REAL TIME SURFACE RECONSTRUCTION FOR MOVING MOLECULAR FRAGMENTS

M.F. SANNER and A.J. OLSON
The Scripps Research Institute, 10666 North Torrey Pines Road, La Jolla, CA 92037, USA

Recently we introduced the Reduced Surface as an efficient tool to built molecular surfaces. We describe here how this geometric construct can be used to efficiently reconstruct the solvent excluded surface of a protein for which the coordinates of a subset of atoms are changing. We show that, the complexity of that operation is not dependant upon the size of the molecule and is in O[$t$log(t)] where t is the maximum of the number of probes and atoms involved in the reconstruction of the surface. The algorithms described here have been implemented and tested on several proteins. The triangulation of the solvent excluded surface of proteins in which a side chain was changing conformation could be updated at rates ranging from 7 to 22 frames per second. We also applied this method to compute the surface area fluctuation of the FIV protease undergoing a constrained molecular dynamics simulation (16 mobile residues). Rate of 6 frames per second were obtained in this case.

1 Introduction

Molecular surfaces are widely used in molecular modeling[1] to describe hydration effects[2-7] and to visualize and study molecular properties and interactions[8-12]. Several algorithm to compute such surfaces have been presented[13-18]. A method to interactively deform a molecular surface using control points was proposed by Klein *et al.*[19]. Recently, we described a program called MSMS[20] to compute a triangulation of the solvent excluded surface with a complexity in time in O[$n$log(n)] for a molecule made of n atoms. This algorithm is based on the use of the *reduced surface*[20,21] which provides a compact way to store the geometric information needed to generate several molecular surfaces: van der Waals (VdW), solvent accessible[22] (SAS) and solvent excluded[23] (SES) surface. The performance of this algorithm allowed us to envision using it in some interactive applications such as “molecular sculpting” where the surface provides interactive feed-back on the impact of conformational changes on shape. The method presented here, also drastically reduces the CPU time required to compute molecular surfaces in the case of singularities badly handled by MSMS.

After briefly defining the reduced surface and its main properties, we present algorithms to update in real time the SES of a molecule when a subset of atoms is moving. Updating of the surface implies identifying and removing faces, edges and vertices in the reduced and solvent excluded surface and reconstruct these surfaces. The complexity of the presented algorithm is discussed and the

timings of several examples of molecular sculpting are reported. We also used these algorithms to update the surface of the FIV protease undergoing constrained molecular dynamics. We report rates for updating the triangulated representation of its solvent excluded surface as well as the time needed to analytically compute the surface area at every step.

1.1 Reduced Surface Definition

Like the SES and the SAS, the reduced surface is defined using a spherical probe of radius rp representing the volume of a solvent molecule. When the probe is simultaneously in contact with three or more atoms it is in a *fixed position* because it cannot roll further without losing contact with at least one of the atoms. The polygon obtained by connecting with line segments the centers of the atoms on which the probe is lying in a fixed position is a *face of the reduced surface* (RS-face). The atom centers are *vertices of the reduced surface* (RS-vertices) and the line segments are *edges of the reduced surface* (RS-edges).

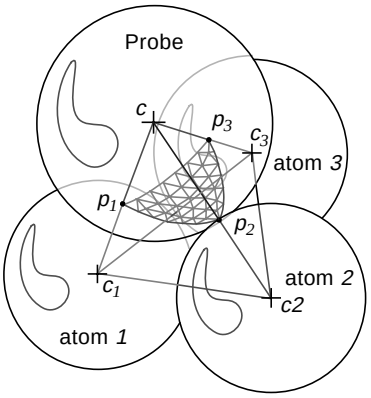

Figure 1: Probe in a fixed position defining an RS-face (c_1,c_2,c_3), an SES spherical reentrant face (p_1,p_2,p_3) and an SAS vertex *c*.

