Efficient molecular surface rendering by linear-time pseudo-Gaussian approximation to Lee–Richards surfaces (*PGALRS*)Herbert J. Bernstein^{a*} and Paul A. Craig^b^aDowling College, 150 Idle Hour Boulevard, Oakdale, NY 11769, USA, and ^bRochester Institute of Technology, 85 Lomb Memorial Drive, Rochester, NY 14623, USA. Correspondence e-mail: yaya@dowling.edu

The *PGALRS* (pseudo-Gaussian approximation to Lee–Richards surfaces) algorithm is discussed. By modeling electron density with unphysical pseudo-Gaussian atoms, the Lee–Richards surface can be approximated by a contour level of that density in time approximately linear in the number of atoms. Having that contour level, the atoms and residues closest to that surface can be identified in average time $O[n^{2/3} \log(n)]$ using a *NearTree*-based nearest neighbor search. If a high-quality Lee–Richards surface is required, then, as a final stage, one of the standard Lee–Richards algorithms can be used but considering only the previously identified surface residues; the typical cost is thereby reduced to $O[n^{2/3} \log(n)]$, making the overall average time for all the steps $O(n)$. For very large macromolecules, such a reduction in computational burden may be essential to being able to render a meaningful molecular surface. This approach extends the feasible range of application for existing molecular surface software, such as *MSMS*, to larger macromolecules, especially to macromolecules with more than 50 000 atoms, and can be used as a starting point for surface-based (as opposed to backbone-based) motif identification, e.g. using *ProMol*.

© 2010 International Union of Crystallography
Printed in Singapore – all rights reserved

1. Finding surface atoms and residues

The identification of surface atoms and residues is helpful in understanding the mechanisms of interactions among molecules. One can approach the problem either as a one-stage problem by examining each atom and asking whether it is exposed to solvent or as a two-stage problem by first generating a surface and then identifying atoms that are in proximity to that surface. Lee & Richards (1971) proposed the identification of molecular surfaces by probing the van der Waals surface with a probe atom. This approach is now generally accepted as the definition of a molecular surface, or ‘solvent-accessible surface’, and is supported by a wide range of specific algorithms (Ivanov *et al.*, 2003).

A Lee–Richards surface is defined as the evolute of the center (the ‘solvent-accessible surface’ or ‘SAS’) or the envelope of the contact points (the ‘solvent-excluded surface’ or ‘SES’ or ‘molecular surface’) of a probe sphere, typically of radius 1.4 Å, rolling on the van der Waals surface of a molecule. The solvent-accessible surface is the same surface that one would obtain by building a van der Waals surface where the atoms have their radii extended by the probe-atom radius. The molecular surface is very similar to the van der Waals surface. Where the probe contacts only one atom, the resulting surface is identical to the van der Waals surface. Where the probe contacts two atoms, a portion of a toroidal surface is generated as the probe rolls around the center of the axis connecting the atom centers (see Fig. 1). Where the probe contacts three atoms, a portion of a spherical triangle is generated. Connolly (1983) rediscovered and popularized the molecular surface. Yeates (1995) approached the problem of identification of surface atoms by finding the largest

radius for a sphere tangent to each atom that would not overlap any other atoms. LeGrand & Merz (1993) provided a very efficient algorithm, *Masker*, based on calculation of the solvent-accessible surface, now most commonly used as modified by Bystroff (2002) to compute the solvent-excluded surface. Sanner *et al.* (1996) provided what has generally been considered the most efficient algorithm, *MSMS*. We introduce here the *PGALRS* (pseudo-Gaussian approximation to Lee–Richards surfaces) algorithm. For very large structures, the *PGALRS* algorithm is more efficient than *MSMS*.

A significant portion of the time for a Lee–Richards surface calculation of necessity requires calculations involving atom pairs. Selection of appropriate pairs of atoms is essentially a bonding

Figure 1
Probe atom (red) rolling on the van der Waals surface. The Lee–Richards surface is the envelope of the contact between the probe atom and the van der Waals surface. Note the toroidal reentrant surface in the middle.

calculation, allowing what might seem to be an $O(n^2)$ calculation to be performed in $O[n \log(n)]$ time by use of Levinthal cubing (Katz & Levinthal, 1972) or Voronoi diagrams (Liang *et al.*, 1998). Because molecules do not admit more than a limited number of atoms per $\text{\AA}^3$, an $O[n \log(n)]$ algorithm can exhibit nearly linear time. However, because of the intrinsic need to work with atom pairs, current algorithms eventually display worse than linear behavior as the number of atoms to be considered becomes large. It is therefore desirable to consider approaches that can approximate a Lee–Richards surface without consideration of any atom pairs so that a truly linear-time algorithm can be achieved.

