ARTICLE OPEN

Coarse-grained molecular dynamics integrated with convolutional neural network for comparing shapes of temperature sensitive bottlebrushes

Soumil Y. Joshi [1], Samrendra Singh[1] and Sanket A. Deshmukh [1] ✉

Quantification of shape changes in nature-inspired soft material architectures of stimuli-sensitive polymers is critical for controlling their properties but is challenging due to their softness and flexibility. Here, we have computationally designed uniquely shaped bottlebrushes of a thermosensitive polymer, poly(N-isopropylacrylamide) (PNIPAM), by controlling the length of side chains along the backbone. Coarse-grained molecular dynamics simulations of solvated bottlebrushes were performed below and above the lower critical solution temperature of PNIPAM. Conventional analyses (free volume, asphericity, etc.) show that lengths of side chains and their immediate environments dictate the compactness and bending in these architectures. We further developed 100 unique convolutional neural network models that captured molecular-level features and generated a statistically significant quantification of the similarity between different shapes. Thus, our study provides insights into the shapes of complex architectures as well as a general method to analyze them. The shapes presented here may inspire the synthesis of new bottlebrushes.

INTRODUCTION

The human brain identifies and organizes everything in the macroscopic world based on an important visual attribute of objects––their shape[1–3]. However, at length scales not visible to the naked eye, shapes of biological and soft matter architectures have also been shown to influence the properties and functions of these structures. While the concave disc-like shape of red blood cells allows for their smooth travel in the bloodstream, the conical shape of the human immunodeficiency viral (HIV) capsid or the bullet-like shape of rabies viral capsid are essential for their infectivity[4–6]. These shapes from nature have served as an inspiration for the scientific community to synthesize soft- and hard-nanomaterials of different shapes[7–11]. Considerable research has been carried out to study the effect of shape on electrical, thermal, and mechanical properties of such macromolecules[12–17], resulting in applications like portable displays, heat exchangers, biological imaging, and drug discovery[18]. However, synthetic chemists face limitations with regards to precisely controlling the shapes and dimensions of these soft-nanomaterials; often representing complex shapes with their easier-to-synthesize simpler analogues[19,20]. Thus, an obvious question is, can these simplified shapes retain the structure and properties of their analogous complex shapes?

Various architectures of soft materials such as self-assembled nanofibers, nanoribbons, and nanocoils of peptide amphiphiles, star-shaped polymers and self-assembled DNA superlattices, have been reported in the literature[21–23]. An emerging class of complex polymer architectures that is inspired by the shapes of proteoglycans is bottle brush polymers (BBPs)[24,25]. The BBPs consist of a long polymeric backbone densely grafted with polymeric side chains. They have been regularly used for applications including lubrication and photonics where the self-assemblies of these structures are essential[24,25]. The conformations of individual BBPs, resulting from structural parameters such as grafting density and side chain length greatly influence their properties as lubricants as

well as their resulting self-assemblies[26–28]. While the conformations of worm-like BBPs have been studied before[29], an in-depth investigation into the conformations of non-cylindrical BBPs has not been reported which could help understand the relationships between BBP shapes and their properties. This forms the motivation for our current work to investigate different shapes of BBPs.

Furthermore, stimuli-sensitive polymers can be grafted around a backbone to create stimuli-sensitive bottlebrush polymers (SS-BBPs)[19,29,30]. In these SS-BBPs, individual side chains can undergo conformational transitions in response to changes in environmental factors like solvent quality, pH, temperature, etc[31–33]. This, in turn, results in a change in the overall shape of SS-BBPs and thus, their properties. For example, thermosensitive polymers like poly(N-isopropylacrylamide) (PNIPAM) have been used to generate worm-like thermosensitive bottlebrush polymers (TS-BBPs)[34,35]. The choice of PNIPAM-like thermosensitive side chains can allow the study of shape change in a given BBP only by changing the surrounding temperature, thus adding only one controlled variable responsible for possible shape change. Specifically, the PNIPAM side chains undergo a coil-to-globule transition in water, when the system temperature is raised above its lower critical solution temperature (LCST = ~305 K), owing to dehydration of PNIPAM chains[36,37]. They have, as a result, been shown to exhibit overall conformational changes depending on their temperature as well as structural parameters including PNIPAM grafting density and side chain length, with potential applications in encapsulation and delivery of substances[27,38]. It is thus intriguing to consider, how TS-BBPs of different shapes, generated by arranging PNIPAM side chains of varying lengths along the backbone, would behave under similar conditions.

Over the years, while experimental research on BBPs has advanced dramatically, the synthesis and characterization of BBPs remains challenging[39]. Furthermore, the inherent complexity of TS-BBPs and limitations in experimentally achievable resolutions,

---

[1]Department of Chemical Engineering, Virginia Tech, Blacksburg, VA 24061, USA. ✉email: sanketad@vt.edu

make it difficult to understand the precise relationships between microscopic parameters, such as the grafting density, individual side chain lengths, structure of the polymer-solvent interface, etc., and the overall TS-BBP conformations[40,41]. Thus, computational techniques like coarse-grained (CG) molecular dynamics (MD) simulations have flourished in studying these systems at the molecular level[42,43]. Furthermore, given the enormous amount of data generated through simulations, the use of innovative methodologies, including machine learning (ML)[44–49], and deep learning (DL)[50–52], for analyzing these complex soft-material architectures including TS-BBPs has further provided insights into their conformations that were previously unknown[29].

