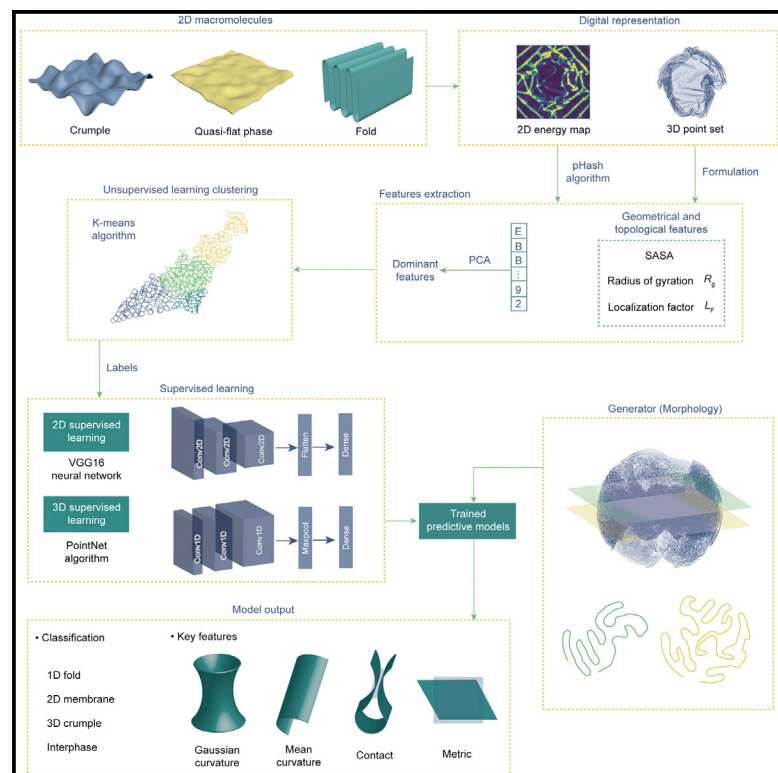


Patterns

Unraveling the morphological complexity of two-dimensional macromolecules

Graphical abstract



Authors

Yingjie Zhao, Jianshu Qin,
Shijun Wang, Zhiping Xu

Correspondence

xuzp@tsinghua.edu.cn

In brief

Morphology of 2D macromolecules defines the microstructures of their condensed phases and the processing-microstructures-properties relationship of their macroscopic assemblies. Zhao et al. integrate molecular simulations and statistical-learning methods to classify the 2D macromolecules into four morphological phases. Beyond the conventional approach using 3D geometry, the physics of lattice distortion and the topology of surface contact are included in the models. The key features correlated with the morphological complexity and material properties of 2D macromolecules are revealed.

Highlights

- Morphology of 2D macromolecules are classified into four phases
- Data-driven models capture physics and topology beyond the geometry
- Condensed-phase properties are understood by the features extracted



Article

Unraveling the morphological complexity of two-dimensional macromolecules

Yingjie Zhao,¹ Jianshu Qin,¹ Shijun Wang,¹ and Zhiping Xu^{1,2,*}¹Applied Mechanics Laboratory, Department of Engineering Mechanics and Center for Nano and Micro Mechanics, Tsinghua University, Beijing 100084, China²Lead contact*Correspondence: xuzp@tsinghua.edu.cn<https://doi.org/10.1016/j.patter.2022.100497>

THE BIGGER PICTURE Resolving morphological complexity of macromolecules is the stepping stone to the design and fabrication of high-performance, multi-functional materials and to understanding the soft matter behaviors in biology and engineering. To extract the physics of lattice distortion and surface contact beyond the conformation is critical, yet challenging. Here, we show that, by labeling the simulation data using the 2D map of potential energies, the 3D geometry, and the topology of contact, morphological classification can be achieved with high accuracy. The well-trained model can be used to decipher the microstructural complexity using simulation or experimental data, which may include the geometrical representation only. This data-driven approach extracts the key geometrical and topological features of 2D macromolecules that are directly responsible for the material performance in relevant applications and can be extended to study other complex surfaces such as red blood cells and the brain.



Proof-of-Concept: Data science output has been formulated, implemented, and tested for one domain/problem

SUMMARY

2D macromolecules, such as graphene and graphene oxide, possess a rich spectrum of conformational phases. However, their morphological classification has only been discussed by visual inspection, where the physics of deformation and surface contact cannot be resolved. We employ machine learning methods to address this problem by exploring samples generated by molecular simulations. Features such as metric changes, curvature, conformational anisotropy and surface contact are extracted. Unsupervised learning classifies the morphologies into the quasi-flat, folded, crumpled phases and interphases using geometrical and topological labels or the principal features of the 2D energy map. The results are fed into subsequent supervised learning for phase characterization. The performance of data-driven models is improved notably by integrating the physics of geometrical deformation and topological contact. The classification and feature extraction characterize the microstructures of their condensed phases and the molecular processes of adsorption and transport, comprehending the processing-microstructures-performance relation in applications.

INTRODUCTION

Polymerized linear and nonlinear molecules with repeating one-dimensional (1D) subunits feature a variety of geometrical forms such as 1D chains, 2D membranes, and 3D dendrimers, of which the morphology is a key determinant of the material properties.¹ 2D macromolecules possess an even richer spectrum of conformational complexity. The competition between the entropy gain caused by thermal corrugation and the enthalpic penalty attrib-

uted to the shear and bending resistance governs their conformational evolution.² The relation between their morphological behaviors and size, bending stiffness, and surface interaction were discussed through the conformational scaling laws.^{3–8} Numerical simulations using the self-penetrable phantom model of tethered membranes confirm the stability of quasi-flat conformation and revealed the crumpling transition at high temperature.^{9–11} Considering the effects of self-avoiding, bending resistance, and surface interaction, simulations of more



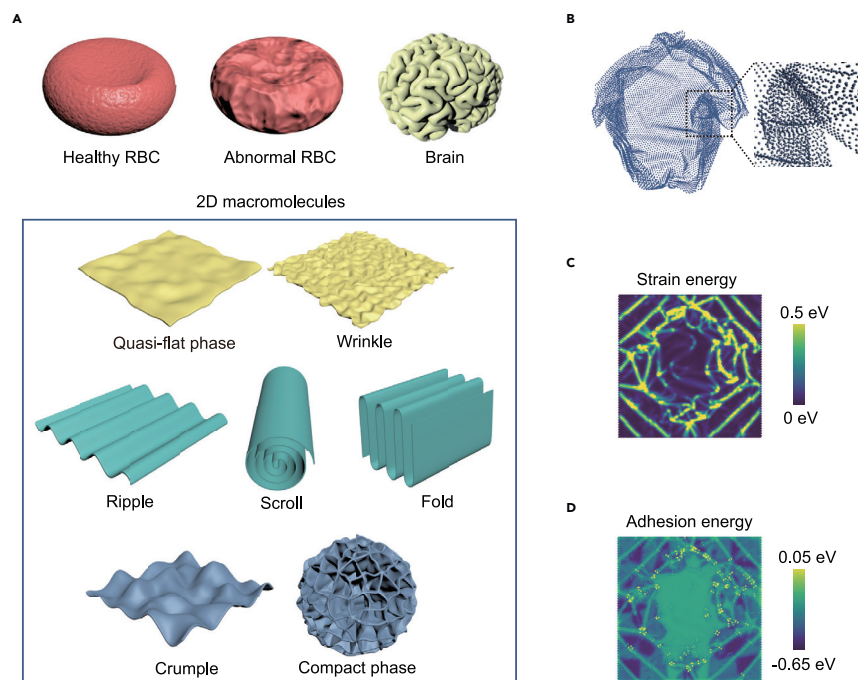


Figure 1. Morphological phases of 2D macromolecules

(A) Phases identified from theoretical and experimental studies, which include the flat,^{6,13,14} quasi-flat,^{9,18} rippled,¹⁹ wrinkled,²⁰ folded,^{13,15,16,23} scroll,^{21,22} crumpled,^{10,12,14,17,18} and compact^{13,14} phases. Similar morphological complexity can be found in the red blood cells (RBCs)²⁴ and the brain.²⁵ These phases are classified into the quasi-flat, folded, and crumpled phases in this work, as indicated by the colors.

(B–D) Digital representations through the 3D map of conformation (B), the 2D map of strain energy (C), and adhesion energy (D). The 2D maps of strain and surface adhesion are strongly correlated due to the physics of mechanical equilibrium (Equation 1).

realistic models predict flat, rippled, wrinkled, crumpled, folded, scrolled, and compact phases,^{6,12–18} which are validated by the experimental studies^{19–23} (Figure 1A).