By modeling electron density with unphysical pseudo-Gaussian atoms, the Lee–Richards surface can be approximated by a contour level of that density in time approximately linear in the number of atoms. It is a well established practice (*e.g.* Im *et al.*, 1998; Jemilawon *et al.*, 2007; Bernstein *et al.*, 2008) to approximate a Lee–Richards surface with such a contour level from an electron density map. In some cases (Vertrees, 2008), such an approximation may be the only way to get any indication of a surface. If all atoms have identical density distributions, the contour level of the density at the common van der Waals radius is often a servicable approximation to the Lee–Richards molecular surface, but the quality of the approximation deteriorates as the atom types vary. There is a solution. Having a contour level, the atoms and residues closest to that contour level can be identified in average time $O[n^{2/3} \log(n)]$ using a *NearTree*-based nearest neighbor search (Andrews, 2001). If a high-quality Lee–Richards surface is required, then, as a final stage, one of the standard Lee–Richards algorithms can be used but considering only the previously identified surface residues. This reduces the typical cost to $O[n^{2/3} \log(n)]$, making the overall average time for all the steps $O(n)$, the time to build a Gaussian on each of n atoms, and removing the inaccuracies caused by use of the density contour.

Despite its unphysical nature, the Lee–Richards surface has long been used as a means to carve out a volume for implicit solvent models in simulations to reduce the computational complexity that would result from modeling the solvent atom-by-atom. Superposition of atomic spherically symmetric functions has been shown to provide a reasonable approximation to the Lee–Richards surface for those applications that are more demanding than simple identification of surface atoms from known sets of structural coordinates (Im *et al.*, 1998). The problem noted by Lee *et al.* (2003) in having to accept bulges in the approximate Lee–Richards surface when choosing functions able to fill interior voids does not, in general, prevent the identification of surface atoms and residues for macromolecules, from which more precise approximations to the Lee–Richards surface may be derived after the initial calculation but working with fewer atoms and many fewer atom pairs.

This approach can then be used as a starting point for surface-based (as opposed to backbone-based) motif identification, *e.g.* by using *ProMol* (Craig *et al.*, 2009).

2. Pseudo-Gaussian atoms versus hard spheres

While unphysical, approximating the electron density for an atom with a three-dimensional Gaussian distribution yields useful experimental models for chemistry and biology. For example, in small-molecule crystallography, an atom of N_e electrons may be modeled with a scaled Gaussian:

$$\rho(X, Y, Z) = \frac{N_e}{\sigma^3 2\pi(2\pi)^{1/2}} \exp\left[-\frac{(X-x)^2 + (Y-y)^2 + (Z-z)^2}{2\sigma^2}\right]. \quad (1)$$

The hard-sphere model, as used in computing van der Waals surfaces and Lee–Richards surfaces, represents each atom as a ball of radius r_{vdw} . We can relate these two models by defining a ‘spread’, s , in radii per σ , so that $\sigma = sr_{\text{vdw}}$, and rewriting the scaled Gaussian density at a point $\mathbf{x}$ in space for an atom at center $\mathbf{c}$ as

$$\rho(\mathbf{x}) = \frac{N_e}{s^3 r_{\text{vdw}}^3 2\pi(2\pi)^{1/2}} \exp\left(-\frac{\|\mathbf{x} - \mathbf{c}\|^2}{2s^2 r_{\text{vdw}}^2}\right), \quad (2)$$

defining ρ_0 for a point $\mathbf{x}_0$ such that $\|\mathbf{x}_0 - \mathbf{c}\| = r_{\text{vdw}}$ as

$$\rho_0 = \frac{N_e}{s^3 r_{\text{vdw}}^3 2\pi(2\pi)^{1/2}} \exp\left(-\frac{1}{2s^2}\right). \quad (3)$$

Then $\eta(\mathbf{x}) = \rho(\mathbf{x})/\rho_0$ has a contour level of 1 at the van der Waals radius of the atom. Using this pseudo-density, η , in place of a ‘real’ electron density, ρ , for all the atoms in a molecule allows a common contour level of 1 to be used to mark an envelope around the van der Waals surface of the molecule that will tend to hug individual atoms in regions that are not in close contact with other atoms and will smoothly bridge the gaps among atoms that are close to one another:

$$\eta(\mathbf{x}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{c}\|^2}{2s^2 r_{\text{vdw}}^2} + \frac{1}{2s^2}\right). \quad (4)$$

The spread, s , provides an additional degree of freedom that can be used to simulate the effect of changing the radius of a Lee–Richards surface-probe atom.

