Supporting Information for

Detecting dynamic domains and local fluctuations in complex molecular systems *via* timelapse neighbors shuffling

Martina Crippa, Annalisa Cardellini, Cristina Caruso and Giovanni M. Pavan

Giovanni M. Pavan
E-mail: giovanni.pavan@polito.it

This PDF file includes:

Figs. S1 to S19
Table S1
Legends for Movies S1 to S3
SI References

Other supporting materials for this manuscript include the following:

Movies S1 to S3

SYSTEM	r_{cut} [Å]	# of g(r)$_m$ peaks	Length of MD[ns]	# of sampled frames	Sampling Δt [ns]	LENS center
DPPC/DIPC Lipids	16	3	10000	1001	10	PO4
F-NP/H Micelle	16	3	3000	3001	1	HEAD (CoM)
DPPC Lipids 293K	16	3	1000	1001	1	PO4
DPPC Lipids 273K	16	3	1000	1001	1	PO4
DPPC Lipids 323K	16	3	1000	1001	1	PO4
TIP4P/Ice Water	7.4	3	50	500	0.1	OW
Cu(210) 700K	4.9	3	150	502	0.3	Cu
Cu(211) 600K	4.9	3	150	502	0.3	Cu
Au-NP 200K	4.4	2	1000	1000	1	Au
BTA	15.8	3	20000	20001	1	BENZ (CoM)

Table S1. Setup details of all the LENS analyses conducted in this work.

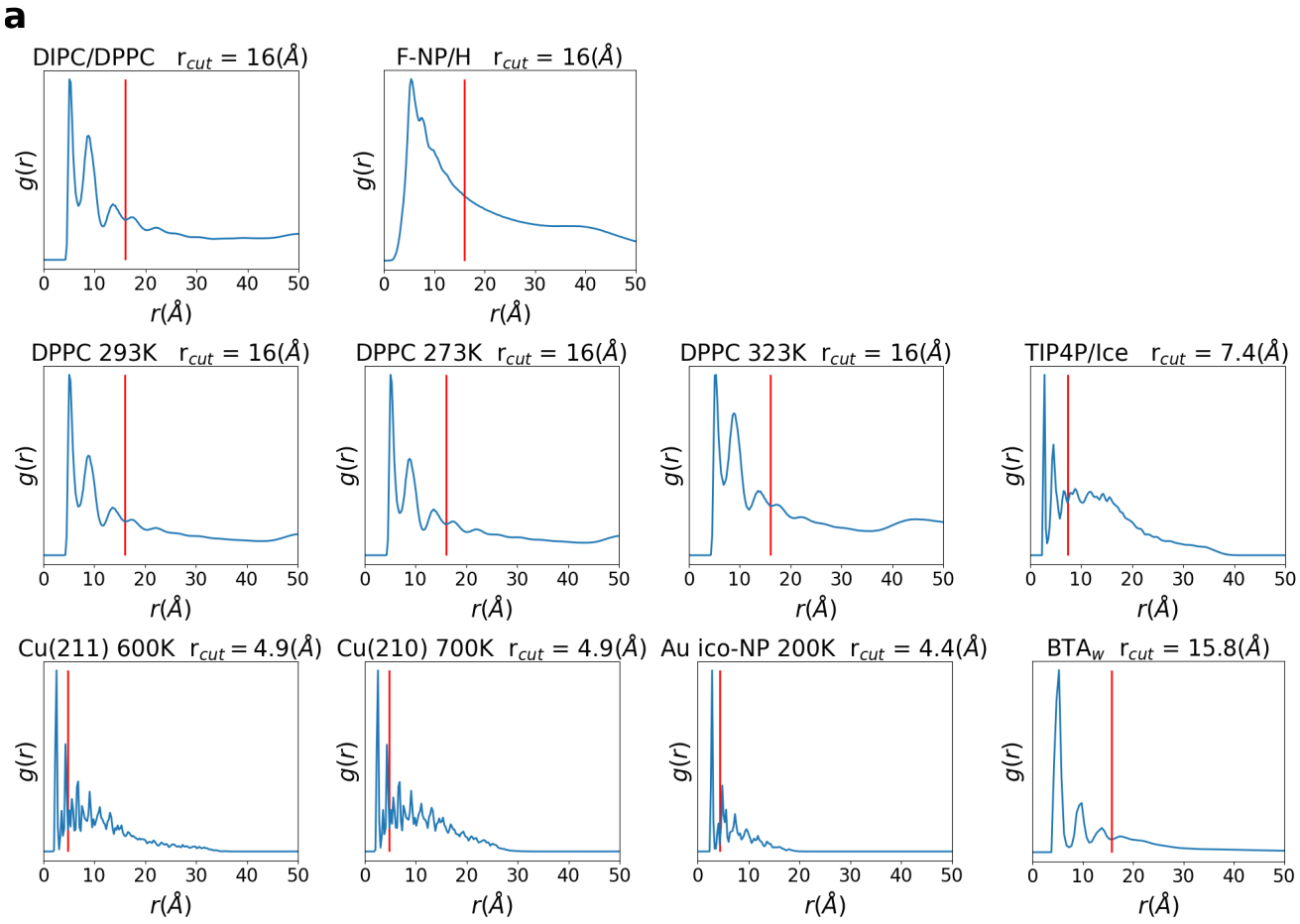


Fig. S1. (a) Radial distribution functions (g(r)) and cut-off radius r_{cut} for all systems.

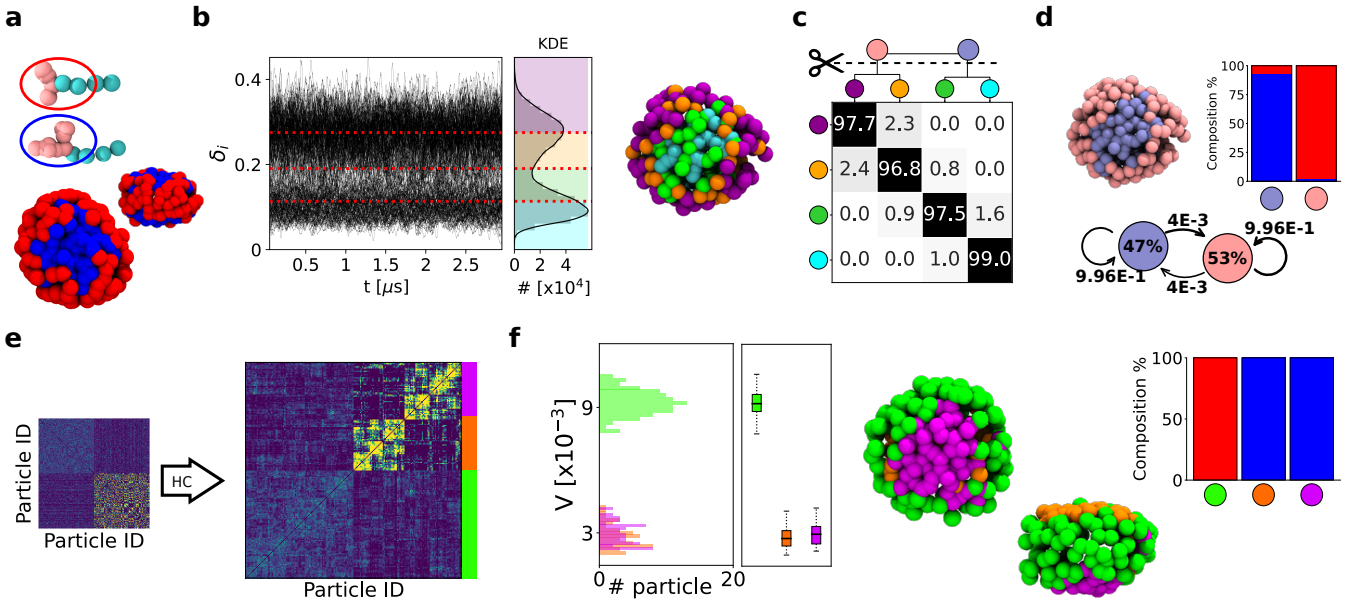


Fig. S2. LENS analysis of fluid-like systems. (a) Bicomponent amphiphile micelle composed of 100 **H** surfactants and 100 **F-NP** surfactants colored in red and blue, respectively. (b) Time-series of LENS signals, $\delta_i(t)$, with the Kernel Density Estimate (KDE) of LENS distribution classified into four clusters (left). MD snapshot of the micelle colored according to their clusters of belonging (right). (c) Inter-clusters normalized transition probability matrix. The p_{ii} and p_{ij} matrix entries indicate the $\%$ probability that molecules with LENS signal typical of a cluster i remain in that dynamical environment or move to another one j (with different dynamics) in Δt. Hierarchical grouping of the dynamically-closer clusters (dendrogram cutting) is reported on top of the matrix, and it provides two macroclusters, merging cyan and green on one hand, and orange and purple on the other hand. (d) MD snapshot of the micelle colored according to macroclusters in (c): light-blue identifying **F-NP** surfactants, pink identifying **H** surfactants (top-left). Cluster composition histogram (top-right) and interconversion diagram (bottom) with the transition exchange probabilities and the cluster population percentages (within colored circle). (e) HC analysis of the D^T matrix identifying three main clusters (green, purple, orange). (f) *Variability*, V, analysis of the clusters: distributions, median (first and third quartiles), maximum and minimum values (whiskers). The green have higher V than the orange and magenta clusters (left). MD snapshot front and lateral view of the micelle colored according the HC clustering of D^T matrix (middle). Cluster composition histogram (top-right): the green cluster is made of **H** surfactants (in red in (a)), while the orange and magenta ones correspond to the **F-NP** surfactants (in blue in (a)).

