putnam, A. S. Arvai, Y. Bourne, and J. A. Tainer *J. Mol. Biol.* **296** (2000) 295.

21. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, *Nucleic Acids Res.* **28** (2000) 235–242.

22. U. Borštnik, M. Hodošček, and D. Janežič, *J. Chem. Inf. Comput. Sci.* **44** (2004) 359–364.

23. U. Borštnik, M. Hodošček, and D. Janežič, *Croat. Chem. Acta* **78** (2005) 211–216.

SAŽETAK

Šetnja po površini molekule

Janez Konc, Milan Hodošček i Dušanka Janežič

Razvijen je novi algoritam za izračun otapalu dostupne površine molekule (MS) i grafički računalni program za šetnju po površini molekule. Površina se ostvaruje kotrljanjem kuglaste sonde, koja predstavlja molekulu otapala, po atomima istraživane molekule. Korištenjem ovako dobivenog prikaza grafički računalni program *MS walk* ostvaruje novi pogled na molekule. Prednost pristupa *MS walk* je u mogućnosti promatranja šupljina u molekuli, koje su samo dijelom vidljive u drugim postupcima. Izvedba ovdje uvedenog algoritma, čija kompleksnost linearno raste s brojem atoma, usporediva je s drugim sličnim algoritmima. Algoritam i program *MS walk* dostupni su na Web stranici: http://www.cmm.ki.si/konc/ms_walk