3. Pseudo-Gaussian approximation to Lee–Richards surfaces

Consider two atoms of common van der Waals radius r_{vdw} with centers d apart upon which atoms a probe atom of radius r_p is rolled to form a Lee–Richards surface. The midpoint, $\mathbf{x}_{\text{mid}}$, of the reentrant surface will be a distance h above the line between the atom centers, so that $(d/2)^2 + (h + r_p)^2 = (r_{\text{vdw}} + r_p)^2$ and $(d/2)^2 + h^2 = w^2$, where w is the distance from either atom center to $\mathbf{x}_{\text{mid}}$.

$$w^2 = r_{\text{vdw}}^2 + 2r_p r_{\text{vdw}} + 2r_p^2 - r_p(4r_{\text{vdw}}^2 + 8r_p r_{\text{vdw}} + 4r_p^2 - d^2)^{1/2}. \quad (5)$$

If we require that $\eta(\mathbf{x}_{\text{mid}}) = 0.5$ and solve for the square of spread, s^2 , then

$$s^2 = \frac{2r_p^2 + 2r_{\text{vdw}} r_p - r_p[4(r_{\text{vdw}} + r_p)^2 - d^2]^{1/2}}{2 \ln(2) r_{\text{vdw}}^2}. \quad (6)$$

The half-distance, $d/2$, can range from the bonding radius, r_{bond} , through the van der Waals radius, r_{vdw} , to the distance at which the probe atom just touches the line between the two atom centers, $r_{\text{max}} = (r_{\text{vdw}}^2 + 2r_{\text{vdw}} r_p)^{1/2}$. The resulting spreads, s_{bond} , s_{vdw} and s_{max} , for these three half-distances and for probe radii of 1.4, 1.6 and 1.8 Å, respectively, are shown in Table 1. For the purpose of identification of surface atoms and residues, use of one of the narrower spreads is sufficient and results in a faster algorithm at the expense of the quality of the surface over larger gaps. The risk of identification of core atoms as surface atoms is minimal owing to the bias against the surface contour dipping into the interior with many overlapping pseudo-Gaussians. Figs. 2 and 3 show the strong similarity between the pseudo-Gaussian contour and the true Lee–Richards surface. In most cases it takes much less time to generate the pseudo-Gaussian contour than to generate a true Lee–Richards surface (see the discussion of *Performance of the approximation* below).

4. The PGALRS algorithm

The PGALRS algorithm can then be summarized as follows:

(1) Choose the value of d (default $2r_{\text{vdw}}$) to be used in calculation of the square of the spread, s^2 (default s_{vdw}^2), according to formula (6).

(2) For each atom, compute the pseudo-Gaussian density $\eta(\mathbf{x})$ according to formula (4) and sum over a uniformly spaced grid. Note that grid points more than 6σ (for double precision, 15+ digits), 4.5σ (for single precision, 8+ digits) or 3σ (for 3+ digits) from the atom center need not be considered.

(3) Contour the grid at level 1.

(4) Select all atoms within slightly more than an atomic radius of a grid point as surface atoms.

The next steps to be taken depend on the use to which the surface will be placed. If an area is needed, the resulting surface atoms can then be processed with MSMS to compute an analytic surface area. If a rendering is needed, a very good approximation to a Lee–Richards surface can be obtained from the superposition of the contour as a surface with the van der Waals surface of the selected atoms. If an even better rendering of an approximation to a Lee–Richards surface is required then a full Lee–Richards surface rendering of the selected atoms can be performed.

One important use of the PGALRS algorithm is to identify the core of a macromolecule, the portion of the molecule that is screened from solvent by the surface atoms. In the first application of the

Figure 2

Each column shows three artificial triangles of three atoms, with the two atoms in the base of each triangle far apart at the top, just touching in the middle and overlapping at the bottom. The left column shows the pseudo-Gaussian approximation to a Lee–Richards surface in grey. The right column shows an accurate rolling-ball rendering in yellow. In most cases it takes much less time to generate the pseudo-Gaussian approximation than to generate the Lee–Richards surface.

