Identifying interfacial molecules in nonplanar interfaces: the generalized ITIM algorithm

Marcello Sega*
Tor Vergata University of Rome, via della Ricerca scientifica 1, I-00133 Rome, Italy

Sofia S. Kantorovich
Sapienza University of Rome, p.le A. Moro 4, I-00188 Rome, Italy and
Ural Federal University, Lenin Ave. 51, 620083 Ekaterinburg, Russia

Pál Jedlovszky
Laboratory of Interfaces and Nanosize Systems, Institute of Chemistry,
Eötvös Loránd University, Pázmány stny 1/a, H-1117 Budapest, Hungary
MTA-BME Research Group of Technical Analytical Chemistry. Szt. Gellért tér 4, H-1111 Budapest, Hungary and
EKF Department of Chemistry, H-3300 Eger, Leányka u. 6, Hungary

Miguel Jorge
LSRE/LCM-Laboratory of Separation and Reaction Engineering,
Faculdade de Engenharia, Universidade do Porto,
Rua Dr. Roberto Frias, 4200-465 Porto, Portugal

We present a generalized version of the ITIM algorithm for the identification of interfacial molecules, which is able to treat arbitrarily shaped interfaces. The algorithm exploits the similarities between the concept of probe sphere used in ITIM and the circumsphere criterion used in the α-shapes approach, and can be regarded either as a reference-frame independent version of the former, or as an extended version of the latter that includes the atomic excluded volume. The new algorithm is applied to compute the intrinsic orientational order parameters of water around a DPC and a cholic acid micelle in aqueous environment, and to the identification of solvent-reachable sites in four model structures for soot. The additional algorithm introduced for the calculation of intrinsic density profiles in arbitrary geometries proved to be extremely useful also for planar interfaces, as it allows to solve the paradox of smeared intrinsic profiles far from the interface.

I. INTRODUCTION

Capillary waves represent a conceptual problem for the interpretation of the properties of liquid-liquid or liquid-vapor planar interfaces, because long-wave fluctuations are smearing the density profile across the interface and all other quantities associated to it. This is usually overcome by calculating the density profile using a local, instantaneous reference frame located at the interface, commonly referred to as the intrinsic density profile, $\rho(z) = \left\langle A^{-1} \sum_i \delta \left(z - z_i + \xi(x_i, y_i) \right) \right\rangle$, where (x_i, y_i, z_i) is the position of the i-th atom or molecule, and the local elevation of the surface is $\xi(x_i, y_i)$, assuming the macroscopic surface normal being aligned with the Z axis of a simulation box with cross section area A. During the last decade several numerical methods have been proposed to compute the intrinsic density profiles at interfaces[1–6]. Despite several differences in these approaches, they are, in general, providing consistent distributions of interfacial atoms or molecules[6] and density profiles[7]. Among these methods, ITIM[4] proved to be an excellent compromise between computational cost and accuracy[6], but it is limited to macroscopically flat interfaces, therefore there is a need to generalize it to arbitrary interfacial shapes.

Before these works, albeit for other purposes, several surface-recognition algorithms have been devised, and will be briefly mentioned below. All of them are possible starting points for the sought generalization under the condition that, once applied to the special case of a planar interface, they lead to consistent results with existing algorithms for the determination of intrinsic profiles.

Historically, the first class of algorithms addressing the problem of identifying surfaces was developed to determine molecular areas and volumes. The study of solvation properties of molecules and macromolecules (usually, proteins) might require the identification of molecular pockets, or the calculation of the solvent-accessible surface area for implicit solvation models[8]. Two intuitive concepts are commonly used to describe the surface properties of molecules, namely, that of solvent-accessible surface[9,10] (SAS), and that of molecular surface[11,12] (MS, also known as solvent excluded surface, or Connolly surface). The MS can be thought as the surface obtained by letting a hard sphere roll at close contact with the atoms of the molecule, to generate a smooth surface made of a connection of pieces of spheres and tori, which represents the part of the van der Waals surface exposed to the solvent. During the process of determining the surface, interfacial atoms can be identified using a simple geometrical criterion. Many approximated[13–24] or analytical[11,12,25–30] methods have been developed to compute the MS or the SAS. In general, these methods are based on discretization or tessellation procedures, requiring therefore the determination of the geometrical

structure of the molecule. Other methods which allow to identify molecular surfaces include the approaches of Willard and Chandler[5] or the Circular Variance method of Mezei[31]. Incidentally, the way the MS is computed in the early work of Greer and Bush[15] resembles very closely the ITIM algorithm[4].

From the late 1970s, the problem of shape identification had started being addressed by a newly born discipline, computational geometry. In this different framework, several algorithms have been actively pursued to provide a workable definition of surface, and in particular the concept of α-shapes[32,33] showed direct implications for the determination of the molecular surfaces[34,35]. The approach based on α-shapes is particularly appealing due to its generality and ability to describe, besides the geometry, also the intermolecular topology of the system.

Prompted by the apparent similarities between the usage of the circumsphere in the alpha shapes and that of the probe sphere in the ITIM method, as we will describe in the next section, we investigated in more detail the connection between these two algorithms. As a result, we developed a generalized version of ITIM (GITIM) based on the α-shapes algorithm. The new GITIM method consistently reproduces the results of ITIM in the planar case while retaining the ability to describe arbitrarily shaped surfaces. To the best of our knowledge, the concept of α-shapes has been employed in the determination of intrinsic densities at fluid interfaces only once before, by Usabiaga and Duque[36], who also noticed the formal similarities betweeen the α-shapes algorithm and ITIM.

In the following we describe briefly the alpha shapes and the ITIM algorithms, explain in detail the generalization of the latter to arbitrarily shaped surfaces, and present several applications.

II. ALPHA SHAPES AND THE GENERALIZED ITIM ALGORITHM

The concept of α-shapes was introduced several decades ago by Edelsbrunner[32,33]. To date the method is applied in computer graphics application for digital shape sampling and processing, in pattern recognition algorithms and in structural molecular biology[37]. The starting point in the determination of the surface of a set of points in the α-shapes algorithm is the calculation of the Delaunay triangulation, one of the most fruitful concepts for computational geometry[38,39], which can be defined in several equivalent ways, for example, as the triangulation that maximizes the smallest angle of all triangles, or the triangulation of the centers of neighboring Voronoi cells. The idea behind the α-shapes algorithm is to perform a Delaunay triangulation of a set of points, and then generate the so-called α-complex from the union of all k-simplices (segments, triangles and tetrahedra, for the simplex dimension k=1,2 and 3, respectively), characterized by a k-circumsphere radius (which is the length of the segment, the radius of the circumcircle and the ra-
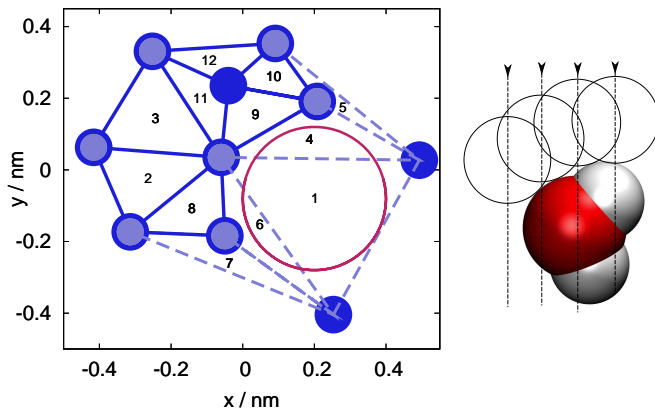