In this *article*, we employ CG MD simulations to study model systems of TS-BBPs resembling six nature-inspired shapes by varying PNIPAM side chain length gradually or abruptly. PNIPAM side chains between 6-mers and 30-mers were attached to a hydrophobic backbone as its persistence length to show an LCST behavior is 10-mer[36]. Gradually changing structures consisted of the (i) cone, (ii) hourglass, and (iii) diamond-like BBPs while the abruptly changing structures consisted of the (i) cake, (ii) dumbbell, and (iii) plus-like BBPs (See Supplementary Fig. 1 in the Supplementary Information (SI)). It should be noted that three of these structures- cone, hourglass, and diamond, consisted of equal numbers of CG beads, and thus equal theoretic molecular weights (MW), with the only difference being the arrangement of side chains around the backbone. The six shapes were particularly chosen to identify if analogous TS-BBPs such as the cone-like and cake-like BBPs (narrow at one end and broad at the other) showed similar thermosensitive behavior[19,20]. More details on structure generation are provided in Supplementary Note 1 of the SI. These BBPs were represented with CG models of PNIPAM and hydrocarbon for the side chains and backbone, respectively, and solvated in a 1-site CG water model, all developed previously in our group[53–56]. Detailed information about these models is presented in Supplementary Note 2 of the SI.

CG MD simulations of solvated TS-BBPs were performed below (300 K) and above (320 K) the LCST of PNIPAM for 500 ns, which was deemed adequate for equilibration based on previous studies on TS-BBPs[29]. For direct comparison with a known shape, a worm-like (cylindrical) BBP was also studied, which consisted of all PNIPAM 30-mer side chains. An account of the system and computational details is presented in Supplementary Note 3 of the SI. Through the conventional analyses of free volume, contacts, asphericity, small-angle neutron scattering (SANS), and solvent accessible surface area (SASA) we found that the TS-BBPs with different shapes exhibited different conformations below and above LCST. We also observed, through an analysis of radii of gyration (R_g) and radial distribution function (RDF), that this behavior was strongly dependent on the length and immediate environment of the grafted PNIPAM side chains. We have additionally developed, a three-dimensional (3D) convolutional neural network (CNN)-based DL model to accurately classify and quantitatively analyze the similarities and differences in these TS-BBPs of different shapes, at the two different temperatures. Our DL approach is very general and can also be similarly used for the processing of MD trajectories of several other complex hybrid architectures of soft materials, demonstrating an application of deep learning in mesoscale modeling.

RESULTS AND DISCUSSION
Compactness of BBPs
Visual inspection of the simulation trajectories generally showed that TS-BBPs at 320 K exhibited more compact and bent structures than at 300 K, when simulated for 500 ns, as presented in Fig. 1. The equilibration of BBPs in the given time frame was verified by analyzing the Root Mean Square Displacement (RMSD) of the

overall BBPs, as well as some randomly selected 30-mers in the structures which is shown for three representative structures in Supplementary Fig. 3. To quantify the compactness of our structures, the free volumes of BBPs and the number of PNIPAM-PNIPAM contacts per monomer in the entire BBP were calculated. The free volume for these BBPs was calculated by building a convex hull around the entire BBP structure (Supplementary Fig. 4) to estimate the overall volumes occupied by these structures and then, subtracting their theoretical volumes[57–59]. Note, similar methodology has been used to calculate free volume in several other polymeric systems[60,61]. Our results showed that the free volumes of all BBPs, irrespective of the temperature and shape, reduced as the simulations progressed, but the observed reduction in values was generally greater at 320 K––indicating that systems at 320 K resulted in more compact structures. It was also observed that the cake, hourglass and diamond-like shapes displayed varying free volume values, irrespective of their same theoretical MW. This suggests that the arrangement of chains with the same number of monomers along the backbone is critical for its free volume. A thorough report of the free volume analysis is presented in Supplementary Note 4 of the SI. The progression of PNIPAM-PNIPAM contacts per PNIPAM monomer throughout the simulation run was calculated by considering bead pairs, having a Euclidean distance lower than the vdW cutoff (12 Å) used in CG MD simulations[62]. These bead pairs, due to their shorter distance, were considered to directly interact and thus defined as being in contact. Our idea behind using this metric was different from the conventional sense of the analysis[63–65], and we only wished to analyze if structures at 320 K were more aggregated and compact as compared to those at 300 K. The logic was that more compact structures would have a higher number of overall contacts. Here, most BBPs at 320 K were observed to have more PNIPAM-PNIPAM contacts as compared to 300 K, indicating that the PNIPAM side chains were closer to each other, and interacted strongly at 320 K––in perfect agreement with the free volume data. The evolution of contacts, throughout the simulation run, for all the studied shapes, as well as additional observations can be found in Supplementary Fig. 5 and Supplementary Note 5 of the SI.