Morphology of 2D macromolecules plays an important role in defining the microstructures of their macroscopic assemblies in forms of fibers and films, as well as their mechanical, transport, and functional properties.^{26–28} For example, plasticization spinning strategy converts the wrinkled conformation of graphene to the flat one and enlarges the crystallite domains, thus enhancing the mechanical properties of graphene fibers.²⁹ Microfluidic design regulates the alignment and orientation of graphene sheets and improves the transport properties of fibers.³⁰ Graphene films assembled from the flat, folded, and crumpled phases exhibit an improved hydrophobicity in sequence.²³ Solvent-mediated topography regulation triggers reversible fusion-fission transitions of graphene fibers, where the solvent modulates morphologies of the shell and interface in the fibers, switching between the wrinkled and spread phases.³¹

Folded or crumpled phases of 2D macromolecules also define the adsorption and transport processes of ions and molecules in these condensed phases, which are the key to develop relevant energy and environment applications. For example, electrodes of folded graphene provide continuous transport pathways with high electron or ion mobility,³² where the crumpled phases demonstrate high specific surface area pore volume and excellent capacitance.³³ Crumpled graphene balls possess meso- and micro-pores as well as stacking-resistant structures, exhibiting efficient micro-pollutant absorption from water.³⁴ Rationalizing the conformational map is thus of critical importance to understand and control the microstructures and functions of 2D macromolecules and their assemblies.^{28,35}

Although ample morphological phases of 2D macromolecules were reported in previous studies, their classification has been limited by visual inspection of the geometry. In this work, we

use machine learning techniques to discriminate the morphological phases of 2D macromolecules. Data-driven statistical-learning techniques developed for pattern recognition and prediction^{36–38} have been applied in the materials sciences.^{39,40} Unsupervised and supervised learning studies identify the distinct polymer states.^{41–43} For 2D macromolecules, unsupervised learning was used to classify graphene oxide (GO) according to the chemistry (the C/O ratio) and morphology (the mean size of flakes), which were determined by X-ray photoelectron spectroscopy and scanning electron microscopy analysis, respectively.⁴⁴ Supervised learning recognizes nanobubbles in graphene from the electronic density of states spectra, and predicts the height and width of nanobubbles.⁴⁵ Statistical-learning methods combining unsupervised and supervised learning have been utilized in the conformational recognition of molecules and polymers, and the determination of phase transition.^{42,46,47} Self-supervised learning was used to embed geometrical features into the graph neural network to assist in the molecular conformational identification and property prediction.⁴⁶ The pre-training process utilizes abundant unlabeled samples to learn and import the geometrical features into the neural network, and the fine-tune process uses a handful of labeled samples to perform the recognition and property prediction tasks.⁴⁶ The confusion scheme trains models with data that are deliberately labeled incorrectly, and the phase transition can be determined according to the performance of the models trained with different labels.⁴⁷ Following this approach, the configurations of the polymers were recognized, and the critical energies of phase transition were determined.⁴² We thus integrate unsupervised and supervised techniques to provide a tool that can be used to discriminate the conformational phases of 2D macromolecules and offer insights into the transition between them.

Notably, in addition to the geometry of 2D macromolecules, the lattice distortion and topology of surface contact are also of vital importance to understand the microstructures-performance relationship. However, this physics cannot be extracted from the 3D conformation obtained from, for example, experimental computed tomography. To address this issue, we perform coarse-grained molecular dynamics (CGMD)

simulations to generate macromolecular structures of graphene and the physics of molecular interaction. Unsupervised learning is conducted based on the features extracted from the energy of strain and adhesion, or the 3D geometry and topology of the contact. The model trained can be used in supervised learning for morphological recognition and classification using simulation or experimental data. This combined approach allows the physics behind the observable geometrical and topological characteristics to be included in the discussion on the morphological complexity, and the assessment of their significance. The results lay the ground for the understanding of processing-microstructures relationships of 2D macromolecules, and the design principles of macroscopic assemblies with outstanding performance and functions.

RESULTS

Digital representation of 2D macromolecules

We generate 2,484 samples of conformation from the CGMD simulations, as well as the potential energy of each atom, bond, dihedrals, and non-bonding interacting pairs (see [experimental procedures](#) for details). The initial configuration of graphene is a flat square sheet with lateral size L of 100 nm. We use isotropic, spherical and anisotropic, and cylindrical confinement, as well as their linear combination, to trigger conformational changes of the 2D macromolecules.^{17,23,48} By further exploring the parameter space spanned by the temperature and the bending stiffness, the simulation results cover a large subspace of the morphological phases ranging from 1D to 3D.

2D macromolecules can be represented directly using the point set from numerical simulations or experimental tomography. This approach captures the full geometrical information including the curvature and a distance map in the 3D Euclidean space. However, although the ridges or vertices can be identified by their geometrical features, the metric changes in the basal plane and out-of-plane bending of 2D macromolecules cannot be represented in the point set without a reference geometry. Moreover, the physics of surface contact cannot be extracted from a point-set representation, where the intramolecular bonding network and surface contact cannot be distinguished ([Figure 1B](#)).

The conformation is also analyzed through the 2D map of potential energy to extract key conformational features,²³ which captures the physics of lattice distortion (metric changes, bending) and surface adhesion implicitly from the energy of the bonded and non-bonding interaction (see [experimental procedures](#)). Mapping the atomic positions into the initial planar configuration of 2D macromolecules, the network structure of ridges and vertices can be visualized from the map of strain energy ([Figure 1C](#)). Ridges are created by out-of-plane bending, while the vertices accommodate in-plane deformation along with bending. Physical contact forms between regions of the 2D macromolecules through the map of surface adhesion ([Figure 1D](#)). However, the topological information is missing in the 2D representation, which could be measured by the distance map in the 2D manifold, which is constructed from the initial configuration. The combination of the 3D point set and the 2D energy map could and should thus be combined to understand the physics behind the geometry and topology of macromolecules.

The conformation of 2D macromolecules is determined by the competition between the resistance to elastic deformation and surface interaction that could be adhesion or steric repulsion. The geometry of the 2D manifold and the topology of contact thus are closely tied to the deformation and surface interaction. Thermal fluctuation also plays a role, especially in the solution environment, in triggering the morphological changes. The potential energy of 2D macromolecules can be modeled through the generalized Helfrich functional²

$$U = \int dS \left[\frac{1}{2} \kappa (2H - c_0)^2 + \bar{\kappa} K \right] + 2\gamma S_c, \quad (\text{Equation 1})$$

where S is the surface area. H and K are the mean and Gaussian curvature, respectively. κ is the bending rigidity, $\bar{\kappa}$ is the Gaussian rigidity, and the extrinsic geometry term $\bar{\kappa}K$ measures the coupling between out-of-plane bending and in-plane deformation. c_0 is the spontaneous curvature induced by topological defects embedded in the 2D macromolecules, which is zero here.⁴⁹ γ is the surface energy density determined by the van der Waals or electrostatic interaction, and S_c is the area of contact. This functional, in combination with the entropy term, defines the geometry and topology of 2D macromolecules in 3D space. The physics behind the conformational phases can thus be extracted from the potential energy of lattice distortion and surface contact, as well as the geometrical and topological measures of the conformation.

Feature extraction

The flowchart of conformational classification and feature extraction in this work is illustrated in [Figure 2](#). The solvent-accessible surface area (SASA) and radius of gyration (R_g) are the two key geometrical features measuring the surface exposure and the compactness of conformation, respectively (see [experimental procedures](#).^{17,48} Surface contact is a topological feature that modulates the transport processes through the open spaces embedded in the condensed phases, such as the folds and crumples. A localization factor (L_F) is defined to characterize the degree of localization for surface contact from the 2D distance map, discriminating the local and long-range modes of contact (see [experimental procedures](#)). The combination of SASA, R_g , and L_F provides multi-resolution characterization of the 3D geometry and topology of contact for 2D macromolecules, which are used for labeling in the unsupervised learning.

Features in the 2D energy map of strain and adhesion are extracted by the perceptual Hash (pHash) algorithm⁵⁰ and principal-component analysis (PCA).⁵¹ The pHash algorithm converts the pixels information of images into a string of fingerprints for comparison. PCA further reduces the dimension of these fingerprints to yield the dominant features, the physics of which can be discussed through comparison with the characteristics evaluated directly from the 3D point set.

