



A tool for the morphological analysis of mixtures of lipids and water in computer simulations

Marc Fuhrmans, Siewert-Jan Marrink

► To cite this version:

Marc Fuhrmans, Siewert-Jan Marrink. A tool for the morphological analysis of mixtures of lipids and water in computer simulations. *Journal of Molecular Modeling*, 2010, 17 (7), pp.1755-1766. 10.1007/s00894-010-0858-6 . hal-00637162

HAL Id: hal-00637162

<https://hal.science/hal-00637162>

Submitted on 31 Oct 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A tool for the morphological analysis of mixtures of lipids and water in computer simulations

Marc Fuhrmans · Siewert-Jan Marrink

Received: date / Accepted: date

Abstract When analyzing computer simulations of mixtures of lipids and water, the questions to be answered are often of a morphological nature. They can deal with global properties, like what kind of phase is adopted or the presence or absence of certain key features like a pore or stalk, or with local properties like the local curvature present in a particular part of the lipid/water interface. While all information for both the global and the local morphological properties of a system is in principle contained in the set of atomic coordinates obtained from a computer simulation, their extraction is tedious and usually involves using a visualization program and doing the analysis by eye. Here we present a tool that employs the technique of morphological image analysis (MIA) to automatically extract the global morphology, as given by Minkowski functionals, from a set of atomic coordinates and creates an image of the system onto which the local curvatures are mapped as a color code.

Keywords Minkowski functionals · Phase determination · Molecular dynamics · Image analysis · Local curvature

1 Introduction

1.1 Motivation

With the development of new models and the steady increase of computational power available, computer sim-

ulations have become more and more valuable in the study of lipid systems. While the exact conformation of individual lipid molecules is of interest for some applications, most of the time the focus is on the behavior of aggregates of lipids as a whole. Recent examples have been reviewed in [1].

Many of these studies have in common that at some point during the analysis of the simulation a morphological property of the system, *i.e.* a property that solely depends on the shape of the lipid aggregate, needs to be characterized. In the case of the more general properties like the adopted phase and presence or absence of stalks or pores the task at hand can be accomplished by loading the obtained coordinates into a visualization program and doing the analysis by eye, but analyzing a large number of simulations in this way can be tedious. In the case of the determination of more specific, quantitative properties like interface area, volume or curvatures this naive approach is even principally impossible.

A possible way to automatize morphological analyses of trajectories generated by computer simulations is to use the technique of morphological image analysis [2] to extract morphological information in the form of Minkowski functionals [3]. This approach has been used to study, *e.g.*, a distribution of pores [4], membrane fusion events during a phase transition [5] and to monitor the self-assembly of vesicles [6]. Another approach is to describe morphological features as persistent voids based on the framework of alpha shapes [7] and persistent homology [8] and has been applied to characterize vesicle fusion [9]. However, no implementation of either method is currently readily available to the majority of researchers, especially not as part of any of the widely used molecular dynamics software packages.

Groningen Biomolecular Sciences and Biotechnology Institute & Zernike Institute for Advanced Materials, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands
Tel.: +31-50-3634457
Fax: +31-50-678910
E-mail: s.j.marrink@rug.nl

Table 1 The relation between volume V , surface area A , mean breadth B , Euler characteristic χ , integrated mean curvature H , integrated Gaussian curvature K , the voxels' edge length ξ and the numbers of cubic voxels n_c and constituting faces n_f , edges n_e and vertices n_v defining the positive space.

morphological property	related property
$V/\xi^3 = n_c$	
$A/\xi^2 = -6n_c + 2n_f$	
$2B/\xi = 3n_c - 2n_f + n_e$	$H = 2\pi B$
$\chi = -n_c + n_f - n_e + n_v$	$K = 4\pi\chi$

Here, we present an extension of the Gromacs software package [10] that enables morphological image analysis of molecular aggregates. In addition, the option to extract local curvatures has been added to the method, which to the authors' best knowledge has not been employed before, at least in the field of lipid aggregates.

1.2 Theory

In three dimensions, there are four Minkowski functionals corresponding to the volume whose morphology is to be determined, the area of the interface separating that volume from the rest of the system and the integrated mean and Gaussian curvature of that interface¹. As such, both geometrical (shape) and topological features (connectivity) are characterized².

For black and white digital, *i.e.* pixelated, images, the extraction of the Minkowski functionals is well-established and can be accomplished by simply counting the pixels and pixel-components of lower dimensionality comprising the image. This means, that for three-dimensional pictures one only needs the number of voxels³ and the number of faces, edges and vertices these voxels consist of, where voxel components shared by several voxels are counted only once. The Minkowski functionals can then be obtained as sums over these numbers as given in Table 1. A way of obtaining the morphology of a set of coordinates is therefore to translate the system into a three-dimensional image composed of black and white voxels [2].

Advantages of this method are the straightforwardness of implementation and its rigorousness in the sense that the resulting numbers are the exact values of the

Minkowski functionals for the image. Its only disadvantage is therefore the approximation introduced by the image itself. The use of voxels entails a limitation to right angles which imposes restrictions on the values for the surface area and integrated mean curvature obtained with this method, causing several structures to share the same value. As an example, removing any voxel from a cube of eight voxels will leave the surface area and integrated mean curvature unchanged, resulting in a general tendency to overestimate these functionals.

However, the Euler characteristic which only requires the connectivity to be identical for the image and the original system can be determined exactly, and the volume can be obtained with only slight errors minimizable by choosing a sufficiently high resolution.

For a broad spectrum of morphological tasks the obtained values are sufficient, even with the restrictions mentioned above. For most applications concerning molecular aggregates, the Euler characteristic and the integrated mean curvature are arguably the most important values. Purely topological analyses, including both phase determination and the detection of stalks or pores, rely primarily on the Euler characteristic which is not affected by the limitations of morphological image analysis. In addition, due to the systematic nature of the error of the integrated mean curvature, the obtained value can still be used to extract morphological information. Absence of mean curvature is accurately recognized as zero mean curvature, and systems with positive can be distinguished from those with negative total mean curvature. In addition, both the integrated mean curvature and the surface area can be used to further characterize structures within families of similar topology, since the lack of absolute values is not detrimental for relative comparisons.

As an extension to this basic application of morphological image analysis, it is also possible to obtain local values of the mean and Gaussian curvature. As has been shown by Hyde et al. [11], every surface vertex can be associated with a certain mean and Gaussian curvature. Again, these values are exact for the image and summation over all surface vertices while taking into account the different surface areas associated with each vertex lead to global (integrated) values of the mean and Gaussian curvatures identical to the ones obtained with the method described above. Mapping the local curvatures onto the image as a color code allows further characterization of the structure at hand, enabling easy detection of areas of different curvature as well as detailed comparison of similar structures.

¹ Alternatively, integrated mean and Gaussian curvature can be replaced by the mean breadth and Euler characteristic, to which they are proportional (see Table 1).

² Note, however, that while a given morphology specifies a specific set of Minkowski functionals, the reverse is not necessarily true.

³ *i.e.* the three-dimensional analog of pixels, essentially small cubes

The rest of this chapter is organized as follows: In Section 2.1 and 2.2 details about the implementation and the user-definable parameters are given, while Section 2.3 describes the parameters used in the simulation that were analyzed to test our program. Section 3 gives the results of these sample applications in addition to tests performed on model systems.