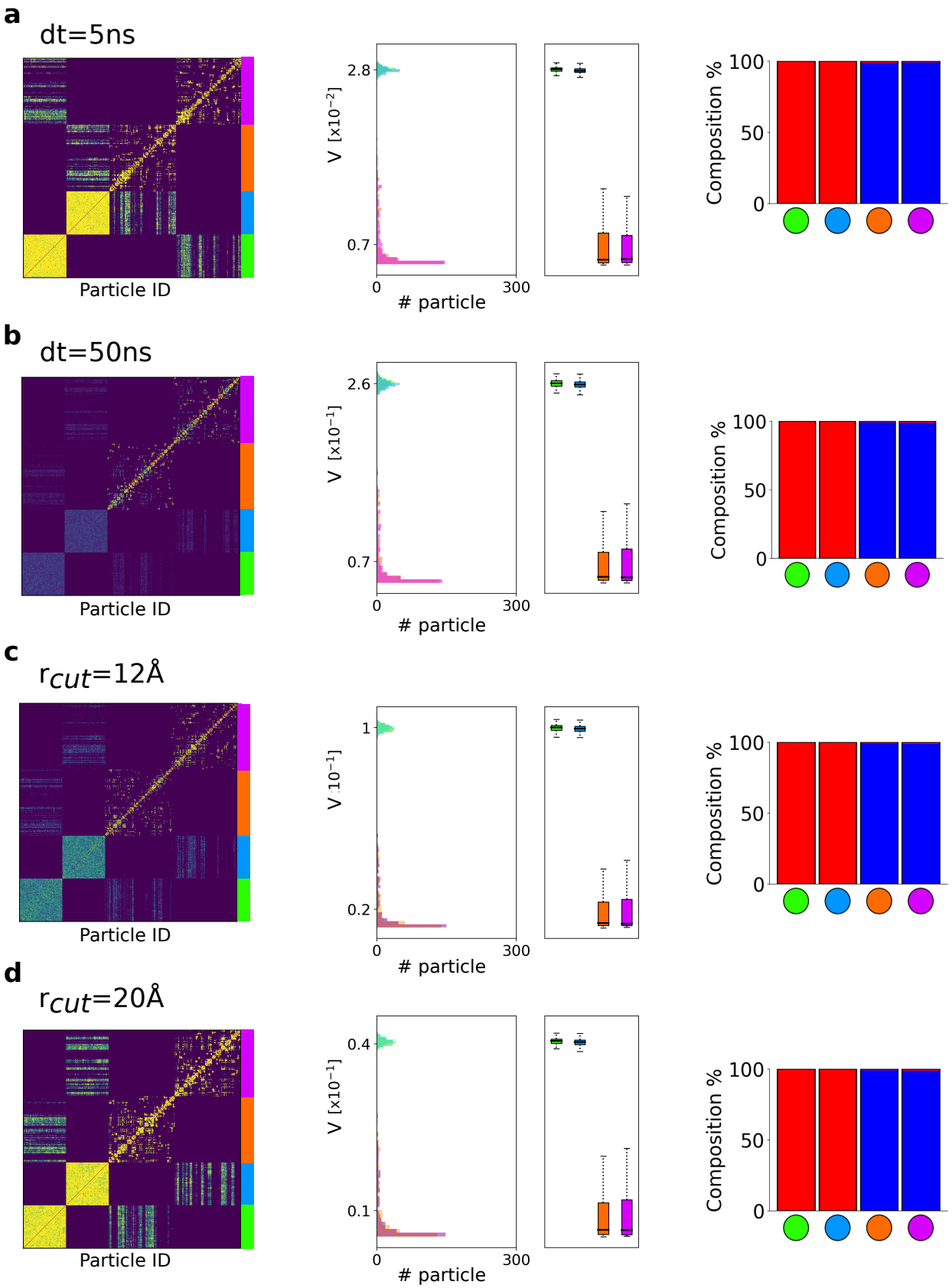


Fig. S3. Statistical analysis for **DIPC**/**DPPC** lipid bilayer, varying sampling step Δt or neighborhood cutoff radius r_{cut} while keeping all the other parameters as reported in Table S1. (a) $\Delta t = 5$ ns, (b) $\Delta t = 50$ ns, (c) $r_{cut} = 12$ Å and (d) $r_{cut} = 20$ Å.

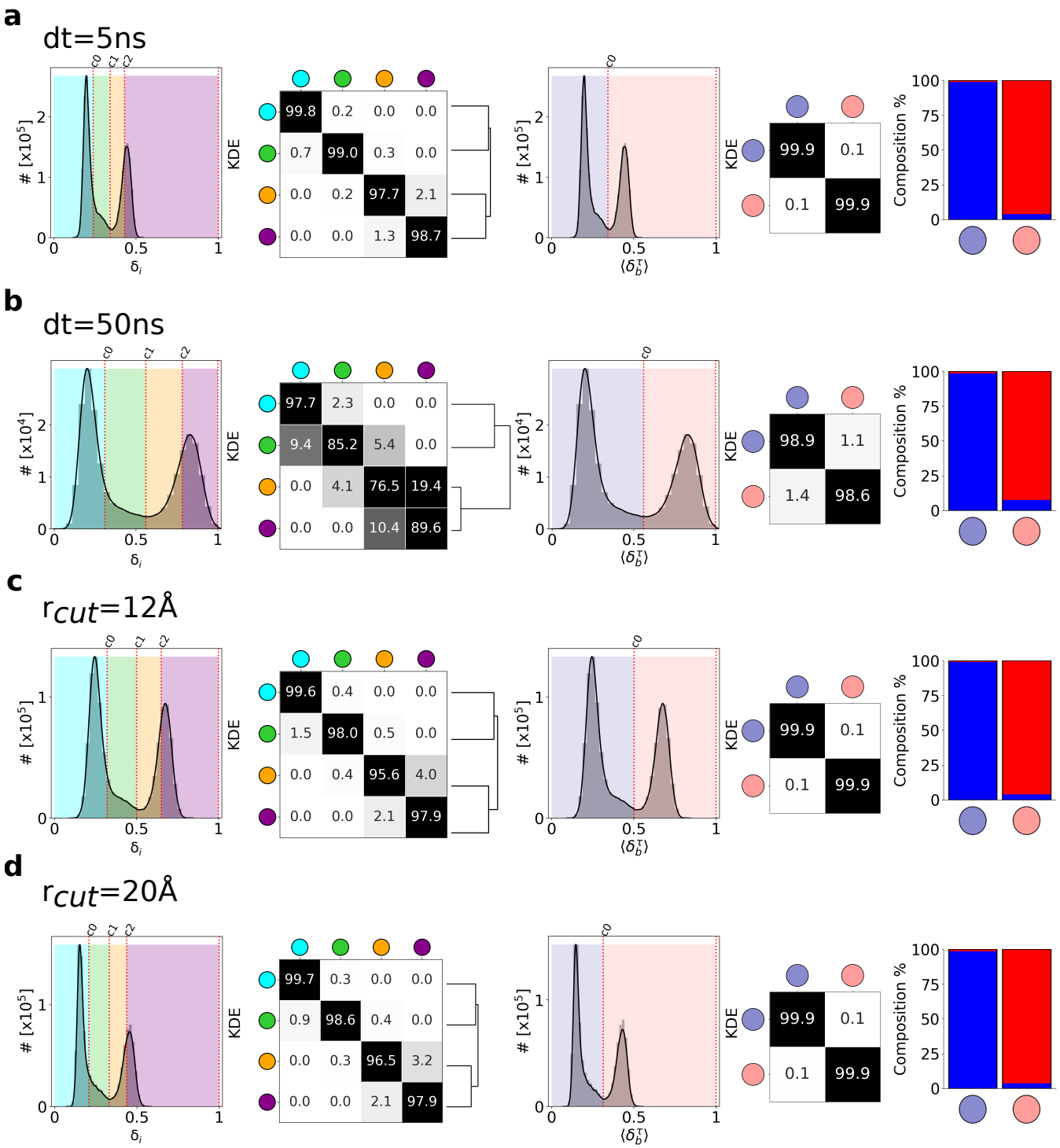


Fig. S4. LENS analysis for **DIPC**/**DPPC** lipid bilayer, varying sampling step Δt or neighborhood cutoff radius r_{cut} while keeping all the other parameters as reported in Table S1. (a) $\Delta t = 5$ ns, (b) $\Delta t = 50$ ns, (c) $r_{cut} = 12$ Å and (d) $r_{cut} = 20$ Å. Comparison of these results demonstrates the robustness of the LENS analysis: while the microscopic information captured by the analysis may change with the Δt or r_{cut}, a "zoom out" via grouping the adjacent microclusters based on the hierarchical dendrograms provide consistent results (right).

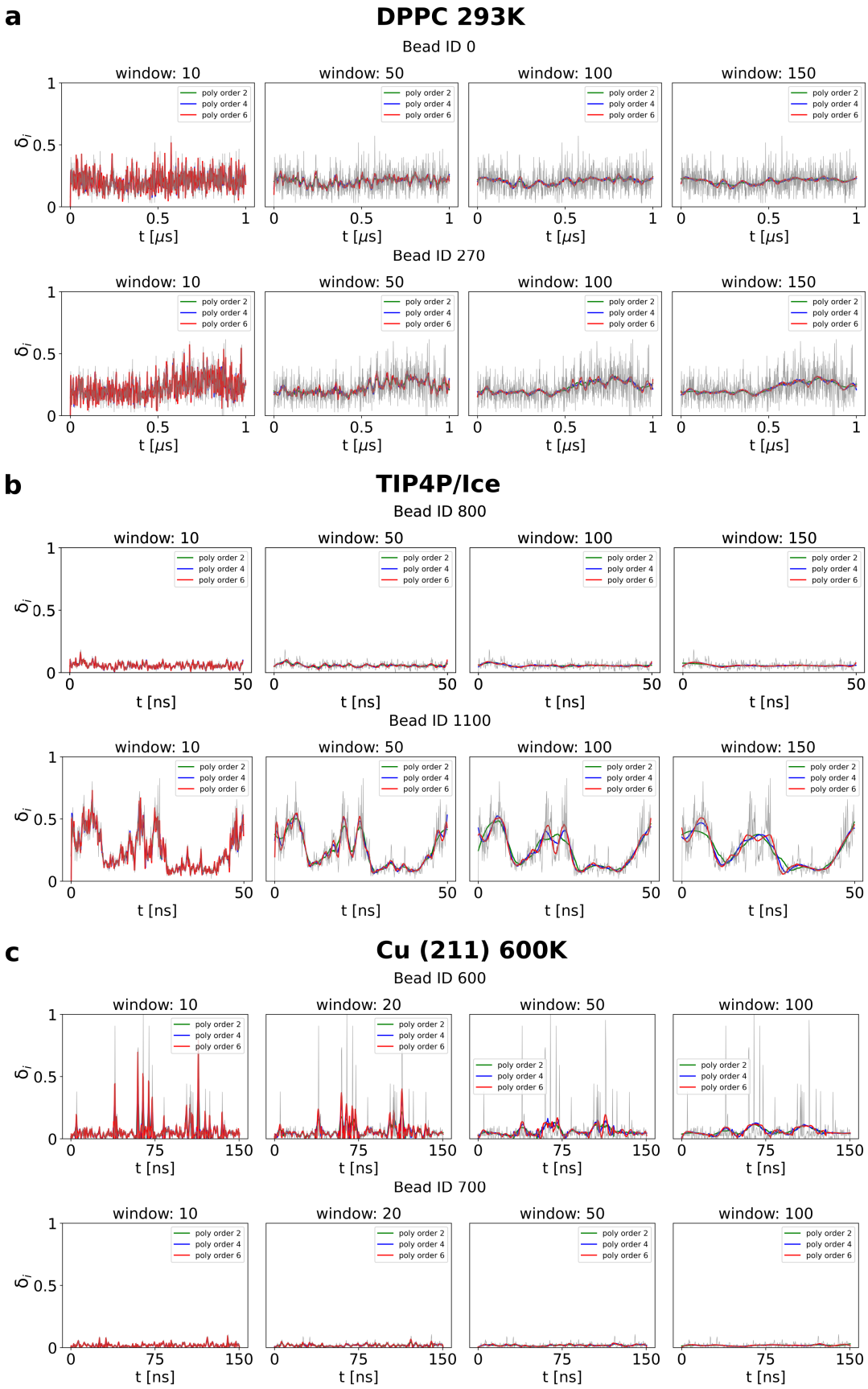


Fig. S5. Parameter study for applying the Savitzky–Golay filter varying both window interval and polynomial order for two bead examples each of (a) **DPPC** $293\ \mathrm{K}$, (b) **TIP4P/Ice** water and (c) **Cu (211)** copper slab.