Conventional analysis of shapes of BBPs
To investigate the effect of increased PNIPAM-PNIPAM contacts and the compact nature of PNIPAM side chains on the overall BBP shape, the asphericity for all the BBPs was calculated based on the approach developed by Yethiraj[66], where asphericity values, ranging between 0 and 1, indicate more spherical and more cylindrical conformations at the two ends of the scale, respectively. The asphericity values (shown in Supplementary Table 3) suggested that most of the BBPs were relatively more cylindrical at 320 K compared to 300 K except for worm and dumbbell-like BBPs. This is interesting because both these structures had several chains of 30-mers at both ends of the backbone unlike the rest of the BBPs. We also found that asphericity values were generally higher for abruptly changing BBPs compared to their gradually changing analogue, further showing the importance of architecture of BBPs. Detailed account of this analysis for individual BBPs can be found in Supplementary Note 6 of the SI. These structural analyses provided a good representation of the general BBP conformation and its shape below and above LCST, which were used to compare the different analogous BBP structures with each other as well as worm-like BBPs (Fig. 2).

Comparing the cake and cone-like, and dumbbell and hourglass-like BBPs based on the 3D plots shown in Fig. 2, it was observed that data points at 300 K formed unique individual clusters, thus indicating little similarities between the shapes being compared. Even the cluster of data points for the worm-like BBP was found to be discrete and did not overlap with the other clusters of the four aforementioned shapes at 300 K. At 320 K,
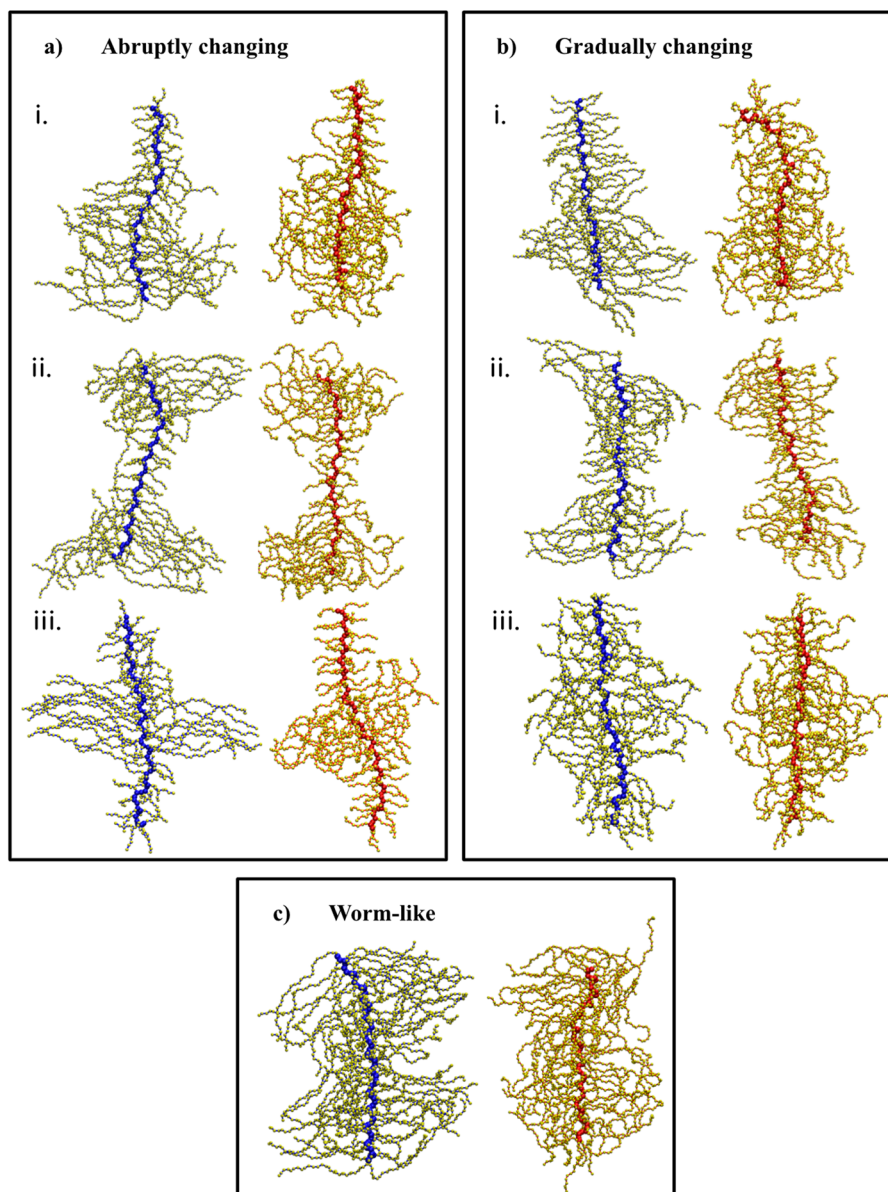
Fig. 1 Final snapshots for the BBPs after 500 ns simulation run. a Abruptly changing, **b** Gradually changing and **c** Worm-like BBPs studied after 500 ns simulation run. Abruptly changing BBPs include: **a**-i Cake, **a**-ii Dumbbell and **a**-iii Plus-like BBPs. Gradually changing BBPs include: **b**-i Cone, **b**-ii Hourglass and **b**-iii Diamond-like BBPs. Simulations at 300 K are presented with a blue BBP backbone and those at 320 K are presented with a red BBP backbone. PNIPAM side chains are represented in yellow with CG beads AMP and ISP hidden to provide a relatively simpler representation of the structures.