FIG. 1: Left: example of the α-shapes algorithm on a set of points on the plane. The lines connecting the points represent the Delaunay triangulation (the triangles are labeled by numbers from 1 to 12). Solid lines mark triangles belonging to the α-complex, and dashed lines those which are not. The lightshaded circles mark those points belonging to the α-shape, which is the border of the α-complex. Two points (in triangle 1) are outside the α-shape, and one (shared by triangles 9-12) is inside the α-shape. A circle with the radius of the probe disc α=0.2, with center in triangle 1, is also shown. Right: schematic representation of the ITIM algorithm, applied to a single water molecule: the probe spheres (circles) are moved down the test lines (dashed lines) until they touch an atom.

dius of the circumsphere for k=1,2 and 3, respectively) smaller than a given value, α (hence the name). The α-shape is then defined as the border of the α-complex, and is a polytope which can be, in general, concave, topologically disconnected, and composed of patches of triangles, strings of edges and even sets of isolated points. In a pictorial way, one can imagine the α-shape procedure as growing probe spheres at every point in space until they touch the nearest four atoms. These spheres will have, in general, different radii. Those atoms that are touched by spheres with radii larger than the predefined value α are considered to be at the surface.

An example of the result of the α-shapes algorithm in two dimensions is sketched in Fig. 1a. The ITIM algorithm is based instead on the idea of selecting those atoms of one phase that can be reached by a probe sphere with fixed radius streaming from the other phase along a straight line, perpendicular to the macroscopic surface. An atom is considered to be reached by the probe sphere if the two can come at a distance equal to the sum of the probe sphere and Lennard-Jones radii, and no other atom was touched before along the trajectory of the probe sphere. In practice, one selects a finite number of streamlines, and if the space between them is considerably smaller than the typical Lennard-Jones radius R_p, the result of the algorithm is practically independent of the location and density of the streamlines. The same is not true regarding the orientation of the streamlines; this is a direct consequence of the algorithm being designed
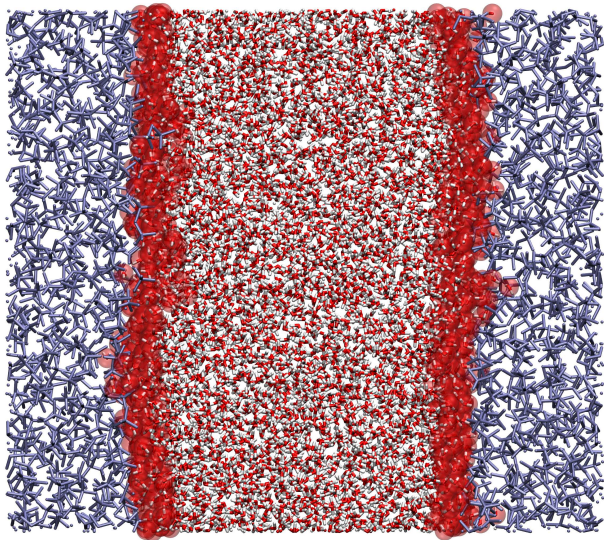
FIG. 2: Simulation snapshot of a H_2O/CCl_4 system. The oxygen atoms at the interface between the H_2O phase (inner) and CCl_4 phase (outer) as recognized by the GITIM algorithm are represented with an additional halo. Unconnected points belong to molecules which cross periodic boundary conditions.

for planar surfaces only. The basic idea behind the ITIM algorithm are sketched in Fig. 1b. A closer inspection reveals that the condition of being a surface atom for the ITIM algorithm resembles very much that of the α-shapes case. Quadruplets of surface atoms identified by the ITIM algorithm have the characteristic of sharing a common touching sphere having the same radius as the probe sphere. In this way, one can see the analogy with the α-shapes algorithm, the R_p parameter being used instead of α. The most important differences in the α-shapes algorithm with respect to ITIM are the absence of a volume associated with the atoms, and its independence from any reference frame. We devised, therefore, a variant of the α-shapes algorithm that takes into account the excluded volume of the atoms.

In the approach presented here the usual Delaunay triangulation is performed, but the α-complex is computed substituting the concept of the circumsphere radius with that of the radius of the touching sphere, thus introducing the excluded volume in the calculation of the α-complex. Note that this is different from other approaches that are trying to mimic the presence of excluded volume at a more fundamental level, like the weighted α-shapes algorithm, which uses the so-called regular triangulation instead of the Delaunay one[33]. In addition, in order to eliminate all those complexes, such as strings of segments or isolated points, which are rightful elements of the shape, but do not allow a satisfactory definition of a surface, the search for elements of the α-complex stops in our algorithm at the level of tetrahedra, and triangles and segments are not checked. In this sense GITIM can provide substantially different results from the original α-shapes

algorithm.

The equivalent of the α-complex is then realized by selecting the tetrahedra from the Delaunay triangulation whose touching sphere is smaller than a probe sphere of radius R_p, and the equivalent of the α-shape is just its border, as in the original α-shapes algorithm. The procedure to compute the touching sphere radius is described in the Appendix.

In the implementation presented here, in order to compute efficiently the Delaunay triangulation, we have made use of the quickhull algorithm, which takes advantage of the fact that a Delaunay triangulation in d dimensions can be obtained from the ridges of the lower convex hull in $d+1$ dimensions of the same set of points lifted to a paraboloid in the ancillary dimension[40]. The quickhull algorithm employed here[41] has the particularly advantageous scaling $\mathcal{O}(N \log(\nu))$ of its computing time with the number N and ν of input points and output vertices, respectively.

A separate issue is represented by the calculation of the intrinsic profiles (whether profiles of mass density or of any other quantity) as the distance of an atom in the phase of interest from the surface is not calculated as straightforwardly as in the respective non-intrinsic versions. For each atom in the phase, in fact, three atoms among the interfacial ones have to be identified in order to determine by triangulation[7] the instantaneous, local position of the interface. This issue will be discussed in Sec. III for the planar, for the spherical or quasi-spherical and for the general case: here we simply note that we turned down an early implementation of the algorithm that searches for these surface atoms, based on the sorting of the distances using $\mathcal{O}(N \log N)$ algorithms like quicksort, in favor of a better performing approach, based on kd-trees[42,43], a generalization of the one-dimensional binary tree, which are still built in a $\mathcal{O}(N \log N)$ time, but allow for range search in (typically) $\mathcal{O}(\log N)$ time.