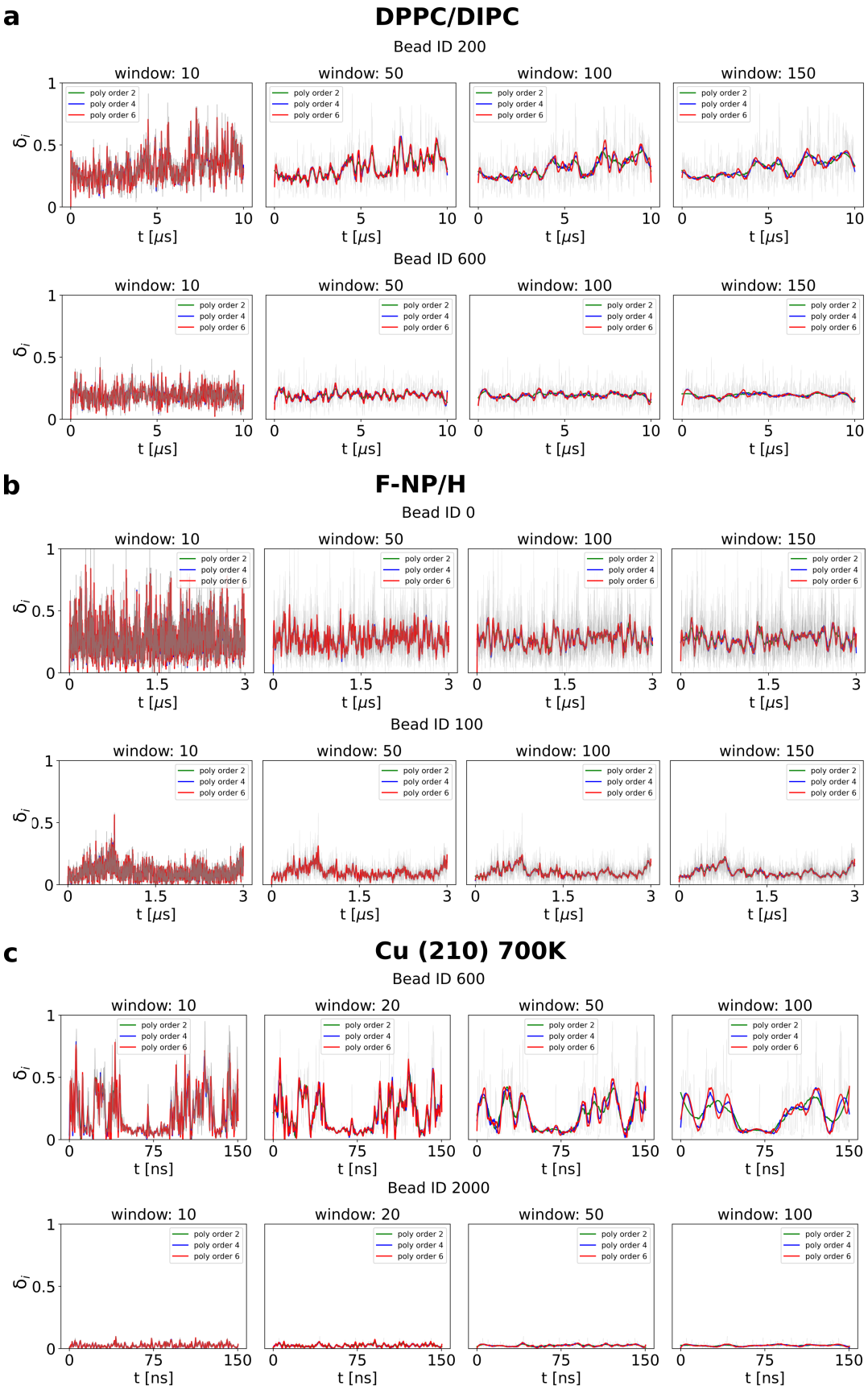


Fig. S6. Parameter study for applying the Savitzky–Golay filter varying both window interval and polynomial order for two bead examples each of (a)**DPPC**/**DIPC** lipid bilayer, (b) **F-NP**/**H** micelle and (c) **Cu (210)** copper slab.

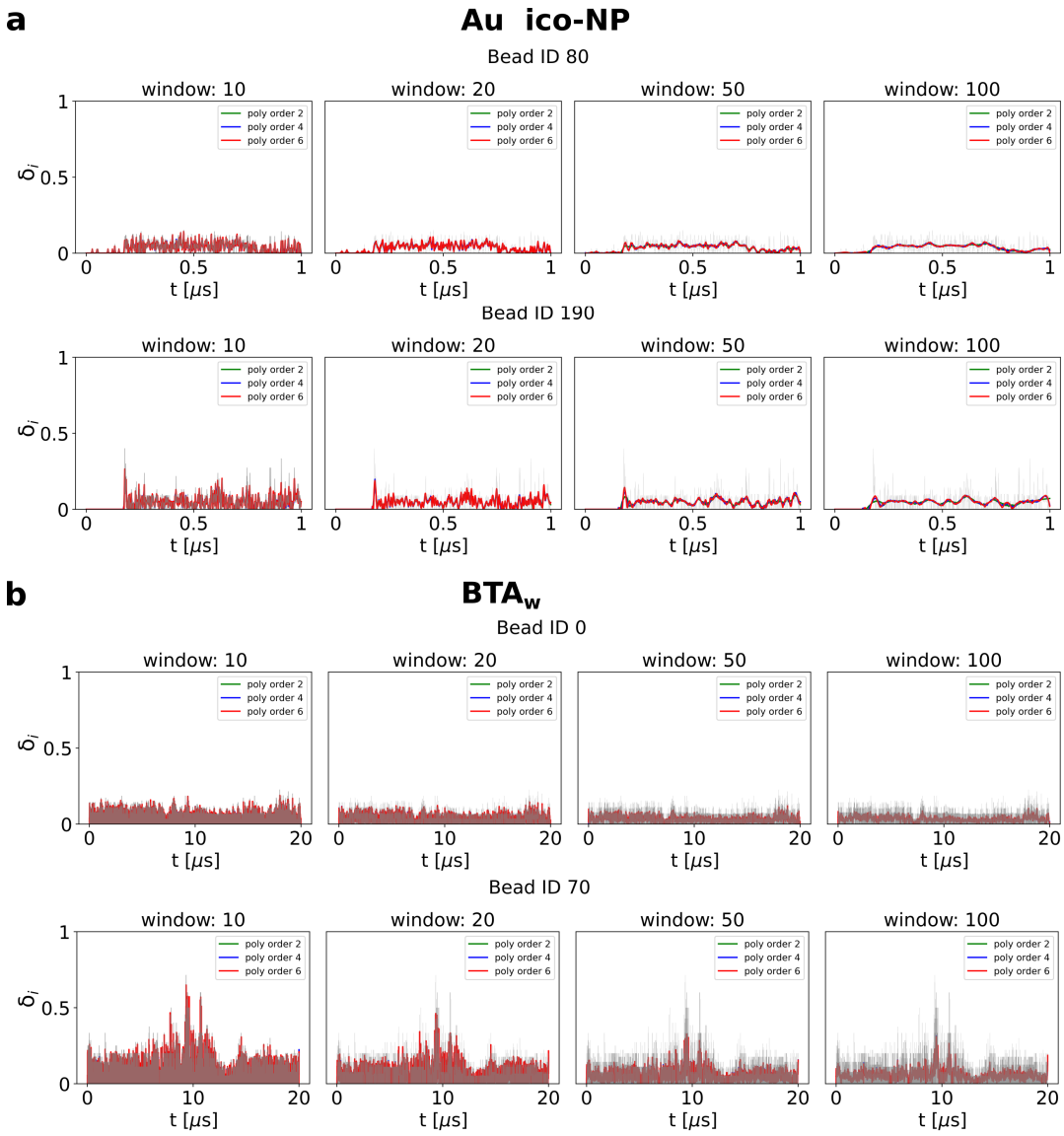


Fig. S7. Parameter study for applying the Savitzky–Golay filter varying both window interval and polynomial order for two bead examples each of (a) **Au-NP** and (b) **BTA**.

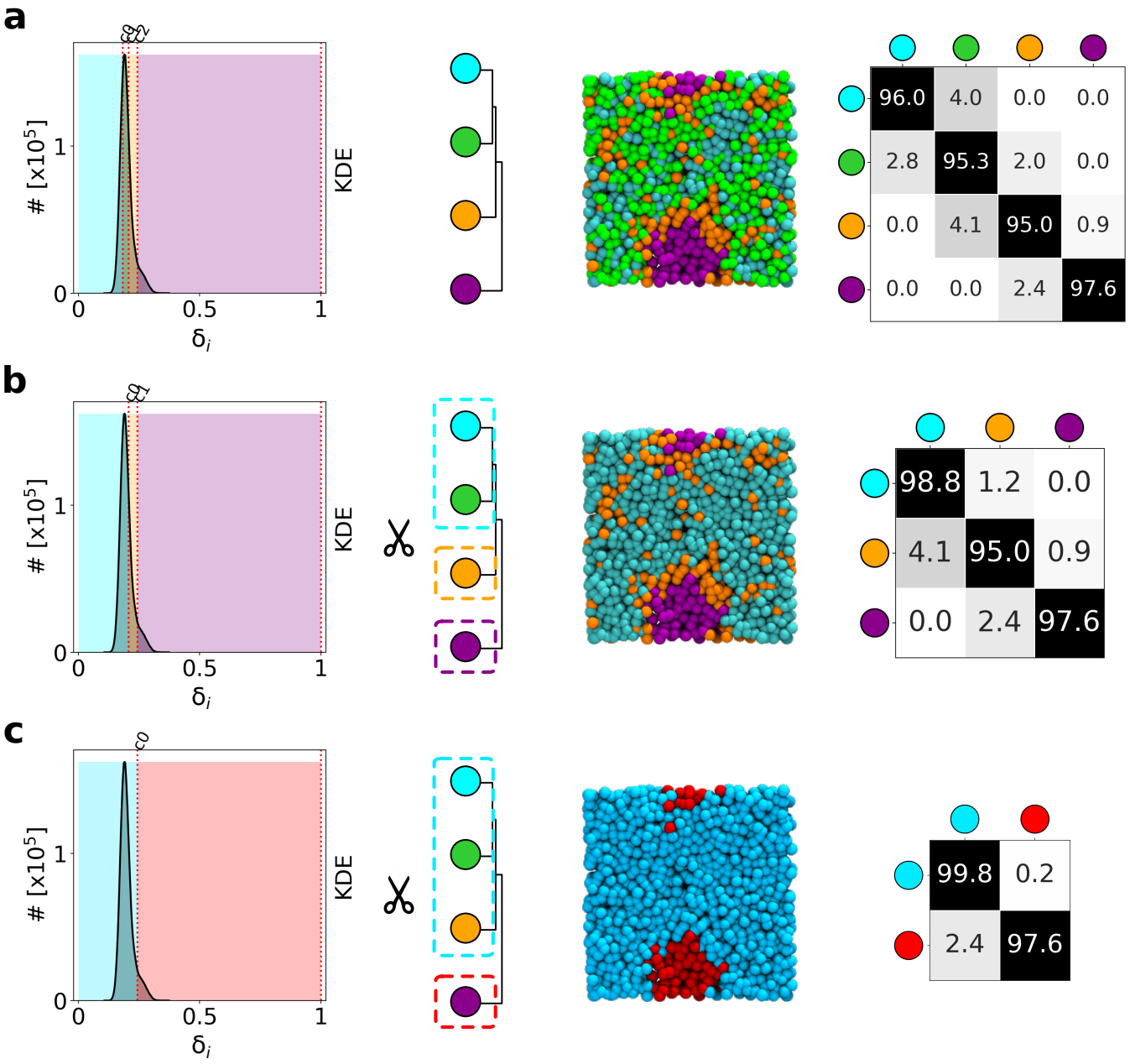


Fig. S8. DPPC Lipid bi-layer at $T = 293$ K: merging LENS clusters at different levels (no merging (a), two clusters merged into one (b) and three clusters merged into one (c)) following the hierarchy given by the dendrogram, example snapshot and transition probability matrix.

