Phase transition of anisotropic materials is ubiquitously observed in physics, biology, materials science, and engineering. Nevertheless, how anisotropy of constituent molecules affects the phase transition dynamics is still poorly understood. Here we investigate numerically the phase transition of a simple model system composed of anisotropic molecules, and report on our discovery of multistep nucleation of nuclei with layered positional ordering (smectic ordering), from a fluid-like nematic phase with orientational order only (no positional order). A trinity of molecular dynamics simulation, machine learning, and molecular cluster analysis yielding free energy landscapes unambiguously demonstrates the dynamics of multistep nucleation process involving characteristic metastable clusters that precede supercritical smectic nuclei and cannot be accounted for by the classical nucleation theory. Our work suggests that molecules of simple shape can exhibit rich and complex nucleation processes, and our numerical approach will provide deeper understanding of phase transitions and resulting structures in anisotropic materials such as biological systems and functional materials.
phase transition has intrigued researchers not only as an interesting academic problem of condensed matter physics and statistical physics but also from a technological viewpoint. For example, a wide variety of daily products such as food, plastics, and alloys are processed by precise control of the transition from liquid mixture to solid. Phase transition has relevance also to biology because some biological systems such as deoxyribonucleic acid (DNA), ribonucleic acid (RNA), and membranes in our bodies function by repeating many kinds of phase transitions. These materials exhibit full of diverseness in structures and functions, and the process of phase transition often determines the performance of products through pattern formation. Hence, in a wide range of research fields including physics, biology, materials science, and engineering, it is an important problem to understand how phase transition occurs and how it can be controlled.

First-order phase transitions occur via the nucleation and growth of the stable new phase from the metastable host phase, and typical examples include the transition between the solid phase and liquid phase, and the crystallization of solute molecules from a solution. Nucleation was initially understood by the classical nucleation theory (CNT) in which the Gibbs free energy of a nucleus is assumed to be the sum of the bulk contribution and the interface free energy between the two phases, and evaluated as a function of the size of the nucleus. The distribution of clusters obeys the Boltzmann statistics, and a nucleus starts to grow when its size exceeds a critical value determined by the balance of the bulk and the surface free energies. Although it provided a comprehensible picture of the nucleation and growth processes, CNT was challenged by subsequent experimental studies because it failed to give a quantitative account for the nucleation rate and other pertinent properties of the nucleation processes. Two-step nucleation provides an alternative scenario that can explain the discrepancies between experimental results and the prediction of CNT. Two-step nucleation scenario has been drawing considerable interest, with a successful application to the crystallization of solute molecules, wherein the formation of dense or highly ordered precursors of solute molecules precedes that of crystalline clusters. Still, the step-wise pathway of this two-step nucleation can be understood within the context of CNT by considering realistic kinetic factors of clusters. An even more complex scenario has been proposed for solute nucleation that does not fall into the category of CNT. In this scenario, the nucleation is typically initiated by the formation of disordered networks or complex frameworks of solute precursors that are thermodynamically stable within the boundary of the mother solution. Such clusters are referred to as pre-nucleation clusters (PNCs), although their existence is required for reliable statistical analyses.

The anisotropy and the resulting additional degrees of freedom of constituent molecules allow a richer possibility of the formation of different ordered phases and thus phase transition behavior, as exemplified in liquid crystals (LCs) and also biological systems. However, this richness renders the investigation of the phase transition behavior even more challenging, although a few studies addressed the formation and kinetic pathway of orientationally ordered clusters from a host isotropic phase. Here we focus on pre-transitional clusters with local layered positional order (smectic clusters) formed from a fluid-like nematic phase with orientational order but no positional or layer order. There have been several experimental studies that corroborated the formation of pre-transitional smectic clusters (also known as cybotactic clusters) by X-ray diffraction. Fewer studies address the question of how such pre-transitional fluctuations and smectic nuclei are formed. In molecular simulation, the difficulty lies in identifying local smectic order from orientationally ordered nematic phase, which is overcome by supervised machine learning (ML) that finds appropriate order parameters for this purpose. In this work, we show the power of the trinity of molecular dynamics (MD) simulations, ML and molecular cluster analysis by investigating the elusive dynamics of the formation of smectic clusters from a nematic phase, and spotting the multistep nature of its kinetic pathways.