2 Methods

2.1 Implementation

The implementation discussed in this publication was realized using the Gromacs-3.3 software package [10], but should in principle compile with any version of Gromacs from 3.0 to date with only minor modifications. The executable is called `gmia` and was written in the C programming language. The source code is available upon request. Acceptable input file formats are the standard formats supported by Gromacs.

2.1.1 Basic algorithm

We treat the image as a three-dimensional cubic grid representing the simulation box, onto which every coordinate is mapped⁴. To avoid artificial empty spaces introduced by representing atoms (or in the case of coarse-grained models groups of atoms) by their center-of-mass only, every coordinate is expanded to a spherical cloud of coordinates each of which is mapped onto the grid individually⁵. Depending on the type and number of particles mapped to it, cells are declared as positive or negative, where positive cells represent the molecular aggregate. The global values of the Minkowski functionals can then be obtained by counting the number of cubes, cube-faces, edges and vertices taking into account the periodic boundaries.

For the local values of the mean and Gaussian curvature, every surface vertex⁶ is identified as one of the possible cases listed in Fig. 1, storing the corresponding local curvatures given by the product of the interface area and the curvature value associated with that type

	positive: $a = 3\xi^2/4$ $h = \pi/2\xi$ $k = 2\pi/3\xi^2$	negative: $a = 3\xi^2/4$ $h = -\pi/2\xi$ $k = 2\pi/3\xi^2$
	positive: $a = \xi^2$ $h = \pi/4\xi$ $k = 0$	negative: $a = \xi^2$ $h = -\pi/4\xi$ $k = 0$
	positive: $a = 3\xi^2/2$ $h = \pi/6\xi$ $k = -2\pi/3\xi^2$	negative: $a = 3\xi^2/2$ $h = -\pi/2\xi$ $k = 2\pi/3\xi^2$
	positive: $a = 5\xi^2/4$ $h = \pi/10\xi$ $k = -2\pi/5\xi^2$	negative: $a = 3\xi^2/4$ $h = -\pi/10\xi$ $k = -2\pi/5\xi^2$
	positive: $a = 7\xi^2/2$ $h = \pi/14\xi$ $k = -6\pi/7\xi^2$	negative: $a = 7\xi^2/2$ $h = -5\pi/14\xi$ $k = 2\pi/7\xi^2$
	positive: $a = \xi^2$ $h = 0$ $k = 0$	negative: $a = \xi^2$ $h = 0$ $k = 0$
	positive: $a = 3\xi^2/2$ $h = 0$ $k = -2\pi/3\xi^2$	negative: $a = 3\xi^2/2$ $h = 0$ $k = -2\pi/3\xi^2$

Fig. 1 Overview of the possible types of surface vertices and the associated local values of surface area a , mean h and Gaussian curvature k in dependence on the edge-length ξ (adapted from [11]). For each pattern, values are given both for the positive (black representing lipids) and the negative case (black representing water). While more patterns are possible in principle, these represent noise and should not occur as long as the resolution used does not exceed the coordinate density.

⁴ Assuming the resolution of the grid is chosen sufficiently high, the distortion introduced by a potential mismatch between grid and simulation box and by ensuring that the periodic boundary conditions for non-rectangular boxes can be mapped is only minor.

⁵ in the current implementation, the cloud is generated via a simple loop over spherical coordinates, generating N^3 coordinates within N equidistant shells containing N^2 coordinates each, corresponding to a density distribution decaying towards the perimeter

⁶ i.e. every vertex that is part of both negative and positive voxels

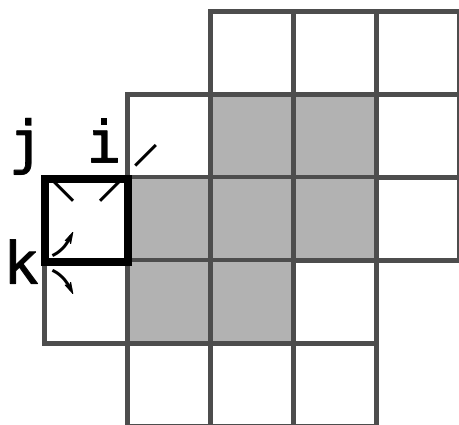


Fig. 2 Mapping of local curvature from surface vertices to voxels. After elimination of non-surface voxels (grey shaded squares), the curvature that has been calculated for the surface vertices is distributed in equal parts to all adjacent surface voxels (arrows). The local curvature C of the highlighted voxel is obtained as $C = \frac{1}{2} i + \frac{1}{2} j + \frac{1}{2} k$, where i , j and k are the curvatures corresponding to the surface vertices adjacent to the highlighted voxel. For simplicity, a two-dimensional example is given, but the method applies equally for three dimensions.

of surface vertex. However, a mapping of the curvature to voxels, not vertices is desired. To that end, non-surface voxels, *i.e.* positive voxels not contributing a single face to the interface are eliminated. The stored curvatures of the surface vertices are then distributed in equal parts to the surface voxels adjacent to that particular vertex, as illustrated in Fig. 2.

For the visualization of the local curvatures, a PyMOL [12] file is generated that represents the image as voxels onto which the curvatures are mapped as a color code. Due to the differences in the range of curvatures encountered, using a fixed color-scale is impossible. We therefore employ a two-color scheme, in which white corresponds to a curvature of zero, while the two colors are used to distinguish negative and positive curvature, with the intensity of the color indicating the value. Full intensity is assigned to the voxel(s) with the maximum absolute curvature encountered in a given system, and the color-range is symmetric in the sense that full intensity indicates the same (absolute) value for both colors. While this causes every image to get its own color-code, it is the most efficient scheme to highlight differences in local curvature.

2.1.2 Optional steps

The data generated can often be improved considerably by performing some image manipulation steps and averaging.

Image manipulation Depending on the particle density in the coordinate file and the desired resolution of the grid, it is possible to include an image manipulation step right after the creation of the image, in which isolated clusters of either positive or negative cells below a certain size are interpreted as noise and removed. As a benefit, performing this step allows the determination of the number of actual isolated clusters above the threshold size at no additional cost, which is useful morphological information in its own right.

Spatial averaging Due to the fixed nature of the grid even aggregates with perfectly homogeneous curvature like a sphere will display different curvatures for different regions, depending on how well the rasterization of the image fits the surface in that region. In general, the curvature tends to be underestimated when the surface is aligned with the grid, and overestimated when it is diagonal to the grid.

Two options of spatial averaging can be employed to reduce this effect. First, the obtained local curvature can be averaged over neighboring surface voxels within a certain distance. In addition, it is possible to further improve the results by performing the determination of local curvatures for multiple grid orientations. For this, the resulting curvature values of each positive surface voxel for every orientation are stored together with the coordinate corresponding to the center of that voxel rotated back to the original orientation. The values of all rotations are then mapped back onto the original grid, averaging the values over the entries mapped onto the same cell. If needed, the resulting values can be averaged over neighboring cells. Since it is not possible to preserve the periodic boundary conditions with a rotated grid, the area of interest is centered in the box and only cells within a certain distance from the center, *i.e.* cells that lie within both the volume of the box and the rotated grid for all rotations, are taken into account.

Time averaging While not included as such in the current version of the presented tool, it can also be useful to average the curvatures over time, *i.e.* over several snapshots of a trajectory. For the global values, this can easily be accomplished post-analysis by taking the floating average of the calculated curvatures. For the local values, a time averaging can be performed on the coordinate level prior to the analysis, effectively yielding time averaged curvatures.