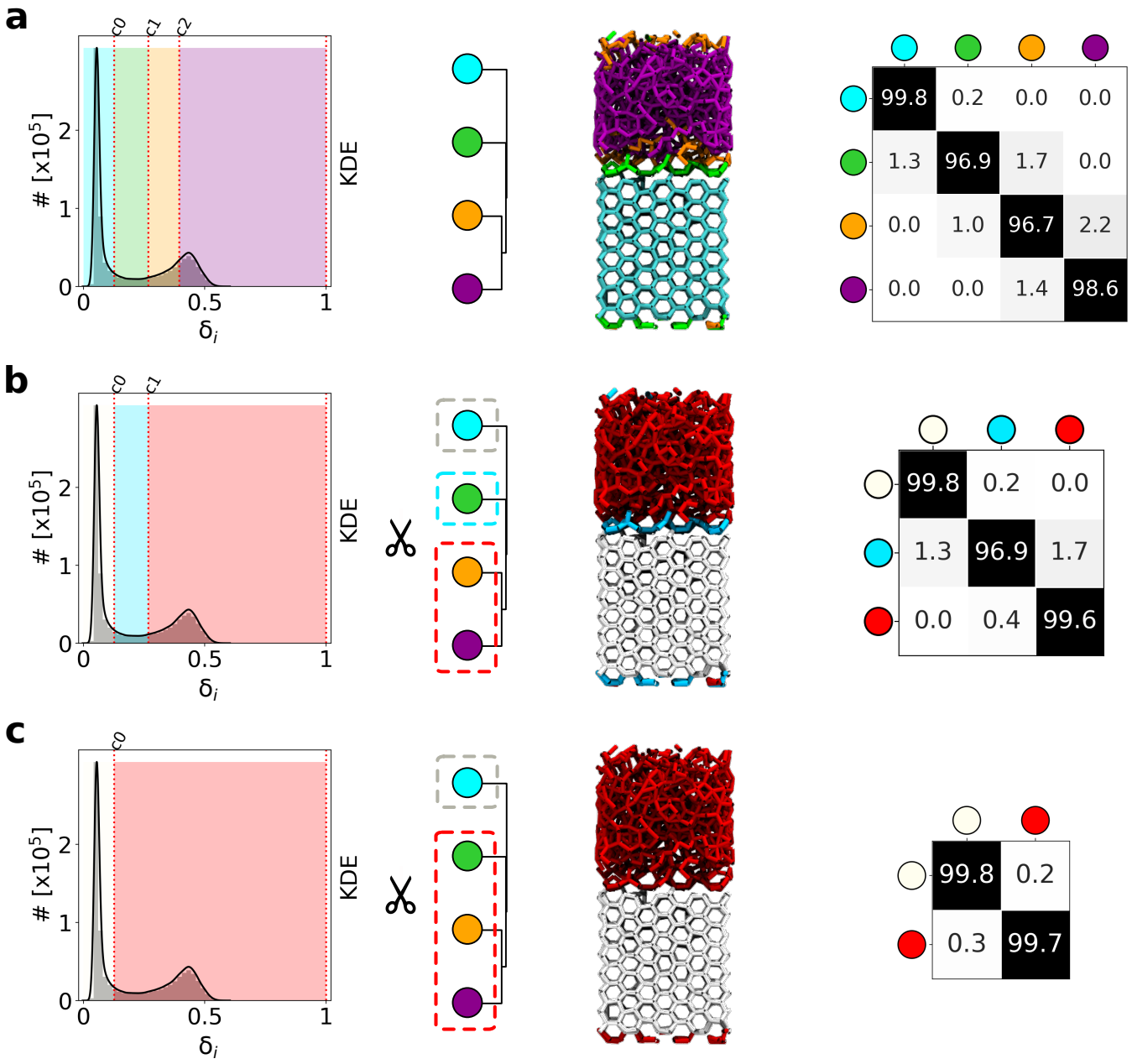


Fig. S9. TIP4P/Ice froze/melted water: merging LENS clusters at different levels (no merging (a), two clusters merged into one (b) and three clusters merged into one (c)) following the hierarchy given by the dendrogram, example snapshot and transition probability matrix.

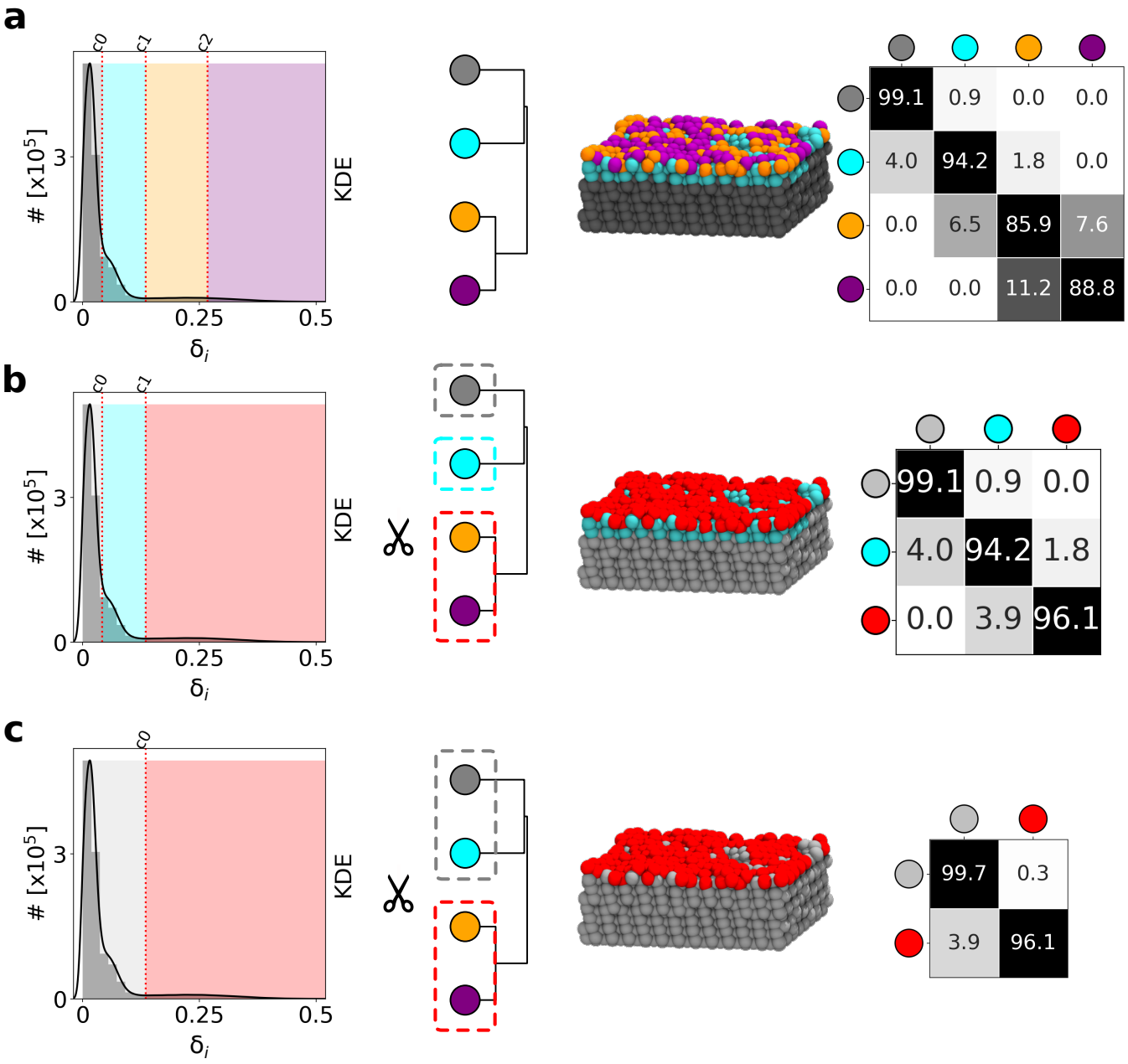


Fig. S10. Cu (210) copper slab at $T = 700$ K: merging LENS clusters at different levels (no merging (a), two clusters merged into one (b) and four clusters merged into two (c)) following the hierarchy given by the dendrogram, example snapshot and transition probability matrix.

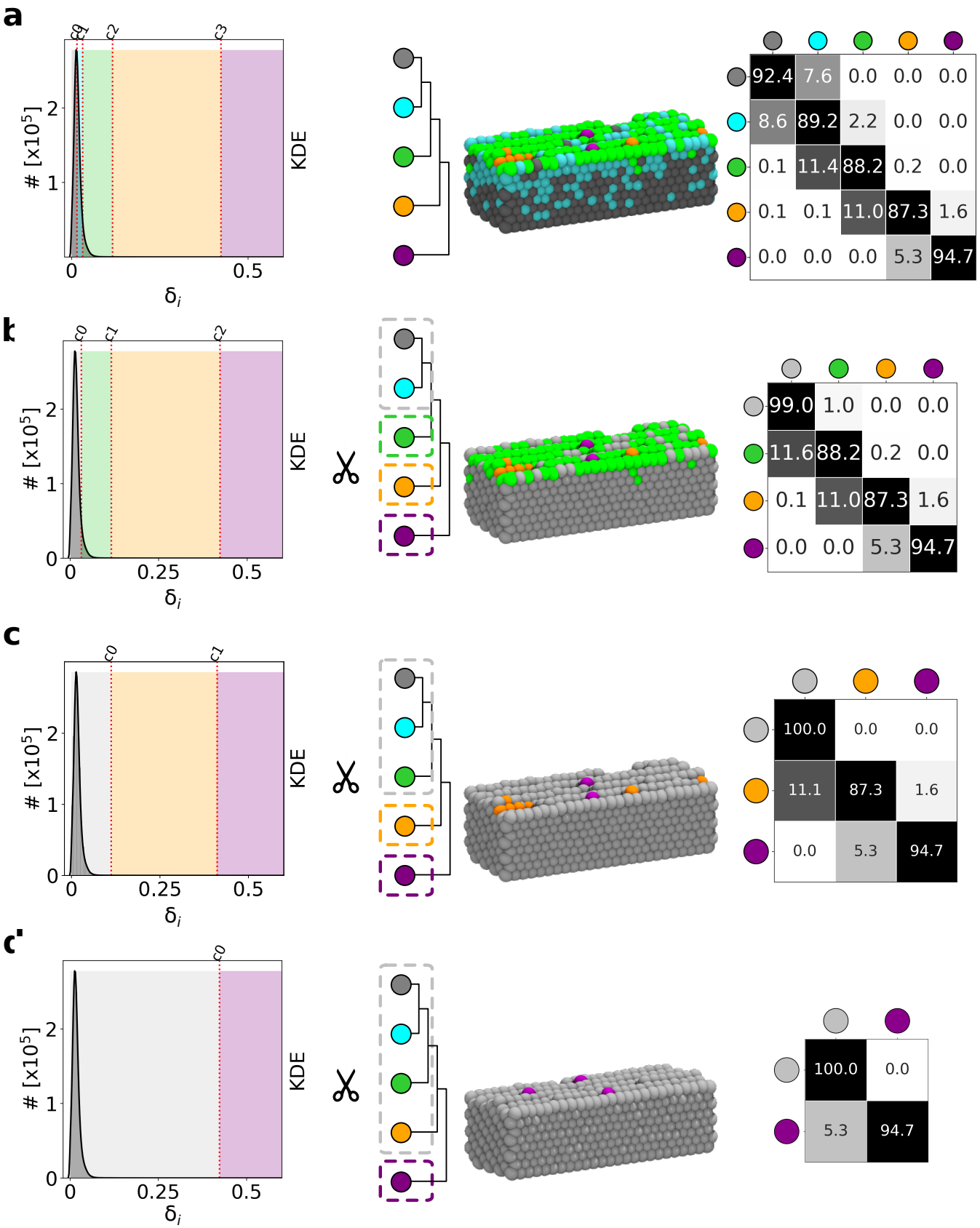


Fig. S11. Cu (211) copper slab at $T = 600$ K: merging LENS clusters at different levels (no merging (a), two clusters merged into one (b), three clusters merged into one (c) and four clusters merged into one (d)) following the hierarchy given by the dendrogram, example snapshot and transition probability matrix.