Correlation between the first and second principal features (P1 and P2, respectively) extracted from the 2D map of strain energy and SASA, R_g , L_F measured from the 3D conformation are summarized in [Figure 3](#). The results show that P1 has the strongest correlation with R_g ([Figure 3B](#)), and P2 is most relevant to L_F ([Figure 3F](#)), which suggests that P1 captures the global shrinkage of

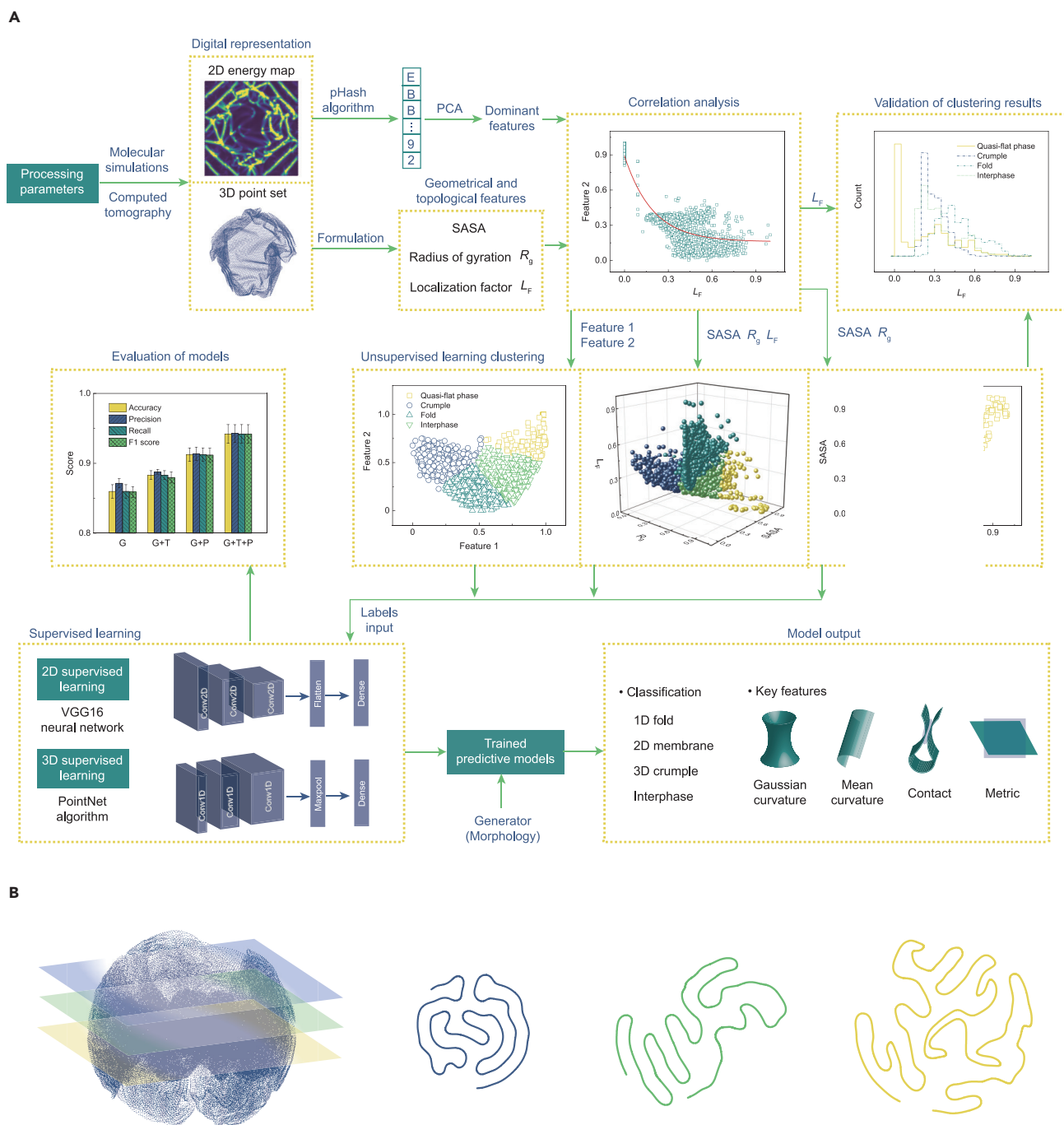


Figure 2. Machine learning procedures

(A) Flowchart of conformational recognition and features extraction for 2D macromolecules.
(B) Illustration of the computed tomography data of 2D macromolecules.

the morphology, while P2 measures the surface contact. SASA is more relevant to P2 than P1 (Figures 3A and 3D). As L_F has the largest correlation factor with the principal features (Figure 3F), we use SASA and R_g for unsupervised classification of the conformation, and L_F for the validation. We also analyze the energy map of surface adhesion, which is highly correlated with the

strain energy map (Figure S1). For this reason, our following discussion is limited to the 2D map of strain energy.

Statistical learning

Unsupervised learning based on the extracted features (SASA, R_g , L_F) are performed to label the data. We first use the

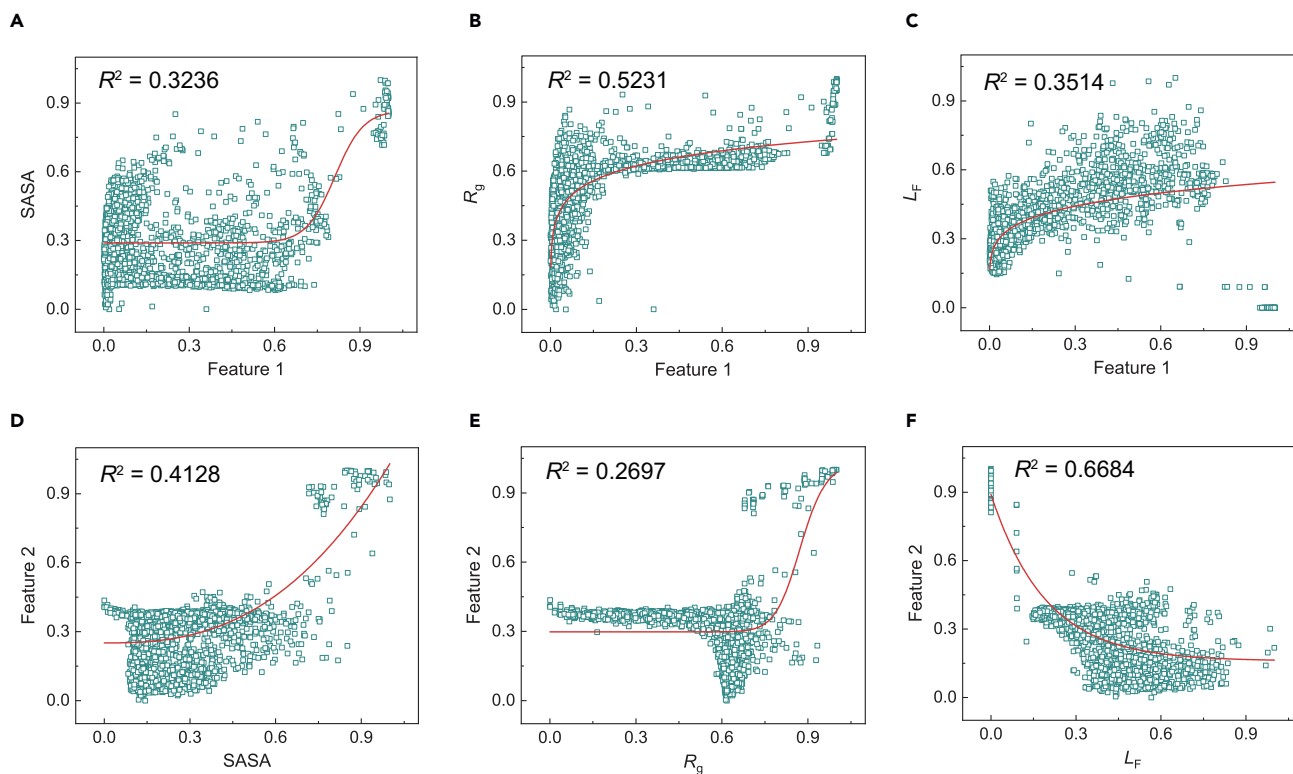


Figure 3. Correlation between features extracted from the 2D map of strain energy and morphological parameters extracted from the 3D point set

(A–C) The correlation between the first principal feature (P1) and nondimensionalized values of SASA, R_g , and L_F .
(D–F) The correlation between the second principal feature (P2) and nondimensionalized values of SASA, R_g , and L_F .

K-means algorithm⁵² to classify the conformation in the parameter space of SASA and R_g only. Four classes are identified as quasi-flat, crumpled, and folded phases, as well as the interphases (Figure 4A). Quasi-flat phases have large SASA and R_g values for high surface exposure and low shrinkage. Folded phases have a relatively larger value of R_g than that of the crumples since anisotropic folds only contract in one direction, while isotropic crumples shrink in all directions. The interphases are located between these three well-defined phases. The distribution of L_F is calculated to validate the clustering-based classification (Figure 4B). In the quasi-flat phases, L_F shows a peak at the characteristic length scale set by the intramolecular bonding network since no contact is formed. The peak of L_F of the folded phase has the largest value, indicating that the contact has a long-range nature, while the crumple has a peak at the intermediate distance, and the contact is local. The value of L_F for the interphases resides between the other three phases, displaying the nature of transition states.