2.2 User-definable options and parameters

It is, in general, not possible to use the same set of parameters for the analysis of all possible structures and

representations. The implementation therefore allows most parameters to be determined by the user. This section describes the parameters and discusses what to consider to achieve the optimal results. The corresponding command line options are given in parentheses.

Input files The tool needs a coordinate or trajectory file (-c) and an index file (-n) in which the particles that correspond to the positive phase are listed.

Imaging options The edge length of the grid (-dim), the radius of the spherical cloud used to expand the coordinates (-sr) and the number of coordinates generated during the expansion (-npts) as well as the minimum number of coordinates mapped onto a grid cell required to count it as positive (-thresh1) need to be specified.

As a general consideration, the resolution needs to be high enough to accurately depict the structure to be analyzed, but is limited by memory requirements, due to several three-dimensional arrays required during the computation⁷. In addition, using a high resolution usually requires expansion of the coordinates to avoid artificial empty voxels caused by the limited coordinate density, which partially offsets the desired high resolution. The radius of the spherical cloud should therefore be chosen as the smallest radius sufficient to avoid noise. (An example of the effects of the chosen resolution for

as an application is given in Section 3.2, Table 2.) It also turns out that, in order to accurately detect flat morphologies as having zero mean curvature, it is required to calibrate the parameters used. Since molecular aggregates usually have a low short-range order, fluctuations of individual molecules from the mean will show as either bumps or dents in the created image. Since a given resolution does not necessarily have the same propensity for producing bumps as for producing dents a net-curvature will be measured. The threshold parameter can be used to adjust the number of "positive" coordinates mapped onto a single grid-cell required to count that cell as positive to, on average, produce an equal number of bumps and dents and therefore not introduce artificial mean curvature to the measurement.

In addition, it is possible to also use the coordinates of the particles corresponding to the negative phase, mapping them onto the grid as has been described above but counting them as negative. If that is desired, the number of phases to consider must be set from 1 to 2 (-np), and the index file needs to contain a second group in which these particles are listed.

If isolated clusters below a certain size are to be removed (see above), the maximal cluster size to be considered as noise must be specified (-cs).

Averaging options The range over which the local curvatures are averaged over neighboring voxels needs to be specified (-ar1 and -ar2), with a value of zero indicating no averaging. Two values are needed, one for the averaging for every single grid orientation (-ar1) and one for the averaging after the values of all grid orientations have been collected (-ar2).

If multiple grid orientations are to be used, the number of rotations around every axis (-nx, -ny and -nz) and the corresponding angle increments (-depsilon, -dphi and -dtheta), as well as the radius around the center of the box within which voxels are considered have to be set (-dr)⁸. In order to achieve the best result, care must be taken to avoid sampling similar orientations.

In addition, a threshold can be specified to only count voxels as positive if a minimum number of local curvatures corresponding to different rotations has been mapped on that voxel (-thresh2). However, unlike the other averaging steps, this option will discard curvature and no longer give the exact results and should therefore be used with care. For the results presented in this work, a threshold of zero has been used, effectively disabling this option.

For the results discussed in Section 3, the grid resolution and the radius used for the expansion of the coordinates will be given along with the number of rotations and the distance used for the averaging of the local values.

2.3 Simulation setup

The simulations shown in this article were performed using the coarse-grained MARTINI model [13] with the Gromacs-3.3 software package [10] using the standard run-parameters for the MARTINI model at a timestep of 40 fs. Both pressure and temperature were coupled to a reference value using the Berendsen scheme [14]. Lennard-Jones and Coulomb interactions were obtained every step for particles within a cut-off of 1.2 nm according to a neighbor list updated every 10 steps. Both the Lennard-Jones and Coulomb potential were modified with a shift function to have the interaction smoothly vanishing at the cut-off. Electrostatic interactions were screened with an effective dielectric constant of 15 (which is the standard value for the MARTINI model).

⁸ This radius needs to be specified as a value between 0 and 1 and will be multiplied with half the smallest box dimension internally.

⁷ no attempt was made to optimize the code in this respect

Three processes were used as sample applications: spontaneous aggregation of lipids into a lipid bilayer, closure of a pore in a membrane, and stalk formation between apposed lipid bilayers (with setups similar to simulations described in [15–17], respectively).

Spontaneous aggregation The system simulated consists of 256 DOPE⁹ lipids with 768 water beads (one bead corresponding to four water molecules) starting from random coordinates. The simulation was carried out at a reference temperature of 315 K with a coupling time constant of 0.5 ps, and with anisotropic pressure coupling with a compressibility of $5 \times 10^{-5} \text{ bar}^{-1}$ for the diagonal elements and $1 \times 10^{-7} \text{ bar}^{-1}$ for the off-diagonal elements of the pressure tensor, coupling time constants of 1.2 ps and reference pressures of 1.0 bar.

Porated membrane The system consists of a bilayer of 128 DPPC¹⁰ lipids with a preformed pore at excess hydration (2653 water beads). After a short equilibration, the simulation was carried out at a reference temperature of 323 K with a coupling time constant of 1.0 ps, and with semi-isotropic pressure coupling with a compressibility of $1 \times 10^{-5} \text{ bar}^{-1}$, a coupling time constant of 1.0 ps and a reference pressure of 1.0 bar for the direction perpendicular to the bilayer and a compressibility of 0 bar^{-1} for the plane containing the bilayer.

Stalk formation The initial configuration was two bilayers of 98 DOPE lipids each, separated by two slabs consisting of 65 water beads each, corresponding to an effective hydration level of 2.65 water molecules per lipid. To induce the formation of stalks, the simulation was carried out at a reference temperature of 375 K with a coupling time constant of 0.5 ps, and with semi-isotropic pressure coupling with a compressibility of $1 \times 10^{-5} \text{ bar}^{-1}$, a coupling time constant of 1.2 ps and a reference pressure of 1.0 bar for all directions.

3 Results

3.1 Model systems

The method was first tested on two artificially constructed model systems with a very high coordinate density: a solid sphere and a toroidal pore. This way, the potential of the method could be assessed by analyzing virtually noise-free structures and the exact values for these ideal geometries were available for comparison. Plots of the coordinates of the model systems used are depicted in Fig. 3 and 4.

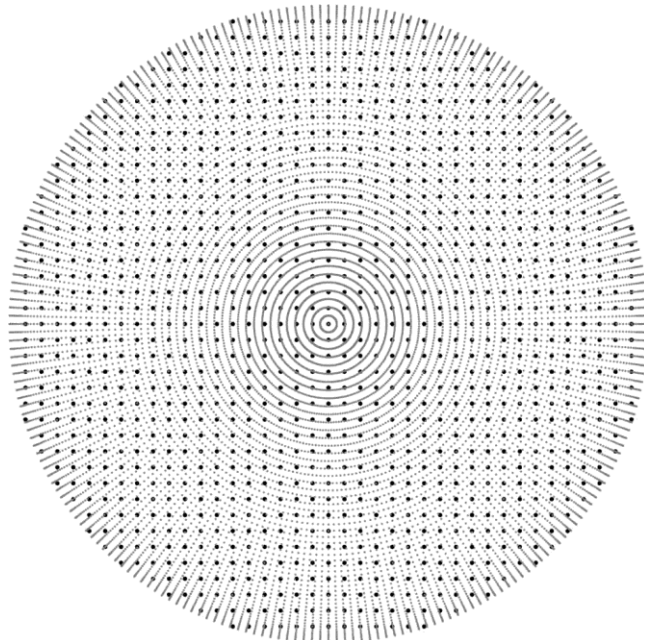


Fig. 3 The spherical model system. The volume of the spheres has been filled with coordinates at a regular distance ($d = r/20$) dependent on the radius r (large black dots). In addition, the surface as the most crucial part has been covered with coordinates at a very high density (small grey dots).