III. COMPARISON BETWEEN THE ITIM AND THE GITIM METHODS

We have compared the results of the ITIM and GITIM algorithms applied to the water/carbon tetrachloride interface composed of 6626 water and 966 CCl_4 molecules. The water and CCl_4 molecules have been described by the TIP4P model[44], and by the potential of McDonald and coworkers[45], respectively. The molecules have been kept rigid using the SHAKE algorithm[46]. This simulation, as well as the others reported in this work have been performed using the Gromacs[47] simulation package employing an integration time step of 1 fs, periodic boundary conditions, a cutoff at 0.8 nm for Lennard-Jones interactions and the smooth Particle Mesh Ewald algorithm[48] for computing the electrostatic interaction, with a mesh spacing of 0.12 nm (also with a cut-off at 0.8 nm for the real-space part of the interaction). All simulations were performed in the canonical ensemble at a tem-
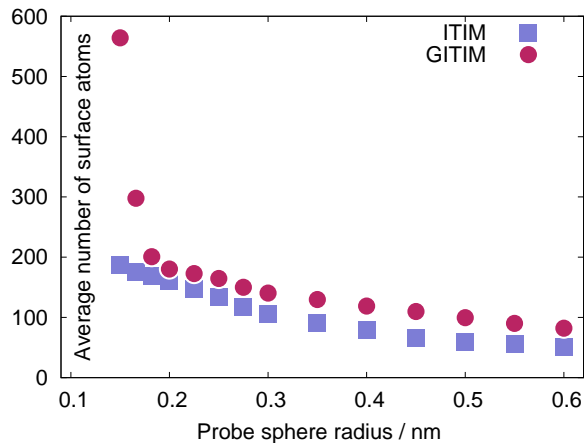
FIG. 3: Average number of surface atoms identified by ITIM (squares) and GITIM (circles) as a function of the probe sphere radius.

FIG. 4: Water surface oxygen atoms in the H_2O/CCl_4 system in one simulation snapshot as recognized by GITIM exclusively (small spheres), ITIM exclusively (large spheres) or by both methods (sphere with halo).

perature of 300K using the Nosé–Hoover thermostat[49,50] with a relaxation time of 0.1 ps. A simulation snapshot of the H_2O/CCl_4 interface is presented in Fig. 2, where the surface atoms identified by the GITIM algorithm using a probe sphere radius of 0.25 nm are highlighted using a spherical halo.

We have used the ITIM and GITIM algorithms to identify the interfacial atoms of the water phase in the system, for different sizes of the probe sphere. In general, GITIM identifies systematically a larger number of interfacial atoms than ITIM for the same value of the probe sphere radius R_p, as it is clearly seen in Fig. 3. Remarkably, for values of the probe sphere radius smaller than about 0.2 nm (compare, for example, with the optimal ITIM parameter $R_p = 0.125$ nm suggested in Ref. 6), the interfacial atoms identified by GITIM show the onset of percolation. The reason for this behavior traces back to the fact that ITIM is unable to identify voids buried in the middle of the phase, as it is effectively probing only the cross section of the voids along the direction of the streamlines. This difference could explain the higher number of surface atoms identified by GITIM, as voids in a region with high local curvature (or, in other words, with a local surface normal which deviates significantly from the macroscopic one) will not be identified as such by ITIM. In GITIM, on the contrary, probe spheres can be thought as inflating at every point in space instead of moving down the streamlines, and this is the reason why the algorithm is able to identify also small pockets inside the opposite phase.

It is possible to make a rough but enlightening analytical estimate of the probability for a probe sphere of null radius in the ITIM algorithm to penetrate for a distance ζ in a fluid of hard spheres with diameter σ and number density ρ. Using the very crude approximation of randomly distributed spheres, the probability p_0 to pass the first molecular layer, at a depth $\zeta = \sigma$ is the effective cross section $p_0 = 1 - \frac{\pi}{4}\rho^{2/3}\sigma^2$, and that
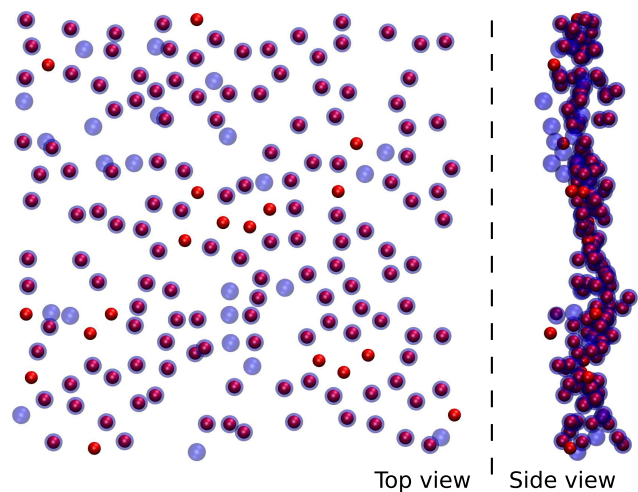
of reaching a generic depth ζ can be approximated as $p(\zeta) = p_0^{\zeta/\sigma}$, where $\kappa = \ln(1/p_0)/\sigma$ defines a penetration depth. Therefore, using a probe sphere with a null radius, ITIM will identify a (diffuse) surface at a depth $1/\kappa$, while GITIM will identify every atom as a surface one. For water at ambient conditions, the penetration is $\kappa^{-1} \simeq 0.186$ nm, a distance smaller than the size of a water molecule itself. This could explain why in Ref. 6, even using a probe sphere radius as small as 0.05 nm, almost only water molecules in the first layer were identified as interfacial ones by ITIM (see the almost perfectly Gaussian distribution of interfacial water molecules in Fig.9 of Ref. 6).

Nevertheless, it is important for practical reasons to be able to match the outcome of both algorithms. It turns out that choosing R_p so that the average number of interfacial atoms identified by both algorithms is roughly the same leads also, not surprisingly, to very similar distributions. The probe sphere radius required for GITIM to obtain a similar average number of surface atom as in ITIM can be obtained by an interpolation of the values reported in Fig. 3. An example showing explicitly the interfacial atoms identified by the two methods ($R_p = 0.2$ nm for ITIM and $R_p = 0.25$ nm for GITIM) is presented in Fig. 4: roughly 85% of surface atoms are identified simultaneously by both methods, demonstrating the good agreement between the two methods once the probe sphere radius has been re-gauged. The condition of identifying the same atoms as interfacial ones is much more strict that any condition on average quantities, like the spatial distribution of interfacial atoms or intrinsic density profiles. Hence, it is expected that a good agreement on such quantities can also be achieved.

The intrinsic density profiles of water and carbon tetrachloride are reported in Fig. 5, as computed by ITIM and
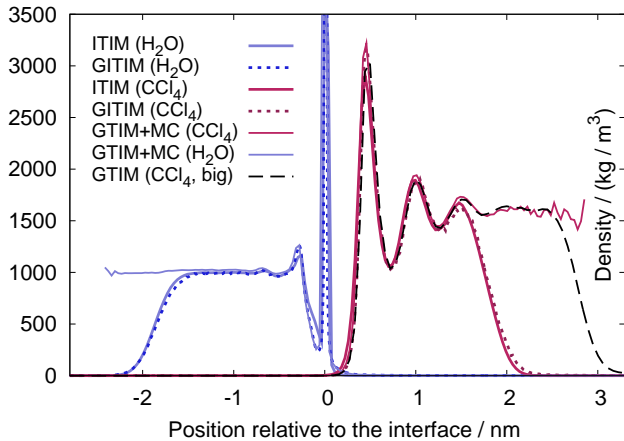
FIG. 5: Intrinsic density profiles of water (curves on the left) and carbon tetrachloride (curves on the right) with respect to the water surface as computed with ITIM (thick, solid lines) or with GITIM (thick, dotted lines). The profile computed using GITIM and the Monte Carlo normalization procedure described in Sec. IV are also shown (thin, solid lines), as well as the one for carbon tetrachloride computed in the bigger system (thin, dashed line) using GITIM and no Monte Carlo normalization.