The effect of contact topology is further assessed by including L_F as one of the labels in addition to SASA and R_g . The results show that the information of topological contact improves the physical significance of clustering, which is of crucial importance for the transport and adsorption processes in the condensed phases or assemblies of 2D macromolecules. For example, we find two samples of the same size with different morphologies but similar SASA and R_g values (Figure 4D). Their L_F values, however, show discrepancy. Geometrical clustering classifies these

two samples as quasi-flat phase (Figure 4A), but the topological consideration by adding L_F in the labeling process corrects the prediction by recognizing one of them with the larger value of L_F to be a fold (Figure 4C).

The energy landscape of morphological phases defines the richness of the morphological phases in thermal equilibrium according to the Boltzmann factor $\exp(-k_B T)$ (Figures 4E and S2). The crumples and folds own high and low potential energies for their strong lattice distortion and surface adhesion, respectively, and those of the quasi-flat phases or interphases are between them. The path of transition between these phases can be inferred from the confusion scheme, where the accuracy of prediction at specific potential energy measures the likelihood of the corresponding morphology as a transitional one (see [experimental procedures](#)).⁴⁷ Comparing the accuracy with the distribution of morphological phases over the energy space suggests that the interphases could bridge the crumples and folds, while the quasi-flat phases and some of the interphases show features of transition states (Figure 4E). This result agrees with the experimental finding that direct transition between the crumples and folds is prohibited by the symmetry and should proceed through intermediate phases such as the quasi-flat one.²³

Supervised learning is carried out using the labels obtained from unsupervised learning to recognize the conformation of 2D macromolecules represented by the 2D map of strain energy or the 3D point set. The labeling using SASA and R_g contains geometrical (G) features only (Figure 4A), and that with L_F

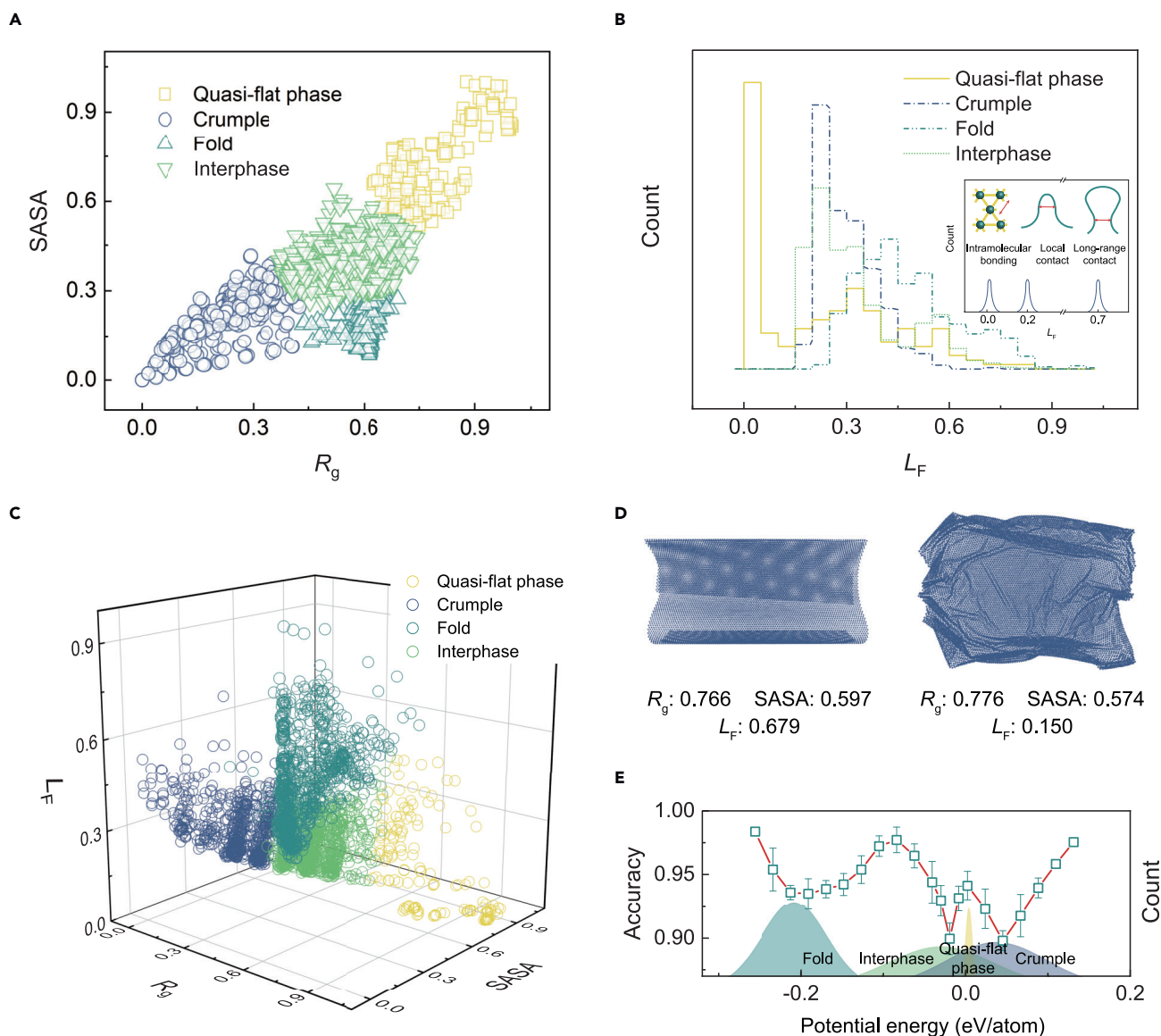


Figure 4. Classification of morphological phases from unsupervised learning

(A) Clustering of samples in the space spanned by the normalized values of SASA and R_g .

(B) Distribution of normalized L_F , which is defined to capture the bimodal feature of the intramolecular bonding network and (local, long range) surface contact (see the inset for details), respectively.

(C) Clustering using SASA, R_g , and L_F labels.

(D) Two samples with similar SASA and R_g values but different L_F .

(E) The distribution of morphological phases in the space of potential energy. The accuracy of prediction in the confusion scheme is used to identify the possible transition states from the peaks, while the valleys are the well-identified phases. The standard deviation is reported in the error bars.

includes both geometrical (G) and topological (T) information (Figure 4C). For the representations of macromolecular conformation, the 3D point set comprises the geometrical features, while the 2D energy map includes the physical characteristics of lattice distortion and deformation (P). The combination of labeling and digital representations thus produces four models, which are: geometrical labeling and 3D point-set representation (the G model), geometrical and topological labeling and 3D point-set representation (the G + T model), geometrical labeling and 2D energy-map representation (the G + P model), geomet-

rical and topological labeling and 2D energy-map representation (the G + T + P model). Using these models, we randomly sample the data at a ratio of 0.7:0.15:0.15 for the training, validation, and test sets. Data in the training and validation sets are used for training, and the validation set is also used to adjust the hyperparameters. Samples in the test set are used to evaluate the performance of the well-trained model.

The 2D map of strain energy is explored by using the VGG16 neural network.^{53,54} After training, samples in the test set are used as the input for prediction. The output includes the classified

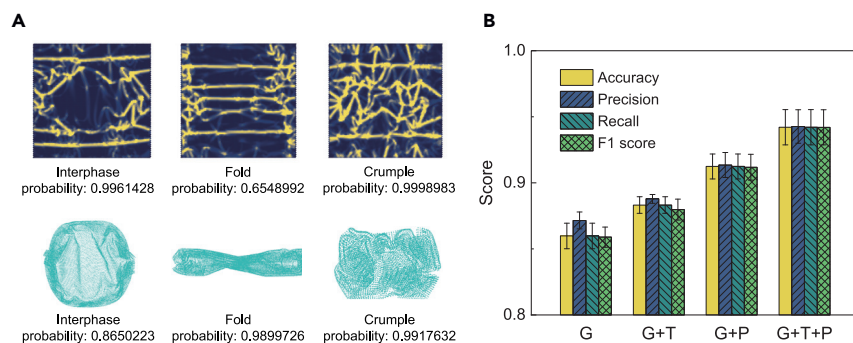


Figure 5. Predictions from statistical learning

(A) Phases predicted by the well-trained supervised model using SASA and R_g as the labels and the corresponding prediction probability. The energy heat images are the maps of strain energy.

(B) Evaluation of the statistical-learning models. The standard deviation is reported in the error bars.

phases and their probabilities (Figure 5A). The confidence of prediction for a specific class is scored by the probability. The 3D morphological map is studied using the PointNet algorithm.⁵⁵ The classification and the corresponding probability of prediction are summarized (Figure 5A). We use the metrics of accuracy, precision, recall, and F1 score to evaluate the models (see [experimental procedures](#)). The metric scores of different models are summarized in Table S1. We find that the geometrical features possess a fundamental contribution to all the models. The topological and physical information notably improves the performance of model (Figure 5B). The G + T + P model with the geometrical, topological, and physical characteristics demonstrates the best performance (Figure 5B). We also label the data using the principal features extracted from the 2D energy map (Figures S3A and S3B). The performance of these models is not as competitive as the G + T + P model (Figure S3C). By further considering that extracting SASA, R_g , and L_F is more convenient than that for the principal features, the usage of direct geometrical and topological labels in unsupervised clustering is preferred in practice.