Spheres Fig. 5 shows the measured and theoretical values of the Minkowski functionals for solid spheres of different radii. The image was constructed with a resolution of 0.4 nm and expansion of the coordinates to spheres of 0.2 nm. As predicted for a solid object, the Euler characteristic is obtained with the exact value of 1. The volume of the image is only slightly higher than that of the original, which is due to the rasterization of the image and the expansion of the coordinates to spheres. The surface area and integrated mean curvature, however, are overestimated to a larger extent. In fact, the obtained values lie between the values of the sphere and a cube with an edge length identical to the diameter of the sphere (see discussion in Section 1.2). Nevertheless, the values are proportional to the values of the original and could therefore in principle be used to distinguish spheres of different size.

For the calculation of the local curvatures, eight rotations around every axis were used and values were averaged over neighboring voxels up to a distance of three grid cells. Looking at the mapping onto the image as shown in Fig. 6, we find that both mean and Gaussian curvature are accurately mapped with positive values. While the mean curvature is correctly mapped almost homogeneously over the whole surface, the distribution of the Gaussian curvature for the larger sphere is less even, even with the performed averaging. This is a symptom of a general difficulty in the mapping

⁹ dioleoylphosphatidylethanolamine

¹⁰ dipalmitoylphosphatidylcholine

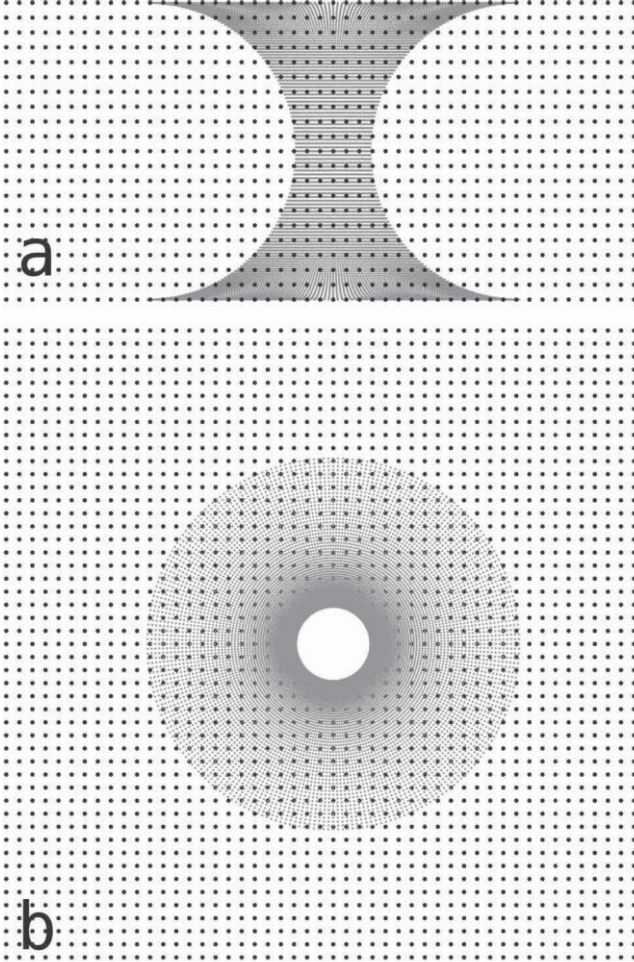


Fig. 4 The model system of the ideal toroidal pore (a: xz plane, b: xy plane). The volume of the porated slab has been filled with coordinates at a regular distance of 0.176 nm in the x and y and 0.2 nm in the z direction (large black dots). In addition, the

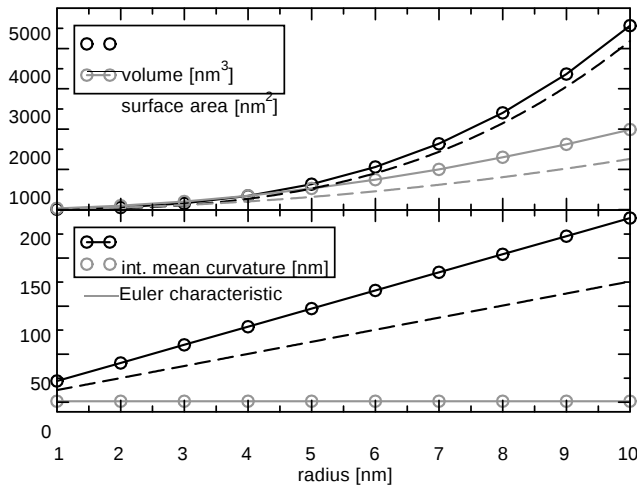


Fig. 5 Morphological image analysis of model systems of solid spheres of different radii (lines between measured points serve to guide the eye). For comparison, calculated values corresponding to the underlying geometry are plotted (dashed lines).

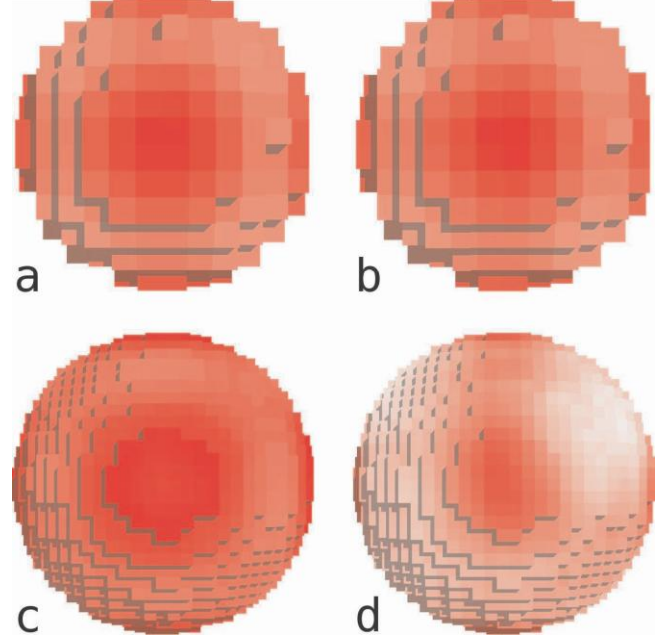


Fig. 6 Mapping of the local values of mean (a, c) and Gaussian (b, d) curvature onto the constructed images of spherical model systems. Shown are the results for spheres with a radius of 2 nm (a, b) and 5 nm (c, d). Positive curvature is depicted as red, zero curvature as white and negative curvature as blue.

of the Gaussian curvature found in most of four measurements for systems which display large areas of homogeneous Gaussian curvature¹¹. However, while this behavior might seem problematic at first, it is partially due to the color-scale employed, which assigns full color intensity to the voxel with the highest absolute curvature (see Sect. 2.1). In the presence of regions with high Gaussian curvature (as in the example of the smaller sphere), these are accurately detected, and artificial fluctuations in regions of lower Gaussian curvature become relatively less important as well as less visible in our depiction.