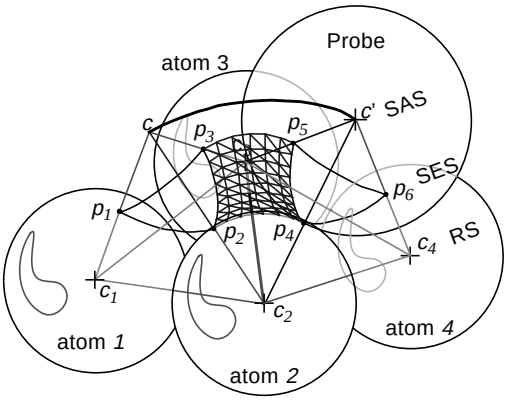

Figure 2: Probe rolling over two atoms defining an RS-edge (c_2,c_3), an SES toroidal face (p_1,p_2,p_3) and an SAS edge (c,c').

Figure1 shows a probe in a fixed position and the RS-face it defines. In some cases, the probe may roll completely around a pair of atoms without colliding with a third atom, the RS-edge joining the centers of these two atoms will be called a *free RS-edge* because it doesn't belong to any RS-face. Similarly, when the probe can roll on an atom without touching any other atom its center will be called a *free RS-vertex*. The polygons associated with all fixed positions of the probe and the free RS-edges and RS-vertices define the reduced surface. This

surface can be made of several components each of them made of several closed surfaces connected by RS-vertices and/or free RS-edges.

1.2 Reduced Surface Properties

Under the assumption that the probe is never in contact with more than three atoms at a time, the faces of the reduced surface are triangles. There is an *outer component* that encloses all components that are not reduced to a single free RS-element. Geometrical properties of the reduced surface have been described previously[20]. In particular, one can show that for a protein made of n atoms the number of RS-faces and RS-edges grow as a linear functions of n and we gave an algorithm to compute the outer component of this surface in $O[n\log(n)]$ operations.

It should be noted that the reduced surface provides all the geometric information necessary to build the SAS and the SES. Figure 1 shows how an RS-face (c_1,c_2,c_3) corresponds to a spherical reentrant SES-face (p_1,p_2,p_3) and an SAS vertex c. Similarly, figure 2 shows the relationship between the RS-edge (c_2,c_3), the SES spherical reentrant face (p_2,p_4,p_5,p_3) and the SAS edge (c,c'). In the same way every RS-vertex corresponds to a contact face of the SES and an SAS face. It is also important to realize that the RS-vertices are located at atom centers providing a direct hint about which parts of the reduced surface need to be updated when some atoms move.

2 Updating the surfaces after a subset of atoms has moved

Let us assume that we have computed, for a molecule made of n atoms and a given probe size, the reduced surface, the analytical model of the SES and its triangulation and that these surfaces are available in the data structures described in our previous work[20].

The following algorithms will update these surfaces for a subset of m atoms which have been assigned new positions and/or radii. In the following we will refer to these atoms as the "*moving atoms*". To update the surfaces means to delete the obsolete parts of the reduced and analytical SES, reconstruct these surfaces and triangulate the newly created SES-faces.

2.1 Updating the Reduced Surface

An RS-face needs to be updated in the two following cases: one of its vertices is the center of a moving atom or secondly a moving atom has an intersection with the probe corresponding to the face. In the latter case, this RS-face will not belong to the updated reduced surface and has to be deleted. In the first case,

this RS-face may or may not be part of the updated reduced surface. To find out, one needs to check the new probe position for intersections with all other atoms. In terms of complexity, this is as expensive as deleting and reconstructing the face. Therefore all these faces are systematically deleted. RS-edges shared by deleted RS-faces are deleted, as well as RS-vertices for which all faces they belong to, have been deleted. Free edges of the reduced surface with at least one vertex located at the center of a moving atom are deleted as well.

As mentioned in section 1.2 every face, edge and vertex of the reduced surface corresponds to a face in the analytical description of the solvent excluded surface. The SES-faces corresponding to the deleted elements of the reduced surface are tagged for deletion.

To reconstruct the reduced surface, we apply the same algorithm as the one used to construct the initial reduced surface[20]. This algorithm starts from an initial face defining a fixed probe position and "rolls" the probe over an RS-edge to the next fixed position defining the neighbor face (see Fig. 2). This operation was referred to as the "*treatment of an edge*". This recursive algorithm is executed as long as there are edges to be treated. In our case, after the out-dated parts of the reduced surface have been deleted, we are left with a number of RS-edges which belong to only one face (Fig. 3). Any such face-edge pair can be used to start this algorithm. Every RS-vertex which belongs to a reconstructed RS-face or RS-edge is then checked for free RS-edges using the algorithm described previously[20].