however, some overlaps were observed for these BBP pairs and the worm-like BBP. This can be attributed to the LCST behavior of the thermosensitive PNIPAM side chains. Interestingly, for the plus and diamond-like BBPs, overlap was observed at 300 K but not at 320 K. This indicates that the position of longer vs shorter chains on the BBP backbone may be important in determining conformations of side chains as these are the only two BBPs for which chain length increases as we get closer to the center of the backbone. The Pearson's correlation coefficients (PCC), used to measure the monotonic association between two variables, were calculated for the three aforementioned properties as shown in Supplementary Table 4[67]. It was observed that the free volume and contacts were negatively correlated with the correlation increasing at 320 K for most structures. This indicated that as the free volume of BBPs decreased, the number of PNIPAM-PNIPAM contacts increased and vice-versa. The correlation between free

volume and asphericity and between asphericity and PNIPAM contacts did not seem to follow any trend, indicating that the asphericity values were largely independent of the other two properties. A negative correlation ($\sim$−0.5) was however observed between the free volume and asphericity for certain cases involving the cake (320 K), hourglass (300 K) and worm-like BBPs (both temperatures). It was also observed that correlation between all three properties was almost the same at both temperatures for the worm-like BBP, indicating that the worm-like structure showed similar trends in behavior irrespective of temperature. These 3D plots were also used for comparing shapes at 300 K and 320 K as shown in Supplementary Fig. 6.

Next, small angle neutron scattering (SANS) was analyzed for the different BBPs at 300 K and 320 K to further compare their shapes and sizes. The SANS spectra was obtained for the last 50 ns of the simulation using the CRYSON program and is presented in
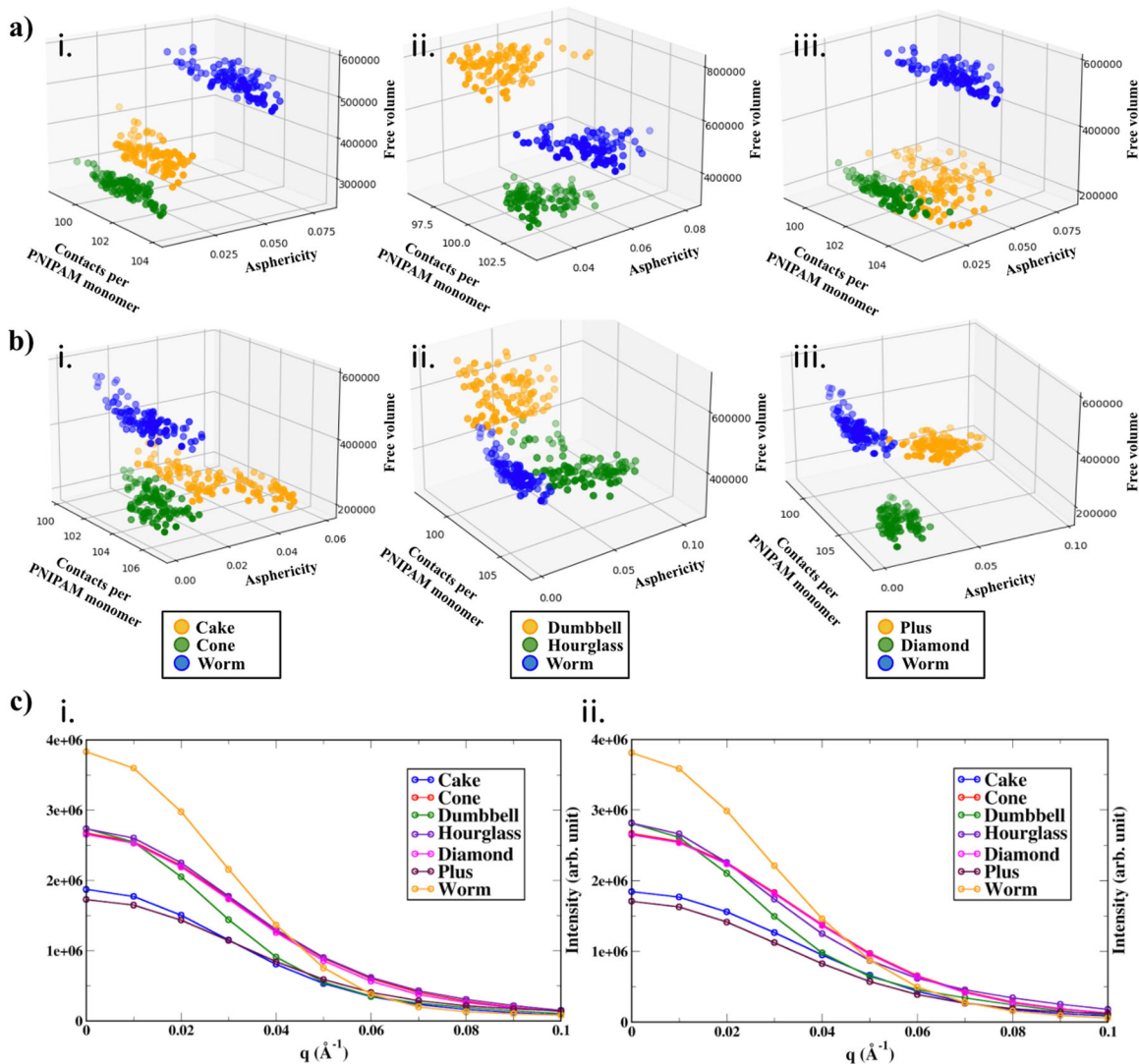
Fig. 2 Conventional analysis of compactness and shapes of BBPs. Three-dimensional (3D) plots plotted using the BBP free volume, contacts per PNIPAM monomer and asphericity data comparing analogous BBP pairs with worm-like BBPs at **a** 300 K, and **b** 320 K. The analogous pairs include (i) Cake-Cone, (ii) Dumbbell-Hourglass, and (iii) Plus-Diamond shaped BBPs. **c** Small angle neutron scattering (SANS) spectra plotted for all BBPs at (i) 300 K and (ii) 320 K. All data generated for the last 50 ns of the simulation run. Figure legends present alongside the figures.

