Electronic Supplementary Material (ESI) for Environmental Science: Atmospheres.

Surface Functionality of Sub- to Full-Monolayer Organic Coverage of Water Aerosols Determined by Molecular Dynamics Simulations

Aisling C. Stewart, Martin J. Paterson, Stuart J. Greaves

Institute of Chemical Sciences, Heriot-Watt University, Edinburgh EH14 4AS, United Kingdom

Supporting Information

Contents

S1 Simulation Details

All simulations were carried out using GROMACS version 2020.2.[1,2] Fatty acids were described using OPLS-AA force field parameters[3,4] generated using the LigParGen server[5] and water molecules were simulated using the SPC/E model[6] and restrained using the SETTLE algorithm.[7] Addition of fatty acid molecules was carried out using a random insertion procedure, described in more detail in section S2. Energy minimisation used the steepest-descent algorithm and proceeded until the total energy of the simulation had converged. Short equilibration steps between the addition of molecules employed the leap-frog algorithm,[8] with a time step of 0.5 fs and a length of 50 ps. After the addition of all required molecules an equilibration step (10 ns) followed by a production step (10 ns) were conducted using a velocity Verlet algorithm[9] and a time step of 0.5 fs.

Simulations were carried out using smooth particle mesh Ewald (PME) electrostatics.[10,11] Short range interactions were measured with a cutoff of 1.3 nm and a Verlet cut-off scheme. The PME algorithm had an order of 4 and a Fourier grid spacing of 0.12 nm. Calculations employed a V-rescale thermostat with a time constant of 0.1 ps. This was set with a target temperature of 298.15 K for all simulations. Pressure coupling was not used as all simulations were carried out under constant volume conditions. For all simulations 3D periodic boundary conditions were employed. A restraining potential was applied throughout all energy minimisation and NVT steps to prevent migration of molecules from the upper to the lower faces of the slab and vice versa. The restraining potential was a flat-bottomed potential, with no force applied within the central 14 nm of the box, and a harmonic potential with a force constant of 10,000 kJ mol^{-1} nm^{-2} beyond this.

S2 Description of the Random Insertion Process

Random insertion of molecules into the boxes was as follows. A fully equilibrated slab of water of dimensions 6.5x6.5x6.5 nm in a periodic box of dimensions 6.5x6.5x50 nm was generated, as discussed in the main text. The water slab was centred in the box and fatty acid molecules were inserted to regions in the box that were between 1.5 and 5.5 nm from one side of the water slab. This was done by using a random number generator of uniform probability to independently select x, y and z coordinates that were within this region of interest. These coordinates were then fed to the gmx insert-molecules algorithm as position coordinates for the centre of mass of the molecule. The gmx insert-molecules algorithm inserted the desired molecule at these coordinates and at a random orientation, which it selected itself from a uniform distribution of polar angles. If the insertion at these atomic coordinates led to an overlap with atoms from molecules that were already present, the coordinates of the molecule being added were discarded and a new set of centre of mass coordinates and orientation was selected. The range 1.5-4.5 nm from the water slab was chosen as this was found to be a compromise between several factors. A larger region or one that was further from the water slab led to organic molecules taking longer to find the aerosol slab. A smaller region meant that once a certain number of fatty acid molecules was reached there would be so little space remaining in the coordinate selection region that it would be difficult to find atomic coordinates that did not led to atom overlap. As molecules attached to the water surface a selection region that started <1.5 nm from the water core also had this effect. When discussing the range relative to the water core it was assumed that the extent of the core remained ±3.25 nm from the centre of the periodic box, with any atomic level differences or changes to the core from the addition of the organic molecules ignored in this discussion of the range of the selection region. The choice of selecting coordinates within given selection regions also ensured that the same number of fatty acid molecules were added to the regions of vacuum above and below the slab, thus allowing for each interface to have exactly the desired molecular coverage. Molecule velocities were selected at the start of each equilibrium or production run stage by random selection from a Boltzmann distribution at 298.15 K, as a standard part of the leapfrog and velocity Verlet algorithms.[8,9] As discussed in the main paper a flat-bottomed restraining potential ensured that molecules did not migrate from one side of the box to the other, even if their initial velocities were directed away from the slab core.