Figure 3

Molecular surface for PDB entry 4ins created from pseudo-Gaussian approximation to the Lee–Richards surface on the left and from a high-resolution pair-wise rolling-ball Lee–Richards surface on the right. The image on the right is identical to the result obtained by rolling a ball on just the 618 atoms identified as surface atoms out of 831 atoms total in the pseudo-Gaussian approximation. Hydrophobic residues are colored red in both versions. The branch points for alternate conformers are marked in light green in the rolling-ball model. The white regions are not hydrophobic.

Table 1

Pseudo-Gaussian spreads to match Lee–Richards surface points for specified probe radii, r_{probe} , and element types at atom center half-distances r_{bond} , r_{vdw} and r_{max} .

r_{probe}	Element	r_{bond}	r_{vdw}	r_{max}	s_{bond}	s_{vdw}	s_{max}
1.4 Å	C	0.68	1.7	2.77	0.23	0.596	1.09
	N	0.68	1.55	2.6	0.258	0.608	1.142
	O	0.68	1.52	2.56	0.265	0.611	1.153
	P	1.05	1.8	2.88	0.332	0.588	1.059
	Fe	1.34	2	3.1	0.373	0.573	1.005
1.6 Å	C	0.68	1.7	2.89	0.238	0.614	1.165
	N	0.68	1.55	2.71	0.267	0.626	1.22
	O	0.68	1.52	2.68	0.274	0.628	1.232
	P	1.05	1.8	3	0.344	0.606	1.132
	Fe	1.34	2	3.22	0.386	0.592	1.074
1.8 Å	C	0.68	1.7	3	0.245	0.629	1.236
	N	0.68	1.55	2.83	0.275	0.641	1.294
	O	0.68	1.52	2.79	0.281	0.643	1.307
	P	1.05	1.8	3.12	0.354	0.622	1.201
	Fe	1.34	2	3.35	0.398	0.608	1.139

PGALRS algorithm all non-solvent atoms, A , are used to allow the set, S , of surface atoms to be selected. The core, C , is then defined by removing the surface atoms from the set of non-solvent atoms, $C = A - S$. The boundary of the core can be visualized by applying PGALRS to C . The results of such a calculation are shown in Fig. 4.

5. PGALRS implementation and initial trials

The PGALRS algorithm has been implemented using *RasMol* (Sayle & Milner-White, 1995; Bernstein, 2000) as a testbed. The algorithm is portable and open source and can be adapted to almost any molecular graphics engine. By modeling electron density with unphysical pseudo-Gaussian atoms, the Lee–Richards surface can be approximated by a contour level of that density in time approximately linear in the number of atoms. The spacing of the contours determines the quality of the approximation. In the *RasMol* test version

```
map resolution 1.4
map resolution 1.0
map resolution 0.7
```

specify grid spacings of 1.4, 1.0 and 0.7 Å, respectively.

Figure 4

Stereo (cross-eyed) illustration of a slice through the molecular surface and the boundary of the internal core for PDB entry 4ins with both sets of atoms identified by PGALRS.

Table 2

Comparison of *MSMS* surface areas for the atoms identified as surface atoms and for all the non-water atoms for PDB entries 4ins (831 non-water atoms), 1w0k (22 826 atoms) and 1n32 (52 275 atoms) for surface atoms within 1.85 and 2.1 Å of the contour.

1.85 Å search.

Entry	Resolution	N_{surf}	A_{surf}	N_{all}	A_{all}	% error
4ins	1.4 Å	583	5185	831	5120	1.27%
	1.0 Å	609	5141	831	5120	0.41%
	0.7 Å	625	5138	831	5120	0.35%
1w0k	1.4 Å	15628	106956	22826	105855	0.88%
	1.0 Å	16407	106577	22826	105855	0.53%
	0.7 Å	16793	106510	22826	105855	0.46%
1n32	1.4 Å	32625	281345	52275	275015	2.30%
	1.0 Å	34298	278873	52275	275035	1.40%
	0.7 Å	35324	277824	52275	275045	1.02%

2.1 Å search.