DISCUSSION

2D versus 3D supervised learning

The 2D map of the potential energy identifies the ridges, vertices, and surface adhesion, but the metric changes and curvature of geometry and the topology of surface contact are only implicitly considered through the energy terms. For the true positives, the higher probability of prediction from 2D supervised learning using the G + P model indicates that physical discrimination through the energy map is more feasible, while geometrical recognition is more convenient in 3D by using the G model (Figure 5A).

Data failing in the prediction from labels in the unsupervised learning is dominated by the interphases with the nature of transition states (Figure S4A). The false negatives of quasi-flat and crumpled phases are less significant in 2D learning, since the lattice distortion can be recognized. The quasi-flat phase has a sparse network of ridges and vertices, while the crumpled phase has a dense one. The false negatives of folds and the interphases are less significant in 3D learning as the geometrical feature is more significant. The fold has a distinct 1D anisotropic feature, while the interphases have mild characteristics in geometry compared with the quasi-flat, folded, and crumpled phases.

Unsupervised and supervised learning

False predictions may be attributed to poor labeling from unsupervised learning, the results of which are used as the input for su-

pervised learning. For example, there is a sample labeled as the interphases, but predicted as a fold by both 2D (G + P) and 3D (G) supervised learning (Figure S4B). The normalized value of L_F for this conformation is 0.48108, which aligns with the feature of folds (Figure S4B). The supervised learning thus can outperform unsupervised learning by correcting poor labeling, although unsupervised learning can also be improved by defining more suitable features for extraction (Figure S5). The maximum score achieved by the G + P + T model is 0.9515, which is limited by the physics of labels chosen (SASA, R_g , and L_F). The graph neural networks that implant the topological information into the structure of neural networks may be used for improvement.⁵⁶

Surface interaction

Surface contact in 2D macromolecules can be regulated by solvent or surface modification. For example, flat GO remains stable in the solution with low concentrations of dimethylformamide (DMF), while folded phases are identified in the solution with divalent Ca^{2+} ions due to the short-range attraction between GO, and crumpling occurs in the hydrazine (N_2H_4) solution as a result of the long-range attraction after reduction.²³ Changes in the surface charge density trigger the transition from nanomembranes to nanoscrolls, which reduce the electrostatic potential barrier of nucleation and electrostatic repulsion during the process of scrolling.²² The reversibility of phase transitions between the flat phase and folds or crumples depends on the nature of surface interaction. Surface adhesion yields an enthalpic penalty for the process of unfolding or uncrumpling, while repulsion can drive these reverse processes as the boundary constraints are released. We explore the effects of surface interaction by tuning the nature of interaction from being attractive to repulsive in the simulations. The values of SASA, R_g , and L_F are nondimensionalized by $2L^2$ and L , which are the surface area and lateral size of the flat square sheet, respectively. The results using the G model show that, SASA and R_g of 2D macromolecules with repulsive surface interaction are similar as those measured with attraction (Figure 6A). However, the L_F value increases from attraction to repulsion, indicating the shift from local to long range (Figure 6B).

The size effects

The data used for morphological classification in this work is limited by the size of simulated models. The size effects can be assessed through a dimensionless measure of L/L_p , where L_p is the 2D persistence length, $L_p = \xi \exp(2\pi\kappa/k_B T)$, ξ is the short distance cutoff (the lattice constant), and k_B is the Boltzmann constant.² For $L \leq L_p$, the 2D macromolecules behave as rigid or elastic sheets. Our simulations are focused on models with $L > L_p$ and $L \gg d$, where d is the spacing of the surface

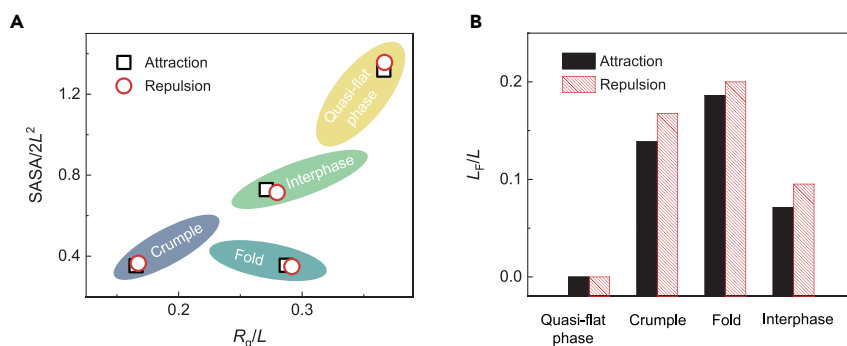


Figure 6. Effects of surface interaction

(A) The distribution of samples with attractive and repulsive surface interaction in the parameter space of SASA and R_g , which shows minor changes. (B) L_F calculated for samples with attractive and repulsive surface interaction, which shows significant dependence.

contact, which is 0.335 nm for graphene and 0.6 – 0.7 nm for GO. By tuning the values of κ and T (see [experimental procedures](#)), we modify the value of L_p , and generate a rich spectrum of conformational phases (Figure S6A). We find that the values of SASA and R_g are bounded by the values of the 1D cylindrical, 2D flat, and 3D spherical phases. Moreover, the bounds obtained for samples with $L = 50, 100,$ and 500 nm are close (Figure S6B), suggesting that this size effect is not a major issue in the context of discussion here. However, the geometrical features, such as the length of folds and the area of contact, are physically limited by the value of L chosen here for the consideration of computational costs.

The completeness of data

The performance of phase recognition here is limited by the space of conformation generated by our molecular simulations where spatial confinement is enforced to trigger the phase changes. The correlation between the resulted phases and the conditions of confinement shows that crumples and folds are mainly generated by the spherical and cylindrical confinement, respectively, while quasi-flat phases and interphases can be obtained under spherical or cylindrical confinement or their combination (Figure S8A). The effect of temperature in the range of 300 – 900K is not significant since the kinetic energy is much lower than the energy corrugation during the phase changes (Figure S8B). At high temperature, changes in the covalent bond network, such as the sp^2 - sp^3 transition, may be activated, which could implant lattice imperfections and crosslinks at the contact. The behaviors of the morphological phases generated by spatial confinement can be considered as enthalpic, while the entropic effects are negligible. In the solvent, however, the entropic contribution to the free energy could be significant, especially for the quasi-flat phases and interphases. These results suggest that our dataset can be expanded, for example, by performing long-time equilibrium simulations, or non-equilibrium ones using the free energy techniques,⁵⁷ which are crucial for accurate identification of the paths and energy barriers of phase transitions, although the procedure could be technically challenging and computationally costly.

Conclusion

To summarize, we utilize machine learning techniques to classify the morphological phases of 2D macromolecules through the 2D map of lattice distortion, surface adhesion, and 3D conformation. SASA, R_g , and L_F are defined as the key conformational measure for the surface exposure, compactness, anisotropy,

morphological complexity. Unsupervised learning clusters the samples based on their geometrical and topological features, and provides the labels needed in subsequent supervised learning. 2D supervised learning identifies a variety of morphological phases from the potential energy of lattice distortion and surface adhesion. 3D supervised learning completes the discrimination by extracting the geometrical and topological information with a distance map in the 2D manifold supplied.

The well-trained models established with the geometrical, topological, and physical information can be used for recognition and classification of the simulation or experimental data, which may consist of geometrical information only, for practical consideration. The model can be applied to the assemblies of multiple 2D macromolecules, and takes the advantage in the identification of defects, which may create localized lattice distortion and modify the interaction between different regions of the 2D macromolecules, resulting in additional features of geometrical deformation and topological contact (Figure S7). This work thus lays the ground for the understanding of the microstructures and material properties of 2D macromolecules in their condensed phases or macroscopic assemblies, and could be extended to other complex surfaces in, for example, the red blood cells²⁴ and the brain.²⁵ Our study also suggests that a theoretical description to characterize the morphology of 2D macromolecules should include topological features, such as the local and long-range contacts, in addition to the geometrical representation, which results from the competition between the deformation and surface interaction.

EXPERIMENTAL PROCEDURES

Resource availability

Lead contact

Request for information and resources used in this article should be addressed to Dr. Zhiping Xu (xuzp@tsinghua.edu.cn).

Materials availability

There were no physical materials associated with this study.

Data and code availability

The data used in this study are generated from molecular simulations. The codes and data used in the paper are available at <https://zenodo.org/badge/latestdoi/452994542>.