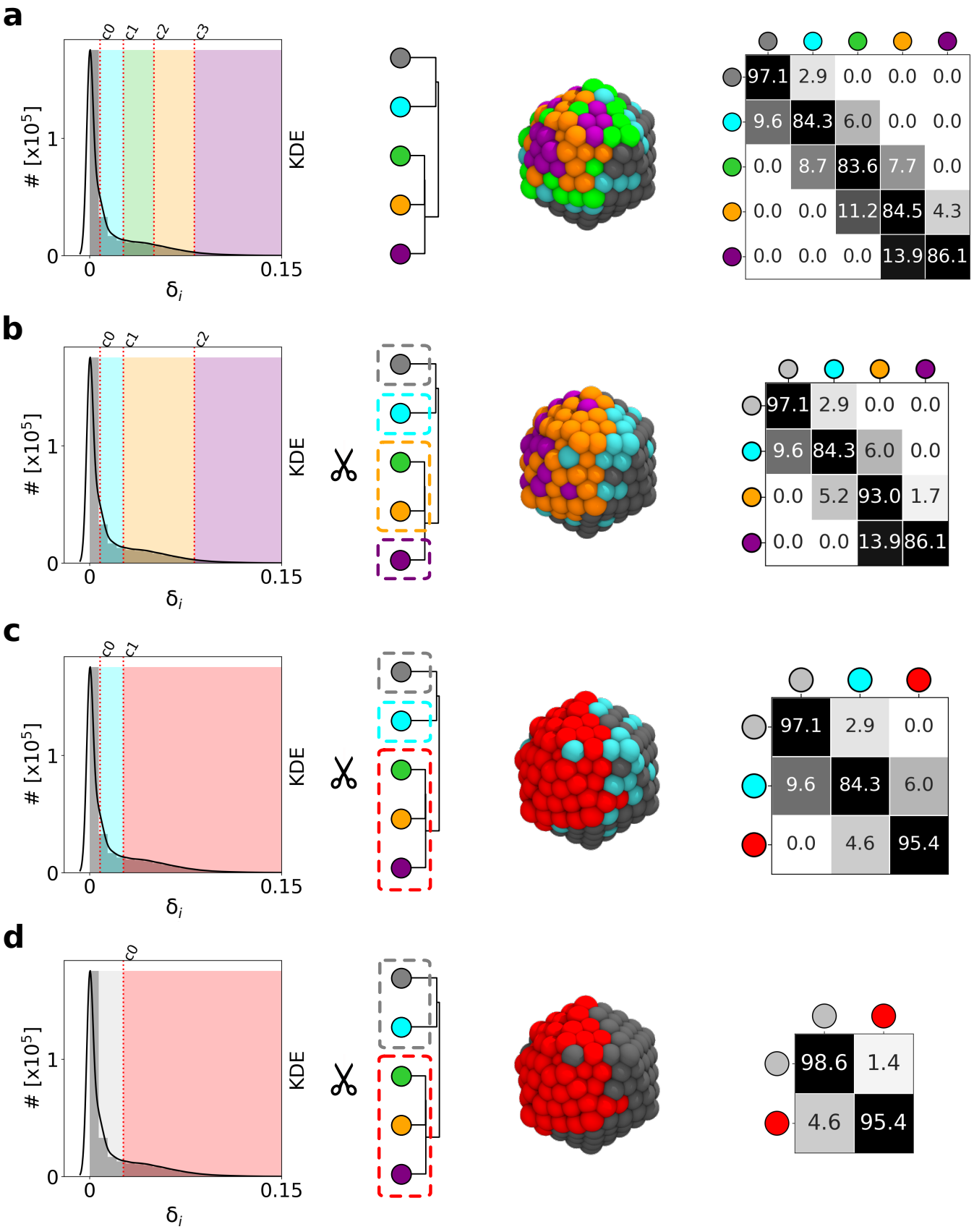


Fig. S12. Au-NP nanoparticle $T = 200$ K: merging LENS clusters at different levels (no merging (a), two clusters merged into one (b), three clusters merged into one (c) and three clusters and two clusters merged into two (d)) following the hierarchy given by the dendrogram, example snapshot and transition probability matrix.

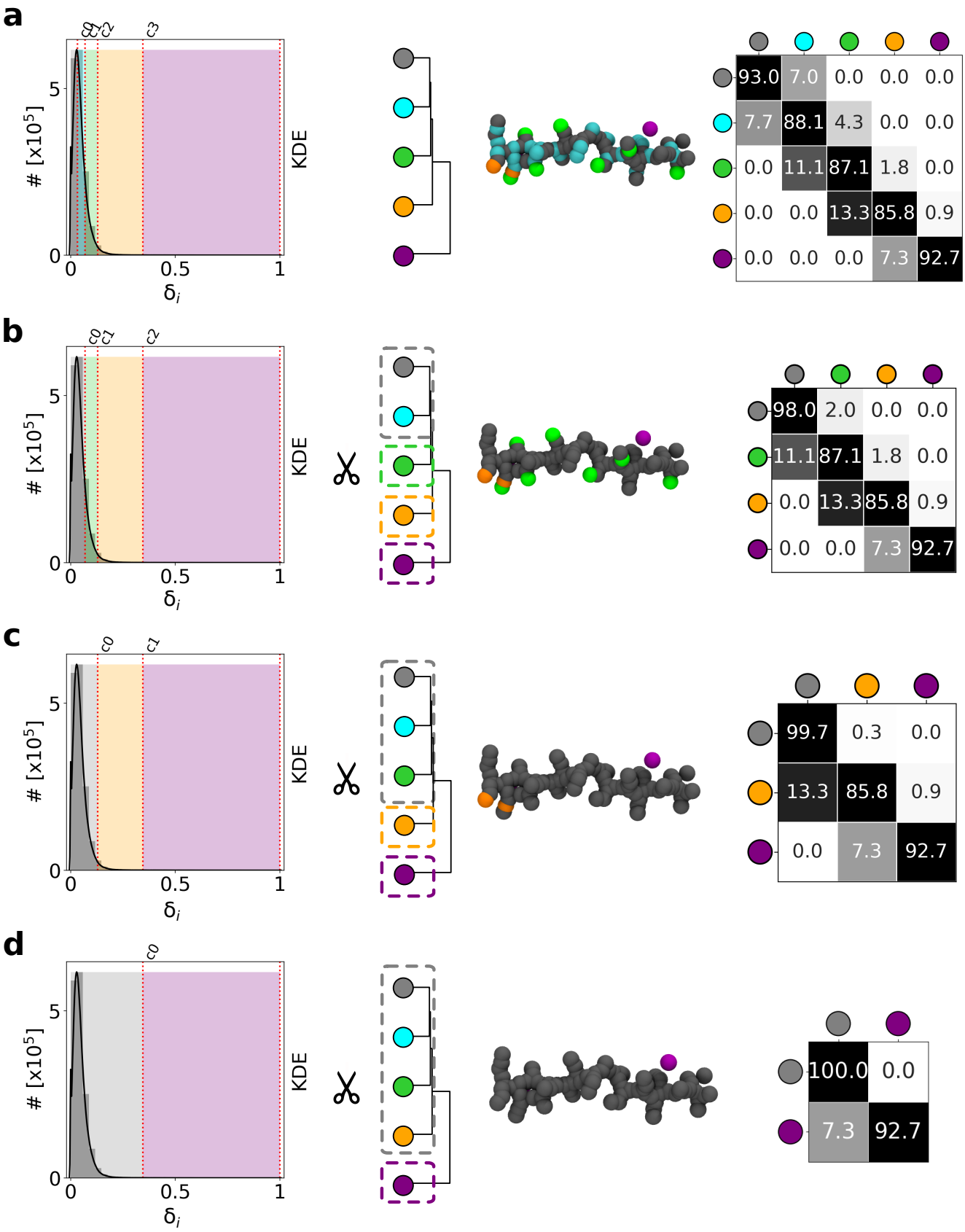


Fig. S13. BTA fiber: merging LENS clusters at different levels (no merging (a), two clusters merged into one (b), three clusters merged into one (c) and four clusters merged into one (d)) following the hierarchy given by the dendrogram, example snapshot and transition probability matrix.

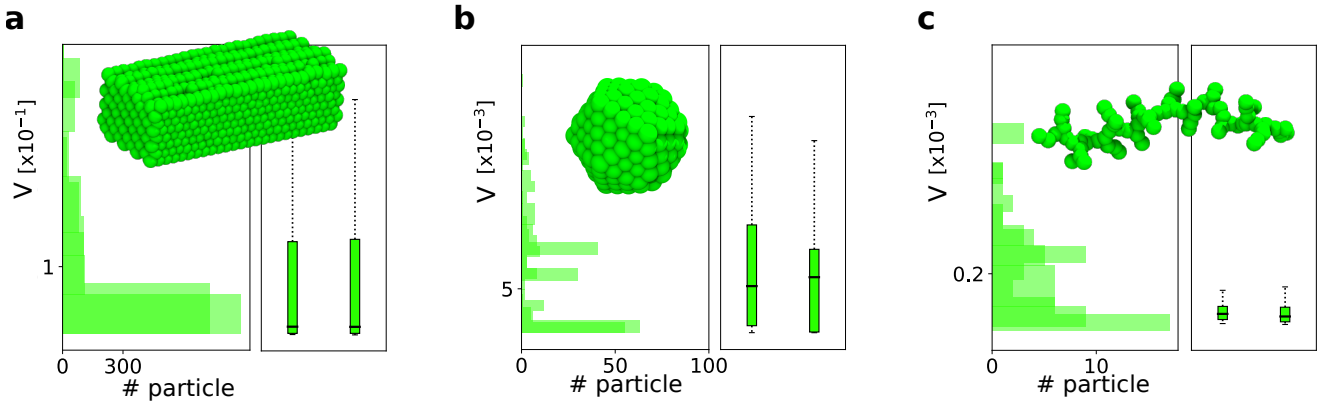


Fig. S14. Time-independent statistical analysis for **Cu (211)** copper slab at $T = 600$ K(a), **Au-NP** nanoparticle at $T = 200$ K(b) and **BTA** fiber (c): when the system is characterized by discrete and fluctuation-like dynamics, a time-indepentend averaged analysis fails to recognize patterns which are not statistically relevant. For example, in the systems reported above, the two clusters identified by HC have identical variability V and they can be classified as the same cluster.