Entry	Resolution	N_{surf}	A_{surf}	N_{all}	A_{all}	% error
4ins	1.4 Å	721	5127	831	5120	0.14%
	1.0 Å	777	5121	83	5120	0.02%
	0.7 Å	798	5121	831	5120	0.01%
1w0k	1.4 Å	19762	106251	22826	105855	0.22%
	1.0 Å	21078	106095	22826	105855	0.07%
	0.7 Å	21752	106072	22826	105855	0.05%
1n32	1.4 Å	40211	276313	52275	275015	0.47%
	1.0 Å	43019	275674	52275	275035	0.23%
	0.7 Å	44925	275574	52275	275045	0.20%

The command used in the *RasMol* test version to generate the Gaussians and plot points on a contour at level 1 is

```
map generate lrsurf dots
```

That provides enough information about the contour to allow atoms close to the contour to be identified. The 1.4 Å spacing is sufficiently fine for this purpose. Alternatively, if one of the commands

```
map generate lrsurf mesh
map generate lrsurf surface
```

were used instead, combined with a 1.0 or 0.7 Å grid spacing for a smoother image, then a servicable approximation to a Lee–Richards surface would have been generated at very low cost, and, for most purposes, such as visualization of a surface, we could stop the algorithm at this point. However, let us assume that a very accurate surface is required.

Having the contour level, the atoms and residues closest to that surface can be identified in average time $O[n^{2/3} \log(n)]$ using a *NearTree*-based nearest neighbor search, applying L. C. Andrews' approach. The command used in the *RasMol* test version to identify surface atoms is

```
map select atom 1.85 within
```

to select the atoms with centers within 1.85 Å of points on the contour. For most macromolecules, search ranges from 1.85 to 2.1 Å seem to be sufficient to pick up all surface atoms. The 'within' modifier ensures that solvent atoms will not be picked up and added to the list.

In the *RasMol* testbed a high-quality Lee–Richards surface for a 1.4 Å probe molecule radius can then be generated with the command

```
surface molecule 1.4
```

Table 3

Deviation of the pseudo-Gaussian contour from the Lee–Richards surface for the atoms for PDB entries 4ins (831 non-water atoms), 1w0k (22 826 atoms) and 1n32 (52 275 atoms).

Entry	Resolution	Deviation			s.u.
		Mean	Minimum	Maximum	
4ins	1.4 Å	−0.04	−1.13	0.64	0.26
	1.0 Å	−0.02	−1.18	0.72	0.26
	0.7 Å	−0.02	−1.18	0.73	0.26
1w0k	1.4 Å	0.05	−1.29	0.93	0.28
	1.0 Å	−0.04	−1.25	0.83	0.27
	0.7 Å	−0.04	−1.26	0.83	0.27
1n32	1.4 Å	0.04	−1.23	1.05	0.24
	1.0 Å	0.05	−1.25	0.92	0.24
	0.7 Å	0.06	−1.26	0.88	0.24

Alternatively, the selected atoms can be exported for processing of the surface by *MSMS* or *PyMOL* (DeLano, 2002).

The *PGALRS* algorithm in the *RasMol* testbed was compared with the Connolly surface code in *PyMOL* and with Sanner *et al.*'s *MSMS* on a 2.1 GHz MacBook Intel Core 2 Duo, using various renderings of the Protein Data Bank (PDB; Bernstein *et al.*, 1977; Berman *et al.*, 2000) entries 4ins (Baker *et al.*, 1988), 1w0k (Kagawa *et al.*, 2004), 1n32 (Ogle *et al.*, 2002) and 1jj2 (Klein *et al.*, 2001). 4ins is a 'small' macromolecule with 831 non-water atoms. 1w0k is a larger macromolecule with 22 826 non-water atoms. 1n32 is a large macromolecule with 52 275 non-water atoms. 1jj2 is one of the largest molecules in the PDB with 90 650 non-water atoms. Various test cases were run on these four molecules with the *RasMol PGALRS* testbed, *MSMS* and *PyMOL*.

5.1. Accuracy of the approximation

The *PGALRS* algorithm allows accurate calculation of the surface areas of molecules with more than 50 000 atoms. The accuracy of the approximation was tested in two ways.

In the first test the *MSMS* analytic solvent-excluded surface-area value was treated as the definitive measure of molecular surface area. The *MSMS* analytic solvent-excluded surface area for the subset of atoms closest to the pseudo-Gaussian contour was compared with the *MSMS* analytic solvent-excluded surface area for all non-water atoms. 1jj2 was too large for *MSMS* to complete either calculation. The comparison based on including atoms within 2.1 Å of the contour points for the other three macromolecules, 4ins, 1w0k and 1n32, is shown in Table 2. The reported areas were computed by *MSMS*, but in the N_{surf} A_{surf} columns the number of atoms and the *MSMS* surface area are reported for only the atoms identified as surface atoms by *PGALRS*, while in the N_{all} and A_{surf} columns the number of atoms and the *MSMS* surface area are reported for all non-solvent atoms. For a 1.85 Å search the errors in the computed area range from 2.30 down to 0.35%. A 2.1 Å search reduced the range of errors to less than 0.50%.