This procedure will re-build a closed RS-component and new free RS-edges if there are any. In the case where a subset of one or more atoms moves far enough from the rest of the molecule to form its own component, this component wouldn't be found. For this to happen, this subset of atoms has to move further than a probe diameter away from any RS-vertex belonging to a re-built RS-face or RS-edge. Another assumption made for this algorithm to work is that there is at least one face of the reduced surface that is not deleted. If this is not the case the problem is equivalent to recomputing the whole surface.

2.2 Reduced Surface Update Complexity

For every RS-vertex we know its associated RS-faces. Therefore, the RS-elements having at least one vertex located at the center of a moving atom can be selected in order m operations for m moving atoms.

To find the RS-faces for which a moving atom collides with the associated probe, one has to test the probes in a fixed position against every moving atom. In fact, only probes close enough to the moving atoms have to be considered. Let us call k the number of probes in fixed positions close to any moving atom during the entire course of motion. These probes can be selected once at the begin-

ning and at every update we can build a BSDtree[20] for these probes and the ones created during the previous update. This tree can be built in O[$k$log(k)] operations and enables the testing of any moving atoms for intersections with fixed probes in O[log(k)] operations.

The algorithm to re-build the reduced surface also uses a BSDtree to select close atoms while treating an edge. This tree has to be re-built for every update to include the new positions of the moving atoms. Let us call l the number of atoms involved in the reconstruction of the reduced surface at any step. These atoms can be selected before hand and the tree can be build in O[$l$log(l)]. Once this tree is built, an edge of the reduced surface can be treated in O[log(l)] operations.

Both, the number k of probes and the number l of atoms involved in the update depend on the number m of moving atoms as well as on the compactness of the set of moving atoms. Their upper bound is the number of atoms n, but for applications like molecular sculpting, both are independent of n and the complexity of updating the reduced surface is in O[$t$log(t)] where t = max(l,k).

2.3 Updating The Analytical Description Of The Solvent Excluded Surface

The SES-faces to be deleted have been tagged during the update procedure of the reduced surface. These faces are now deleted along with all SES-edges shared by two deleted SES-faces and SES-vertices shared only by deleted SES-faces (Fig. 3). SES-edges shared by a spherical reentrant face and a deleted toroidal face are deleted as well. This is done in order to be able to re-construct the SES using the same procedure as the one used to build it initially.

The procedure to build the SES starts by creating a spherical reentrant face corresponding to an initial face of the reduced surface. For every edge of this initial RS-face, the spherical reentrant triangle corresponding to the neighbor RS-face is created along with the toroidal reentrant face connecting these two spherical reentrant faces (Fig. 2). This recursive algorithm stops when all the edges and faces of the reduced surface have been treated at which point all the reentrant faces of the analytical SES have been created.

To re-build the SES we start this algorithm on any reconstructed RS-face. Once, all reentrant faces are created the contact faces are found by building cycles of convex edges belonging to toroidal reentrant faces. This analytical surface description is then submitted to the same singularity handling algorithm as the one used for the initial SES. This algorithm will correct the self intersecting parts of the SES by checking singular edges that appeared during the SES reconstruction.

The complexity to update the analytical description of the solvent excluded surface is a linear function of the number of vertices, edges, and faces deleted in

the reduced surface. These numbers are in O[m] for m moving atoms.

2.4 Updating The Triangulation Of The Solvent Excluded Surface

In MSMS the triangulation of an SES-component was stored in dynamically allocated arrays, one for the vertices and one for the faces. To enable an efficient update of the triangulation we have modified this data structure. Now, the triangulation is stored with the analytical surface description. Every SES-vertex is its own triangulation vertex. Every SES-edge holds an array of triangulation vertices used for its discrete representation and finally, every SES-face stores an array of internal triangulation vertices (not on the edge or at the vertices) and an array of triplets of vertices pointers describing the triangular faces.

When an SES-face is deleted, its triangulation is deleted too. Updating the triangulation of the SES simply requires the triangulation of the newly created faces using the same, template based, procedure as the one used for the initial triangulation.