Results

We performed MD simulations of 1 million particles of Soft-Core Gay-Berne (SCGB) model. Well equilibrated nematic systems were quenched to temperature $T = 1.80$, below the smectic transition temperature $T_{N-Sm} = 2.25$. The time series of quenched coordinates was then analyzed using the above-mentioned ML scheme that precisely determines whether a certain molecule belongs to a nematic-like or a smectic-like local structure (for details, refer to Methods, and Supplementary Fig. 1). A molecular cluster analysis was applied to the time series of extracted smectic-like local structures.

Behavior of the whole system. To observe how the smectic ordering evolves in the whole system during phase transition, the time evolution of the number of “smectic molecules” was traced, as shown in Fig. 1a. The number of smectic molecules belonging to the largest smectic cluster is also plotted. Smectic nucleation must precede the drastic increase of molecules belonging to the largest smectic cluster around $t ≈ 1.0t$ (the time unit associated with a single particle, described in Methods). The percolation of the smectic phase progresses at $1.0t ≤ t ≤ 1.5t$, until almost all the smectic molecules belong to the largest cluster at $t ≥ 1.5t$.

Previous X-ray scattering experiments suggest the formation of pre-transitional fluctuations, known as cybotactic clusters, at a relatively early period of the transition to the smectic phase. In Fig. 1b we show the time evolution of numerical X-ray scattering intensity that can be directly compared with experiments (for calculation details and scattering intensity profiles, refer to Supplementary Fig. 2). The scattering intensity is plotted also as a function of the number of smectic molecules in the system. Figure 1b clearly indicates that the scattering intensity depends almost linearly on both the time ($t ≤ 1.0t$), and the number of smectic molecules in the system ($t > 1.7 \times 10^5$).
suggestive of a strong correlation between the X-ray scattering intensity and the number of smectic molecules determined by our ML scheme. As noted above, smectic nucleation dominates the kinetic process of phase transition at \( t \lesssim 1.0 \tau \). Hence, the pre-transitional local structures or cybotactic clusters involved in the smectic nucleation process are indeed responsible for the growth of the X-ray scattering intensity observed in experiments.

**Free energy landscape.** To obtain more detailed information on the static and dynamic properties of pre-transitional local structures, we constructed the free energy landscape as a function of the cluster size \( N \) and the order parameter quantifying the degree of liquid crystalline order. The order parameter \( Q \) was provided by our ML scheme\(^{51,54} \), and the free energy landscape was calculated by the transition probability approach based on the previous work of Mochizuki and co-workers\(^{55} \) (for calculation details, refer to Methods). The latter is presented in Fig. 1c (the reference (zero) of the free energy is taken to be that of the nematic phase), and the saddle point corresponds to the critical nucleus that is made up of approximately 420 smectic molecules. Furthermore, a pocket region at \( N \approx 150 \) was discovered, indicating the existence of metastable clusters. This estimate of \( t \) h e s i z e o f t h e c r i t i c a l n u c l e i \( N \approx 420 \), and the presence of metastable clusters at \( N \approx 150 \) agree well with the result of a conventional threshold method\(^{56} \), as shown in Supplementary Fig. 4.

We carried out further systematic calculations to elucidate how the variation of the degree of supercooling \( \Delta T \equiv T_{N,Sm} - T \) influences the free energy landscape and the resulting major pathway and metastable clusters. The same qualitative features described above are observed regardless of \( \Delta T \) (see Supplementary Fig. 5). In Fig. 2a, b we show the \( \Delta T \) dependence of the CN size and the height of the energy barrier from nematic to CN, respectively. Figure 2 clearly demonstrates the non-classical nature of the nematic-smectic phase transition; the CN size is
Fig. 2 Scaling behavior of smectic nucleation. $\Delta T$ dependence of $a$ the CN size and $b$ the height of the energy barrier from nematic to CN.