In the second test of accuracy, each sampling point of the contour was tested against the analytic Lee–Richards surface of the two nearest atoms belonging to the molecule. The mean, minimum and maximum deviations of the contour points from the analytic Lee–Richards surface are shown in Table 3. The mean deviation is less than 0.10 Å in all cases. The contours dip as much as 1.29 Å into the Lee–Richards surface and rise as much as 1.05 Å above the Lee–Richards surface, with an estimated standard deviation of less than 0.30 Å in all cases. The results reported here were for a 6σ cutoff in

summing pseudo-Gaussians. The 4.5σ cutoff produced essentially the same results.

5.2. Performance of the approximation

The performance of the *PGALRS* algorithm was compared in two ways with the performance of *MSMS*, and, to a limited extent, with the performance of *PyMOL*'s implementation of Connolly's algorithm. *PyMOL* cannot generate a Lee–Richards surface for the largest macromolecules. Therefore an approximation similar to *PGALRS*, but without the scaling from ρ to η , is used in *PyMOL* to handle the largest macromolecules.

In the first test, two approaches to the computation of an *MSMS* triangulation of the molecular surface at a density of 2 points per $\text{\AA}^2$ were compared:

- (1) compute the triangulation directly with *MSMS* using all atoms and a 1.4 $\text{\AA}$ probe radius; *versus*
- (2) find the pseudo-Gaussian density contour on a 1.4 $\text{\AA}$ grid spacing, identify the atoms within 2.1 $\text{\AA}$, export that subset of atoms to *MSMS* and compute the triangulation on just that subset.

This comparison was tested on subsets of sizes of approximately 5000 atoms through 50 000 atoms taken from 1jj2 and 1n32. The results are shown in Fig. 5. The *PGALRS* time for generation of the density and selection of the atoms near the contour is included. The lines marked *MSMS* are purely *MSMS* calculations, while the lines marked *PGALRS+MSMS* are the combined calculation. Note that the currently available version of *MSMS* is unable to handle more than 50 000 atoms, while using *PGALRS* followed by *MSMS* (marked as *PGALRS+MSMS*) allows *MSMS* to produce a surface for 65 000 atoms because *PGALRS* significantly reduces the number of atoms input into *MSMS*.

In the second test, the times for rendering of molecular surfaces in three different ways were compared.

- (1) The *PGALRS* algorithm was applied using a 1.4 $\text{\AA}$ grid spacing, a 4.5σ Gaussian cutoff and a search range of 2.1 $\text{\AA}$ for surface atoms. The resulting contour was rendered as a triangle-based surface and superimposed on the van der Waals surface of the same atoms (*PGALRS+CPK*).
- (2) The *PGALRS* algorithm was applied using a 1.4 $\text{\AA}$ grid spacing, a 4.5σ Gaussian cutoff and a search range of 2.1 $\text{\AA}$ for surface atoms. After the surface atoms were identified, the contour was discarded and a true Lee–Richards surface rendered on the surface atoms (*PGALRS+MS*).

Figure 5

Comparison of CPU times for triangulation of surfaces of subsets of atoms from PDB entries 1jj2 and 1n32 in steps of 5000 atoms, using *MSMS* on all atoms (solid lines) and using *MSMS* only on the atoms within 2.1 $\text{\AA}$ of the level 1 contour of the *PGALRS* pseudo-Gaussian density (dashed lines).

Table 4

Comparison of time to calculate a surface and divide it into triangles for the *RasMol PGALRS* testbed with a 3σ , 4.5σ and 6σ cutoff, *MSMS*, and *PyMOL*'s contoured density approximation of a Lee–Richards surface and *PyMOL*'s implementation of Connolly's algorithm for PDB entries 4ins (832 non-water atoms), 1w0k (22 663 atoms), 1n32 (52 053 atoms) and 1jj2 (90 418 atoms).