GITIM, respectively, with the interfacial water molecules as reference. The procedure for identifying the local distance of an atom from the surface is in its essence the same as described in Ref. 7. Starting from the projection $P_0 = (x, y)$ of the position of the given atom to the macroscopic interface plane, the two interfacial atoms closest to P_0 are found (their position on the interface plane being P_1 and P_2, respectively). The third closest atom with projection P_3 has then to be found, with the condition that the triangle $P_1 P_2 P_3$ contains the point P_0. A linear interpolation of the elevation of P_0 from those of the other points is eventually performed, and employed to compute the distance $z - \xi(x, y)$ which is used to compute the intrinsic density profile. Efficient neighbor search for the P_1, P_2 and candidate P_3 atoms is implemented using kd-trees[43] as discussed before. The two pairs of profiles are very similar, besides a small difference in the position and height of the main peak of the CCl_4 profile (curves on the right in Fig. 5) and in the minimum of the water profile (curves on the left in Fig. 5) next to the surface position, which are anyway compatible with the differences observed between various methods for the calculation of intrinsic density profiles[7]. The delta-like contribution of the water molecules at the surface is included in the plot in Fig. 5, and defines the origin of the reference system. Negative values of the signed distance from the interface correspond to the aqueous phase.

IV. THE PROBLEM OF NORMALIZATION OF DENSITY PROFILES

Before applying GITIM to non-planar interfaces, one important issue has still to be solved, namely that of the proper calculation of intrinsic density profiles in non-planar geometries. In general, one uses one-dimensional density profiles (intrinsic or non-intrinsic) when the system is, or is assumed to be, invariant under displacements along the interface, so that the orthogonal degrees of freedom can be integrated out. When the interface has a non-planar shape, one needs to use a different coordinate system.

For the sake of simplicity we will refer now to the spherical or quasi-spherical case, but the following considerations apply to any other coordinate system. To compute the non-intrinsic density profile with respect to an object whose surface is fluctuating but is on average spherical, one can use the spherical coordinate system and normalize each bin by the integral of the Jacobian determinant, that is the volume of the shell at constant distance from the origin. In the intrinsic case, however, the istantaneous volume of the shell at constant distance from the intrinsic surface of the object is different from the spherical shell volume. Atoms at the same distance from the intrinsic surface might be associated to different spherical shells, with correspondingly different values of the Jacobian, thus introducing unwanted artefacts. An example of how this normalization affects the calculation of the density profile will be presented in Sec. V A.

To avoid these problems, one needs to provide a proper normalization by calculating at every frame the volume of shells at constant intrinsic distance. In principle, this could be calculated by ordinary numerical integration, but this would require a large computing time and storage overhead. Here, instead, we propose to employ an approach based on simple Monte Carlo integration: in parallel with the calculation of the histograms for the various phases, we compute also that of a random distribution of points, equal in number to the total atoms in the simulation. The volume of a shell can then be estimated as the ratio of the number of points found at a given distance and the total number of random points drawn, times the volume of the simulation box. We are following the heuristic idea that for each frame j one does not need to know the volume of the shell $V_j(r)$ with a precision higher than that of the average number of atoms in it, $n_j(r)$. In addition, we assume that the surface area of the interface is large enough for the shell volume variations $\delta V_j(r)$ to be small with respect to its average value $\hat{V}(r) = \sum_j^T V_j(r)/T$.

The average intrinsic number density profile

$$\rho(r) = \frac{1}{T} \sum_{j=1}^{T} \frac{n_j(r)}{V_j(r)} \tag{1}$$

can therefore be approximated using a Taylor expansion

as

$$\rho(r) \simeq \frac{1}{T} \frac{1}{\hat{V}(r)} \sum_{j=1}^{T} \left[n_j(r) - n_j(r) \frac{\delta V_j(r)}{\hat{V}(r)} \right]$$

$$= \frac{\hat{n}(r)}{\hat{V}(r)} \left[1 + \mathcal{O}\left(\delta V_j(r)/\hat{V}(r) \right) \right]. \quad (2)$$

When the relative volume changes $|\delta V/V|$ are small, one can therefore simply normalize the histogram $\hat{n}(r) = \sum_{j}^{T} n_j(r)/n$ by the average volume $\hat{V}(r)$ obtained by the Monte Carlo procedure, disregarding the terms of order $\mathcal{O}(\delta V/V)$.

The correctness of our assumption is demonstrated incidentally by the application of this normalization once again to the planar case. The thin lines in Fig. 5 represent the ITIM intrinsic mass density profile of water and carbon tetrachloride, using the Monte Carlo normalization scheme instead of the usual normalization with box cross sectional area and slab width. Close to the interface, the Monte Carlo normalization gives results which are fully compatible with the usual method, showing that the accuracy of the volume estimate is adequate. On the other hand one can see that far from the interface the two profiles behave quite differently.

The profile computed with usual normalization decays slowly to zero when approaching the distance of roughly half a box length. This is because the minimum image convention is used, in conjunction with periodic boundary conditions, to determine the distance of one particle from the (nearest) interface. The maximum distance from the surface that a point can attain depends therefore on the local thickness of the slab. Where the slab is thicker the maximum allowed distance is smaller than in regions where the slab is thinner. Because of this purely geometric artefact, particles at large distances (close to half a simulation box minus the average slab thickness) will become more unlikely with increasing distance of the point from the surface, leading to a smoothly vanishing density profile, even though on physical grounds one would expect a constant one.

The case with Monte Carlo normalization, on the contrary, shows that it is possible to reach the expected constant density profile at large distances. The use of Monte Carlo normalization appears to change some features of the profile, such as the heigth of the thrid peak at about 1.5 nm, and reveals new ones such as a small fourth peak around 2 nm. To check that the Monte Carlo normalization is revealing indeed physical properties and not some artefacts, we performed a new simulation with larger width of both the water and the carbon tetrachloride slabs (9746 and 1566 water and carbon tetrachloride molecules, respectively), and we calculated the density profile without the Monte Carlo normalization. The resulting profile agrees with the one computed with the aid of the Monte Carlo normalization in the smaller system (see 5), confirming the correctness of the Monte Carlo normalization and its ability to extract relevant information in poorly sampled regions.
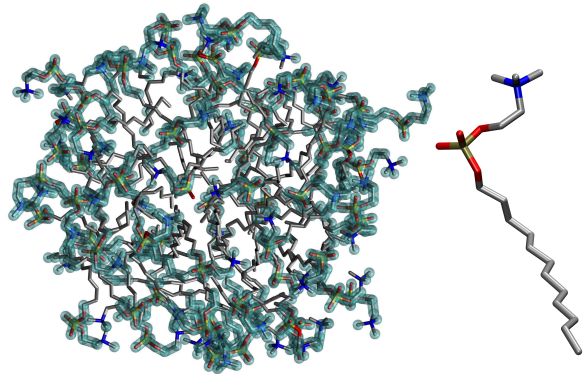

FIG. 6: Right: schematic structure of a DPC molecule. Left: snapshot of a DPC micelle in water. Only the DPC constituents are shown for the sake of clarity. Atoms with a halo are those recognized by GITIM as surface ones.