Toroidal pores Fig. 7 shows the values of the Minkowski functionals for a toroidal¹² pore through an 8.8×8.8 nm² layer of 4.0 nm thickness in dependence on the pore radius¹³, obtained using a grid-size of 0.2 nm and expanding the coordinates to a radius of 0.1 nm. In addition, the analytical values for the volume V , surface area A and integrated mean curvature H are plotted¹⁴:

$$V = V_{\text{slab}} - V_{\text{cyl}} + \pi^2 d^2 (d + r) - 4\pi d^3 \quad (1)$$

¹¹ To a lesser extent this is also true for the mean curvature; however, the values of the mean curvature tend to be higher and therefore less sensitive to artificial fluctuations caused by the rasterization.

¹² i.e. a pore in which the curvature of the surface varies smoothly in a manner identical to the “inner” part of a torus

¹³ defined as the radius of the actual opening

¹⁴ note that periodic boundary conditions apply

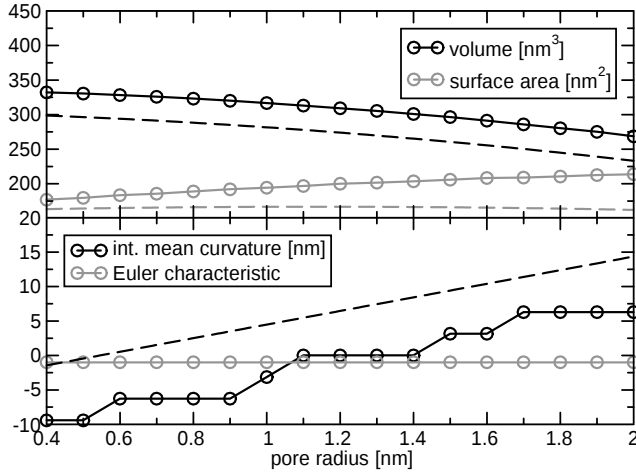


Fig. 7 Morphological image analysis of model systems of ideal toroidal pores of different radii (lines between measured points serve to guide the eye). For comparison, calculated values corresponding to the underlying geometry are plotted (dashed lines).

$$A = 2(A_{\text{rec}} - A_{\text{circ}}) + 2\pi^2 d(d+r) \quad (2)$$

$$H = \pi^2(d+r) - 4\pi db \quad (3)$$

In these expressions d is half the thickness of the slab, r is the radius of the pore at its smallest extension, A_{rec} is the area of the bottom or top of the unporated slab, A_{circ} is the area of the circle with radius $d+r$, V_{slab} is the volume of the unporated slab and V_{cyl} is the volume of the cylinder with a height of $2d$ and a radius of $d+r$. As before, the Euler characteristic is obtained with the exact value of -1 and the volume of the image is higher, but proportional to that of the original. The surface area is overestimated to a larger extent, again showing how the area of curved surfaces is increased by the rasterization of the image. The fact that the surface area of the image is actually found to be increasing over the whole range of radii in disagreement with the values calculated for the original reflects the increasing percentage of the total surface that is curved for larger pore radii. This causes the slight decrease of surface area in the original geometry to be overshadowed by the overestimation of areas of curved surfaces in the image.

The integrated mean curvature has the same general trend for both image and original, but the amount of negative curvature is higher in the image for the measured range of radii. This causes small pores to display negative values for radii up to 1 nm, while the actual cross-over point for the original geometry is at approximately 0.5 nm. In addition, it becomes apparent that the values obtained by morphological image analysis are discrete and not continuous¹⁵, causing small changes in

curvature in the original geometry to go unnoticed in the image.

The local curvatures were calculated using four orientations for each axis and averaging over neighboring voxels up to a distance of five grid cells. Looking at the mapping onto the image as shown in Fig. 8, the dominance of negative mean curvature for pores of small radii found in the global values is also visible. The mean curvature is accurately found to be minimal in the midsections of the pores, reflecting the highest negative principal curvature located in that region, and maximal close to the rim, reflecting the lowest negative principal curvature in that region¹⁶ and is in fact accurately found to be approximately zero in the midsection of the pore of radius 2.0 nm (for this radius and a layer-thickness of 4.0 nm the two principal curvatures cancel in this region). In addition, it becomes overall more positive for higher pore radii, in accordance with the lower negative principal curvature. The Gaussian curvature is also found to be accurately mapped, with the maximum (negative) curvature found in the midsection and gradually decreasing to zero the closer one gets to the rim for the two bigger pores. Only for the smallest radius the minimum for the Gaussian curvature in the midsection is not detected due to the pore size being close to the limit of the resolution used. This problem could, in principle, be avoided by using a higher resolution.

It is worth mentioning that the negative space of the images of the ideal toroidal pores is an image of a stalk. The corresponding stalks will therefore have identical Gaussian curvature and surface area as the pores, and the mean curvature will only invert its sign. For the global values it can therefore be deduced that stalks are accurately characterized as having negative mean curvature, if one considers that stalks have a certain minimum radius given by the length of the lipid tails (approximately 2.0 nm for a typical lipid tail of 16–18 carbon atoms).

3.2 Applications

Next we tested our method with trajectories and snapshots taken from actual simulations of lipids. For these, it proved advantageous to define the positive phase as only the atoms or beads corresponding to the lipid tails. That way, details like pores are amplified and stalks can be distinguished from configurations in which two membranes are merely close without contact of the hydrophobic cores.

¹⁵ with the distance between the discrete levels depending on the grid resolution

¹⁶ the second, positive, principal curvature is constant across the whole pore

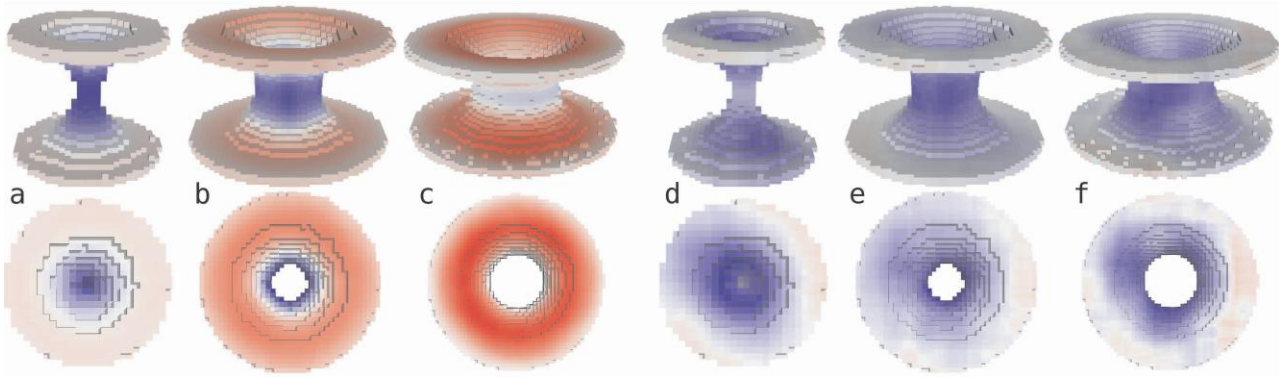


Fig. 8 Mapping of the local values of mean (a, b, c) and Gaussian (d, e, f) curvature onto the constructed images of model systems of ideal toroidal pores. Shown are the results for pores with a radius of 0.4 nm (a, d), 1.2 nm (b, e) and 2 nm (c, f). Positive curvature is depicted as red, zero curvature as white and negative curvature as blue.