1. Voronoi analysis

The data reported in Figure S15 show a comparison between LENS- vs. Voronoi-based clustering in case of the lipid-species compartmentalization seen in a DIPC-DPPC lipid bilayer at 280K (which is expected experimentally (1)). Figure S15a (left) shows a top view of the lipid bilayer where it is clear how the DIPC (red lipids) and DPPC (blue lipids) are self-segregated in this system (see Figure 2). The LENS-based clustering identifies two main dynamic domains: a light violet cluster (smallest LENS values), fitting almost exactly with the blue DPPC lipids, and a pink cluster (largest LENS values), fitting essentially with the red DIPC lipids (see the composition bar plot in Figure S15a). Figure S15a shows the Voronoi analysis estimated with APL@Voro v3.0 (2) where each Voronoi area/tassel (i.e. A_v), is colored on the basis of its size. Visually, such a comparison shows how the Voronoi analysis of the lipid qualitatively correlates with the LENS results. Small values of A_v - and hence, a more compact aggregation of the lipids – qualitatively correspond to the less dynamic light-violet LENS cluster (with the blue DPPC lipids, in gel-phase in such conditions). On the other hand, large Voronoi areas roughly corresponds to the more dynamic pink LENS cluster (red DIPC lipids, in liquid phase in these conditions). A systematic study of the effect of setting the lipid threshold areas (A_t) for the Voronoi classification of lipids in liquid vs. in gel phase has revealed how the detection of different dynamical domains based on structural factors (Voronoi) is less robust than the one that can be obtained by directly monitoring the microscopic dynamics in the neighborhood of each lipid in these systems (via LENS). Figure S15b reports four reference-choices of the A_t and the consequently obtained results: Top-left: $A_t = 0.561\text{nm}^2$ is the average Area Per Lipid (APL) computed on the entire lipid bilayer under study. The representation of clusters in violet and yellow and their composition (Figure S15b top left) demonstrates a reliable classification in terms of cluster size (lipids population in %) and an overall good qualitative matching with the lipid species reported above in blue and red. The violet cluster mostly matches with the light violet LENS-cluster (blue lipids), while the yellow domain mainly corresponds to the pink LENS-domain (red lipids). Obviously, the same happens to the correlation between the yellow (liquid) and violet (gel) Voronoi clusters, which correlate in good approximation with the red DIPC and blue DPPC lipids, but less precisely than what obtained with LENS (panel a). As additional cases, we also report the results obtained with an $A_t = 0.574\text{nm}^2$. This is threshold area value providing the best agreement with the LENS analysis. As it is clear, the results are very similar to those obtained with $A_t = 0.561\text{nm}^2$: between the two analysis there is correlation, but not perfect agreement. As the last two demonstrative cases, we show what results are obtained by decreasing or increasing over the A_t threshold value (bottom-left and -right respectively). The results of these analyses demonstrate how in such cases the Voronoi analysis becomes less accurate in capturing the two liquid and gel phases present in this system (the bilayer appearing more and more liquid/gel while using lower-and-lower/larger-and-larger A_t threshold values).

Figure S16 shows the same comparison between LENS- and Voronoi-based clustering in the case of the liquid phase nucleation and liquid-gel phase coexistence in the DPPC lipid bilayer at 293K. Figure S16a left shows the LENS-based clustering of DPPC lipids at 293K. As described in the main text of our paper, the red and cyan LENS clusters identify well the liquid and gel phases (large and small LENS signals/variability respectively). This is further validated in the table of Figure S16c, reporting the Mean Square Displacement (MSD) of the DPPC lipids in the cyan (gel) and red (liquid) LENS clusters at 293K (2nd row). By comparing the first two rows of the table it is clear the MSD of the lipids in the cyan and red LENS domains are found in the same order of magnitude of the MSD expected for lipids in the gel phase (cyan LENS lipids' MSD similar to the MSD of DPPC lipids at 273K, where they are fully in gel phase) and in the liquid phase (red LENS lipids' MSD in the same order of the MSD of DPPC lipids at 323K, where they are fully in the liquid phase). On the other hand, a 2D Voronoi analysis is found less efficient in discriminating the two phases, and the microscopic nucleation and coexistence of one phase into the other. Figure S16a (right) shows the Voronoi tessellation estimated with APL@Voro v3.0 (2). As in new Figure S15, in order to quantitatively distinguish the two phases, and thus to classify each Voronoi polyhedron area as liquid or gel, in Figure S16 we compare the results obtained using four different threshold Voronoi area values, namely $A_t = 0.478\text{nm}^2$, $A_t = 0.510\text{nm}^2$, $A_t = 0.400\text{nm}^2$, $A_t = 0.600\text{nm}^2$, and the relative classification of DPPC lipids into liquid ($A_v > A_t$) or gel ($A_v < A_t$) phases, colored in yellow and violet respectively. $A_t = 0.478\text{nm}^2$ corresponds to the average Area Per Lipid (APL) computed for this DPPC bilayer at 293K (very close to the experimental one of $APL = 0.473\text{nm}^2$ for DPPC at 293K, see Figure S16b, top left). In this case, as demonstrated by the composition bar plot, the yellow and violet Voronoi clusters are equally populated, and both mostly composed of cyan LENS lipids (in the gel phase). This is also demonstrated by the MSD values obtained for the two yellow and violet clusters (3rd row in the table of Figure S16c), showing how the two detected Voronoi clusters do not fit with a liquid vs. gel environments (MSD close in both cases to that of a gel system). As done for the previous case, we also optimized the choice of the threshold Voronoi area to maximize the correspondence between the LENS and Voronoi clusters – case with $A_t = 0.510\text{nm}^2$ in Figure S16b (top right). Despite slightly improved results, also in this case we clearly observe how a structural-based, Voronoi analysis is less efficient and accurate than our LENS analysis in detecting the two distinct phases. The last results of Figure S16 demonstrate also how increasing/decreasing further the At generates worse results, tending to detect one single phase. These outcomes demonstrate how, in such a case – where, in particular, the reduced statistical presence of the nucleating phase (compared to the statistical weight of the dominant phase) makes it even more difficult to detect it – LENS demonstrates a remarkable efficiency in achieving this goal, while a structural-based analysis (such as e.g. the 2D Voronoi tessellation used herein) is in comparison less efficient.

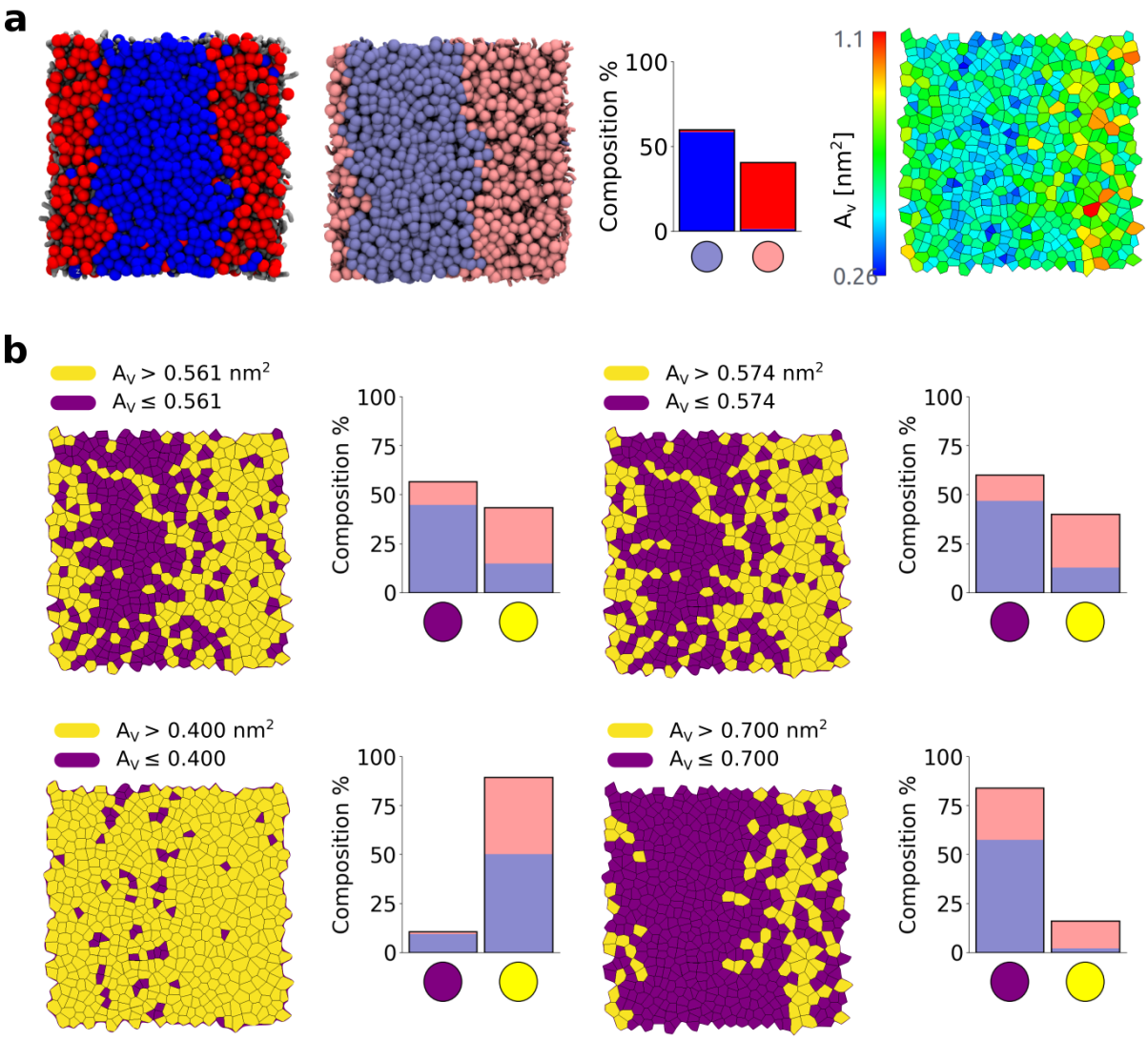
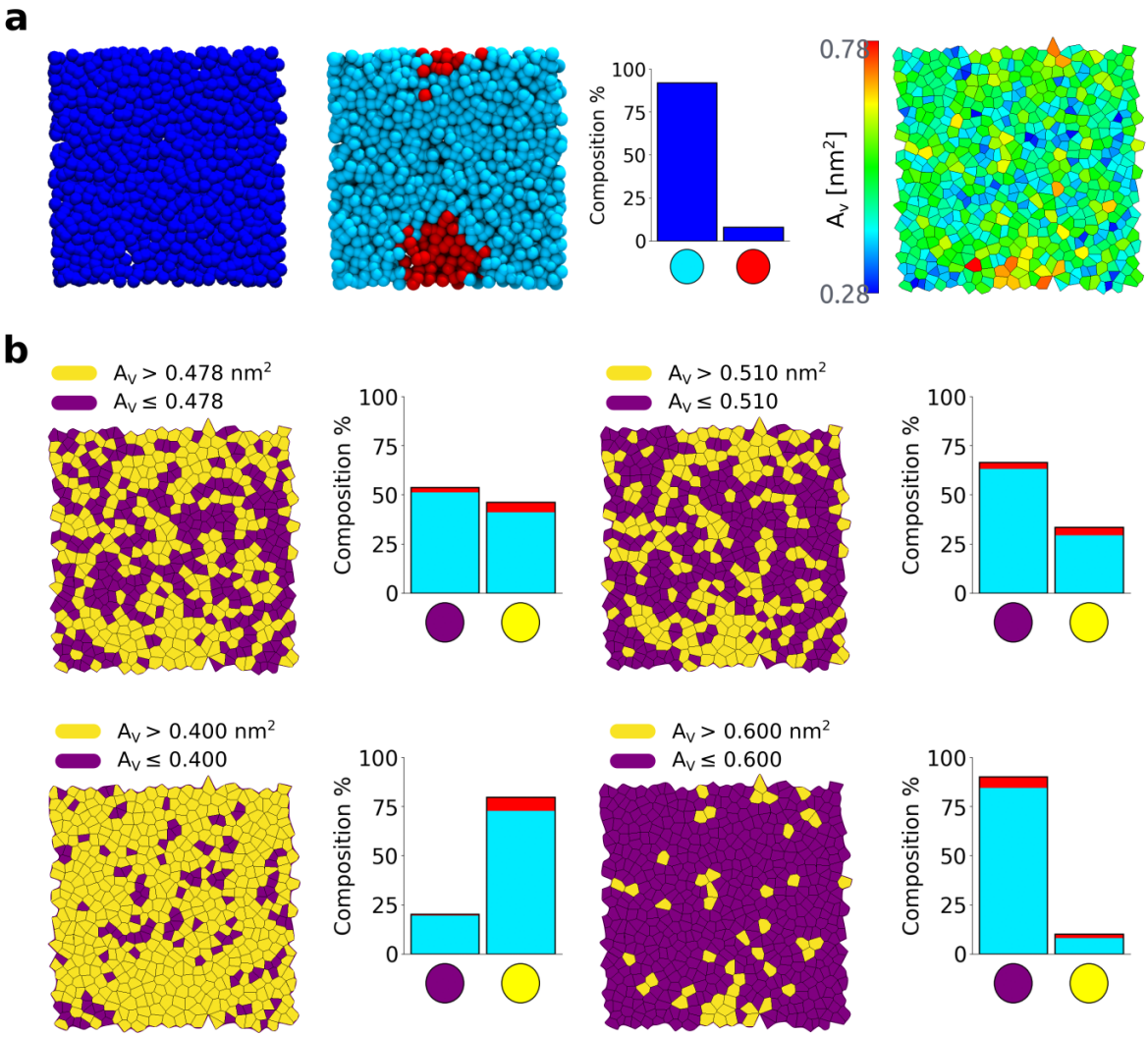