This shows that the use of a proper normalization is of fundamental importance also for macroscopically planar interfaces. The calculation of the Monte Carlo normalization factors does not change the typical scaling of the algorithm, as it consists in calculating the histogram for an additional phase of randomly distributed points (which effectively behaves as an ideal gas). Since by using the Monte Carlo Normalization smaller systems are required in order to extract the same information, the Monte Carlo normalization procedure could even be beneficial in terms of performance. In the following examples, therefore, the presented density profiles will always be the ones obtained using the Monte Carlo normalization, if not stated otherwise.

V. EXAMPLES OF NON-PLANAR INTERFACES

A. DPC micelle

Dodecylphosphocholine (DPC) is a neutral, amphiphilic molecule with a single fatty tail that can form micelles in solution: these play a relevant role in biochemistry, especially for NMR spectroscopy investigations aiming at understanding the structure of proteins or peptides bound to an environment that is similar to the biological membrane[51–54]. The molecular structure of DPC is shown in Fig.6. We have simulated for 500 ps a micelle of 65 DPC and 6305 water molecules using the force field and configurations from Tieleman and colleagues[55], and have calculated the intrinsic mass density profiles of both phases (DPC and water) using GITIM and the Monte Carlo normalization procedure, with a probe sphere radius $R_p = 0.25$. The result of the interfacial atoms identification on the DPC micelle for a single frame is shown in Fig. 6, where water molecules have been removed for the sake of clarity, and interfacial atoms are
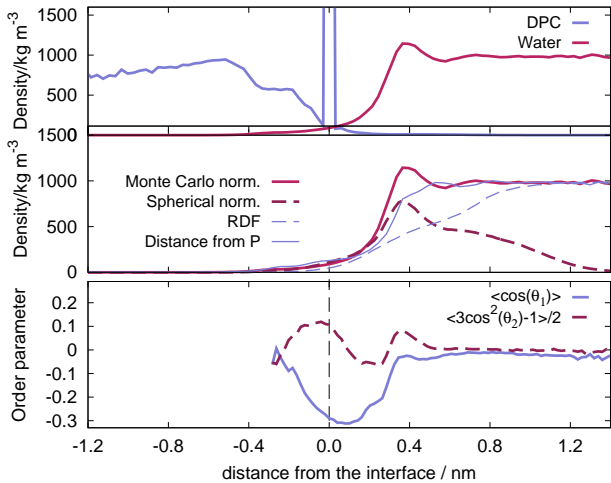
FIG. 7: Upper panel: intrinsic density profiles of water (right) and of DPC (left). Central panel: intrinsic density profile of water (thick, solid line) computed using GITIM and the Monte Carlo normalization; density profile obtained by weighting each count with the volume of the correponding spherical shell (thick, dashed line); intrinsic density profile obtained using the phosphorous atoms as surface points and the Monte Carlo normalization (thin, solid line) ; the radial distribution function of water with respect to the center of mass of the micelle. Lower panel: intrinsic profile of the orientational order parameters (S_1, solid line, S_2, dashed line). The vertical dashed lines marks the position of the interface.

highlighted as usual with a halo.

In order to compute the intrinsic density profile, the local distance from the surface of a generic atom located at $R_0 = (x, y, z)$ can be calculated in a way that resembles the planar case. Starting from the two surface atoms R_1 and R_2, closest to R_0, one can find a third surface atom R_3 so that the line connecting the center of mass of the micelle R_C and R_0 passes inside the triangle $R_1R_2R_3$. The distance from the surface can be then defined as the length of the segment uniting R_0 and the intersection of the line R_CR_0 with the triangle surface.

The intrinsic mass density profiles, calculated relative to the DPC surface, is reported in the top panel of Fig. 7, with the DPC mass density profile shown on the left, and the water profile on the right. As usual, the delta-like contribution at $r = 0$ identifies the contribution from interfacial DPC atoms. In order to show the importance of selecting the truly interfacial atoms, and the effect of the Monte Carlo normalization procedure, in the middle panel of Fig. 7 we reported the density profiles of water molecules calculated using three different approaches, namely, (a) the profile obtained by substituting the Monte Carlo normalization with a procedure that assign to each count in the histogram at intrinsic distance r the weigth $1/(4\pi|R_0 - R_C|^2)$ (thick, dashed line), (b) the intrinsic profile obtained by using the phosphorous atoms as surface atoms instead or relying on the GITIM approach (thin, solid line), and (c) the radial distribution function of water calculated with respect to the center of

mass of the DPC micelle (thin, dashed line).

The intrinsic profile computed using GITIM (the same data as in the upper panel, thick solid line) is also presented, for the sake of comparison. The procedure (a) is an attempt to normalize the histogram without making use of the Monte Carlo procedure, but adds, to the shortcomings seen in the planar case at large values of r, the problem of mixing absolute and intrinsic distances (points at different absolute distance $|R_0 - R_C|$ can have the same intrinsic distance r, and *viceversa*). The resulting profile does not show the expected features of a typical intrinsic density profile, namely, the saturation to the bulk density value far away from the susrface, and a density peak higher than the bulk value next to the surface. Exploiting the amphiphilic nature of the DPC molecules to identify *a priori* surface atoms (in this case, phosphorous atoms) instead of performing the GITIM procedure does also lead to a density profile which is does not reproduce the peak close to the intrinsic surface. Finally, the radial distribution function of water atoms with respect to R_C shows how markedly the correlations are suppressed by the surface fluctuations.

In addition to the density profiles, we have calculated the intrinsic profiles of the orientational order parameters S_1 and S_2 of the water molecules around the DPC micelle. The two parameters are defined as $S_1 = \langle \cos(\theta_1) \rangle$ and $S_2 = \langle 3\cos^2(\theta_2) - 1 \rangle /2$, where θ_1 and θ_2 are the angles between the water molecule position vector (with respect to the micelle center), and the water molecule symmetry axis and molecular plane normal, respectively. The orientation is taken so that $\theta_1 < \pi/2$ when the hydrogen atoms are farther from the micelle than the corresponding oxygen. The complete picture of the orientation of water molecules would be delivered by the calculation of the probability distribution $p(\theta_1, \theta_2)$[56,57], but here we limit our analysis to the two separate order parameters and their intrinsic profiles.

Note that, since these quantities are computed per particle, there is no need to apply any volume normalization. The polarization of water molecules, which is proportional to S_1, appears to be different from zero only very close to the micellar surface. In particular, S_1 has a correlation with the main peak of the intrinsic density profile in the proximity of 0.4 nm. Water molecules located closer to the interface show a first change in the sign of the polarization and a subsequent one when crossing the interface. Farther than 0.25 nm inside the micelle, not enough water molecules are found to generate any meaningful statistics. Also the S_2 order parameter is practically zero beyond 0.6 nm, and again a correlation is seen with the main peak of the intrinsic density profile, and the maximum in the orientational preference is found just next to the interface, where $S_1 \simeq 0$, showing that water molecules are preferentially laying parallel to the interfacial surface.
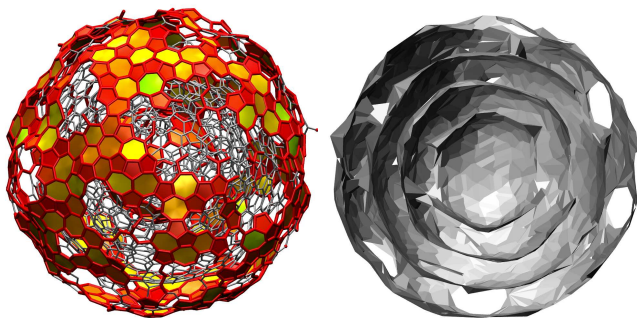
FIG. 8: The S_1^I soot model[58] represented in section (right, triangulated surface) and in whole (left, wireframe) with the atoms identified by GITIM as surface ones highlighted using thicker, red elements. Besides surface atoms, also chemical bonds between surface atoms are highlighted, as well as five, six and seven membered rings (filled surfaces).