S3 Curvature Effects

All calculations discussed in this work use periodic boundary conditions in the x, y and z directions. This was needed in order to generate a slab that was infinite in the x and y directions. This was used to mimic a larger aerosol particle.

In the atmosphere, aerosols cover a wide range of sizes. Smaller aerosol particles have been traditionally simulated as droplets, however, when looking at larger radius aerosol particles this becomes computationally prohibitive. However, larger aerosols have greater radii and therefore lower curvature values and are thus better modelled using a flat slab mimic.

The curvature of a sphere of a given radius and its comparison to a flat slab can be assessed by looking at the how far the circumference of a circle of that radius will have deviated from a tangent to the circle at a given distance from the point where the tangent and the circle meet. Fig. S1 shows this for a distance of 6.5 nm, chosen as this is equal to the x and y dimensions of the slabs we are investigating. It can be seen that for circles of radius >0.21 µm there is a deviation of <1 Å, which is less than the deviations in surface location due to atomic level roughness of the slab. Therefore, it is concluded that

Fig S1

Fig. S1 The deviation of the z coordinates of the circumference of a circle from those of a tangent to it, at points on the circumference of the sphere that are Δx = 6.5 nm from the point at which the tangent meets the circle, for circles of varying radii. For circles of radius >0.21 µm this deviation is less than 1 Å and therefore less than the deviations in surface position caused by atomic level surface roughness. Inset: diagram showing the main parameters discussed in the main graph. r represents the radius of the circle. The tangent meets the circle at x=0. A flat surface represents the tangent to this circle, meeting the circle at point x = 0. At x = 6.5 nm the surface of the circle now has z coordinates that differ from those of the tangent (flat slab) by Δz. If Δz is less than the atomic level roughness then curvature effects are considered to be not greatly significant over the length of the periodic box.

an aerosol of radius >0.21 µm can reasonably be approximated using the infinite slab method. For further details, see the ESI of our previous work.[12]

S4 Top-Down Views of Slabs: By Molecule Type

An understanding of how molecules pack together at the surface of an aerosol and which species and functional groups are present at the aerosol-atmosphere interface is important when considering how these aerosols might interact with gaseous species, as it is these interfacial atoms that will be the first that are encountered by incoming species. One of the simplest ways that the surface of a simulated aerosol can be visualized is by looking directly at the interface using visualization software, such as the Visual Molecular Dynamics (VMD)[13] package. Such images are shown in Fig. 2 of the main text, which gives selected top-down views for the extreme surface coverages for stearic, oleic and linoleic acid, presented using VMD. In Figures S2-4, below, extended versions of these are shown for all the surface coverages studied and for each of the acids investigated. For each surface coverage one side of one slab is shown (out of the four slabs that were simulated for that coverage) and the images are a snapshot from the final frame at the end of the 10 ns production run, showing an example of typical equilibrium positions of the atoms. Water molecules are shown in cyan and acid molecules in red, with each of these displayed using their van der Waals volumes. Where lower numbers of acid are present at the interface there are significant amounts of water present, which indicates that these slabs are still at a significantly sub-monolayer coverage of fatty acid molecules, as there is clearly not a complete layer of acid molecules present on the water surface. As the number of acid molecules increases this water becomes less visible, with hardly any water molecules being seen at number concentrations of 130+ for stearic acid and around 110+ for oleic and linoleic acid. The reason that this number is higher for stearic acid is likely to be a result of the greater capacity of stearic acid molecules to pack more closely together, as discussed in the main paper. The preference for acid molecules in partially filled monolayers to clump together in groups is also clearly displayed in the slabs of lower acid coverage.

Fig S2

Fig. S2 Top-down views of slabs of stearic acid, for each of the surface area coverages studied. Acid molecules are shown in red and water molecules in cyan. Views are from the final frame of a 10 ns production run. Atoms are shown using their van der Waals volumes.

Fig S3

Fig. S3 Top-down views of slabs of oleic acid, for each of the surface area coverages studied. Acid molecules are shown in red and water molecules in cyan. Views are from the final frame of a 10 ns production run. Atoms are shown using their van der Waals volumes.

Fig S4

Fig. S4 Top-down views of slabs of linoleic acid, for each of the surface area coverages studied. Acid molecules are shown in red and water molecules in cyan. Views are from the final frame of a 10 ns production run. Atoms are shown using their van der Waals volumes.