3 Results

To evaluate the performance of the algorithms described here, we timed two types of surface reconstruction experiments. First we modified the conformation of a single side chain in proteins of different sizes, next we progressively increased the number of moving atoms for the same protein. All times reported in that section are given in seconds and have been obtained on a Dec Alpha 250 running at 233 Mhz.

3.1 Protein Side Chain rotation

We have chosen a set of seven proteins with evenly distributed sizes ranging from 327 to 3263 atoms. For each of these proteins, shown in table 1, we selected a solvent accessible side chain allowing a 360 degree rotation about its CHI1 or CHI2 angle while remaining at the surface of the protein. Steric repulsion was not taken into account and some moving atoms can get very close to some fixed atoms during the rotation. We incremented the value of the rotatable torsion angle in steps of 5 degrees and updated the SES triangulated at a density of 1.0 vertex/Å^2.

We report in table 1 the average CPU time per update and the average number of RS-faces deleted per update. The update rate ranges from 22/sec. for CRAMBIN to 7/sec. for LIVER ALCOHOL DEHYDROGENASE and CYTOCHROME P450. These values show that our implementation still is molecule size dependant. This dependency was expectable because our implementation

relies to a large extent on the use of existing procedures written for MSMS and our data structure doesn't yet allow storage of lists of reconstructed faces, edges and vertices, sometimes forcing procedures to check all elements of a surface to find out which ones need attention. Nevertheless, the rates achieved still enable real time manipulation of the side chains in fairly large proteins. The times indicated here include reading the new coordinates of the moving atoms from a different file at every step. As expected, the number of RS-faces deleted at each step correlates well with the number of moving atoms.

Table 1: Timing for 71 surface updates after modifying a single side chain conformation in proteins of different sizes.

Protein	Number of atoms	Rotatable side chain	Modified torsion angle	Average update time (sec.)	Average removed RSF face	Moving atoms
1crn	327	TYR-29	CHI 2	0.045	27.75	6
1fd2	848	ARG-106	CHI 1	0.058	23.07	6
2lzm	1309	LYS-34	CHI 1	0.068	14.72	4
3pgm	1854	TRP-75	CHI 1	0.125	32.20	9
7abp	2338	GLU-42	CHI 1	0.104	15.86	4
5adh	2787	GLN-299	CHI 1	0.145	16.56	7
3cpp	3263	GLN-343	CHI 1	0.145	14.73	4

3.2 Constrained Dynamic Trajectory

The trajectory of the FIV protease mutant D30N,D30'N undergoing a constrained molecular dynamics (MD) simulation was used to evaluate the update rate dependency upon the number of moving atoms. During this 55 picosecond MD simulation, made to monitor side conformations in the active side, most of the protein was held fixed. Only 16 residues corresponding to 100 atoms of the protease were allowed to move. The simulation was done with Biosym's program Discover under the following conditions: a temperature of 350 degree Kelvin and a time step of 10^{-15} second. It started with 5000 steps equilibration

Table 2: List of the moving residues in the FIV protease MD simulation.

1	ILE 1037	5	LEU 2029	9	LEU 1029	13	ILE 2037
2	VAL 1059	6	ASN 2030	10	ASN 1030	14	VAL 2059
3	GLY 1060	7	THR 2031	11	THR 1031	15	GLY 2060
4	GLY 1061	8	GLY 2032	12	GLY 1032	16	GLY 2061

followed by 50,000 steps of simulation. Every hundredth step the conformation

was saved. We took every tenth conformation (a conformation every picosecond) out of the 500 ones saved during the simulation. For these 50 conformations we measured the time to update the surface according to the number of atoms whose positions were taken from the simulation. We made 13 runs progressively increasing the number of residues with moving atoms. The column "Residues" in table 3 reports the range of residues (as reported in table 2) taken as mobile in every run. We report the CPU time to reconstruct the SES triangulated at a density of 1.0 vertex/$Å^2$ and the number of RS-faces deleted over 50 updates.

Table 3: Timing for updating 50 times the SES of FIV for an increasing number of moving atoms.