Fig. S15. Voronoi analysis for **DIPC-DPPC** lipid bilayer at $T = 280\mathrm{K}$. (a) Left: Top view of the bicomponent lipid bilayer colored by component (DIPC in red, DPPC in blue) and by LENS cluster assignment (light violet and pink clusters identify small and high LENS values: respectively, gel and liquid phases). The overlap of DIPC-DPPC components with LENS clusters and their composition percentages are shown in the histogram. Right: example of the Voronoi tessellation where each Voronoi area A_v, is colored based on its size. (b) Voronoi clustering based on a selected threshold area At equal to: the average Area Per Lipid in the system (top left: $A_t = 0.561\mathrm{nm}^2$), the value of A_t maximizing the match between Voronoi and LENS analyses (top right), and results obtained with smaller and higher At values (bottom left and right respectively). Cluster color code: yellow for Voronoi area $A_v > A_t$ (liquid) and violet for $A_v \leq A_t$ (gel). The histograms show the LENS/Voronoi clusters matching for all cases. While the results show that an optimized Voronoi analysis matches qualitatively well with the LENS one, the results show how LENS detects in more accurate and robust way the fact that the two gel and liquid domains correspond to segregated DPPC and DIPC domains, consistently with the experimental evidence.(1)



c

	DPPC 273K (gel)	DPPC 323K (liquid)	DPPC 293K
MSD [Å^2]	25.96	391.66	40.88
	DPPC 293K LENS cyan cluster	**DPPC 293K LENS red cluster**	
MSD [Å^2]	35.93	100.00	
	DPPC 293K $A_v \leq 0.478$Å^2	**DPPC 293K** $A_v > 0.478$Å^2	
MSD [Å^2]	38.26	43.93	

Fig. S16. Voronoi analysis for liquid phase nucleation in the gel phase and liquid/gel phases coexistence in a **DPPC** lipid bilayer at $T = 293\mathrm{K}$. (a) Left: Top views of the lipid bilayer colored according to the DPPC species (blue lipids) and by LENS cluster assignment (cyan and red clusters identify small and high LENS values, respectively). Right: example of the Voronoi tessellation where each Voronoi area A_v, is colored based on its size. (b) Voronoi clustering based on a selected threshold area A_t equal to: the Area Per Lipid of the system (top left), optimized area to obtain the best match between Voronoi and LENS clusters (top right), and too small and too high A_t values (bottom). Cluster color code: yellow for Voronoi area $A_v > A_t$ and violet for $A_v \leq A_t$. The histogram shows LENS/Voronoi clusters overlapping and their composition percentages for each At threshold. (c) Table reporting the Mean Squared Displacement (MSD) analysis of DPPC lipids in a bilayer configuration at $273\mathrm{K}$ (gel), $323\mathrm{K}$ (liquid) and $293\mathrm{K}$ (phases coexistence) – computed with a stride $10\mathrm{ns}$ – comparing the MSD of the lipids in the LENS or Voronoi clusters with, e.g., the MSD expected for lipids in full gel and liquid phase. The obtained results demonstrate how LENS capture well the presence of liquid and gel environments in the lipid bilayer (e.g., MSD values of cyan and red LENS clusters in the same order of magnitude of those expected for gel or liquid DPPC lipid bilayers) and how, on the contrary, a standard Voronoi analysis is inefficient in this sense (similar MSD for violet and yellow clusters, close to that of gel bilayers).

2. Steinhardt and SOAP analysis

In Figure S17a, we have plotted the Steinhardt (3) parameters *q4* and *q6* related to each atom of **Cu(211)** slab at $T = 600$K (see the MD snapshot reported on the right). HDBSCAN clustering (4) has been carried out on the cloud of order parameter data and two main domains are detected. As it is clear in Figure S17, a blue cluster corresponds to the atoms belonging to the topmost edges of the ideal (0K) Cu(211) surface, while all other atoms (bulk plus the other surface atoms) correspond to an orange cluster. Upon thermalization at 600K, a larger part of the surface atoms turns blue, meaning that they become less coordinated and ordered, and more dynamic (surface atoms that were orange in the ideal surface become more similar to the edge ones in terms of structure of their local neighborhood). This is sensible and not surprising, and it is in a sense a lower-resolution version of what it has been seen recently using a SOAP-based analysis.(5)

As an additional comparison, we also performed additional data-driven analyses on the same Cu surface based on SOAP (6) that allow in principle for a richer structural analysis of the atomic motifs that populate the surface (see Figure S17b). Such SOAP analysis has been conducted following to the same procedure recently used for the study of similar systems.(5) In this case, clustering of the SOAP data extracted from the MD trajectory of the **Cu(211)** surface at 600K shows many more colors, and a richer distinction of the different atomic environments that constitute the surface. As also seen in the Steinhardt analysis, it is clear that upon thermalization the surface becomes more "disordered/dynamic" than at 0K (the number of colors – i.e., number of different SOAP environments – increases).

A very similar result is obtained for the **Au-NP** at $T = 200$K. We have also computed the Steinhardt bond order parameters for the **Au-NP** at 200K, as shown in Figure S18a, left. The cluster representation based on *q4* and *q6* parameters demonstrates that the analysis is extremely accurate to reconstruct the geometrical environments in the Au NP, detecting e.g. edges, faces, vertexes of the icosahedral NP, and showing how upon heating to 200K the surface environments intermix while the surface becomes dynamic (see Figure S18). Also in this case, this result is very similar to that obtained recently with a SOAP classification.(7)