proportional to $\Delta T^{-1.0}$, and the barrier from nematic to CN is proportional to $\ln(\Delta T^{-1.4})$ (In CNT, they should behave as $\propto \Delta T^{-3}$ and $\propto \Delta T^{-2}$, respectively). The MC1 and MC2 sizes are almost independent of $\Delta T$ (respectively 7 and 168 ± 20, see Supplementary Fig. 6a). The barrier from MC1 to MC2 decreases weakly with the increase of $\Delta T$ and falls within the range [1.60k_B T, 2.22k_B T] (see Supplementary Fig. 6b). The barriers from nematic to MC1 and from MC2 to CN are almost constant ([9.68 ± 0.04] k_B T and [0.77 ± 0.08] k_B T, respectively). The depth of the local minimum of MC1 and MC2 is only weakly dependent of $\Delta T$ and increases with the increase of $\Delta T$ (see Supplementary Fig. 6c). The high barrier ≈9.68 k_B T clearly indicates that MC1 is metastable. From the low energy barrier from MC1 (~2.0k_B T), and the size insensitivity to $\Delta T$, MC2 might be regarded as “metastable PNC” formed in a mixture of host nematic phase and MC1 clusters. In view of the insensitivity of the features of metastable clusters to $\Delta T$, in the following sections $T$ is set to 1.80 except where specifically noted.

Real-space density profiles of metastable clusters and critical nuclei. After the identification of metastable clusters (MC1 and MC2) and CN (More details of the identification criteria are presented in Methods, and Supplementary Fig. 7), their real-space geometrical information can be deduced. Figure 3 shows the averaged density profiles of MC1, MC2, and CN. The abc orthogonal coordinate system is such that its origin is at the center of the molecule that is closest to the center of mass of the cluster, and the $a$-axis is parallel to the average orientation of the constituent molecules. The void at the origin indicates that the molecule at the cluster center excludes the other molecules sterically. MC1 shown in Fig. 3a is composed of a simple hexagonal column and has a small tail in the direction perpendicular to the $a$-axis. The shape of MC2 shown in Fig. 3b is oblate, although it may depend on the interaction parameters as suggested in a previous study on nematic droplets. Figure 3b clearly indicates that at least three distinct smectic layers are involved, with additional 1 or 2 layers being recognized as the tail along the $a$ direction. The shape of CN shown in Fig. 3c, involving 5 distinct smectic layers and additional 1 or 2 layers as the tail, is highly similar to that of MC2. We emphasize that the real-space profile of pre-transitional fluctuations towards the smectic phase, also known as cybotactic clusters, has been presented only schematically. Our work, with the aid of the trinity of MD simulation, ML, and molecular cluster analysis, provides the first real-space density profiles of metastable clusters and critical nuclei in a clear manner.

Dynamics of metastable clusters, critical nuclei and supercritical nuclei in the transition. For further understanding of the smectic nucleation process, it is important not only to construct the free energy landscape as discussed above, but also to observe the actual dynamics of smectic clusters associated with the major pathway of the free energy landscape. We have already presented how MC1, MC2 and CN are identified, and we also define a supercritical nucleus (ScN) as a cluster larger than CN (namely, $N > 443$, as the size of CN is $410 \leq N \leq 442$. See Supplementary Fig. 7c). To monitor the history of the clusters, we further introduce the labelling tMC1, tMC2, and tCN; We label as tMC1 a cluster or a molecule which was formerly MC1 but does not belong to MC1, MC2, CN, nematic phase or ScN at the time of the identification. The labelling tMC2 and tCN are defined similarly. The size distribution of the clusters labeled as MC1, MC2, CN, tMC1, tMC2, or tCN shown in Supplementary Fig. 8, whose vertical axis is the logarithm of the existence probability, agrees well with minus the free energy landscape in Fig. 1e. Recalling that Fig. 1e is the major pathway of the nucleation process, and that minus the logarithm of the size distribution represents the free energy as a function of the cluster size, our identification of characteristic molecules or clusters (MC1, MC2, CN, tMC1, tMC2 or tCN) successfully samples the major pathway of the nucleation process.