Run	1	2	3	4	5	6	7	8	9	10	11	12	13
Residues	1	1-2	1-4	1-5	1-6	1-7	1-8	1-9	1-10	1-11	1-13	1-14	1-16
Nb. atoms	8	15	23	31	39	46	50	58	66	73	85	92	100
RS-Faces	384	1796	2798	2798	3248	3333	3973	3973	4077	4115	5045	6121	6809
Time (sec.)	3.8	6.5	8.8	8.3	9.0	9.3	10.3	10.4	10.4	10.6	12.1	13.7	14.6

In this experiment the dependency upon the size of the protein is a constant and the increase in time corresponds to the increase of the number of RS-elements and SES-elements that have to be updated. Figure (4.A) shows that the time to reconstruct the surfaces grows as a linear function of the number of RS-faces deleted. Since all side chains do not contribute equally to the surface the dependency upon the number of moving atoms (figure 4.B) is not as straight.

When all 16 residues are allowed to move approximately 8% of the reduced and solvent excluded surfaces is reconstructed and triangulated at each update at a rate of 3 updates per second. This number should improve as the data structures will allow to make the reconstruction molecule size independent.

3.3 Analytical surface area calculation

The 50 conformations extracted from the MD simulation presented in 3.2 were used to evaluate the time needed to analytically compute the area of the SES. All 16 residues (100 atoms) were considered mobile. For every conformation we updated the reduced surface and the analytical representation of the SES and re-computed the surface area. The average time per update was 0.17 second which is equivalent to almost 6 frames per second.

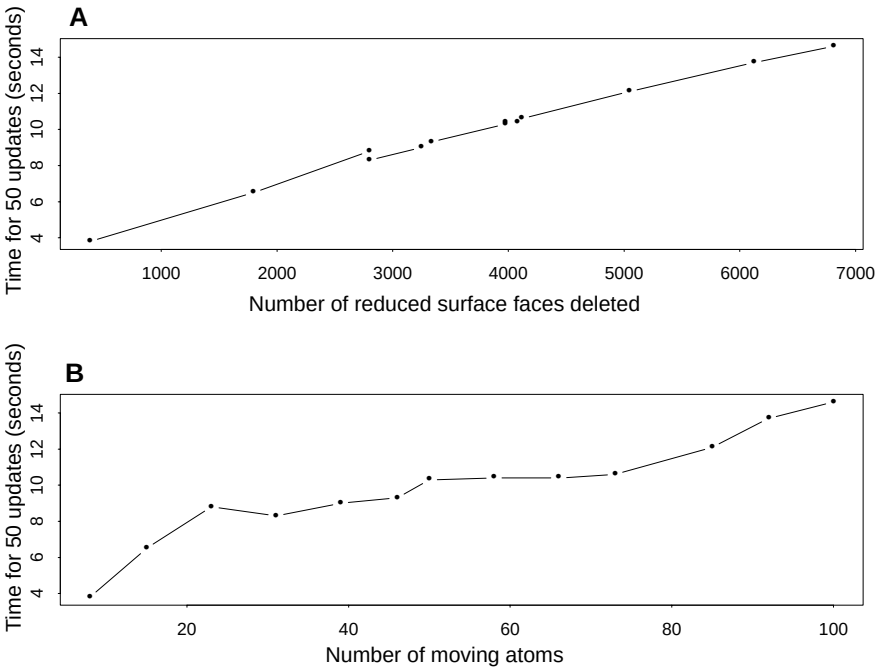

Figure 4: Time for 50 updates of the triangulated SES of the FIV protease during the MD simulation as a function of the number of moving atoms (A) and deleted RS-faces(B).

4 Discussion and conclusion

We have presented an algorithm to reconstruct the reduced and solvent excluded surface of a molecule when a subset of atoms is moving. The performance of this algorithm enables real-time *sculpting* of molecular surface. A user can select a side chain of a protein, modify its conformation and see the surface move along with atoms as the torsion angles are modified. This enables the use of molecular surfaces as a visual cue showing the influence of such conformational changes on the molecule's shape. The only condition for the described algorithms to work is that for every update, at least one RS-face of the previous surface remains unchanged. The complexity of an update operation is in O[$t$log(t)] where t is the maximum of the number of probes and the number of atoms involved in the update operation. This number t is a function of the number of moving atoms and of the amplitude of the motion.