Entry	Resolution	3σ	4.5σ	6σ	<i>MSMS</i>	<i>PyMOL</i>
4ins	1.4 $\text{\AA}$	0.07	0.12	0.19	0.23	1.0
	1.0 $\text{\AA}$	0.15	0.23	0.41	0.25	1.1
	0.7 $\text{\AA}$	0.36	0.53	0.97	0.3	1.2
	Hi-res	0.50			13.3	
1w0k	1.4 $\text{\AA}$	1.9	2.4	6.4	3.2	2.7
	1.0 $\text{\AA}$	4.5	5.6	13.6	3.6	5.1
	0.7 $\text{\AA}$	11.9	14.7	32.4	4.5	11.0
	Hi-res	19.0			566.2	
1n32	1.4 $\text{\AA}$	5.1	6.2	12.3	23.0	5.6
	1.0 $\text{\AA}$	12.7	15.4	28.2	24.3	11.6
	0.7 $\text{\AA}$	34.7	41.4	70.6	26.6	27.9
	Hi-res	40.6				
1jj2	1.4 $\text{\AA}$	6.2	14.1	20.1	N/A	8.3
	1.0 $\text{\AA}$	14.7	25.1	45.6	N/A	17.2
	0.7 $\text{\AA}$	38.0	55.1	111	N/A	39.8
	Hi-res	68.73				

- (3) A true Lee–Richards surface was rendered on all atoms (*MS* alone).

The comparison was tested on subsets of approximately 5000 atoms through 90 000 atoms selected from 1jj2. The results are shown in Fig. 6. The ratios of performance for (*PGALRS+CPK*):(*PGALRS+MS*):(*MS*) were approximately 3:2:1.

An attempt was made to use *PyMOL*'s implementation of Connolly's algorithm in the comparison, but *PyMOL* was not able to complete execution for the larger molecular fragments and molecules. Therefore, following Vertrees' suggestion, a comparison was made between *PGALRS* alone in *RasMol* as an approximation to a molecular surface and a contour of a generated electron density in *PyMOL*, and then both were compared with the surface triangulation time in *MSMS*. In all cases, the setup time for inputting the coordinates of the molecule was subtracted. Table 4 shows the results for the *RasMol* testbed for a 3σ , 4.5σ and 6σ cutoff, *MSMS*, and *PyMOL* for triangulations at 1.4, 1.0 and 0.7 $\text{\AA}$ resolution, and, in the cases where a full high-resolution surface could be calculated, the 'hi-res' times.

The *RasMol PGALRS* 3σ cutoff times and the *PyMOL* density approximation times are very similar, which is to be expected inasmuch as they are both calculating and contouring densities, differing

Figure 6

Comparison of CPU times for calculation of rendering information for selections of subsets of atoms from PDB entry 1jj2 in steps of 5000 atoms, using *PGALRS+CPK*, *PGALRS+MS* and *MS* alone to render molecular surfaces.

primarily in the choices of scaling and spreads for the individual Gaussians. The *PyMOL* density approximation times were faster than the *PGALRS* 4.5σ and 6σ cutoff times, except for the smallest test case, 4ins. This suggests that the 3σ cutoff should be considered for use of *PGALRS* for very large macromolecules.

6. Future plans

Doucette *et al.* (2009) are reviewing the accuracy of the pseudo-Gaussian-based algorithm on a significant random subset of protein structures from the PDB. Rosa *et al.* (2009) are exploring the division of the problem into one of per-residue molecular surface evaluation, in which the van der Waals surfaces dominate, and one of inter-residue reentrant surface evaluation. Rosa *et al.*'s approach should provide major speed improvements by itself, and, when coupled with the identification of surface residues by use of the pseudo-Gaussian-based algorithm, should allow further speed improvements.

The identification of an atom or even of a residue as being on a surface of a domain in a complex is not completely specified by the three-dimensional coordinates as determined by a crystallographic study of an assembled complex and even less so in assemblies created computationally from independently determined domain structures. Therefore, surface identification algorithms need to be adjustable and the computations may need to be repeated. For a Lee–Richards algorithm, control is provided by adjusting the radii used as van der Waals radii of molecular atoms and by adjusting the radius of the probe atom. For the pseudo-Gaussian contour approach used in this project, the adjustments are achieved by changing the choice of heights and spreads of the pseudo-Gaussians for various atom types. Future investigations will explore the trade-offs involved in those choices.

The *PGALRS* algorithm will be implemented in *PyMOL*.