Now we focus on how ScNs are formed; more specifically, whether a given ScN emerges through the major pathway of the nucleation process. We refer to ScNs formed through the major pathway as “induced ScN” (IScN), and those through other minor pathways as “non-induced ScN” (NScN). The identification of IScN and NScN is as follows: We count the number of characteristic molecules included in newly born ScNs during nucleation process. We denote by $M_t$ the sum of the numbers of characteristic molecules included in an ScN. From the bipolar form of the generation probability of ScNs as a function of $M_t$ (Supplementary Fig. 9), we identify an ScN as IScN if $M_t \geq 200$, and as NScN otherwise. Note that in the following analyses the characteristic molecules in newly born IScNs or NScNs are immediately relabeled as IScN or NScN. More specific rules for the classification of molecules are shown in Supplementary Fig. 10.
IScNs become the largest component of the system along with the progress of the nucleation process (Supplementary Fig. 11), and therefore monitoring the time evolution of the number of IScNs is highly important for the understanding of the smectic nucleation process. Figure 4a clearly demonstrates that the nucleation of IScNs involves three processes: The first process is at \(0.50\tau < t < 0.60\tau\) with the nucleation rate of \(J_{1st} = 2.26 \times 10^{-5} \sigma^{-3} r^{-1}\), followed by the 1st plateau. The second process is at \(0.64\tau < t < 0.69\tau\) with the nucleation rate of \(J_{2nd} = 6.28 \times 10^{-5} \sigma^{-3} r^{-1}\), followed by the 2nd plateau. The latter is conceivable as the lag time to the third process at \(0.75\tau < t < 0.80\tau\) with the nucleation rate of \(J_{3rd} = 7.61 \times 10^{-5} \sigma^{-3} r^{-1}\), again followed by the 3rd plateau. Note that the nucleation rate increases with the evolution of the nucleation processes. In the following, we refer to these three processes as the "first (second or third) nucleation process".

Let us consider the origin of three-step nucleation. The first nucleation rate is clearly smaller than that for the other processes, and is close to the nucleation rate for the minor pathway (see Supplementary Fig. 12). Hence the 1st process is a nucleation barrier forming the efficient pathway. The time evolution of the number of tCNs and tMC2s shown in Fig. 4b explains the second and third processes. The number of tCNs has two peaks; it steadily increases until the end of the 2nd nucleation process and then starts to decrease, followed by the second increase during the 2nd plateau. Its second peak is at the end of the 3rd nucleation process. Hence these two peaks obviously mark the 2nd and 3rd nucleation processes of IScNs. Note that tCNs are the largest component of characteristic molecules included in IScNs, and clearly dominate the formation of IScNs (Supplementary Fig. 13). The number of tMC2s exhibits one clear peak at the beginning of the 2nd nucleation process followed by the decrease. The recovery of the number of tCNs during the 2nd plateau is clearly attributed to the consumption of tMC2s to form CNs and thus tCNs. Therefore, the lag time between 2nd and 3rd nucleation arises from the competition between the consumption of tCNs due to fast nucleation and the supply of tCNs from tMC2s.

The importance of tMC2 in the nucleation process is demonstrated also by monitoring the time evolution of the formation of characteristic clusters in real space. Figure 4c, d shows the snapshots of tMC2s (yellow), tCNs (red) and IScNs (blue) at \(t = 0.50\tau\) and \(0.80\tau\), respectively. The positional overlap between tMC2s at \(t = 0.50\tau\) and IScNs at \(t = 0.80\tau\) clearly indicates that the former induces the nucleation of the latter. In fact, the overlap rate is consistently high, and steadily increases at \(t > 0.50\tau\) (Supplementary Fig. 14). The low mobility of tMC2 is clearly seen in Supplementary Fig. 15 and Supplementary Movie 1, which reflects the trapping effect owing to the local minimum of the free energy. Hence low-mobility regions of smectic molecules constituting tMC2 form the backbone for the nucleation of IScNs, which highly resembles the formation of crystal ice from the low-mobility regions of supercooled liquid water\(^{40}\).

The second and third nucleation processes on the major pathway have been shown to proceed in the order of MC2, CN, and IScN, and therefore cannot be regarded as one- or two-step processes. MC2 and tMC2 distinguish the multistep nucleation process of the nematic-smectic transition from conventional ones, and as mentioned above, play an important role in that they form the backbone for efficient pathway, and that tMC2 clusters act as the precursors for the nucleation of IScNs. We therefore conclude that MC2 and tMC2 are unambiguously identified as the pre-transitional fluctuations that dominate the dynamics of nematic-smectic transition by inducing CN and tCN, and thus IScN.
have shown62, and we hope our study will promote further studies, which can be far richer and more complex than previous studies of simple systems, not restricted to anisotropic molecules. The study of multistep nucleation in a simple one-component system of model LCs has a profound meaning because nucleation phenomena involving two different types of metastable clusters with properties distinct from those predicted by the CNT. The origin of the three-step nucleation was explained by clarifying the free energy landscape and the major pathway of the nucleation process, and also by tracking the dynamics of metastable clusters and CNs. Our demonstration of multistep nucleation by our numerical techniques might enable the prediction of where pattern formation would be initiated, without monitoring the whole nucleation processes. Our demonstration of multistep nucleation in a simple one-component model anisotropic molecules has a profound meaning because nucleation processes of simple systems, not restricted to anisotropic ones, can be far richer and more complex than previous studies62, and we hope our study will promote further studies towards deeper understanding of the complex nature of nucleation phenomena.