Although our implementation still shows a dependency upon the size of the

molecule, this can be overcome. This dependency occurs mainly because we implemented these algorithms as an extension to the existing program MSMS[20]. Many procedures used to build the initial surfaces are used by the surface updating algorithm. These procedures often run over all faces, edges and vertices of the surfaces creating a size dependency which is reflected by lower update rates for larger molecules. However, the results show that even with this limitation one can achieve real time surface update on fairly large molecules. In many applications the set of moving atoms is compact and the amplitude of their motion is limited. Knowledge of these conditions helps reducing the complexity of the problem. A known restriction is that, if atoms move far enough from the rest of the molecule to form their own component, they will be omitted and the algorithm will simply close the main component of the reduced surface. Typically, this will not happen when only covalent torsions are modified. In other applications where covalency is not maintained, this component can be identified in O[$m\log(t)$] operations for m moving atoms. The complexity is partly due to the use of BSDtrees which require, in the current implementation, to be reconstructed at every update.

The times and update rates presented here were obtained for a triangulation vertex density of 1.0 vertex / $Å^2$. Higher density will multiply the required CPU time by a constant but it is straightforward to implement a scheme where the surface would be updated with a low triangulation density during the motion and with a higher density as the motion stops.

Analytical surface area can be computed without having to triangulate the solvent excluded surface. The speed achieved in this case would enable inclusion of such a term in a minimization process. Method for rapid calculation of buried surface areas are currently under development.

This type of application requires a very robust program because of the large number of molecular surfaces computed. MSMS was tested against a large set of proteins and proved to be robust. Still, in some cases it restarts the calculation after modifying the radii of some atoms. This happens mainly because of numerical instability in handling singularities. The algorithms presented here are useful in such cases because MSMS no longer has to recompute the whole surface, but simply updates it after modifying some atomic radii.

Acknowledgments

The authors thank Garrett Morris for providing the HIV protease trajectory coordinates and Jean Claude Spehner for his contributions to the definition and study of the reduced surface. This work was supported by NIH grants number P01GM38794 and P41 RR08065-02. This is publication number xxxx-MB from The Scripps Research Institute.

References

1. M.L. Connolly, T.J. O'Donnell, S. Warde. *Network Science*. April issue (1996). http://edisto.awod.com/netsci/Issues/Apr96/toc.html.
2. J. Vila, R.L. Williams, M. Vásquez and H.A. Scheraga, *Proteins: Structures, Functions and Genetics*, **10(3)**, 199 (1991).
3. K.A. Palmer and H.A. Scheraga, *J. Comp. Chem.* **12**, 505 (1991).
4. W.C. Still, A. Tempczyk, R.C. Hawley and T. Hendrickson, *J. Am. Chem. Soc.* **112**, 6127 (1990).
5. C. Zheng, C.F. Wong and J.A. McCammon, *Biopolymers*, **29(14)**, 1877 (1990).
6. I. Tuñòn, E. Silla and J.L. Pascual-Ahuir, *Protein Engineering*, **5(8)**, 715 (1992).
7. R.M. Jackson, M.J.E. Sternberg, *Nature*, 366 (1993).
8. C.L. Fisher, J.A. Tainer, M.E. Pique and E.D. Getzoff. *J. Mol. Graphics*, 8, 125 (1990).
9. C.D. Zachmann, S.M. Kast, J. Brickman. *J. Mol. Graphics*. **13**, 89 (1995)
10. F.E. Blaney, C. Edge and R.W. Phippen. *J. Mol. Graphics*. **13**,165 (1995)
11. B.S. Duncan and A.J. Olson, *Proceedings of the 13th Molecular Graphics Society Meeting* (1994).
12. R.N. Norel, S.L. Lin, H.J. Wolfson and R. Nussinov. *J. Mol. Biol*., **252**, 263 (1995).
13. R.J. Zauhar and R.S. Morgan, *J. of Comp. Chem.*, **11(5)** 603 (1990).
14. E. Silla and F. Villar, O. Nilsson, J.L. Pascual-Ahuir and O. Tapia, *J. Mol. Graphics*, **8**, 168 (1990)..
15. G. Perrot,B. Cheng,K.D. Gibson,J. Villa, K.A. Palmer, A. Nayeem, B. Maigret and H.A. Scheraga, *J. Comp. Chem.*, **13(1)**, 1 (1992).
16. M.L. Connolly, *J. Mol. Graphics*, **11**, 139 (1993).
17. A. Varshney, W.V. Wright and F.P. Brooks Jr., *IEEE Computer Graphics and Applications*. **15(5)**, 19 (1994).
18. F. Eisenhaber, P. Lijnzaad, P. Argos, C. Sanders and M. Scharf, *J. Comp. Chem.*, **16(3)**, 273 (1995).
19. T.E. Klein, C.C. Huang, E.F. Pettersen, G.S. Couch, T.E. Ferrin and R. Langridge, *J. Mol. Graphics*, **8**, 16 (1990).
20. M.F. Sanner, A.J. Olson and J.C. Spehner. *Biopolymers*, **38**, 305 (1996).
21. Sanner M.F. *Ph.D. dissertation thesis: Modeling and applications of molecular surfaces*. Université de Haute-Alsace, France (1992).
22. B. Lee and F.M. Richards, *J. Mol. Biol.*, **55**, 379 (1971).
23. F.M. Richards, *Ann. Rev. Biophys. Bioeng*, **6**, 151 (1977).