Molecular simulations

We use a hexagonal lattice to construct the coarse-grained (CG) models of 2D macromolecules.⁵⁸ The atoms are clustered into beads with equal masses. The bonding interaction between bonded beads is modeled as a linear elastic spring with stiffness $k_s = (3^{1/2}/2)Yt$, where Y is the Young's modulus and t is the thickness. The in-plane elastic energy is thus $U_s = k_s(r - r_0)^2/2$,

where r is the bond length with an equilibrium value of r_0 . The bending resistance is modeled as a harmonic dihedral with stiffness $k_b = (2/\sqrt{3})\kappa$, where κ is the bending stiffness.⁵⁹ The out-of-plane bending energy is $U_b = k_b(1 + \cos\varphi)$, where φ is the angle of dihedral with an equilibrium value of $\varphi_0 = \pi$. The non-bonded interaction between the beads is modeled by the Lennard-Jones 12 – 6 potential $U_c = 4\varepsilon[(\sigma/r)^{12} - (\sigma/r)^6]$ with the 1 – 2, 1 – 3, and 1 – 4 neighbor exclusion enforced, where the parameters σ and ε are fitted through the spacing at the contact and the cohesive energy.⁶⁰ We use the force field parameters developed for GO as the reference, and tune k_b and the cutoff distance r_c in evaluating U_c for the bending stiffness and surface interaction ($r_c = 2.5\sigma$) from attraction to repulsion ($r_c = 2^{\frac{1}{2}}\sigma$), respectively. The parameters of the CG force field used in this study are summarized in Table S2, which can be further modified for generation to other for 2D macromolecules.

A large-scale atomic/molecular massively parallel simulator was used to perform all CGMD simulations.⁶¹ The initial conformation of 2D macromolecules is flat, and the lateral size L is 100 nm if not specified otherwise. Conformational changes are driven by applying spherical, cylindrical, or combined linear elastic constraints. The constraints with a harmonic spring move slowly to interact with the beads representing the 2D macromolecules, triggering the conformational transition. The spring stiffness is set to 20 kcal/mol, and the constant speed of constraints is defined by the compression ratio R and duration τ . We generate 2,484 morphological phases by adjusting the temperature in the range of 300 – 900 K, the bending stiffness between 1 and 60 kcal/mol, and the conditions of constraint (mode, speed, ratio of compression, and duration). The compression ratios of spherical and cylindrical constraints are $R = 0.4 - 0.9$ and $0.15 - 0.9$, respectively, where the lower values correspond to the compact limits. The duration of compression is set to $\tau = 0.5 - 2$ ns. A Langevin thermostat is used for temperature control and to include the implicit solvent effect. The time step is 1 fs, which assures the stability of the numerical integration.

SASA, R_g , and L_F

SASA measures the surface area of graphene that is accessible to a solvent, which is calculated using the Shrake-Rupley algorithm,⁶² where a bead of probe with a radius of 2.5 nm is chosen.⁶³

The radius of gyration tensor S is defined as

$$S = \frac{1}{N} \sum_{i=1}^N \begin{bmatrix} (x_i - x_c)^2 & (x_i - x_c)(y_i - y_c) & (x_i - x_c)(z_i - z_c) \\ (y_i - y_c)(x_i - x_c) & (y_i - y_c)^2 & (y_i - y_c)(z_i - z_c) \\ (z_i - z_c)(x_i - x_c) & (z_i - z_c)(y_i - y_c) & (z_i - z_c)^2 \end{bmatrix}, \quad (\text{Equation 2})$$

where N is the number of CG beads, $r_i = (x_i, y_i, z_i)$ is the Cartesian coordinates of the i -th bead, and $r_c = (x_c, y_c, z_c)$ is that of the center of mass. The scalar radius of gyration is $R_g^2 = \frac{1}{N} \sum_{i=1}^N (r_i - r_c)^2$.

To quantitatively measure the surface contact, we count the number of contact N for each bead i in the 3D conformation by using a distance cutoff of $2.5\sigma = 3.7$ nm for the pairs of interacting beads. This distance displays a bimodal feature originating from the bonded and non-bonded interaction. The distance between beads i and j in the reference 2D lattice is defined as D_{ij} , and then the averaged 2D distance of contact is

$$\alpha_i = \frac{\sum_j D_{ij}}{N_i}, \quad (\text{Equation 3})$$

where the summation is taken over all N_i beads in contact with i . The value of α_i is related to the bond length for the planar phase, or the interlayer distance for the contact. The number of beads with contact ($\alpha_i \neq 0$) is counted as M_i . A localization factor L_F is then defined as

$$L_F = \frac{\sum_i \alpha_i}{M_i}, \quad (\text{Equation 4})$$

where the summation is taken over M_i . The distribution of L_F thus includes the contributions from the intramolecular bonding network and surface contact in the local and long-range modes (Figure 4B).

Machine learning methods

For unsupervised learning, the K-means algorithm⁵² is used for unsupervised clustering. The number of clusters is set to four considering the 1D cylindrical,

2D flat, and 3D spherical limits, as well as the interphase characteristics. We conduct unsupervised clustering with four to six classes according to the 1D-3D features recognized in the simulation and experimental results. Clustering with more than four classes results in sub-division of the crumpled phase (Figure S9A) or the interphase (Figure S9B), which are named as severe and mild sub-classes. However, these sub-divisions do not show essential difference in the geometry and topology. For example, both severe and mild crumples display features of isotropy and local contact, and the two interphases both demonstrate the nature of transition states with only minor difference in the degree of shrinkage. Therefore, our discussion in this study is focused on the results using four clusters to avoid redundant sub-division or over-refined classification.

2D supervised learning uses the results of unsupervised learning for labeling. The model contains 13 convolutional layers, 1 flattening layer, and 2 fully connected layers. The rectified linear unit (RELU) is used as the activation function except for the last layer, which uses the softmax function. The VGG16 neural network⁵³ implemented in TensorFlow⁵⁴ is used to construct the architecture of the convolution layers. The root-mean-square prop (RMSprop) algorithm is used as the optimizer. The learning rate is 2×10^{-5} without further specification, and the cross-entropy is chosen as the loss function.

3D supervised learning uses the PointNet algorithm implemented in Pytorch for morphological classification.⁵⁵ The model consists of three convolution layers, one maximum pooling layer, and three fully connected layers. The activation function is RELU except for the last layer, which uses the softmax function. The adaptive moment estimation (Adam) and StepLR are used as the optimizer and learning rate scheduler. The negative log likelihood is used as the loss function.

The confusion scheme is used to explore the possible path of transition between the morphological phases.⁴⁷ In this scheme, a value of potential energy (E) in the range of $[E_{\min}, E_{\max}]$ is specified to discriminate the morphological data into two classes ($[E_{\min}, E]$ and $[E, E_{\max}]$). The accuracy of this binary classification through supervised learning is then calculated. The critical energy (E_c) is determined at the local maxima of the accuracy, which may correspond to the transition state between the morphological phases with potential energies lower or higher than E_c . On the other hand, the local minima correspond to the well-recognized classes.

The true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) of the morphological phases are counted to calculate the metrics of accuracy ($\frac{\sum_i TP_i}{\sum_i TP_i + FN_i}$), precision ($\sum_i \text{weight}_i \times \text{precision}_i$), recall ($\sum_i \text{weight}_i \times \text{recall}_i$), and F1 score ($\sum_i \text{weight}_i \times \text{F1score}_i$), where i is the index of morphological phases, and $\text{weight}_i = \frac{TP_i + FN_i}{\sum_i TP_i + FN_i}$, $\text{precision}_i = \frac{TP_i}{TP_i + FP_i}$, $\text{recall}_i = \frac{TP_i}{TP_i + FN_i}$, $\text{F1score}_i = \frac{2 \times \text{precision}_i \times \text{recall}_i}{\text{precision}_i + \text{recall}_i}$.⁶⁴ The weighted

average is introduced to account for the class imbalance. The accuracy and recall are the same in this condition. Accuracy reflects the overall predictive power of the model; that is, the proportion of correctly identified phases. Precision measures the exactness of the model predictions through the ratio of correct recognition in the prediction of a certain phase. Recall characterizes the effectiveness of the model to identify positive labels; that is, the ratio of the identified phases in the actual class of certain phases. F1 score is a comprehensive metric considering contributions from both precision and recall. A high F1 score of a model indicates high-precision prediction of morphological phases and complete recognition of conformation. Three independent experiments are performed by varying the training epoch. The mean and standard deviation of metrics of different models are calculated.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.patter.2022.100497>.

ACKNOWLEDGMENTS

This study was supported by the National Natural Science Foundation of China through grants 11825203, 11832010, 11921002, and 52090032. The computation was performed on the Explorer 100 cluster system of Tsinghua National Laboratory for Information Science and Technology.