Our approach is particularly promising for the investigation of phase transition behavior of biological LCs that widely exist in nature as a rich set of soft materials with anisotropic structures. By tracking metastable clusters, sophisticated functions related to phase transition might be revealed for in vivo or in vitro structures of biological building blocks2,6,45,63. Our approach also has a great potential for understanding not just phase transition of LCs but a much broader range of phase transition phenomena in solutions and polymers. More broadly, suitable molecular design that can optimize the size and shape of metastable clusters may enable the control of nucleation rates and pattern formation in a wide variety of anisotropic materials. This is significant not only for the control of basic properties, but also for the design of emergent properties of advanced materials such as self-healing64. Further investigations of phase transition phenomena with detailed analysis of energetic stability, structure and dynamics of clusters will open the door to new technology for designing highly advanced materials.

**Discussion**

We investigated numerically how anisotropic molecules attain their positional order, or more specifically, how smectic nuclei are formed in the course of the nematic-smectic phase transition. Our state-of-the-art numerical techniques, namely, the trinity of molecular dynamics simulation, machine learning, and molecular cluster analysis, discovered three-step nucleation involving two different types of metastable clusters with properties distinct from those predicted by the CNT. The origin of the three-step nucleation was explained by clarifying the free energy landscape and the major pathway of the nucleation process, and also by tracking the dynamics of metastable clusters and CNs. Our analysis also revealed the positional overlap between metastable clusters and ensuing IScNs, and the observation of metastable clusters by our numerical techniques might enable the prediction of where pattern formation would be initiated, without monitoring the whole nucleation processes. Our demonstration of multistep nucleation in a simple one-component model anisotropic molecules has a profound meaning because nucleation processes of simple systems, not restricted to anisotropic ones, can be far richer and more complex than previous studies have shown62, and we hope our study will promote further studies towards deeper understanding of the complex nature of nucleation phenomena.

**Methods**

**Molecular dynamics simulations.** The MD simulations of GB and SCGB particle systems were performed using an open source program called the Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS)65.

For the pair interactions of ellipsoidal GB particles, the following pairwise interaction potential, \( U_{\text{GB}} \), was calculated for each pair of particles:

\[
U_{\text{GB}} = 4\epsilon_a \left[ \frac{\sigma_e}{r - \sigma_e + \sigma_e} \right]^{12} - \left[ \frac{\sigma_e}{r - \sigma_e + \sigma_e} \right]^{6},
\]

where \( \epsilon_a \) denotes the anisotropic energy for an ellipsoidal pair, \( r \) the distance between the centers of mass for a pair of particles, \( \sigma_e \) the anisotropic length for the ellipsoidal pair, and \( \sigma_s \) the length for the side-by-side configuration of the ellipsoids.

By introducing parameter \( \kappa = a_0/s_0 \), in which \( a_0 \) denotes the length for the end-to-end configuration of the ellipsoids, the anisotropic energy \( \epsilon_a \) is then written

\[
\epsilon_a = \epsilon_a' a_0^\kappa,
\]

where \( \epsilon_a' \) and \( \epsilon_a'' \) denote the contributions corresponding to the well depth and configuration anisotropies, and \( \mu \) and \( \nu \) are multipliers for determining these two contributions for the pair potential. Factor \( \kappa' = \epsilon_a'/\epsilon_a'' \), where \( \epsilon_a' \) and \( \epsilon_a'' \) denote energy contributions from the end-to-end and
side-by-side configurations of ellipsoids, respectively. Factor $\epsilon_3$ is characterized by $\kappa$. Therefore, the detailed shape of UGB is determined from the values of four parameters, $\kappa$, $\mu$, and $\nu$. For the characteristic length, energy, and mass of the GB systems, $a_i = a$, $\epsilon = k_BT$, and $m$ are each set to 1; here $m$ is the mass of one GB particle.