Spontaneous aggregation The first application is the determination of the phase adopted by a mixture of DOPE and water starting from random coordinates in a spontaneous aggregation approach [15]. The Minkowski functionals obtained using a grid size of 0.5 nm and expanding the coordinates to a radius of 0.4 nm are shown in Fig. 9. Looking at the Euler characteristic, the most significant morphological indicator, one can see that the system quickly adopts a metastable phase in which it remains for about 0.5 ns before it adopts its final configuration, in which it remains for the rest of the simulation. The metastable state has a negative Euler characteristic of -2, which compared to a lamellar state suggests the presence of pores or stalks. Taking into account the amount of negative mean curvature, the state most likely is a stalk phase, since pores display a higher mean curvature. The stable phase adopted for the rest of the simulation has an Euler characteristic of 0, suggesting the lamellar or inverted hexagonal phase. The fact that the integrated mean curvature remains negative rules out the lamellar phase, leaving the inverted hexagonal phase. Visual inspection confirms these findings.

Porated membrane As a second application, we looked at a closing pore in a DPPC membrane. The Minkowski functionals shown in Fig. 10 were obtained using a grid size of 0.5 nm and expanding the coordinates to a radius of 0.4 nm. The closure of the pore can be detected and is clearly reflected by the Euler characteristic, the integrated mean curvature as well as the surface area, with the observed changes in accordance with our earlier measurements for the model pores (see above). The presence of noise is especially in the integrated mean curvature and the surface area stems from natural fluctuations in the coordinates as well as translational movements of the system and fluctuations of the box size, all

of which cause changes in the image. However, averaging over time, as demonstrated for the integrated mean curvature in Fig. 10, reduces the noise significantly.

In addition, we chose this application to demonstrate the effects of the chosen grid resolution and the radius used for the coordinate expansion on the results obtained. Table 2 shows the average values of the Minkowski functionals for the open pore (i.e. the first 30 nm of the trajectory analyzed in Fig. 10) using a higher resolution or a smaller expansion radius. The deviations of the Euler characteristic from the correct value of 1 indicate the presence of noise in the analysis. Considering the nature of the changed parameters, both the reduced expansion radius and the higher grid resolution potentially can lead to empty grid cells in regions occupied by the molecular aggregate. As a result, both additional tunnels (which would lower the value by one in that particular frame of the trajectory) as well as cavities (which would raise the Euler characteristic

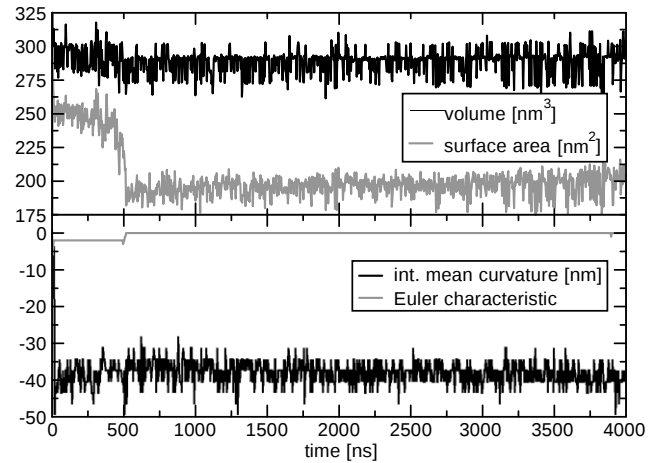


Fig. 9 Morphological image analysis of the spontaneous aggregation of a random mixture of lipids and water into an inverted hexagonal phase.

of which cause changes in the image. However, averaging over time, as demonstrated for the integrated mean curvature in Fig. 10, reduces the noise significantly.

In addition, we chose this application to demonstrate the effects of the chosen grid resolution and the radius used for the coordinate expansion on the results obtained. Table 2 shows the average values of the Minkowski functionals for the open pore (i.e. the first 30 nm of the trajectory analyzed in Fig. 10) using a higher resolution or a smaller expansion radius. The deviations of the Euler characteristic from the correct value of 1 indicate the presence of noise in the analysis. Considering the nature of the changed parameters, both the reduced expansion radius and the higher grid resolution potentially can lead to empty grid cells in regions occupied by the molecular aggregate. As a result, both additional tunnels (which would lower the value by one in that particular frame of the trajectory) as well as cavities (which would raise the Euler characteristic

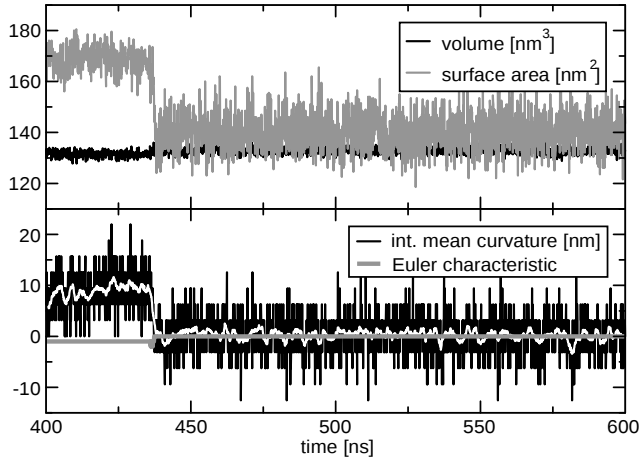


Fig. 10 Morphological image analysis of the trajectory of a closing pore. In addition to the properties indicated by the legend, the running average of the integrated mean curvature is plotted in white.

Table 2 Average values of the volume V , the surface area A , the integrated mean curvature H and the Euler characteristic χ obtained in dependence on the resolution (identified by the edge length d of the grid) and the radius of the spherical cloud r_s used for the expansion of the spheres. Note that in order not to mask any effects, no cluster filtering was applied to the images.

r_s [nm]	d [nm]	V [nm ³]	A [nm ²]	H [nm]	χ
0.4	0.5	131.47(6)	168.8(2)	8.8(2)	-1.000(0)
0.4	0.4	117.4(1)	169.7(2)	11.4(2)	-0.89(2)
0.4	0.3	100.76(3)	247.2(3)	-205(1)	71.4(4)
0.3	0.5	126.00(6)	169.7(2)	11.3(2)	-0.997(3)
0.2	0.4	120.25(6)	170.1(3)	12.2(2)	-0.71(3)

in that respective frame by one) are possible. However, the fact that the Euler characteristic increases indicates that cavities are the dominant artifact. The observed decrease of volume and the increase of surface area both corroborate this interpretation. The changes of the integrated mean curvature are less clear (except for the drastic change to negative values at a grid resolution of 3 nm), but, as has been pointed out in Section 2, the obtained mean curvature depends on the system's propensity to produce bumps and dents in the image which in turn depends on the parameters used, making a comparison of curvatures obtained with different parameters difficult.