The authors acknowledge the invaluable assistance of Frances C. Bernstein. This work has been supported in part by NIH NIGMS grants 1R15GM078077-01 and 3R15GM078077-01S, DOE grant ER63601-1021466-0009501, and earlier NSF grants. The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies. The authors gratefully acknowledge the assistance of current and former students who have worked at Dowling College and at RIT on the SBEVSL project. The people at Dowling College working with HJB have been Isaac Awuah Asiamah, Darina Boycheva, Georgi Darakev, Nikolay Darakev, John Jemilawon, Nan Jia, Petko Kamburov, Greg McQuillan, Daniel O'Brien, Georgi Todorov and Elena Zlateva. The people at RIT working with PAC have been Eno Akpovwa, Abdul

Bangura, Anthony Corbett, Luticha Doucette, Chanelle Frances, Brett Hanson, Katrina Henry, MaryEd Kenney, Desirée Matthews, Scott E. Mottarella, Mario Rosa, Charles Westin and Corey Wischmeyer.

References

- Andrews, L. (2001). *C/C++ Users J.* **19**, 40–49.
- Baker, E. *et al.* (1988). *Philos. Trans. Biol. Sci.* **319**, 369–456.
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.
- Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F. Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *J. Mol. Biol.* **112**, 535–542.
- Bernstein, H. J. (2000). *Trends Biochem. Sci.* **25**, 453–455.
- Bernstein, H. J., Awuah Asiamah, I., Darakev, G., Darakev, N., Jemilawon, J., Jia, N., Kamburov, P., McQuillan, G. & Todorov, G. (2008). 2008 Meeting of the American Crystallographic Association, 31 May–5 June 2008, Knoxville, TN, USA.
- Bystroff, C. (2002). *Protein Eng.* **15**, 959–965.
- Connolly, M. L. (1983). *Science*, **221**, 709–713.
- Craig, P. A., Hanson, B., Westin, C., Wischmeyer, C. & Bernstein, H. (2009). American Crystallographic Association Meeting, 25–30 July 2009, Toronto, Canada.
- DeLano, W. L. (2002). *The PyMol Molecular Graphics System*, <http://www.pymol.org>.
- Doucette, L., Matthews, D., Craig, P. A. & Bernstein, H. J. (2009). Rochester Institute of Technology 18th Annual Undergraduate Research Symposium, 7 August 2009, Rochester Institute of Technology, NY, USA.
- Im, W., Beglov, D. & Roux, B. (1998). *Comput. Phys. Commun.* **111**, 59–75.
- Ivanov, P. S., Jain, P., Osei-Kuffuor, D., Stanev, V. S., Tripathi, R., Zhivkov, P. N. & Bernstein, H. J. (2003). Proceedings of the National Conference of Undergraduate Research, 13–15 March 2003, University of Utah, Salt Lake City, UT, USA.
- Jemilawon, J., Awuah Asiamah, I., Bernstein, H. J., Darakev, G., Darakev, N. & Kamburov, P. (2007). 9th International Conference on Biology and Synchrotron Radiation, 13–17 August 2007, Manchester, UK.
- Kagawa, R., Montgomery, M., Braig, K., Leslie, A. & Walker, J. (2004). *EMBO J.* **23**, 2734–2744.
- Katz, L. & Levinthal, C. (1972). *Annu. Rev. Biophys. Bioeng.* **1**, 465–504.
- Klein, D., Schmeing, T., Moore, P. & Steitz, T. (2001). *EMBO J.* **20**, 4214–4221.
- Le Grand, S. & Merz, K. (1993). *J. Comput. Chem.* **14**, 349–352.
- Lee, B. & Richards, F. M. (1971). *J. Mol. Biol.* **55**, 379–400.
- Lee, M., Feig, M., Salsbury, F. & Brooks, C. (2003). *J. Comput. Chem.* **24**, 1348–1356.
- Liang, J., Edelsbrunner, H. & Woodward, C. (1998). *Protein Sci.* **7**, 1884–1897.
- Ogle, J., Murphy, F., Tarry, M. & Ramakrishnan, V. (2002). *Cell*, **111**, 721–732.
- Rosa, M., Craig, P. A. & Bernstein, H. J. (2009). Rochester Institute of Technology 18th Annual Undergraduate Research Symposium, 7 August 2009, Rochester Institute of Technology, NY, USA.
- Sanner, M. F., Olson, A. J. & Spehner, J. C. (1996). *Biopolymers*, **38**, 305–320.
- Sayle, R. A. & Milner-White, E. J. (1995). *Trends Biochem. Sci.* **20**, 374–376.
- Vertrees, J. (2008). *Huge surfaces*, http://pymolwiki.org/index.php/Huge_surfaces.
- Yeates, T. (1995). *J. Mol. Biol.* **249**, 804–815.