For the intermolecular interactions of SCGB particles, the core repulsion of the original GB potential is replaced by a weaker linear repulsion. The soft-core potential energy $U_{SC}$ is written

$$U_{SC} = a(r - a),$$

where $a$ is the potential slope for soft repulsive energy barrier. The UGB and USC are merged using a sigmoidal logistic function $f$ so that the resulting potential reads $U_{SCGB} = (1 - f)U_{GB} + fU_{SC}$. Calculation of free energy landscape

$$f = \frac{1}{1 + \exp[b(r - a)],}$$

where $b$ is the steepness of sigmoidal logistic seaming function.

To be able to compare the results of GB and SCGB systems directly, we used exactly the same parameter settings of the GB potential for the two models. Specifically, $\kappa = 3$, $\mu = 5$, $\nu = 1$, and $\nu = 3$. Note that the above parameter set has been traditionally used because the physical properties of nematic and smectic liquid crystal phases are well displayed. For SCGB, the terms $\kappa$, $\mu$, and $\nu$ were set to $-70\kappa$ $-70\kappa$ $-100\kappa$, respectively, on the basis of previous studies. Using the above parameter settings, the nematic-smectic transition is guaranteed to be observed for both GB and SCGB systems by quenching from temperature $T = 2.4$ to 1.8 at the density of $3.0\rho - 3$. Therefore, an ensemble was used having a constant number of particles at density $3.0\rho - 3$ contained within a cube of constant volume and with all $x$- and $y$-periodic boundary conditions imposed. The initial configurations for the nematic-smectic phase transition were prepared in a careful manner, being cooled gradually from the isotropic phase at $T = 6.0$ to the nematic phase at $T = 2.4$. The configurations were then quenched to temperature below $T_{\text{N-lnc}}$. To observe the fast nucleation during the first-order phase transition, the velocity Verlet integrator with finite timestep $\Delta t = 6.0 \times 10^{-5}$ and $2.0 \times 10^{-5}$ was used for GB and SCGB, respectively, where $t = (n\Delta t)^{1/2}$ is a time unit. The temperature was controlled using a Nosé–Hoover chain thermostat. The velocity-scaling method was also confirmed by yielding consistent results. To compute precisely the intermolecular interactions during phase transition, the GB and SCGB potentials were truncated at $8.0\kappa$.

For precise computations of the free energy landscape (for details, refer to Calculation of free energy landscape), a large number of smectic clusters must be sampled. Therefore, additional MD simulations of the GB and SCGB systems were performed using the smaller timesteps of $\Delta t = 1.5 \times 10^{-5}$ and $0.5 \times 10^{-5}$, respectively, while maintaining other simulation settings described above fixed.

To explore the $\Delta T$ dependence of the free energy landscape of SCGB systems, $T$ was varied from 1.65 to 1.95 (i.e., $\Delta T$ from 0.30 to 0.60).

**Machine learning.** The time series of quenched configurations was analyzed using the Machine Learning-aided Local Structure Analyzer (ML-LSA)

$$p(t_{i+1}|s_{i}) = \sum_{s_{i+1}} p(t_{i+1}|s_{i+1}) p(s_{i+1}|s_{i}),$$

where $p(s_{i+1}|s_{i})$ is the transition probability from state $s_i$ at time $t_i$ to state $s_{i+1}$ at time $t_{i+1}$. Let $p(t_{i+1}|s_{i})$ denote the probability of a cluster being in a state $s_i$ at time $t_i$. Then

$$p(s_{i+1}|s_{i}) = \sum_{s_{i+1}} p(t_{i+1}|s_{i+1}) p(s_{i+1}|s_{i}),$$

where $p(s_{i+1}|s_{i})$ is the transition probability from state $s_i$ at time $t_i$ to state $s_{i+1}$ at time $t_{i+1}$, which can be obtained from cluster statistics. Hence the stationary distribution $p(s)$ should satisfy

$$p(s) = \sum_{s_{i+1}} p(s_{i+1}|s) p(s_{i+1}).$$

When $p(s_{i+1}|s_i)$ is given, $p(s)$ can be determined from the iteration of Eq. (11). The free energy landscape $\Delta\mathcal{F}(N, Q)$ was obtained from $\Delta\mathcal{F}(N, Q) = -k_B T \ln p(N, Q)$, where $p(N, Q) = p(s)$. The transition probability approach requires constant transition probabilities among cluster states, except those considerably larger than the critical nucleus. For the quality of the statistics, the time interval $\Delta t_{\text{sample}}$ for the sampling should be as large as possible. The optimum $\Delta t_{\text{sample}}$ was determined by comparing the results with different $\Delta t_{\text{sample}}$.