The mapping of the local curvatures obtained for a snapshot of the open pore is depicted in Fig. 11. Data from eight grid orientations for each axis at a resolution of 0.3 nm while expanding the coordinates to 0.3 nm and averaging over neighboring voxels up to a distance of three grid cells were used. The mean and Gaussian curvature associated with the pore are accurately mapped as positive and negative, respectively. While the distribution is less homogeneous com-

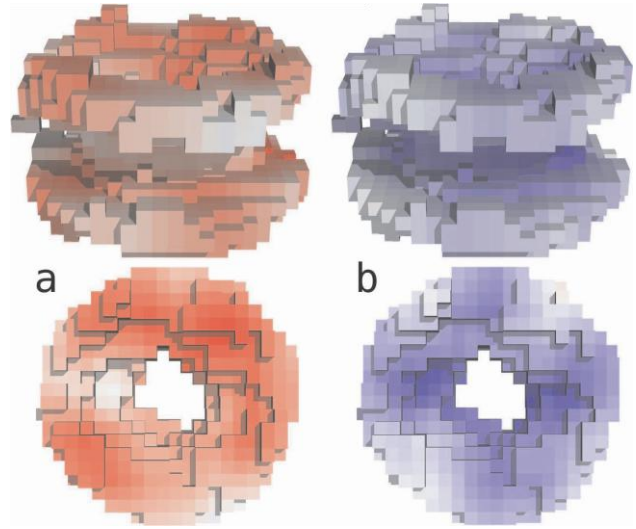


Fig. 11 Visualization of the local distribution of mean (a) and Gaussian (b) curvature for a snapshot of a pore in a DPPC membrane. Positive curvature is depicted as red, zero curvature as white and negative curvature as blue.

pared to the ideal model systems, this is not an artifact of the method, but the accurate detection of features present in the analyzed coordinates. Regions with groups of atoms protruding from the mean are correctly displayed as having high mean and Gaussian curvatures¹⁷, whereas regions with a saddle-splayed surface are detected as having a low mean and high (negative) Gaussian curvature. The general trend of the location of the highest local mean curvature close to the rim of the pore and the highest Gaussian curvature in the mid-section found for the model systems is preserved also in the simulated pores, but slightly modified by superposition of effects due to deviations from the ideal toroidal shape.

For comparison, a snapshot of the underlying structure is shown in Fig. 12.

Stalk formation Fig. 13 shows the Minkowski functionals obtained for the formation of stalks between two DOPE bilayers at low hydration and high temperature. Data was obtained using a grid resolution of 0.2 nm and expanding the coordinates to a radius of 0.34 nm. Starting from a lamellar configuration, the formation of the stalks is reflected by a change of the Euler characteristic from 0 for the two bilayers to -2, indicating the simultaneous formation of two stalks. At the same time, the integrated mean curvature drops from 0 to negative values, also indicating the formed stalks.

¹⁷ In fact, the Gaussian curvature in these regions should be positive, however, this is likely masked by the surrounding negative Gaussian curvature being carried into the relatively small area of positive curvature by the averaging procedure.

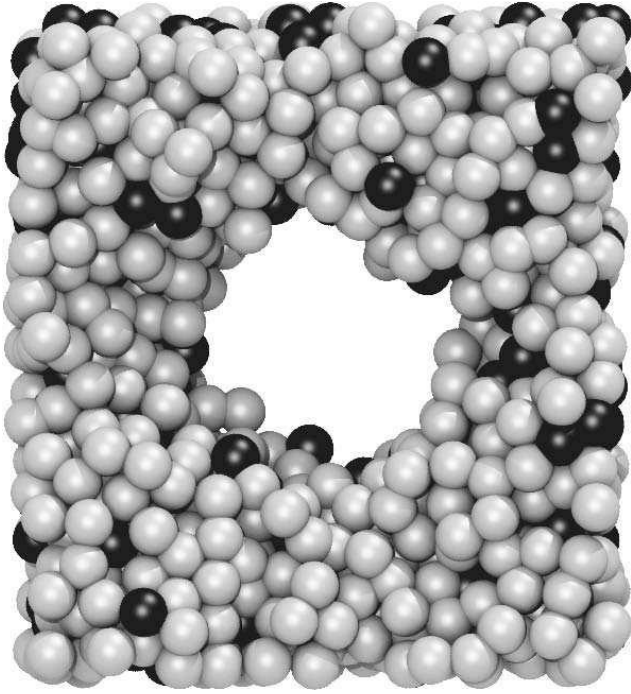


Fig. 12 Snapshot taken from the trajectory of a closing pore in a DPPC membrane. For clarity we show only the beads representing the lipid carbon tails (the terminal beads are shown in black, the remaining beads in grey), which are the beads used to

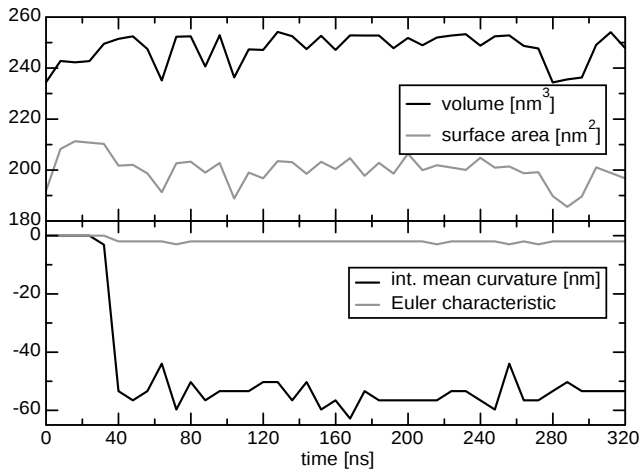


Fig. 13 Morphological image analysis for the simulation of stalk formation starting from two isolated bilayers.

Shown in Fig. 14 is the mapping of the local curvatures of an isolated stalk using data from four rotations of the grid around each axis at a resolution of 0.17 nm while expanding the coordinates to a radius of 0.34 nm and averaging over neighboring voxels up to a distance of three grid cells. As for the simulated pore, the detected curvature is not homogeneous, but the general trend of the relative distribution of curvature between foot (the counterpart to the rim of a pore) and midsec-

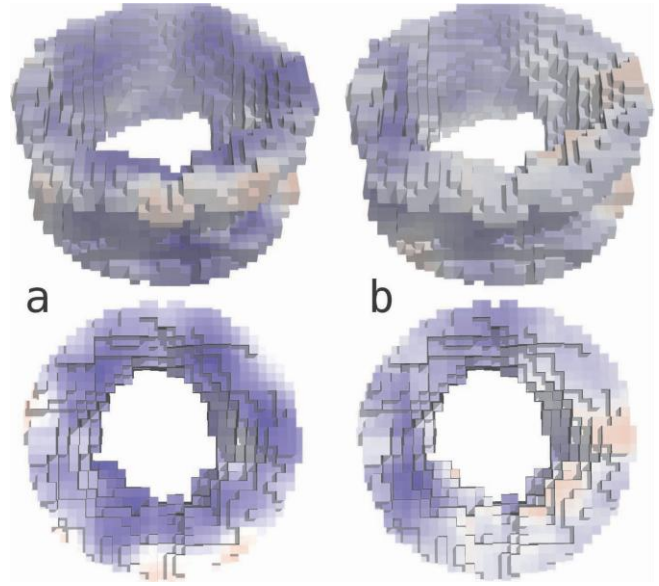


Fig. 14 Visualization of the local distribution of mean (a) and Gaussian (b) curvature for a snapshot of a stalk between two DOPE membranes. Positive curvature is depicted as red, zero curvature as white and negative curvature as blue.

tion is preserved also here: A higher (negative) mean curvature is observed close to the foot of the stalk and a higher (negative) Gaussian curvature close to the mid-section. While part of the observed inhomogeneities, especially in regions of relatively low curvature, are likely caused by artificial noise introduced by insufficient averaging, most of the detected curvature again can be attributed to actual morphological properties contained in the analyzed coordinates.