Fig. 2c[68,69]. Two regions were identified using the obtained SANS spectra where the low q region (<0.01 Å^{-1}) gives information on the overall size and conformation of the BBP, and the mid q region (0.02–0.06Å^{-1} gives information about the cross-sectional size and stiffness[70]. Clear plateaus were observed for all BBPs in the low q regions indicating their overall spherical conformations in agreement with our asphericity data. The intensity of the SANS spectra in the low q region was found to be the highest for the worm-like BBP indicating the largest and most spherical conformation of this type of BBP[38]. The sizes and conformations of the cone, dumbbell, diamond and hourglass-like BBPs were all found to be similar at 300 K as compared to 320 K, where similarity was observed in the cone-diamond and dumbbell-hourglass pairs. The cake and plus-like shapes showed similar overall sizes but not conformations as were determined from Kratky plots for these structures[38].

The sphericity of the BBPs was verified through the Kratky plots generated using the SANS spectra as shown in Supplementary Fig. 7[71,72]. Kratky plots, have been used to identify the flexibility and/or degree of unfolding in protein samples, where compact, globular proteins show a bell-shaped (Gaussian) peak, while more

cylindrical structures show a plateau in the high q region[71]. It was observed that while all BBPs exhibited a bell curve indicating somewhat spherical conformations, the worm, cone, diamond and hourglass-like shapes displayed a more distinct bell curve as compared to the other structures indicating more spherical nature[73]. The Kratky plot for the worm-like BBP was found to be the most narrow and well defined, in agreement with previous results of highest sphericity at 320 K[73]. The hourglass-like BBP displayed a more spherical conformation at 300 K due to the bending of its backbone as compared to at 320 K. Not much difference was observed in the Kratky plots for the cone and diamond-like BBPs at the two temperatures, which is also consistent with the asphericity data we previously discussed. The mid q regions showed that although the worm-like BBP was structurally very different from the others, its spectra intersected that of the cone, diamond, and hourglass-like BBPs indicating that these structures had qualitatively similar cross-sectional areas and stiffnesses. Similarly, it was observed that the cake and plus-like BBPs exhibited similar areas and stiffnesses at 300 K but not at 320 K[70]. The dumbbell-like BBP was found to be very different in

Table 1. Solvent accessible surface area (SASA) values for the overall BBPs.

Shape	Temp (K)	Initial SASA ($Å^2$)	Final SASA ($Å^2$)
Cake	300	146795.63 ± 605.35	143100.22 ± 404.58
Cake	320	146240.60 ± 425.48	142797.75 ± 722.35
Cone	300	151578.27 ± 375.70	147377.48 ± 468.67
Cone	320	150685.88 ± 643.07	146704.68 ± 436.98
Dumbbell	300	183818.45 ± 703.57	178930.51 ± 520.97
Dumbbell	320	182578.07 ± 699.30	178882.42 ± 528.01
Hourglass	300	151224.35 ± 798.36	146962.68 ± 389.91
Hourglass	320	150242.02 ± 436.46	147483.73 ± 599.41
Plus	300	120116.86 ± 448.77	116454.66 ± 513.08
Plus	320	119733.15 ± 539.27	117430.07 ± 332.29
Diamond	300	150123.69 ± 585.77	146890.61 ± 488.77
Diamond	320	150152.68 ± 599.42	145848.39 ± 438.64
Worm	300	246661.66 ± 860.96	238580.06 ± 598.68
Worm	320	245948.40 ± 966.13	238101.54 ± 705.50

terms of these properties from all other shapes studied. To further obtain thermodynamic insights into the behavior of different BBP shapes, configurational entropy for the backbones of randomly selected 30-mer and 7-mer side chains from abruptly changing BBPs was plotted throughout the simulation run using the Quasi-harmonic analysis[74–79], as shown in Supplementary Fig. 8. It was found that the configurational entropy of the PNIPAM backbone increased as the simulation progressed and attained a near stable state in most cases indicating equilibration of the chains[80,81]. It was also observed that 7-mer side chains exhibited similar configurational entropy irrespective of the temperature, whereas 30-mers exhibited a large difference in entropy values at the two temperatures. This indicates that shorter PNIPAM chains do not show much difference in entropic behavior below and above LCST, explaining their inability to undergo coil-to-globule transition.