**Molecular cluster analysis.** To observe smectic nucleation in a nematic-smectic phase transition, a molecular cluster analysis was applied in a time-series analysis of smectic-like local structures extracted from the ML scheme. In the analysis, a cluster was defined as a group of mutually connected molecules within the region containing the first to twelfth neighbor molecules defined in the ML scheme. For the order parameter of the smectic clusters, the averaged local order parameter $Q$ was calculated from $Q_{GB}(i)$ where $Q_{GB}(i)$ is the local order parameter of the GB system. Therefore, the scheme is sufficiently reliable for our present purpose as well. For further details of ML-LSA and the scheme using it, we refer to our previous work.

**Calculation of free energy landscape.** The free energy landscape as a function of $N$ and $Q$ was calculated using the transition probability approach, which considers a network of states of clusters connected by transition paths with certain transition probabilities. The state of a cluster is defined by $s = [N, Q]$. Let $p(t_{i+1}|s_{i})$ denote the probability of a cluster being in a state $s_i$ at time $t_i$. Then

$$p(s_{i+1}|s_i) = \sum_{s_{i+1}} p(t_{i+1}|s_{i+1}) p(s_{i+1}|s_i),$$

where $p(s_{i+1}|s_i)$ is the transition probability from state $s_i$ at time $t_i$ to state $s_{i+1}$ at time $t_{i+1}$, which can be obtained from cluster statistics. Hence the stationary distribution $p(s)$ should satisfy

$$p(s) = \sum_{s_{i+1}} p(s_{i+1}|s) p(s_{i+1}).$$

When $p(s_{i+1}|s_i)$ is given, $p(s)$ can be determined from the iteration of Eq. (11). The free energy landscape $\Delta\mathcal{F}(N, Q)$ was obtained from $\Delta\mathcal{F}(N, Q) = -k_B T \ln p(N, Q)$, where $p(N, Q) = p(s)$. The transition probability approach requires constant transition probabilities among cluster states, except those considerably larger than the critical nucleus. For the quality of the statistics, the time interval $\Delta t_{\text{sample}}$ for the sampling should be as large as possible. The optimum $\Delta t_{\text{sample}}$ was determined by comparing the results with different $\Delta t_{\text{sample}}$.

**Identification of MC1, MC2, and CN.** Here, we describe how metastable clusters (MC1 and MC2) and critical nuclei (CN) were identified using the free energy landscape in terms of the cluster size and the order parameter. The state of a specific cluster was limited by the specified limits in range for free energy, cluster size, and order parameter values. Here we determined the range in free energy and cluster size using the free energy minimum curve illustrated in Fig. 1e, and therefore the range of the order parameter was automatically determined. For MC1 or MC2, the corresponding local minimum in Fig. 1e was selected as a reference state, and the half-height of the nearest and steepest well in Fig. 1e defines the threshold value for energy. States below the threshold were regarded as being from the same metastable cluster. For CN, all points near the lowest energy value with no varying trend
with respect to cluster size were regarded as the CNs. Supplementary Fig. 7 shows the free energy contour map of all the states of MC1, MC2, and CN, selected by the above identification criteria.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability
The data that support the findings of this study are available from the corresponding author upon reasonable request, based on the publication protocol of the research data as permitted by a project (JPNP16010) commissioned by the New Energy and Industrial Technology Development Organization (NEDO).

Code availability
ML-LSA and other codes for analysis are available from the corresponding author upon reasonable request, based on the publication protocol of the developed codes as permitted by JPNP16010 commissioned by NEDO.