AUTHOR CONTRIBUTIONS

Conceptualization, Z.X.; methodology, Y.Z. and J.Q.; investigation, Y.Z., J.Q., and S.W.; writing, Y.Z. and Z.X.; funding acquisition, Z.X.; supervision, Z.X.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: January 27, 2022

Revised: March 5, 2022

Accepted: March 28, 2022

Published: April 22, 2022

REFERENCES

- Guo, Q. (2016). *Polymer Morphology: Principles, Characterization, and Processing* (John Wiley & Sons).
- Nelson, D., Piran, T., and Weinberg, S. (2004). *Statistical Mechanics of Membranes and Surfaces* (World Scientific). <https://doi.org/10.1142/5473>.
- Li, P., Wang, S., Meng, F., Wang, Y., Guo, F., Rajendran, S., Gao, C., Xu, Z., and Xu, Z. (2020a). Conformational scaling relations of two-dimensional macromolecular graphene oxide in solution. *Macromolecules* 53, 10421–10430. <https://doi.org/10.1021/acs.macromol.0c01425>.
- Cates, M. (1984). Statics and dynamics of polymeric fractals. *Phys. Rev. Lett.* 53, 926–929. <https://doi.org/10.1103/PhysRevLett.53.926>.
- Kantor, Y., Kardar, M., and Nelson, D.R. (1987). Tethered surfaces: statics and dynamics. *Phys. Rev. A* 35, 3056–3071. <https://doi.org/10.1103/PhysRevA.35.3056>.
- Plischke, M., and Boal, D. (1988). Absence of a crumpling transition in strongly self-avoiding tethered membranes. *Phys. Rev. A* 38, 4943–4945. <https://doi.org/10.1103/PhysRevA.38.4943>.
- Liu, D., and Plischke, M. (1992). Monte Carlo studies of tethered membranes with attractive interactions. *Phys. Rev. A* 45, 7139–7144. <https://doi.org/10.1103/PhysRevA.45.7139>.
- Kantor, Y., and Kremer, K. (1993). Excluded-volume interactions in tethered membranes. *Phys. Rev. E* 48, 2490–2497. <https://doi.org/10.1103/PhysRevE.48.2490>.
- Paczuski, M., Kardar, M., and Nelson, D.R. (1988). Landau theory of the crumpling transition. *Phys. Rev. Lett.* 60, 2638–2641. <https://doi.org/10.1103/PhysRevLett.60.2638>.
- Abraham, F.F., Rudge, W., and Plischke, M. (1989). Molecular dynamics of tethered membranes. *Phys. Rev. Lett.* 62, 1757–1760. <https://doi.org/10.1103/PhysRevLett.62.1757>.
- Gompper, G., and Kroll, D.M. (1997). Network models of fluid, hexatic and polymerized membranes. *J. Phys. Condens. Matter* 9, 8795–8834. <https://doi.org/10.1088/0953-8984/9/42/001>.
- Åström, J., Timonen, J., and Karttunen, M. (2004). Crumpling of a stiff tethered membrane. *Phys. Rev. Lett.* 93, 244301. <https://doi.org/10.1103/PhysRevLett.93.244301>.
- Abraham, F.F., and Kardar, M. (1991). Folding and unbinding transitions in tethered membranes. *Science* 252, 419–422. <https://doi.org/10.1126/science.252.5004.419>.
- Wen, X., Garland, C.W., Hwa, T., Kardar, M., Kokufuta, E., Li, Y., Orkisz, M., and Tanaka, T. (1992). Crumpled and collapsed conformation in graphite oxide membranes. *Nature* 355, 426–428. <https://doi.org/10.1038/355426a0>.
- Radzihovsky, L., and Toner, J. (1995). A new phase of tethered membranes: Tubules. *Phys. Rev. Lett.* 75, 4752–4755. <https://doi.org/10.1103/PhysRevLett.75.4752>.
- Bowick, M., Falcioni, M., and Thorleifsson, G. (1997). Numerical observation of a tubular phase in anisotropic membranes. *Phys. Rev. Lett.* 79, 885–888. <https://doi.org/10.1103/PhysRevLett.79.885>.
- Cranford, S.W., and Buehler, M.J. (2011). Packing efficiency and accessible surface area of crumpled graphene. *Phys. Rev. B* 84, 205451. <https://doi.org/10.1103/PhysRevB.84.205451>.
- Xu, Y., Pospisil, M.J., and Green, M.J. (2016). The effect of bending stiffness on scaling laws for the size of colloidal nanosheets. *Nanotechnology* 27, 235702. <https://doi.org/10.1088/0957-4484/27/23/235702>.
- Bao, W., Miao, F., Chen, Z., Zhang, H., Jang, W., Dames, C., and Lau, C.N. (2009). Controlled ripple texturing of suspended graphene and ultrathin graphite membranes. *Nat. Nanotechnol.* 4, 562–566. <https://doi.org/10.1038/nnano.2009.191>.
- Li, Z., Kinloch, I.A., Young, R.J., Novoselov, K.S., Anagnostopoulos, G., Parthenios, J., Galiotis, C., Papagelis, K., Lu, C.-Y., and Britnell, L. (2015). Deformation of wrinkled graphene. *ACS Nano* 9, 3917–3925. <https://doi.org/10.1021/nn507202c>.
- Xu, Z., Zheng, B., Chen, J., and Gao, C. (2014). Highly efficient synthesis of neat graphene nanoscrolls from graphene oxide by well-controlled lyophilization. *Chem. Mater.* 26, 6811–6818. <https://doi.org/10.1021/cm503418h>.
- Tang, B., Gao, E., Xiong, Z., Dang, B., Xu, Z., and Wang, X. (2018). Transition of graphene oxide from nanomembrane to nanoscroll mediated by organic solvent in dispersion. *Chem. Mater.* 30, 5951–5960. <https://doi.org/10.1021/acs.chemmater.8b02083>.
- Wang, Y., Wang, S., Li, P., Rajendran, S., Xu, Z., Liu, S., Guo, F., He, Y., Li, Z., Xu, Z., et al. (2020). Conformational phase map of two-dimensional macromolecular graphene oxide in solution. *Matter* 3, 230–245. <https://doi.org/10.1016/j.matt.2020.04.023>.
- Lim, G.H., Wortis, M., and Mukhopadhyay, R. (2002). Stomatocyte-discocyte-echinocyte sequence of the human red blood cell: evidence for the bilayer-couple hypothesis from membrane mechanics. *Proc. Natl. Acad. Sci.* 99, 16766–16769. <https://doi.org/10.1073/pnas.202617299>.
- Tallinen, T., Chung, J.Y., Rousseau, F., Girard, N., Lefèvre, J., and Mahadevan, L. (2016). On the growth and form of cortical convolutions. *Nat. Phys.* 12, 588–593. <https://doi.org/10.1038/nphys3632>.
- Gupta, A., Sakthivel, T., and Seal, S. (2015). Recent development in 2D materials beyond graphene. *Prog. Mater. Sci.* 73, 44–126. <https://doi.org/10.1016/j.pmatsci.2015.02.002>.
- Akinwande, D., Brennan, C.J., Bunch, J.S., Egberts, P., Felts, J.R., Gao, H., Huang, R., Kim, J.-S., Li, T., Li, Y., et al. (2017). A review on mechanics and mechanical properties of 2D materials: graphene and beyond. *Extreme Mech. Lett.* 13, 42–77. <https://doi.org/10.1016/j.eml.2017.01.008>.
- Zheng, B., Zheng, Z., and Gu, G.X. (2021). Uncertainty quantification and prediction for mechanical properties of graphene aerogels via Gaussian process metamodelling. *Nano Futures* 5, 045004. <https://doi.org/10.1088/2399-1984/ac3c8f>.
- Li, P., Liu, Y., Shi, S., Xu, Z., Ma, W., Wang, Z., Liu, S., and Gao, C. (2020b). Highly crystalline graphene fibers with superior strength and conductivities by plasticization spinning. *Adv. Funct. Mater.* 30, 2006584. <https://doi.org/10.1002/adfm.202006584>.
- Xin, G., Zhu, W., Deng, Y., Cheng, J., Zhang, L.T., Chung, A.J., De, S., and Lian, J. (2019). Microfluidics-enabled orientation and microstructure control of macroscopic graphene fibres. *Nat. Nanotechnol.* 14, 168–175. <https://doi.org/10.1038/s41565-018-0330-9>.
- Chang, D., Liu, J., Fang, B., Xu, Z., Li, Z., Liu, Y., Brassart, L., Guo, F., Gao, W., and Gao, C. (2021). Reversible fusion and fission of graphene oxide-based fibers. *Science* 372, 614–617. <https://doi.org/10.1126/science.abb6640>.
- Wang, B., Ryu, J., Choi, S., Song, G., Hong, D., Hwang, C., Chen, X., Wang, B., Li, W., Song, H.-K., et al. (2018). Folding graphene film yields high areal energy storage in lithium-ion batteries. *ACS Nano* 12, 1739–1746. <https://doi.org/10.1021/acsnano.7b08489>.
- Yoon, Y., Lee, K., Baik, C., Yoo, H., Min, M., Park, Y., Lee, S.M., and Lee, H. (2013). Anti-solvent derived non-stacked reduced graphene oxide for