S5 Top-Down Views of Slabs: By Functional Group Type

Figures S2-4 show top-down views of slabs of different surface coverages for all the different acids investigated in this work. In Figures S5-7 (below) the same images are shown, however, here the atoms have been highlighted by functional group. The colour scheme is the same as in Fig. 3 of the main paper. The surfaces of the organic components of the aerosol can be seen to be dominated by methyl groups (red) and CH_2 (green), with very little COOH (blue) at this interface. It can be seen that there is significantly more HC=CH (black) present at the surfaces of the linoleic acid slabs than at the oleic acid ones. This is primarily due to the presence of two alkene groups within a linoleic acid molecule compared with only one within oleic acid.

Fig S5

Fig. S5 Top-down views of slabs of stearic acid (STC), for each of the surface area coverages studied. Atoms are colour-coded by functional group (see key). Views are from the final frame of a 10 ns production run. Atoms are shown using their van der Waals volumes.

Fig S6

Fig. S6 Top-down views of slabs of oleic acid (OLE), for each of the surface area coverages studied. Atoms are colour-coded by functional group (see key). Views are from the final frame of a 10 ns production run. Atoms are shown using their van der Waals volumes.

Fig S7

Fig. S7 Top-down views of slabs of linoleic acid (LLC), for each of the surface area coverages studied. Atoms are colour-coded by functional group (see key). Views are from the final frame of a 10 ns production run. Atoms are shown using their van der Waals volumes.

S6 Extended Views of Selected Slabs

Fig 3 of the main paper shows zoomed in VMD images of the upper sections of the extreme coverages of slabs of stearic (STC), oleic (OLE) and linoleic (LLC) acids, each at the end of a 10 ns production run. In the figure below (S8) we present images of the entire slab, showing the upper and lower interfaces, as well as the periodic box (marked on in blue). These images are from the same slabs as for those shown in Fig.3 of the main paper and are also from the final frames of the 10 ns production runs. Atoms have been shown as their van der Waals radii and have been colour-coded based on their functional group (see key). The images in this figure again clearly show the greater degree of order in going from a lower to a higher surface coverage, as well as the greater degree of ordering within the stearic acid slabs, as compared to the other two acids. In the bottom left-hand side of the image the 150 oleic acid slab is shown once more, but with the periodic box marked on in blue. It can be seen that the height of the box is relatively large compared to the size of the slab. This serves two purposes.

Firstly, it allows for a restraining potential to be used that is large enough at the edges of the box to allow no molecules to cross the periodic boundary and transfer to the opposite interface, whilst still having a large enough potential-free region in the centre of the box for this to not affect the dynamics of molecules attached to or close to the slab itself. Secondly it means that there are no interactions between the slab and its image at the centre of the periodic boxes located immediately above and below the one shown.

Fig S8

Fig. S8 Middle and Right: Side-on views of the slabs from Fig. 3 from the main paper (stearic acid=STC, oleic acid=OLE and linoleic=LLC). Atoms are colour-coded by functional group, as per the key to the right. The number above each slab refers to the number of molecules per interface. The preference for COOH groups (blue) to be positioned closest to the water core (white) and the methyl groups (red) at the slab-atmosphere interface can be seen as well as the greater degree of ordering of stearic acid molecules compared to the other two acids. Bottom left: Slab containing 150 molecules of oleic acid per interface, positioned within the 6.5 x 6.5 x 50 nm periodic box (blue).

S7 Tilt Angle Gaussian Fit Standard Deviations

The angular distribution plots for individual coverages of the different species shown in Figures 4a-c of the main paper give an indication of how the spread of the data across the angular ranges differs between species and coverages, however, this can be quantified using the same Gaussian fits that were used to produce Figure 4d of the main paper. Below (Fig. S9) are shown the standard deviations from each of these fits, colour-coded by species in the same manner as the main paper. It can be seen that these standard deviations are much lower and barely change with coverage for the stearic acid samples. This is an indication that ordered packing is already present in stearic acid at low concentrations and supports the idea that for stearic acid there are islands of well-ordered acid molecules in the partially filled monolayers and that these islands primarily just grow in size on approaching full monolayer coverage.