B. Soot

One of the main byproducts of hydrocarbon flames, soot is thought to have a relevant impact on atmospheric chemistry and global surface warming[59,60]. Electron, UV, and atomic force microscopy have revealed the size and structure of soot particles from different sources at different scales[61–65]. In particular, soot emitted by aircraft is found to be made of several, quasi-spherical, concentric graphitic layers of size in the range from 5 to 50 nm[61]. We have used four model structures (S_1^I, S_2^I, S_4^I and S^{II} from Ref. 58) to demonstrate the ability of GITIM to identify surface atoms in complex geometries. In Fig. 8, the S_1^I model is represented in section as a triangulated surface (right), showing the four concentric layers, and in whole (left) showing the surface atoms as detected by GITIM using $R_p = 0.25$ nm. The histograms of the total number of atoms and of the surface ones, as a function of the distance from the center of the soot particles, are shown in Fig. 9 for the four different models, where it is seen how particles of the size of a water molecule have mostly access only to the inner and outer parts of the innermost and outermost shell, respectively, and cover them almost completely. This finding is in a clear accordance with the results of the void analysis and adsorption isotherm calculations presented in Ref. 58.

C. Secondary cholic acid micelle

Bile acids, such as cholic acid are biological amphiphiles built up by a steroid skeleton and side groups attached to it. The organization of these side groups is such that hydrophilic and hydrophobic groups are located at the two opposite sides of the steroid ring. Thus, bile acids have a hydrophilic and a hydrophobic face (often referred to as the α and β side, respectively) rather than a polar head and an apolar tail, as in the case of other surfactants like, for example, DPC. The unusual molec-
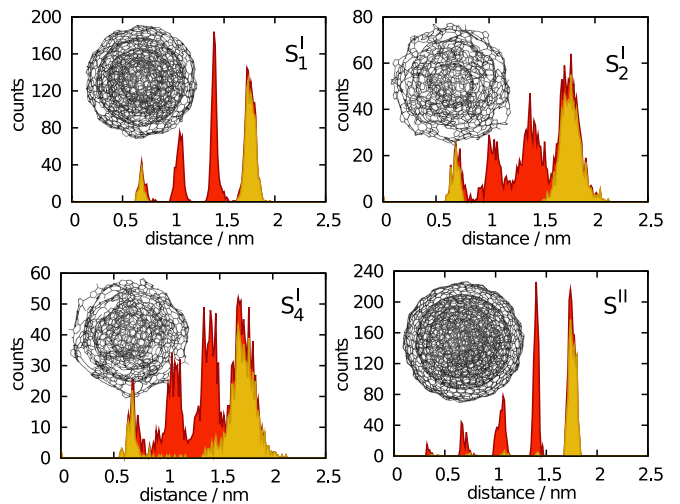

FIG. 9: Histograms of the atoms in the four soot models taken from Ref. 58. Each panel refers to a different structure (depicted with wireframe), and presents the distribution of all atoms (filled, darker area) and of surface atoms identified by GITIM (filled, lighter area), as a function of the distance from the center.

ular shape leads to peculiar aggregation behavior of bile acids. At relatively low concentrations they form regular micelles with an aggregation number of 2-10, while above a second critical micellar concentration these primary micelles form larger secondary aggregates by establishing hydrogen bonds between the hydrophilic surface groups of the primary micelles[66,67]. These secondary micelles are of rather irregular shape,[67,68] which makes them an excellent test system for our purposes.

Here we analyze the surface of a secondary cholic acid micelle composed of 35 molecules, extracted from a previous simulation work[67] and simulated for the present purposes for 500 ps in aqueous environment. An instantaneous snapshot of the micelle is shown in Fig.10 (water molecules are omitted for clarity) together with a schematic structure of the cholic acid molecule.

We calculated the density profile of water as well as of cholic acid relative to the intrinsic surface of the micelle by the GITIM method. In the two previous examples (carbon tetrachloride/water mixture and DPC micelle), the definition of a macroscopic axis (the Z axis and the radial direction, respectively) was used to identify triplets of surface atoms and compute, in turn, the distance from the intrinsic surface. If the interface has not a simple shape, as for the cholic acid micelle, it seems natural to define the distance d of a point $R_0 = (x, y, z)$ from the surface as the minimal distance of R_0 from the triangular surface defined by the three surface atoms R_1, R_2 and R_3, closest to R_0. This means that if the projection of R_0 on the triangular plane falls within the triangle, then d is equal to the distance from the surface, otherwise, d is equal to the distance to the closest among the three atoms.

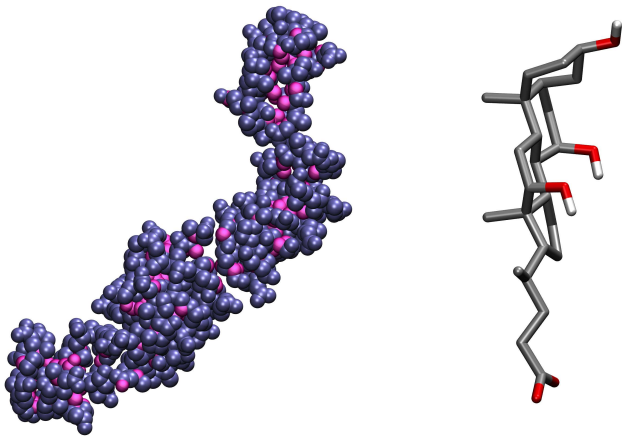The resulting profiles are shown in Fig. 11. The mi-

FIG. 10: Left: simulation snapshot of a secondary cholate micelle, with surface atoms highlighted. Right: the structure of the cholic acid molecule.
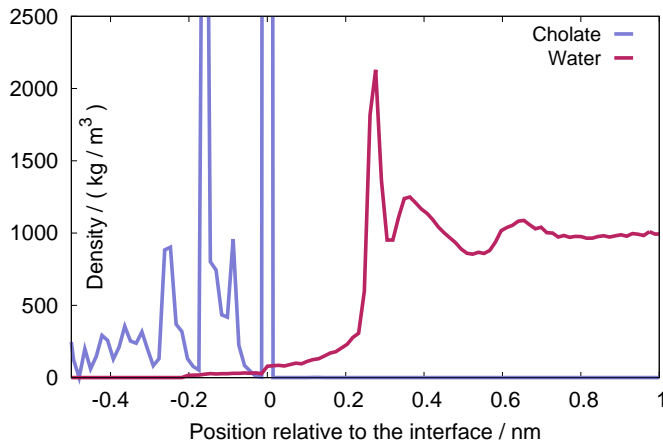

FIG. 11: Density profile of water (right) and cholic acid (left) in the secondary micelle.

celle has a characteristic elongated shape, which exposes a large part of its components to the solvent, so that roughly 80% of the micelle atoms are identified as surface ones. The small volume to surface ratio of the micelle is at the origin of the rather noisy intrinsic density profile for the micelle itself. The profile, in addition to the delta-like contribution at the surface, presents another very sharp peak located at a distance of about 0.18 nm inside the surface, due to the rather rigid structure of the bile molecule.