Solvation dynamics of BBPs

To understand the effect of shape and LCST on overall solvation dynamics, we also calculated the solvent accessible surface areas (SASA) for all the BBPs studied, which are presented in Table 1. Values were calculated for the initial and final stages of the production run, as for several other analyses. It was observed that the SASA values for all structures reduced as the BBPs became more compact during simulation at both 300 K and 320 K. For the plus-like structure, it was again observed that the final SASA values at 320 K were greater than at 300 K, although the initial values displayed a reverse trend. This is consistent with the observations previously made during the analysis of contacts and free volume. Only considering the gradually changing shapes, which had the same molecular weights, it was observed that while the initial SASA values for all these structures was comparable, diamond-like BBP exhibited lower final values as compared to the other two structures. This could be attributed to all longer side chains being located at the center of the structure, resulting in relatively more compact structures owing to higher PNIPAM-PNIPAM interactions. This further shows the importance of shape of BBPs resulting from the arrangement of side chains with different numbers of monomers on the backbone. The worm-like BBPs exhibited the highest SASA values, more than double than those of the plus-like structure. This can be attributed to its much larger size as shown by the SANS analysis. The SASA for the dumbbell-like BBP was lower than the worm-like BBP, but greater than all remaining structures. Thus, a pattern was observed, which was absent in the previous analyses, where the SASA was directly proportional to the theoretical MWs of the BBPs. The reduction in SASA was also found to be greater for systems like the cake, cone and diamond-like BBPs at 320 K, indicating that BBPs with shorter chains at one or both ends of the backbone formed more compact structures compared to those with the short chains in the middle.

Conformations of the BBP backbone

Apart from their compactness, several BBPs exhibited bending in their backbones at 320 K, resulting in overall bent conformations. This bending of the backbones was confirmed by calculating the backbone R_g and end-to-end distances for different BBPs as shown in Fig. 3a. It was observed that the backbone R_g for the cake, cone, diamond, and worm-like BBPs was significantly lower at 320 K than 300 K, indicating that these structures were more bent at 320 K. Also, the backbone R_g for the hourglass and plus-like BBPs, was found to be lower at 300 K similarly indicating more bent structures at 300 K. A coherent trend was observed for the backbone end-to-end distances. The end-to-end distances for the initial and final stages of the simulation for these structures are also provided in Supplementary Table 5. While these values were similar during the initial 50 ns at 300 K and 320 K, considerable differences cropped up as the simulation progressed. The backbone end-to-end distances for final 300 K structures were observed to either be very similar to, or greater than the initial values, indicating that the backbones at 300 K were stretched as compared to at 320 K. The origin of this bending of the backbone was analyzed by calculating the number of contacts between side chains of different lengths as shown in Supplementary Note 7 of the SI. In general, this bending was due to the interactions between smaller (<10-mer) and longer (>10-mer) PNIPAM chains located on the backbone of the BBP. To study the effect of bending on the structure of solvent around the backbone, comparison of the RDFs between water and backbone beads for the abruptly changing shapes was also carried out. This indicated that the backbone of the worm-like BBP was comparatively most hydrated at 300 K followed by the plus, cake, and dumbbell-like structures. This mostly remained the same at 320 K except for the cake-like BBP, which exhibited dehydration due to bending as shown in Supplementary Fig. 13. The evolution of overall R_g of all BBPs is presented in Supplementary Fig. 14 and it shows that the overall R_g of the cake, cone, and diamond-like BBPs significantly reduced at 320 K as compared to 300 K due to the bending of backbones in these structures. It is thus clear that the shapes of TS-BBPs greatly influence not only their overall conformations but also the bending of their backbone.