The oleic acid and linoleic acid molecules, on the other hand, show larger spreads of their angular distributions, indicating that the presence of the kinks in the molecular chains caused by the double bonds leads to a lower degree of order in the packing of these molecules. The trend is for there to be a decrease in these spreads with increasing coverages, indicating a move towards tighter more ordered packing. For linoleic acid this decrease is sharper than for oleic acid.

Fig S9

Fig. S9 Standard deviations for the Gaussian fits for the solid angular distributions shown in Figure 4 of the main paper. Colour-coding is by species and is as in the main paper.

S8 Tilt Angle Analysis of the Upper Interfaces only

The tilt angle analysis that was carried out in order to make Fig. 4 of the main paper was symmetrised, so that it was only the alignment of the C1-C18 vector relative to the surface normal that was investigated and not its sign. This means that these graphs do not differentiate between whether the COOH was pointing upwards or downwards, but only on whether the C1-C18 vector was parallel or perpendicular to the surface normal. This symmetrisation was carried out, as it was assumed that the overwhelming majority of fatty acid molecules would be positioned with their COOH group pointing inwards and their CH_3 groups pointing outwards and very few molecules would be ‘upside-down’. Symmetrisation was carried out by measuring the distribution of the tilt angles of all fatty acid molecules (from both interfaces) simultaneously and across the range 0-180°, before averaging over tilt angle bins that were the same number of degrees from 90°.

In order to investigate the validity of this assumption in the plots below only the fatty acid molecules that were inserted in the areas of volume above the slabs are investigated. The angular distributions of these relative to the surface normal are presented in the range 0-180° without any symmetrisation and without division by $\sin(\vartheta)$, for the extreme coverages of each of the species investigated. Each point is an average over four slabs of that composition with error bars representing one standard error of the mean. 0° represents a molecule that is directed along the surface normal with its COOH pointed downwards towards the water core and 180° represents a molecule also along the surface normal but with its methyl group pointing down. As for the graph in Fig. 4 analysis has only been carried out at times >4 ns in the final 10 ns production run.

It can be seen that for all species and coverages shown in Fig. S10 there are very few molecules present with tilt angles >90°. For the oleic and linoleic acid molecules, as well as the lower coverage of stearic acid molecules, the only amplitude for molecules with angles >90° is found just above the 90° mark and is part of the tail of the broader distribution of angular values that has been attributed in the main paper to the less ordered packing of molecules within layers that are less tightly packed, or that have filled monolayers with free organic molecules in an overlayer. For the stearic acid samples the main peak is much sharper and centred at ~45°. A secondary, much smaller peak can also be seen above the x axis, centred at ~135°.This corresponds to molecules that have a similar alignment to those in the main peak but with their COOH pointing away from the water core instead of towards it. It is likely that these molecules have become trapped in this position as the monolayer packs together with increasing density, with this packing forcing them to line up with the other fatty acid molecules and steric hindrance from other molecules preventing them from rotating to a position that would orientate them with their acid groups closer to the water core. It can be seen by eye, however, that the number of molecules that this affects is relatively small, and numerical analysis of the numbers of molecules contributing to each of the peaks shows that around 2% of molecules are included in the smaller peak. This supports the use of symmetrisation when reaching the conclusions discussed in the main paper.

Fig S10

Fig. S10 Tilt angle distributions for molecules on the upper interfaces only of slabs that have 70 molecules (left hand column) and 150 molecules (right hand column) per interface. Tilt angles have been binned into one degree bins. Points are averages of times >4 ns of the 10 ns production runs of four independently grown slabs with error bars representing one error of the mean. No symmetrisation has been carried out.

S9 C1-C18 Distance Full Results

Figures S11-13 show the full C1-C18 distance distributions for all surface coverages of stearic, oleic and linoleic acids respectively, expanding on the selected coverages shown in Fig. 5 of the main text. As with Fig. 5 of the main text, each graph represents an average over four separately grown slabs, with analysis carried out on each slab at times >4 ns during the 10 ns production runs. The error bars represent one standard error of the mean over the eight interfaces for a given acid coverage. The number and type of acid molecules added to the simulation is shown above each plot. C1-C18 distances, as calculated using the gmx distance function of GROMACS, were binned into 0.05 nm bins. The most populated bin for each acid concentration is shown in the lower part of Fig. 5 of the main paper.