Received: 25 February 2021; Accepted: 18 August 2021; Published online: 06 September 2021

References
1. Chaikin, P. M., Lubensky, T. C. & Witten, T. A. Principles of Condensed Matter Physics Vol. 10 (Cambridge University Press, 1995).
2. Strzelecka, T. E., Davidson, M. W. & Rill, R. L. Multiple liquid crystal phases in biology. Annu. Rev. Cell Dev. Biol. 30, 39–58 (2014).
3. Saw, T. R., Xi, W., Ladoux, B. & Lim, C. T. Biological tissues as active nematic liquid crystals. Adv. Mater. 30, 1802579 (2018).
4. Ilica, S. L. et al. Multiple liquid crystalline geometries of highly compacted nucleic acid in a dsRNA virus. Nature 570, 252–256 (2019).
5. Zhao, J. et al. Advances in biological liquid crystals. Small 15, 1900019 (2019).
6. Leslie, M. Sloppy science or groundbreaking idea? theory for how cells organize contents divides biologists. Science https://doi.org/10.1126/science.abg6495 (2021).
7. Kalikmanov, V. I. in Nucleation Theory (eds. Englert, B.-G. et al.) vol. 860, 17–41 https://link.springer.com/book/10.1007/978-3-642-4364-8 (Springer, 2013).
8. Erdemir, D., Lee, A. Y. & Myerson, A. S. Nucleation of crystals from solution: classical and two-step models. Acc. Chem. Res. 42, 621–629 (2009).
9. Veklov, P. G. Nucleation. Cryst. Growth Des. 10, 5007–5019 (2010).
10. Veklov, P. G. The two-step mechanism of nucleation of crystals in solution. Nanoscale 2, 2346–2357 (2010).
11. Qi, W., Peng, Y., Han, Y., Bowles, R. K. & Dijkstra, M. Nonclassical nucleation in a solid-solid transition of confined hard spheres. Phys. Rev. Lett. 115, 185701 (2020).
12. Salvaglio, M., Perego, C., Giberti, F., Mazzotti, M. & Parrinello, M. Molecular-dynamics simulations of urea nucleation from aqueous solution. Proc. Natl Acad. Sci. 112, E6–E14 (2015).
13. Sosso, G. C. et al. Crystal nucleation in liquids: open questions and future challenges in molecular dynamics simulations. Chem. Rev. 116, 7078–7116 (2016).
14. Lutsko, J. F. How crystals form: a theory of nucleation pathways. Sci. Adv. 5, eaav7399 (2019).
15. Kashchiev, D. Classical nucleation theory approach to two-step nucleation of crystals. J. Cryst. Growth 350, 123500 (2020).
16. De Yoreo, J. in crystallization via Nonclassical Pathways Vol. 1: Nucleation, Assembly, Observation & Application (ed Zhang, X.)1–17 (ACS Publications, 2020).
17. Demichelis, R., Raiteri, P., Gale, J. D., Quigley, D. & Gebauer, D. Stable prenucleation mineral clusters are liquid-like ionic polymers. Nat. Commun. 2, 590 (2011).
18. De Yoreo, J. More than one pathway. Nat. Mater. 12, 284–285 (2013).
19. Gebauer, D. & Cölfen, H. Prenucleation clusters and non-classical nucleation. Nano Today 6, 564–584 (2011).
20. Gebauer, D., Kellermeyer, M., Gale, J. D., Bergström, L. & Cölfen, H. Pre-nucleation clusters as solute precursors in crystallisation. Chem. Soc. Rev. 43, 2348–2371 (2014).
Acknowledgements

The authors are grateful to Dr. Hideo Doi of AIST for supporting the use of ML-LSA. We also thank Dr. Ryugi Sakamaki and Mr. Yuki Furukawa of X-Ability Co., Ltd. for computational optimization of GB and SCGB models in molecular dynamics simulations. This work is based on results obtained from JPNP16010 commissioned by NEDO. J.F. is in part supported by JSPS KAKENHI (Grant Numbers JP17H02947 and JP21H01049).

Author contributions

K.Z.T. and J.F. designed the research; K.Z.T. carried out molecular dynamics simulations, machine learning, and molecular cluster analysis. K.Z.T. and J.F. drafted the paper. T.A. contributed to the interpretation of the part of results and to the writing.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-021-25586-4.

Correspondence and requests for materials should be addressed to K.Z.T.

Peer review information Nature Communications thanks the anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2021, corrected publication 2021