A snapshot of the formed stalk is shown in Fig. 15.

3.3 Computational costs

To give a rough indication of the time required to perform the presented analyses, we will state the CPU-time¹⁸ used for some of the performed calculations. This is not meant as an extensive analysis of scaling and computational efficiency, but rather as a help for readers interested in using the method. So far, the program has not been optimized for computational efficiency; however, even so, the time required for performing the presented morphological analysis is in the order of only a few minutes.

For the computation of the global morphology for the trajectory of the closing pore shown in Fig. 10, a total CPU-time of 120 s was used for all 2,500 frames, corresponding to approximately 50 ms per frame. Reducing the grid resolution by a factor of two, the CPU-time required for the whole trajectory drops to 104 s.

¹⁸ on an Intel Core 2 DUO 6700 2.66 GHz CPU

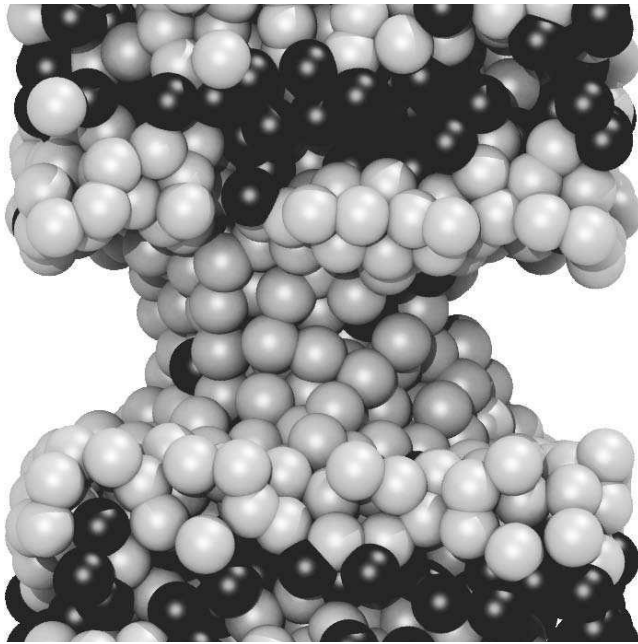


Fig. 15 Snapshot of a stalk formed between two apposed DOPE membranes. For clarity we show only the beads representing the lipid carbon tails (the terminal beads are shown in black, the remaining beads in grey), which are the beads used to define the positive space in our analysis.

For the computation of the local curvatures, the time required for the analysis of the largest of the model pores presented in Fig. 8 is 187s, corresponding to approximately 7s per orientation used in the averaging. As before, reducing the grid resolution increases the performance, with a reduction by a factor of two dropping the required time to 132s. In comparison, the analysis of the local curvature for a snapshot of the pore taken from the simulation referred to in the previous paragraph is much faster, due to the lower coordinate density and system size. The results shown in Fig. 11 required a CPU-time of 57s, corresponding to 167ms per orientation used for the averaging.

4 Conclusions

Our tool uses the technique of morphological image analysis to analyze sets of coordinates describing aggregates of soft matter. The implementation aims at the analysis of mixtures of lipids and water obtained from molecular dynamics simulations, but is in principle applicable to all kinds of coordinates describing binary mixtures and independent of the model and the method used to generate the coordinates.

It has been demonstrated to be helpful in a range of morphological tasks, including phase detection and monitoring of dynamic processes like stalk-formation

and pore-closure. While the obtained global values of the Minkowski functionals are subject to the limitations inherent in the technique of morphological image analysis, *i.e.* an overestimation of surface area and integrated mean curvature, the most significant value, the Euler characteristic, is obtained correctly and virtually free of noise. The systematic nature of the error in the total mean curvature, on top of that, allows its use for comparative analyses and to generally detect presence or absence of mean curvature and distinguish systems with negative from those with positive values, enabling a broad spectrum of applications. Only in complex systems that contain both stalk- and pore-like structures simultaneously, one might run into problems in the interpretation. In such ambiguous cases, using an analysis of the number of separate clusters for both components of the binary mixture, that is also part of our implementation, can help.

In addition, by mapping the local values of mean and Gaussian curvature onto an image representing the system, morphological features inaccessible to the naive analysis by eye were able to be visualized. While there remain some difficulties to avoid artificial inhomogeneities in the detection of local curvature in larger areas of low homogeneous curvature due to the rasterization of the underlying image, the general trend towards a specific curvature in a given area is preserved and areas that show an actual prevalence of either positive or negative curvature in comparison to their neighborhood are accurately visualized.

So far, our implementation only analyzes coordinate sets as they are generated by the simulation, with no option of time averaging over multiple frames of a trajectory. If analysis of average structures is desired, it is therefore necessary to create an averaged coordinate set pre-analysis.

We hope that with our program which is compatible with the popular Gromacs package, a wide range of users may benefit from using morphological image analysis for their research.

Acknowledgements The authors would like to thank H.J. Riselada and T.A. Wassenaar for contributing routines to the code and H.A. De Raedt for useful discussions.

References

1. S.J. Marrink, A.H. de Vries, D.P. Tieleman, *Biochim. Biophys. Acta, Biomembr.* 1788(1), 149 (2009)
2. K. Michielsen, H. De Raedt, *Phys. Rep.* 347(6), 461 (2001)
3. H. Hadwiger, *Vorlesungen über Inhalt, Oberfläche und Isoperimetrie* (Springer, 1957)
4. C. Loison, M. Mareschal, F. Schmid, *J. Chem. Phys.* 121(4), 1890 (2004)

-
5. S.J. Marrink, D.P. Tieleman, *Biophys. J.* **83**(5), 2386 (2002)
 6. G.J.A. Sevink, A.V. Zvelindovsky, *Macromol.* **38**(17), 7502 (2005)
 7. H. Edelsbrunner, E. Mücke, *Assoc. Comput. Mach. Trans. Graphics* **13**(1), 43 (1994)
 8. A. Zomorodian, G. Carlsson, *Discrete Comput. Geom.* **33**(2), 249 (2005)
 9. P.M. Kasson, A. Zomorodian, S. Park, N. Singhal, L.J. Guibas, V.S. Pande, *Bioinformatics* **23**(14), 1753 (2007)
 10. D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A.E. Mark, H.J.C. Berendsen, *J. Comput. Chem.* **26**(16), 1701 (2005)
 11. S.T. Hyde, I.S. Barnes, B.W. Ninham, *Langmuir* **6**(6), 1055 (1990)
 12. W.L. Delano. The PyMOL molecular graphics system (2002)
 13. S.J. Marrink, H.J. Risselada, S. Yefimov, D.P. Tieleman, A.H. de Vries, *J. Phys. Chem. B* **111**(27), 7812 (2007)
 14. H.J.C. Berendsen, J.P.M. Postma, A. DiNola, J.R. Haak, *J. Chem. Phys.* **81**, 3684 (1984)
 15. S.J. Marrink, E. Lindahl, O. Edholm, A.E. Mark, *J. Am. Chem. Soc.* **123**(35), 8638 (2001)
 16. H. Leontiadou, A.E. Mark, S.J. Marrink, *Biophys. J.* **86**(4), 2156 (2004)
 17. S.J. Marrink, A.E. Mark, *Biophys. J.* **87**(6), 3894 (2004)