Fig S11

Fig. S11 C1-C18 distances for different coverages of stearic acid. Populations are averages over four slabs, and over the last 6 ns of production runs on these slabs. Error values represent one standard error of the mean over the eight interfaces of a given coverage.

Fig S12

Fig. S12 C1-C18 distances for different coverages of oleic acid. Populations are averages over four slabs, and over the last 6 ns of production runs on these slabs. Error values represent one standard error of the mean over the four slabs of a given coverage.

Fig S13

Fig. S13 C1-C18 distances for different coverages of linoleic acid. Populations are averages over four slabs, and over the last 6 ns of production runs on these slabs. Error values represent one standard error of the mean over the eight interfaces of a given coverage.

S10 ASA Results for All Functional Groups

Fig. 7 of the main text shows how the surface coverage of selected groups varies with the number of molecules present per interface. Fig. S14 (below) shows an expanded version of this, covering all functional groups within the molecules. As for the main text, each point represents the average over both interfaces of four independently generated slabs, and over the last 6 ns of the 10 ns production run for each of these, with the error bars representing one standard error of the mean. Note the difference in the y axis scales for different functional groups and the different ranges of coverage of each of the groups. Methyl group fractional coverages range from 0.35-0.75 and CH_2 coverage from 0.2 to 0.55 whereas COOH coverage is always below 0.02. HC=CH coverage is higher for linoleic (0.05-0.15) compared to oleic acid (0.01-0.08) as a result of linoleic acid having twice as many double bonds per molecule.

Fig S14

Fig. S14 Variation of surface area coverages of different groups with numbers of fatty acid molecules per interface, for all functional groups investigated. Each point represents the mean over four separately grown slabs, with the error bars representing one standard error of the mean. Only times >4 ns were included in the analysis. Note the difference in scales for each of the functional group types. This figure is an extension of Fig. 7 of the main text.

Fig. S14 Variation of surface area coverages of different groups with numbers of fatty a
functional groups investigated. Each point represents the mean over four separately
representing one standard error of the mean. Only times >4 ns were included in the analy
each of the functional group types. This figure is an extension of Fig. 7 of the main text.

S11 Hydrogen Bonding Analysis

The gmx hbond algorithm included in the GROMACS software package can be used to calculate the number of hydrogen bonds within a sample. It does this based on the number of H -- O -- H and H – N -- H interactions where the each of the hydrogen atoms is within a set distance (<0.35 nm) of the oxygen or nitrogen atom and the angular range between the three atoms is within the limits (180±30°) that would be expected when hydrogen bonding is present. These hydrogen bonds can be investigated for specific species, and here we present how the number of hydrogen bonds between acid molecules and between acid molecules and the water core is affected by increasing coverage of organic molecules.

Fig. S15 shows how the number of hydrogen bonds between acid molecules and the water core changes with increasing numbers of fatty acid molecules. Each point on these graphs represents an average of over four independently built slabs of that fatty acid coverage, with analysis carried out at times >4 ns into the 10 ns production runs. Error values associated with these averages are also plotted, however, are generally too small to be seen on top of the points. The number of fatty acid molecules per interface is on the x axis and the number of hydrogen bonds between these fatty acid molecules and the water core is on the y axis. It can be seen that typically there are just over two hydrogen bonds per carboxylic acid molecule. This is related to how the acid molecules can most effectively be oriented in order to maximise the interactions between the water and the carboxylic acid groups, whilst still maintaining close to the 180° H -- O -- H angle required between the for a hydrogen bond to form. For each of the acid species there is initially a near-linear positive relationship between the number of acid molecules and the number of hydrogen bonds per interface at lower coverages, however at the higher coverages the graphs start to reach a plateau. This suggests that the maximum number of acid molecules that can directly interact with the water core has been reached, meaning that monolayer coverage has been achieved. The fact that this plateauing effect is more prominent for linoleic and oleic acids than it is for stearic acid is, as with other effects mentioned in the main paper, likely to be a result of the greater ability for stearic acid molecules to pack more closely together and thus to fit more molecules into the monolayer.