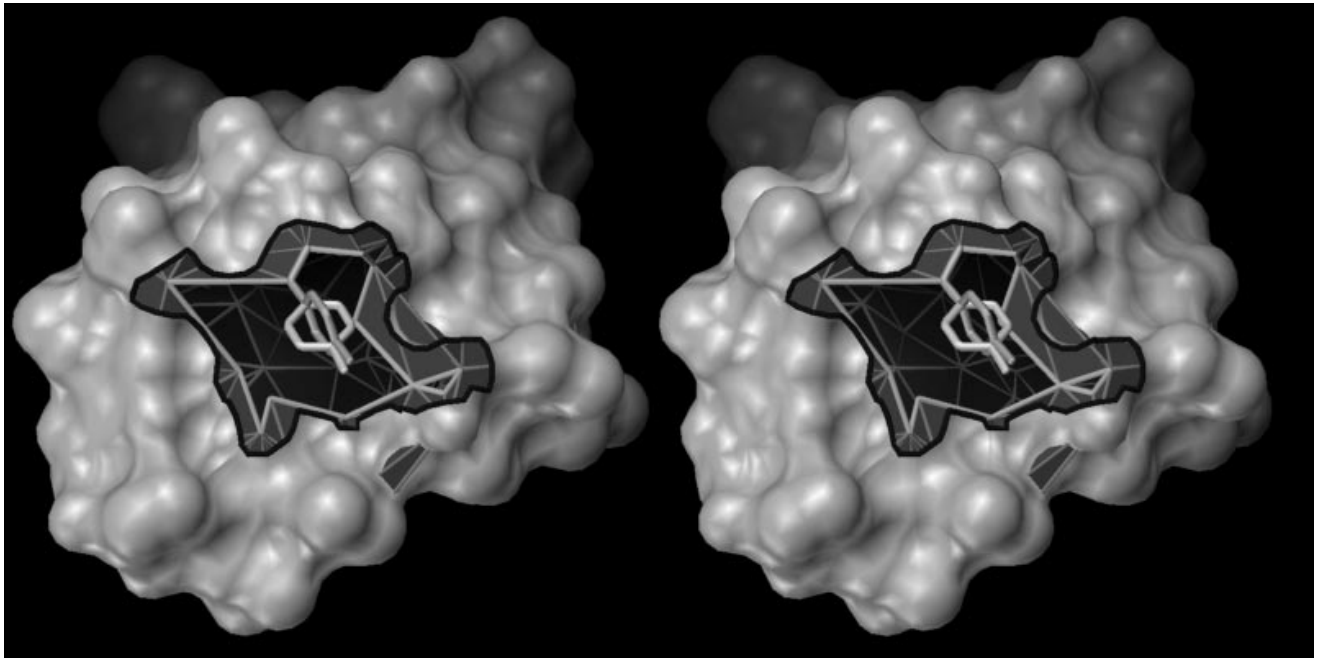

Figure 3: Unchanged solvent excluded and reduced surface for a side chain rotation. The CHI 2 angle of tyrosine 29 in Crambin (1crn) was incremented by 90 degrees. The obsolete parts of the reduced and solvent excluded surface were deleted. The contours of the remaining parts of these surfaces are shown in bold. This is a cross-eyed stereo-diagram