VI.   CONCLUSIONS

In this paper we presented a new algorithm that combines the advantageous features of both the ITIM method[4] and the α-shapes algorithm[32,33] to be used in determining the intrinsic surface in molecular simulations. Thus, unlike the original variant, this new, generalized version

of ITIM, dubbed GITIM, is able to treat interfaces of arbitrary shape and, at the same time, to take into account the excluded volumes of the atoms in the system. It should be emphasized that the GITIM algorithm is not only able to find the external surface of the phase of interest, but it also detects the surface of possible internal voids inside the phase. The method turned out to provide results which are practically identical to the original ITIM analysis for planar interfaces. Further, its applicability to non-planar interfaces was shown for three systems previously simulated, namely, a quasi-spherical micelle of DPC[55], several molecular models of soot[58], and a secondary micellar aggregate of irregular shape composed of cholic acid molecules[67].

In addition, we proposed how to compute density profiles relative to intrinsic interfaces, irrespective of the latter being macroscopically planar or not, by using a simple Monte Carlo-based algorithm, which allows to estimate the volume of slabs at constant distance from the intrinsic surface, and normalize correctly the density histograms. This issue has been shown to be relevant not only in presence of complex interfaces, but also for macroscopically flat ones, because of the combined effect of capillary waves and of the finite width of fluid slabs. We demonstrated that the artificial smearing of the intrinsic density profiles far from the intrinsic interface can be overcome, and that relevant physical information – completely masked without the proper normalization – can be successfully extracted by using this normalization.

Two computer programs that implement, respectively, an optimized version of ITIM and the new GITIM algorithm, as well as the calculation of intrinsic density and order parameters profiles, are made available free of charge at `http://www.gitim.eu/`. The programs are compatible with the trajectory and topology file formats of the Gromacs molecular simulation package[47].

VII.   APPENDIX

Here, following Ref. 70 we derive the expressions for the radius R and position $\mathbf{r} = (x, y, z)$ of the center of the sphere which is touching four other ones, having given

radii and center positions R_i and $\mathbf{r}_i = (x_i, y_i, z_i)$ ($i = 1, 2, 3$ or 4), respectively. The conditions of touching can be expressed with the following nonlinear system of four equations:

$$|\mathbf{r} - \mathbf{r}_i|^2 = (R + R_i)^2. \qquad (3)$$

By subtracting one of them from the other three (without loss of generality we subtract the one with $i = 1$), the quadratic term, $\mathbf{r}^2$, will be eliminated and the system Eq.(3) would become linear with respect to $\mathbf{r}$:

$$\mathbf{Mr} = \mathbf{s} - R\mathbf{d}, \qquad (4)$$

where the matrix $\mathbf{M}$ and the vectors $\mathbf{d}$ and $\mathbf{s}$ are defined as

$$\mathbf{M} = \begin{pmatrix} \mathbf{r}_1 - \mathbf{r}_2 \\ \mathbf{r}_1 - \mathbf{r}_3 \\ \mathbf{r}_1 - \mathbf{r}_4 \end{pmatrix}, \ \ \mathbf{d} = \begin{pmatrix} R_1 - R_2 \\ R_1 - R_3 \\ R_1 - R_4 \end{pmatrix}, \qquad (5)$$

and

$$\mathbf{s} = \frac{1}{2} \begin{pmatrix} \mathbf{r}_1^2 - \mathbf{r}_2^2 - R_1^2 + R_2^2 \\ \mathbf{r}_1^2 - \mathbf{r}_3^2 - R_1^2 + R_3^2 \\ \mathbf{r}_1^2 - \mathbf{r}_4^2 - R_1^2 + R_4^2 \end{pmatrix}. \qquad (6)$$

Equation (4) has a unique solution if matrix $\mathbf{M}$ is non-singular (the singularity of $\mathbf{M}$ corresponds to the case when all 4 spheres are co-planar, which means that the unknown sphere either does not exist, or is not unique):

$$\mathbf{r} = \mathbf{M}^{-1}\mathbf{s} - R\mathbf{M}^{-1}\mathbf{d} \equiv \mathbf{r}_0 - R\mathbf{u}, \qquad (7)$$

where $\mathbf{M}^{-1}\mathbf{s} = \mathbf{r}_0$ and $\mathbf{u} = \mathbf{M}^{-1}\mathbf{d}$. Once Eq.(7) is substituted into the first of the constraints Eq.(3), it leads to the quadratic algebraic equation with respect to R:

$$\left(1 - |\mathbf{u}|^2\right) R^2 + 2\left(R_1 - \mathbf{u} \cdot \mathbf{v}\right) R + \left(R_1^2 - |\mathbf{v}|^2\right) = 0, \quad (8)$$

where $\mathbf{v} = \mathbf{r}_1 - \mathbf{r}_0$. The solution of Eq. (8) can be found in the following form:

$$R_\pm = \frac{-\left(R_1 - \mathbf{u} \cdot \mathbf{v}\right) \pm |R_1\mathbf{u} + \mathbf{v}|}{1 - |\mathbf{u}|^2}. \qquad (9)$$

If $|\mathbf{u}|^2$ is not equal to unity (which corresponds to the case when the 4 spheres are tangential to one plane), then Eq.(9) provides two different solutions, and the positive one expresses the radius R of the touching sphere as a function of the centre position $\mathbf{r}$. Eventually, the positions of their centres can be obtained by inserting R into Eq.(7). In the present implementation, when two solutions are found, the sphere with minimum, positive radius is chosen.