Fig S15

Fig. S15 Plots showing how the number of hydrogen bonds between acid molecules and the water core is related to the number of acid molecules that are present per interface. Points are an average of four independently grown slabs of the same fatty acid coverage and measurements are taken at times >4 ns in a 10 ns production run.

Fig. S16 shows how the number of hydrogen bonds between molecules of acid increases with the total number of acid molecules per interface. As with Fig. S15 each point is the average of four independently grown slabs and times above 4 ns of the 10 ns production runs. For all species and coverages there are far fewer acid-acid hydrogen bonds than the number of acid molecules per interface. This is likely to be a result of the large number of hydrogen bonds between the acid molecules and the water core, leading to fewer potential acid hydrogen bonding sites for interaction with other acid molecules. In opposition to the water-acid hydrogen bonds, the number of acid-acid hydrogen bonds increases more rapidly as the surface coverage of fatty acid molecules increases. This is because as the monolayer approaches completion there are a greater number of acid molecules that are not bonded to the water core and that therefore are available to form hydrogen bonds with other acid molecules.

There are a greater number of acid-acid hydrogen bonds for the higher surface coverages of linoleic and oleic acid than for the equivalent stearic acid molecules. This is again down to the greater packing of stearic acid molecules into the monolayer at higher coverages, which leaves less free acid available to create hydrogen bonds with other acid molecules.

Fig S16

Fig. S16 Plots showing how the number of hydrogen bonds between acid molecules and other acid molecules is related to the number of acid molecules that are present per interface. Points are an average of four independently grown slabs of the same fatty acid coverage and measurements are taken at times >4 ns in a 10 ns production run.

References

1. H. J. C. Berendsen, D. Van der Spoel and R. Van Drunen, GROMACS: A Message-Passing Parallel Molecular Dynamics Implementation. Comput. Phys. Commun. 1995, 91, 43–56.

2. D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. C. Berendsen, GROMACS: Fast, Flexible, and Free. J. Comput. Chem. 2005, 26, 1701–1718.

3. D. L. Beveridge and W. L. Jorgensen, The OPLS (Optimized Potentials for Liquid Simulations) Potential Functions for Proteins, Energy Minimizations for Crystals of Cyclic Peptides and Crambin. Annu. Rev. Biophys. Bioeng. 1988, 110, 39.

4. W. L. Jorgensen, D. S. Maxwell and J. Tirado-Rives, Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. J. Am. Chem. Soc. 1996, 118, 11225–11236.

5. L. S. Dodda, I. C. De Vaca, J. Tirado-Rives and W. L. Jorgensen, LigParGen Web Server: An Automatic OPLS-AA Parameter Generator for Organic Ligands. Nucleic Acids Res. 2017, 45, W331–W336.

6. H. J. C. Berendsen, J. R. Grigera and T. P. Straatsma, The Missing Term in Effective Pair Potentials. J. Phys. Chem. 1987, 91, 6269–6271.

7. S. Miyamoto and P. A. Kollman, Settle: An Analytical Version of the SHAKE and RATTLE Algorithm for Rigid Water Models. J. Comput. Chem. 1992, 13, 952–962.

8. R. W. Hockney, The Potential Calculation and Some Applications. Methods Comput. Phys. 1970, 9, 136.

9. W. C. Swope, H. C. Andersen, P. H. Berens and K. R. Wilson, A Computer-Simulation Method for the Calculation of Equilibrium-Constants for the Formation of Physical Clusters of Molecules: Application to Small Water Clusters. J. Chem. Phys. 1982, 76, 637–649.

10. T. Darden, D. York, and L. Pedersen, Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems, J. Chem. Phys. 1993, 98, 10089–10092.

11. U. Essmann, L. Perera, M.L. Berkowitz, T. Darden, H. Lee, and L.G. Pedersen, A smooth particle mesh ewald potential, J. Chem. Phys. 1995, 103, 8577–8592.

12. A. C. Stewart, M. J. Paterson and S. J. Greaves, Chemical Functionality at the Liquid Surface of Pure Unsaturated Fatty Acids, Environ. Sci. Atmos. 2021, 1, 498-507.

13. W. Humphrey, A. Dalke, K. Schulten, VMD - Visual Molecular Dynamics. J. Molec. Graph. 1996, 14, 33–38.