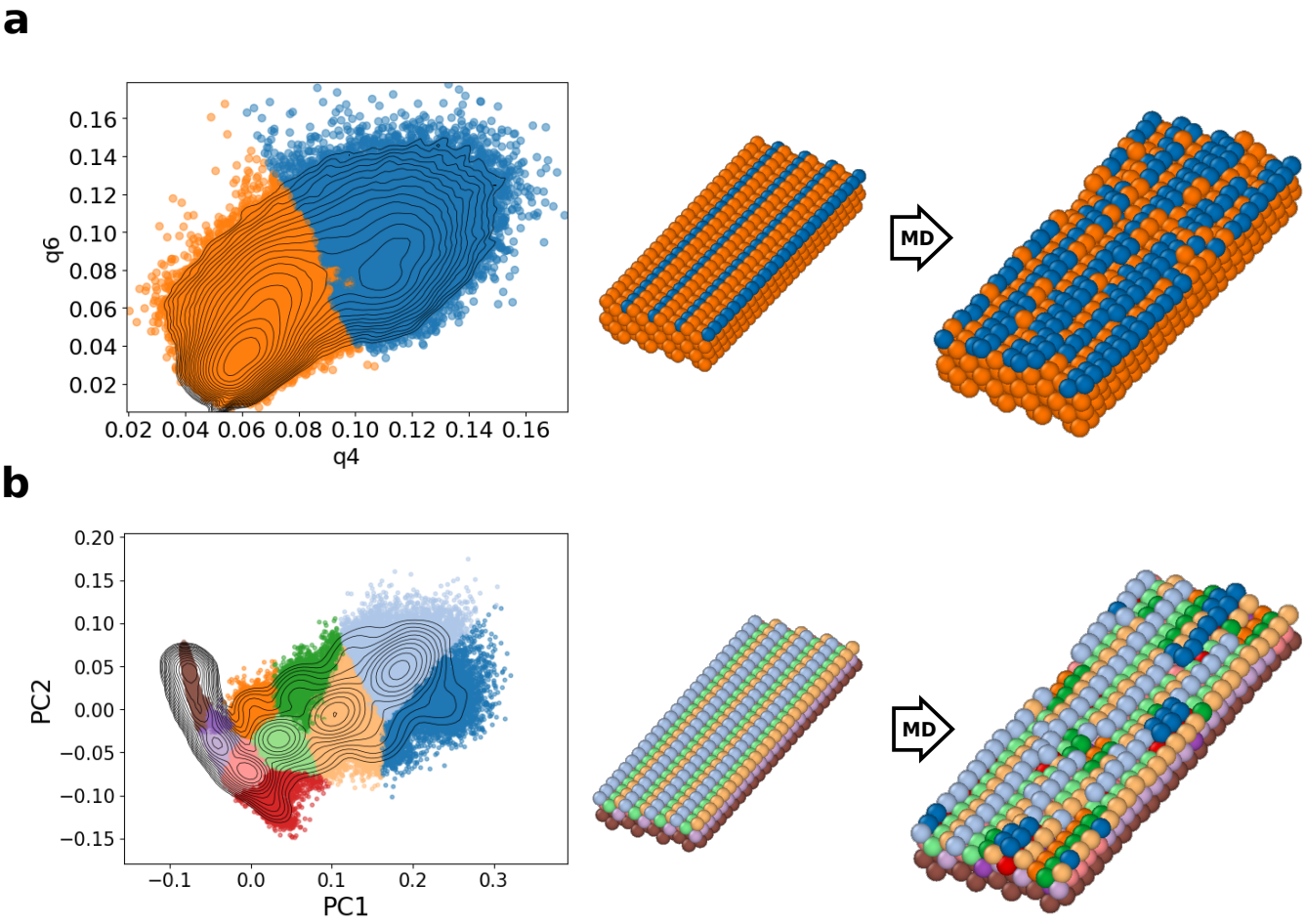


Fig. S17. (a) Steinhardt analysis for **Cu(211)** copper surface slab at $T = 600\mathrm{K}$: *q4* and *q6* order parameters are computed for each atom considering the environment within r_{cut} reported in Table S1; then HDBSCAN clustering (`min_cluster_size`=700,`min_samples`=20 with noise assignment (7)) is applied identifying two main structural environments: bulk and sub-surface (orange), surface (blue). (b) SOAP and clustering analysis for **Cu(211)** copper slab at $T = 600\mathrm{K}$. The high dimensional SOAP spectrum is computed for each atom considering the environment within r_{cut} reported in Table S1 and a Principal Component Analysis (PCA) is applied to reduce the high dimensional spectrum to four dimensions (cumulatively 99.7 % of the information is kept in the four PC, in Figure b left are reported the first two PC). Then HDBSCAN clustering (`min_cluster_size`=250, `min_samples`=2 with noise assignment (7)) is applied identifying eleven structural environments characterizing surface, sub-surface, bulk and structural deviation on the surface. While such analyses can capture a high-level of structural details, the dynamics information – obtained via, e.g., averaging the transitions between the detected atomic environments populating the surface (5) – makes it very difficult to detect sparse rare fluctuations that are important to understand the dynamical properties of such systems (see also LENS Movie S2).

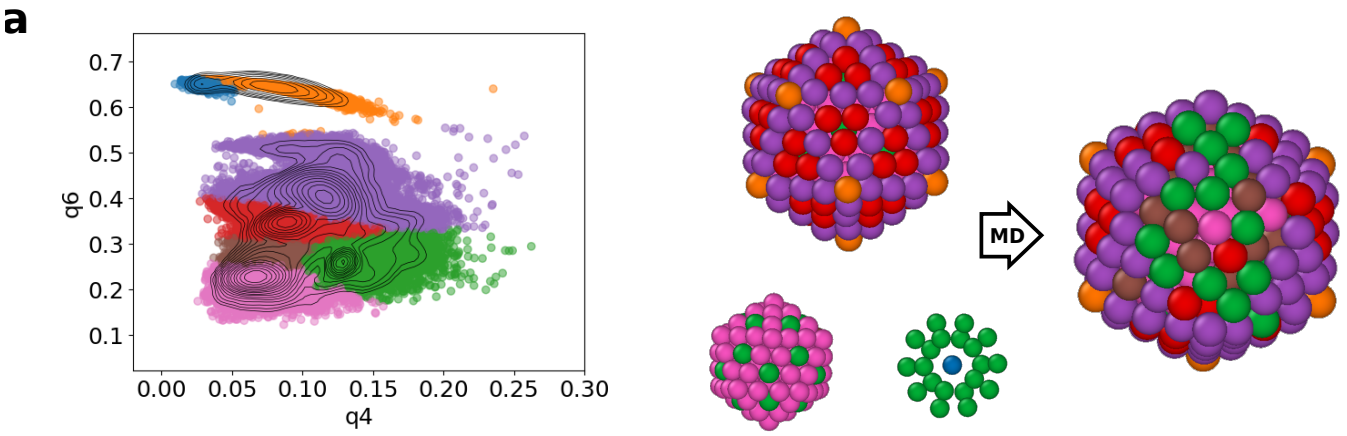


Fig. S18. (a) Steinhardt analysis for **Au-NP** nanoparticle at $T = 200\mathrm{K}$: *q4* and *q6* order parameters are computed for each atom considering the environment within r_{cut} reported in Table S1; then HDBSCAN clustering (`min_cluster_size`=700, `min_samples`=1 with noise assignment (7)) is applied identifying seven structural environments characterizing the surface (vertices, edges and faces) and bulk of the particle. While such structural analyses (e.g. Steinhardt or SOAP (6) can capture a high-level of structural details, the dynamics information reconstructed from them – obtained via, e.g., averaging the transitions between the detected atomic environments populating the surface (5)– makes it very difficult to detect local fluctuations and the effect that these have on the whole system dynamics. On the other hand, LENS shows that half of the Au NP surface becomes dynamic following to the transformation of one vertex into a rosette, while the other half preserves its reduced, crystalline-like vibrational behavior (see also LENS Movie S3).

3. Dynamical Propensity analysis

We have carried out a systematic comparison between LENS and the dynamical propensity (DP) descriptor developed in the group of Michaelides. (8) In order to apply the dynamical propensity to our ice-liquid water system, we computed for each water molecule the parameter DP:

$$DP_i = \left\langle \frac{\|r_i(t+\Delta t) - r_i)t)\|^2}{MSD} \right\rangle_{MDtraj} \quad [1]$$

where $r_i(t)$ is the position vector of molecule i at time t, Δt is the sampling time in our MD trajectory, and the MSD is the mean-square displacement of all oxygen atoms. It is worth noticing that in our case the ensemble average is estimated over the instantaneous displacements collected along the MD trajectory. After computing the DP values for all water molecules, we have estimated the probability density distribution P(DP) (see Figure S19), where two distinct peaks are clearly notable: $DP = 0.5$ and $DP = 2$, highlighting a low and high dynamical propensity of water molecules, respectively. Such result is evidently close to the LENS distribution in Figure 3, indicating a sort of correlation between the DP and LENS, and consequentially validating our descriptor. The resulting DP values are then classified selecting the thresholds both on the minimum of the P(DP) (a) and -+30% from the minimum of the P(DP)(b), obtaining two and three clusters respectively, as reported in Figure S19a and S19b, respectively. As evident from the MD snapshots, the DP-based clustering enables an explicit identification of ice (gray cluster), liquid (blue cluster) phases, and eventually the ice-liquid water interface (red clusters).

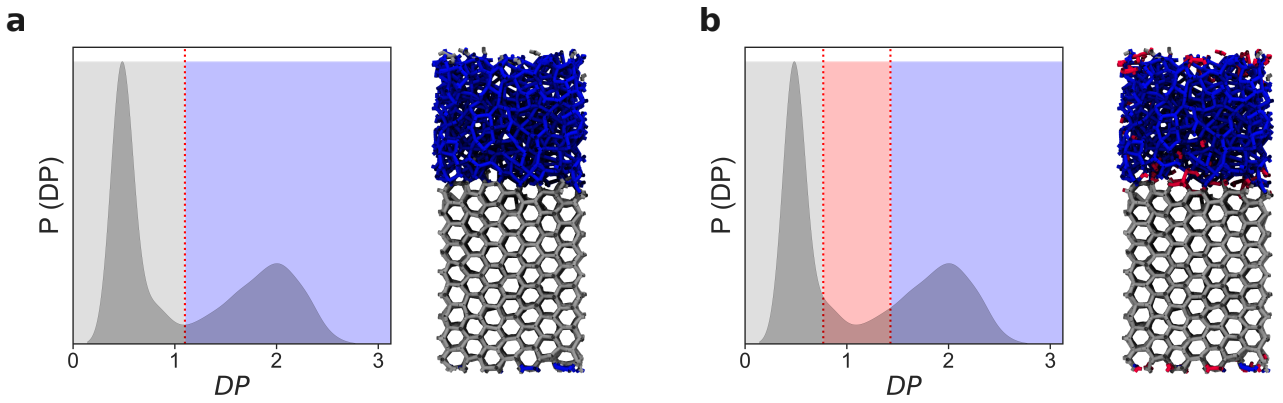


Fig. S19. Comparison with other state-of-the-art benchmark analyses: Probability density distribution of the Dynamical Property (DP)(8) computed for each water molecules included in the ice-liquid phase transition system. The resulting DP values are then classified selecting the thresholds both on the minimum of the P(DP) (a) and -+30% from the minimum of the P(DP)(b), obtaining two and three clusters respectively. The MD snapshots of water report the gray (ice phase), blue (liquid phase), and red (ice-liquid interface) clusters. These DP distributions are consistent with the averaged KDE distributions obtained from the LENS signals in Figure 3e in the main paper.

Movie S1. LENS analysis of gel-liquid phase coexistence in a DPPC lipid bilayer at 293 K of temperature. Atoms are colored based on their main LENS environment of belonging: liquid-phase lipids in red, gel-phase lipids in cyan.

Movie S2. LENS analysis of local dynamic transitions in a Cu(211) surface at 600 K of temperature. Atoms are colored based on their main LENS environment of belonging: static solid-phase atoms in gray, more dynamic surface edge atoms in orange, and fast-diffusing atoms in violet.

Movie S3. LENS analysis of a local sharp transition in a icosahedral Au-NP at 200 K of temperature. Atoms are colored based on their main LENS environment of belonging: crystalline/ordered domains in gray, solid but more dynamic atomic environments in cyan, increasingly dynamic local environments in orange and violet respectively. The movie shows how the LENS analysis detects the local transformation event of one icosahedron vertex (having 5-neighbor atoms) into a concave "rosette" (with 6-neighbor atoms).

References

1. S Baoukina, D Rozmanov, DP Tieleman, Composition fluctuations in lipid bilayers. *Biophys. J.* **113**, 2750–2761 (2017).
2. G Lukat, J Krüger, B Sommer, Apl@voro: A voronoi-based membrane analysis tool for gromacs trajectories. *J. Chem. Inf. Model.* **53**, 2908–2925 (2013).
3. PJ Steinhardt, DR Nelson, M Ronchetti, Bond-orientational order in liquids and glasses. *Phys. Rev. B* **28**, 784 (1983).
4. L McInnes, J Healy, S Astels, hdbscan: Hierarchical density based clustering. *J. Open Source Softw.* **2**, 205 (2017).
5. M Cioni, et al., Innate dynamics and identity crisis of a metal surface unveiled by machine learning of atomic environments. *J. Chem. Phys* **158**, 124701 (2023).
6. AP Bartók, R Kondor, G Csányi, On representing chemical environments. *Phys. Rev. B* **87**, 184115 (2013).
7. D Rapetti, et al., Machine learning of atomic dynamics and statistical surface identities in gold nanoparticles. ChemRxiv [Preprint] (2022) https://chemrxiv.org/engage/chemrxiv/article-details/63642e6aac45c7a2a9a45332.
8. M Fitzner, GC Sosso, SJ Cox, A Michaelides, Ice is born in low-mobility regions of supercooled liquid water. *Proc. Natl. Acad. Sci.* **116**, 2009–2014 (2019).