- high performance supercapacitors. *Adv. Mater.* 25, 4437–4444. <https://doi.org/10.1002/adma.201301230>.
34. Fu, H., Huang, J., and Gray, K. (2021). Crumpled graphene balls adsorb micropollutants from water selectively and rapidly. *Carbon* 183, 958–969. <https://doi.org/10.1016/j.carbon.2021.07.081>.
 35. Fang, W., Peng, L., Liu, Y., Wang, F., Xu, Z., and Gao, C. (2021). A review on graphene oxide two-dimensional macromolecules: from single molecules to macro-assembly. *Chin. J. Polym. Sci.* 39, 267–308. <https://doi.org/10.1007/s10118-021-2515-1>.
 36. Giordani, T., Suprano, A., Polino, E., Acanfora, F., Innocenti, L., Ferraro, A., Paternostro, M., Spagnolo, N., and Sciarrino, F. (2020). Machine learning-based classification of vector vortex beams. *Phys. Rev. Lett.* 124, 160401. <https://doi.org/10.1103/PhysRevLett.124.160401>.
 37. Hobold, G.M., and da Silva, A.K. (2018). Machine learning classification of boiling regimes with low speed, direct and indirect visualization. *Int. J. Heat Mass Transfer* 125, 1296–1309. <https://doi.org/10.1016/j.ijheatmasstransfer.2018.04.156>.
 38. Meng, N., Lam, E.Y., Tsia, K.K., and So, H.K.-H. (2018). Large-scale multi-class image-based cell classification with deep learning. *IEEE J. Biomed. Health Inform.* 23, 2091–2098. <https://doi.org/10.1109/JBHI.2018.2878878>.
 39. Hsu, Y.-C., Yu, C.-H., and Buehler, M.J. (2020). Using deep learning to predict fracture patterns in crystalline solids. *Matter* 3, 197–211. <https://doi.org/10.1016/j.matt.2020.04.019>.
 40. Zhang, Z., and Gu, G.X. (2021). Physics-informed deep learning for digital materials. *Theor. Appl. Mech. Lett.* 11, 100220. <https://doi.org/10.1016/j.taml.2021.100220>.
 41. Wei, Q., Melko, R.G., and Chen, J.Z. (2017). Identifying polymer states by machine learning. *Phys. Rev. E* 95, 032504. <https://doi.org/10.1103/PhysRevE.95.032504>.
 42. Xu, X., Wei, Q., Li, H., Wang, Y., Chen, Y., and Jiang, Y. (2019a). Recognition of polymer configurations by unsupervised learning. *Phys. Rev. E* 99, 043307. <https://doi.org/10.1103/PhysRevE.99.043307>.
 43. Sun, L.-W., Li, H., Zhang, X.-Q., Gao, H.-B., and Luo, M.-B. (2020). Identifying conformation states of polymer through unsupervised machine learning. *Chin. J. Polym. Sci.* 38, 1403–1408. <https://doi.org/10.1007/s10118-020-2442-6>.
 44. Amadei, C.A., Arribas, P., and Vecitis, C.D. (2018). Graphene oxide standardization and classification: methods to support the leap from lab to industry. *Carbon* 133, 398–409. <https://doi.org/10.1016/j.carbon.2018.02.091>.
 45. Song, T., Myoung, N., Lee, H., and Park, H.C. (2021). Machine learning approach to the recognition of nanobubbles in graphene. *Appl. Phys. Lett.* 119, 193103. <https://doi.org/10.1063/5.0065411>.
 46. Fang, X., Liu, L., Lei, J., He, D., Zhang, S., Zhou, J., Wang, F., Wu, H., and Wang, H. (2022). Geometry-enhanced molecular representation learning for property prediction. *Nat. Mach. Intell.* 4, 127–134. <https://doi.org/10.1038/s42256-021-00438-4>.
 47. Van Nieuwenburg, E.P., Liu, Y.-H., and Huber, S.D. (2017). Learning phase transitions by confusion. *Nat. Phys.* 13, 435–439. <https://doi.org/10.1038/nphys4037>.
 48. Chang, C., Song, Z., Lin, J., and Xu, Z. (2013). How graphene crumples are stabilized? *RSC Adv.* 3, 2720–2726. <https://doi.org/10.1039/C2RA21563B>.
 49. Xu, Z. (2015). Defects in two-dimensional materials: topological and geometrical effects. *Chin. Sci. Bull.* 61, 501. <https://doi.org/10.1360/N972015-00383>.
 50. Monga, V., Banerjee, A., and Evans, B.L. (2006). A clustering based approach to perceptual image hashing. *IEEE Trans. Infor. Forensics Secur.* 1, 68–79. <https://doi.org/10.1109/TIFS.2005.863502>.
 51. Wold, S., Esbensen, K., and Geladi, P. (1987). Principal component analysis. *Chemometr. Intell. Lab. Sys.* 2, 37–52. [https://doi.org/10.1016/0169-7439\(87\)80084-9](https://doi.org/10.1016/0169-7439(87)80084-9).
 52. Hartigan, J.A., and Wong, M.A. (1979). Algorithm AS 136: a K-means clustering algorithm. *J. Roy. Stat. Soc. C* 28, 100–108. <https://doi.org/10.2307/2346830>.
 53. Simonyan, K., and Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. In 3rd Int. Conf. Learn. Represent., ICLR 2015, San Diego, CA, USA, May 7–9, 2015, Conference Track Proceedings (IEEE), pp. 1–14. <https://dblp.org/rec/journals/corr/SimonyanZ14a.bib>.
 54. Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., et al. (2016). Tensorflow: a system for large-scale machine learning. In 12th USENIX Symp. Operat. Sys. Design Implement. OSDI 16 (USENIX Association), pp. 265–283. <https://www.usenix.org/conference/osdi16/technical-sessions/presentation/abadi>.
 55. Qi, C.R., Su, H., Mo, K., and Guibas, L.J. (2017). PointNet: deep learning on point sets for 3D classification and segmentation. In Proc. IEEE Conf. Comput. Vis. Pattern Recognit (IEEE), pp. 652–660. https://openaccess.thecvf.com/content_cvpr_2017/html/Qi_PointNet_Deep_Learning_CVPR_2017_paper.html.
 56. Xu, K., Hu, W., Leskovec, J., and Jegelka, S. (2019b). How powerful are graph neural networks? In 7th Int. Conf. Learn. Represent., ICLR 2019, New Orleans, LA, USA, May 6–9, 2019, Conference Track Proceedings (IEEE), pp. 1–17. <https://cs.stanford.edu/people/jure/pubs/gin-iclr19.pdf>.
 57. Bussi, G., and Laio, A. (2020). Using metadynamics to explore complex free-energy landscapes. *Nat. Rev. Phys.* 2, 200–212. <https://doi.org/10.1038/s42254-020-0153-0>.
 58. Yllanes, D., Bhabesh, S.S., Nelson, D.R., and Bowick, M.J. (2017). Thermal crumpling of perforated two-dimensional sheets. *Nat. Commun.* 8, 1381. <https://doi.org/10.1038/s41467-017-01551-y>.
 59. Poulin, P., Jalili, R., Neri, W., Nallet, F., Divoux, T., Colin, A., Aboutalebi, S.H., Wallace, G., and Zakri, C. (2016). Superflexibility of graphene oxide. *Proc. Natl. Acad. Sci.* 113, 11088–11093. <https://doi.org/10.1073/pnas.1605121113>.
 60. Koenig, S.P., Boddeti, N.G., Dunn, M.L., and Bunch, J.S. (2011). Ultrastrong adhesion of graphene membranes. *Nat. Nanotechnol.* 6, 543–546. <https://doi.org/10.1038/nnano.2011.123>.
 61. Plimpton, S. (1995). Fast parallel algorithms for short-range molecular dynamics. *J. Comp. Phys.* 117, 1–19. <https://doi.org/10.1006/jcph.1995.1039>.
 62. Shrake, A., and Rupley, J.A. (1973). Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *J. Mol. Biol.* 79, 351–371. [https://doi.org/10.1016/0022-2836\(73\)90011-9](https://doi.org/10.1016/0022-2836(73)90011-9).
 63. Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. *J. Mol. Graph.* 14, 33–38. [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5).
 64. Sokolova, M., and Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Inf. Process. Manag.* 45, 427–437. <https://doi.org/10.1016/j.ipm.2009.03.002>.