* Electronic address: sega@roma2.infn.it
[1] E. Chacón and P. Tarazona, Phys. Rev. Lett. **91**, 166103 (2003).
[2] J. Chowdhary and B. M. Ladanyi, J. Phys. Chem. B **110**, 15442 (2006).
[3] M. Jorge and M. N. D. S. Cordeiro, J. Phys. Chem. C **111**, 17612 (2007).
[4] L. B. Pártay, G. Hantal, P. Jedlovszky, A. Vincze, and G. Horvai, J. Comput. Chem. **29**, 945 (2008).
[5] A. P. Willard and D. Chandler, J. Phys. Chem. B **114**, 1954 (2010).
[6] M. Jorge, P. Jedlovszky, and M. N. D. S. Cordeiro, J. Phys. Chem. C **114**, 11169 (2010).
[7] M. Jorge, G. Hantal, P. Jedlovszky, and M. N. D. S. Cordeiro, J. Phys. Chem. C **114**, 18656 (2010).
[8] J. Tomasi, B. Mennucci, and R. Cammi, Chem. Rev. **105**, 2999 (2005).
[9] B. Lee and F. M. Richards, J. Mol. Biol. **55**, 379 (1971).
[10] F. M. Richards, Annu. Rev. Biophys. Bioeng. **6**, 151 (1977).
[11] M. L. Connolly, J. Appl. Cryst. **16**, 548 (1983).
[12] M. L. Connolly, Science **221**, 709 (1983).
[13] A. Shrake and J. Rupley, J. Mol. Biol. **79**, 351371 (1973).
[14] F. Richards, J. Mol. Biol. **8**, 2114 (1974).
[15] J. Greer and B. L. Bush, Proc. Natl. Acad. Sci. USA **75**, 303 (1978).
[16] T. Richmond and F. Richards, J. Mol. Biol. **119**, 537555 (1978).
[17] C. Alden and S.-H. Kim, J. Mol. Biol. **132**, 411434 (1979).
[18] S. Wodak and J. Janin, Proc. Natl. Acad. Sci. USA **77**, 17361740 (1980).
[19] J. Muller, J. Appl. Cryst. **16**, 7482 (1983).
[20] M. Pavlov and B. Fedorov, Biopolymers **22**, 15071522 (1983).
[21] J. Pascual-Ahuir and E. Silla, J. Comput. Chem. **11**, 10471060 (1990).
[22] H. Wang and C. Levinthal, J. Comput. Chem. **12**, 868871 (1991).
[23] S. Grand and K. J. Merz, J. Comput. Chem. **14**, 349352 (1993).
[24] J. Weiser, P. S. Shenkin, and W. C. Still, J. Comp. Chem. **20**, 217 (1999).
[25] M. Connolly, J. Am. Chem. Soc. **107**, 11181124 (1985).
[26] T. Richmond, J. Mol. Biol. **178**, 6389 (1984).
[27] K. Gibson and H. Scheraga, Mol. Phys. **62**, 12471265 (1987).
[28] K. Gibson and H. Scheraga, Mol. Phys. **64**, 641644 (1988).
[29] C. Kundrot, J. Ponder, and F. Richards, J. Comp. Chem. **12**, 402409 (1991).
[30] G. Perrot, B. Cheng, K. Gilson, K. Palmer, A. Nayeem, B. Maigret, and et al., J. Comp. Chem. **13**, 111 (1992).
[31] M. Mezei, J. Mol. Graphics. and Modell. **21**, 463 (2003).

[32] H. Edelsbrunner, D. Kirkpatrick, and R. Seidel, IEEE Trans. on Information Theory **29**, 551 (1983).

[33] H. Edelsbrunner and E. Mücke, ACM T. Graphic. **13**, 43 (1994).

[34] H. Edelsbrunner, M. Facello, P. Fu, and J. Liang, Measuring proteins and voids in proteins, in *Proc. 28th Annu. Hawaii Intl. Conf. System Sciences*, volume 5, p. 256264, Los Alamitos, California, 1995, IEEE Computer Society Press.

[35] J. Liang, H. Edelsbrunner, P. Fu, P. V. Sudhakar, and S. Subramanian, Proteins **33**, 1 (1998).

[36] F. Usabiaga and D. Duque, Physical Review E **79**, 046709 (2009).

[37] H. Edelsbrunner, *Handbook of Discrete and Computational Geometry, edited by J. E. Goodman and J. O'Rourke*, chapter 63, pp. 1395–1412, CRC Press, Boca Raton, Florida, 2004.

[38] B. N. Delaunay, Izv. Akad. Nauk SSSR, Otdelenie Matematicheskii i Estestvennyka Nauk **7**, 793800 (1934).

[39] A. Okabe, B. Boots, K. Sugihara, and S. N. Chiu, *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*, Wiley, Chichester, U.K., 2000.

[40] D. Brown, Inf. Process. Lett. **9**, 223228 (1979).

[41] C. Barber, D. Dobkin, and H. Huhdanpaa, ACM T. Math. Software **22**, 469 (1996).

[42] J. L. Bentley, Cmmun. ACM **18**, 509 (1975).

[43] J. Bentley and J. Friedman, ACM Compu. Sur. **11**, 397 (1979).

[44] W. L. Jorgensen, J. Chandrashekar, J. D. Madura, R. Impey, and M. L. J. Klein, Chem. Phys. **79**, 926 (1983).

[45] I. R. McDonald, D. G. Bounds, and M. L. Klein, Mol. Phys. **45**, 521 (1982).

[46] J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, J. Comput. Phys. **23**, 327 (1977).

[47] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl, J. Chem. Theory Comput. **4**, 435 (2008).

[48] U. Essmann, L. Perera, M. Berkowitz, T. Darden, H. Lee, and L. Pedersen, J. Chem. Phys. **103**, 8577 (1995).

[49] S. Nosé, Mol. Phys. **52**, 255 (1984).

[50] W. Hoover, Phys. Rev. A **31**, 1695 (1985).

[51] A. Rozek, C. Friedrich, and R. Hancock, Biochemistry **39**, 15765 (2000).

[52] J. Gesell, M. Zasloff, and S. Opella, J. Biomol. NMR **9**, 127 (1997).

[53] D. Schibli, R. Montelaro, and H. Vogel, Biochemistry **40**, 9570 (2001).

[54] D. Kallick, M. Tessmer, C. Watts, and C. Li, J. Magn. Reson. Ser. B **109**, 60 (1995).

[55] D. Tieleman, D. Van der Spoel, and H. Berendsen, J. Phys. Chem. B **104**, 6380 (2000).

[56] P. Jedlovszky, Á. Vincze, and G. Horvai, J. Chem. Phys. **117**, 2271 (2002).

[57] P. Jedlovszky, Á. Vincze, and G. Horvai, Phys. Chem. Chem. Phys. **6**, 1874 (2004).

[58] G. Hantal, S. Picaud, P. Hoang, V. Voloshin, N. Medvedev, and P. Jedlovszky, J. Chem. Phys. **133**, 144702 (2010).

[59] J. Quaas, Nature **471**, 456 (2011).

[60] S. van Renssen, Nature Clim. Change **2**, 143 (2012).

[61] O. Popovitcheva, N. Persiantseva, M. Trukhin, G. Rulev, N. Shonija, Y. Buriko, A. Starik, B. Demirdjian, D. Ferry, and J. Suzanne, Phys. Chem. Chem. Phys. **2**, 4421 (2000).

[62] L. Sgro, G. Basile, A. Barone, A. D'Anna, P. Minutolo, A. Borghese, and A. D'Alessio, Chemosphere **51**, 1079 (2003).

[63] Y. Chen, N. Shah, A. Braun, F. Huggins, and G. Huffman, Energ. Fuel **19**, 1644 (2005).

[64] A. Abid, N. Heinz, E. Tolmachoff, D. Phares, C. Campbell, and H. Wang, Combust. Flame **154**, 775 (2008).

[65] A. Abid, E. Tolmachoff, D. Phares, H. Wang, Y. Liu, and A. Laskin, P. Combust. Inst **32**, 681 (2009).

[66] D. Small, *Chemistry; The Bile Acids, edited by P. P. Nair, D. Kritchevsky*, volume 1, chapter 8, Plenum Press, New York, 1971.

[67] L. Pártay, P. Jedlovszky, and M. Sega, J. Phys. Chem. B **111**, 9886 (2007).

[68] L. Pártay, M. Sega, and P. Jedlovszky, Langmuir **23**, 12322 (2007).

[69] W. Humphrey, A. Dalke, and K. Schulten, J. Molec. Graphics **14**, 33 (1996).

[70] R. Penfold, A. D. Watson, A. R. Mackie, and D. J. Hibberd, Langmuir **22**, 2005 (2006).