Conformations of individual side chains

To study the structure of individual side chains of PNIPAM in BBPs with different shapes, we calculated their R_g at both temperatures for the last 50 ns of the total 500 ns run. As a reference, the mean chain-wise R_g values at both temperatures for the cake-like BBP are presented in Fig. 3b, along with the normalized distribution of R_g for each side chain length in Supplementary Fig. 15. It was observed that the R_g of most 30-mer chains, at 320 K, was lower than their respective values at 300 K, indicating the presence of globule-like conformations at 320 K[53]. A coil-to-globule transition was also observed for PNIPAM 18-mers. However, the difference in mean R_g values for 18-mers at the two temperatures was lower than observed for 30-mers[53]. It was also observed that for 6-mer side chains, the difference in R_g values was negligible, indicating that short side chains did not exhibit LCST behavior[29,53]. Similar results for PNIPAM chains below 10-mers have been reported in previous studies[36,82,83]. For gradually changing cone, hourglass and diamond-like BBPs, given the wide range of side chain lengths, R_g distribution was studied in smaller groups of chain lengths. Four groups of side chains were created (7-mer to 11-mer,
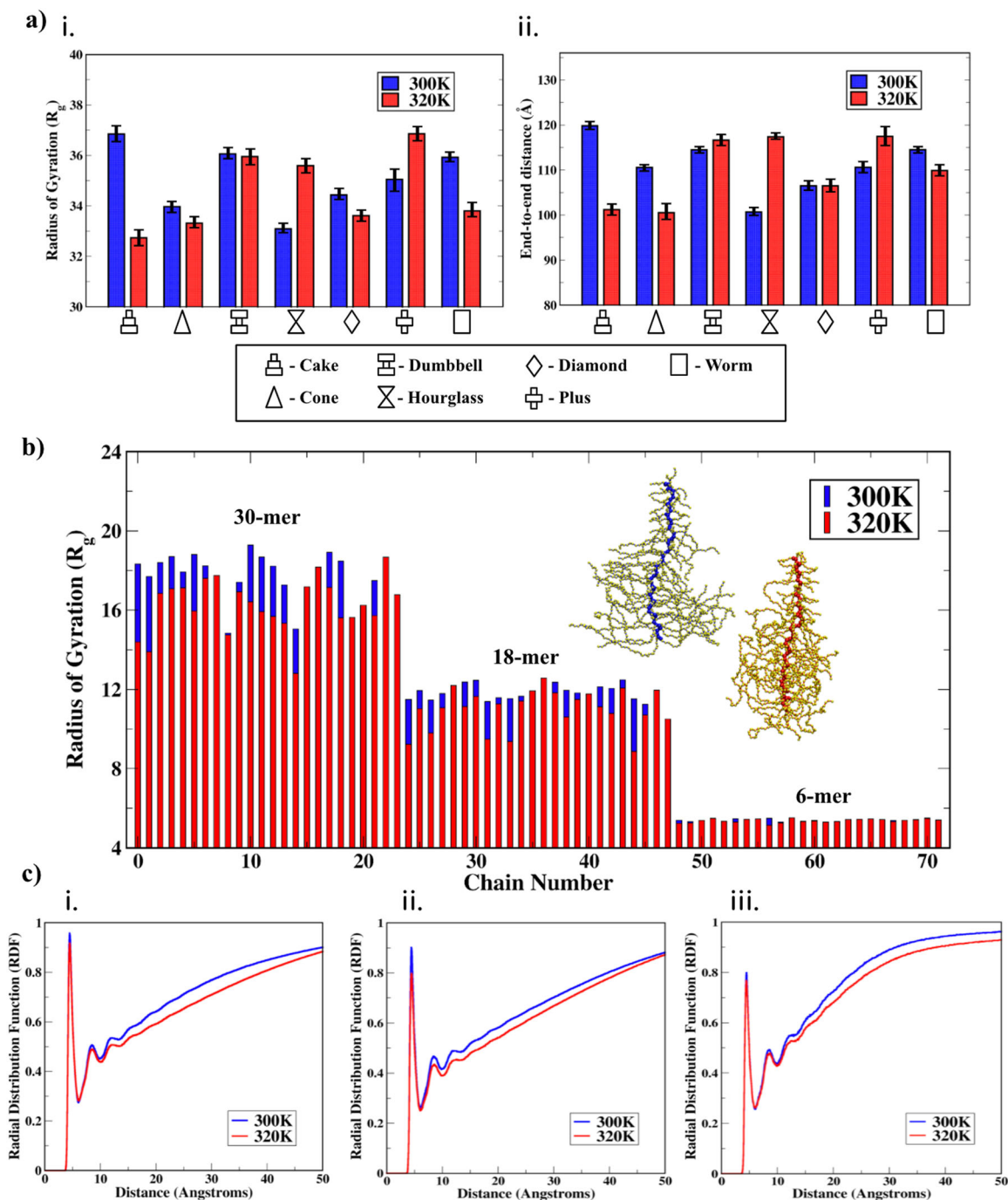
Fig. 3 Configurational and solvation analysis of individual polymer chains. a Mean values of (i) backbone R_g and (ii) backbone end-to-end distance for all BBPs. Legend available below the plots. **b** Mean R_g value for each PNIPAM side chain of the cake-like BBP calculated for the last 150 ns of the simulation run. Side chain lengths are provided in the plot along with the final BBP snapshots below and above LCST. The snapshots follow the scheme mentioned in Fig. 1. **c** Radial Distribution Function (RDF) profiles between water and (i) 30-mer, (ii) 18-mer and (iii) 6-mer long side chains. 300 K data is presented in blue, and 320 K data is presented in red.

12-mer to 17-mer, 18-mer to 24-mer and 25-mer to 30-mer) and normalized distribution of R_g was plotted for the last 150 ns. It was again observed that all the side chain groups except 7-mer to 11-mer exhibited partial globule-like conformations at 320 K with the extent of collapse increasing with increase in chain length. Analysis of the worm-like BBP also showed partial globule-like conformations of several side chains at 320 K. These data can be found in Supplementary Figs. 16–21 and Supplementary Tables 6 and 7.

The availability of solvent around PNIPAM chains plays an important role in their LCST behaviour and is often believed to initiate the process[36]. Thus, we next studied the structure of solvent at the solvent-polymer interface by plotting the RDF between water and PNIPAM side chains. The peaks in the RDF indicate short-range structure *via* positional correlation[84]. While the side chains exhibited dehydration at both temperatures as the simulations progressed, those at 320 K were generally more dehydrated than their 300 K counterparts as seen through decrease in the peak heights. The RDF plots for the representative

cake-like BBP are presented in Fig. 3c. These plots show clear dehydration of polymer chains at 320 K, irrespective of their lengths, with the extent of dehydration being the greatest for 18-mer side chains. It should be noted that the height for the first peak in RDF decreases with decrease in the PNIPAM chain length, indicating that the structure of water might be different near these chains. The evolution of the RDF between the PNIPAM backbone and water, throughout the simulation run was also plotted for the cake-like BBP as shown in Supplementary Fig. 22 and Supplementary Table 8. This plot shows that the structure of solvent significantly changed for the first 200–250 ns of the run and eventually attained a more stable value. PNIPAM-water RDFs and an in-depth analysis for the remaining BBPs can be found in Supplementary Note 8 of the SI, which showcase similar behavior to the cake-like BBP. To further emphasize the extent of dehydration in these systems, data for the mean number of water molecules per PNIPAM monomer, in the first, second and third hydration shells are provided in Supplementary Table 9. These data also show that dehydration was greater at 320 K in all cases except one––the plus-like BBP. The effect of the side chain environment on the LCST behavior of 30-mers in the abruptly changing shapes was also analyzed, which is presented in Supplementary Note 9 of the SI. The conformations of PNIPAM 30-mers in three distinct environments (Supplementary Fig. 28)–– (i) edge (chains at two ends of the backbone), (ii) center (chains at the center of 30-mer blocks) and (iii) interface (chains at the ends of 30-mer blocks, adjacent to a different side chain block) were analyzed, which showed that PNIPAM chains displayed different conformations in all three environments depending on the shape of the BBP studied. PNIPAM side chains at the edge and the center showed appreciable reduction in R_g at 320 K as compared to 300 K. At the interface, 30-mers were observed to fall onto the adjacent shorter side chains, thus showing little change in R_g values at the two temperatures. RDFs between PNIPAM 30-mers and water also showed that chains at the edge were most hydrated, followed by chains at the interface and the center at 320 K. These data indicate that apart from the PNIPAM chain length, their immediate environments also play a big role in determining the conformations of individual side chains in the BBPs.

Need for convolutional neural network (CNN) model
Following these conventional analyses, we developed a 3D CNN-based classifier to further analyze these complex BBP structures for much more minute similarities and differences[85–89]. CNNs, having previously demonstrated great success in the fields of image processing, computer vision, etc[90–96], are easy to understand, fast to implement and can detect important features with little human supervision[97]. CNNs have also recently found way into diverse areas such as material science and biochemistry[98–101], where they have been utilized for defect classification, microstructure identification and reconstruction, heterogeneous material homogenization,[102–105] as well as to identify protein functions, RNA splicing sites and predict subcellular protein localization[106–109]. CNNs have also been shown to perform better than rule-based classifiers for flexible objects due to their ability to analyze shapes irrespective of their orientation[110–112]. The use of DL to analyze our systems was thus warranted for two main reasons––(i) Studying the shapes of dynamic soft-material architectures in the presence of solvents using existing experimental methods or traditional geometry-based computational analysis is very challenging given their flexible nature. Use of DL methods such as CNNs have been shown to analyze flexible objects very well due to their ability to identify shapes irrespective of their orientation––something essential for soft-matter architectures[113,114]. This method, which differentiates/classifies based on the microscopic features of these systems, was thus deemed as an appropriate and powerful tool to solve this problem. (ii) Features obtained from CNN based models are frequently used for image recognition and segmentation, image and video classification, natural language processing etc. However, to the best of our knowledge, this approach has not been used to handle problems concerning nanoscale soft material architectures in the literature. Thus, this provided us the opportunity to demonstrate the applicability of such DL methods in soft-material research.

CNN model development and data processing
The first and most crucial step here was generating easily processable data points out of the available CG MD trajectories without losing out on important structural information. This was achieved by voxelization of the entire BBP structure into a binary 3D grid of $50 \times 50 \times 50$ voxels using the cartesian coordinates of each bead and their corresponding bead radii (see Fig. 4a and Supplementary Fig. 29). Specifically, from the later stage of the production run (last 150 ns), 250 randomly chosen frames were used as individual data points for each of the seven BBP shapes. Models were also trained using different number of data points per BBP shape, and it was observed that with a higher number of frames per shape, the model was trained in fewer epochs (Supplementary Fig. 30) but required more time and memory. This number of 250 frames was chosen as an optimum between the model accuracy, training time and memory usage of our CNN model. The rationale behind using cartesian coordinates was that they allowed us to capture the spatial information of the BBPs, including their overall shape and the minute details regarding their internal structure. The size of 50 voxels (~4 Å) was selected to cover most microstructural details of the BBP structures, with one CG bead generally occupying one unit in the 3D grid. We also attempted to use 75 voxels per side during the grid generation which provided an even more detailed representation of the structure (Supplementary Fig. 31). However, this too required much greater amount of time and memory. Trajectories for 300 K and 320 K were analyzed separately, which provided us with equally distributed 1750 data points (250 frames × 7 shapes) at each temperature.

This dataset of 1750 points was labelled and randomly shuffled before segregation into training, validation, and testing data with a ratio of 70:15:15. Once the data was ready, the next step involved developing a 3D CNN architecture to process the data and classify the input into one of the seven shapes. This was developed using the Keras library with four convolutional layers, the Rectified Linear Unit (ReLU) activation function and max pooling in each layer[115,116]. A kernel size of $3 \times 3 \times 3$, stride length of 1 and no-padding (i.e. no additional rows or columns were added to the existing data while carrying out convolutions) were configured on the convolutional layers. A rigorous hyperparameter tuning process was carried out to identify the best parameters for our CNN model. The number of convolutional layers, batch size, and kernel dimensions were systematically varied to study the performance of the CNN model. While all models with one, two, three and four convolutional layers were capable of accurately classifying the shapes, the features each model considered to do so were not the same. The number of convolutional layers was chosen as four to extract a rich representation of the microscopic features of each shape during convolution. Kernels that operate directly on the input data usually only learn low-level features, such as lines or planes. However, as the depth of the network is increased, the features are abstracted at a higher order. Thus, features become more complex as more convolutional layers are added[117,118]. Batch sizes including 32, 64, 96 and 128 were studied and comparable results were obtained for all the cases. Kernel dimensions for the first convolutional layer were also varied from 3 to 7, which showed comparable results. It was generally observed that as the shapes
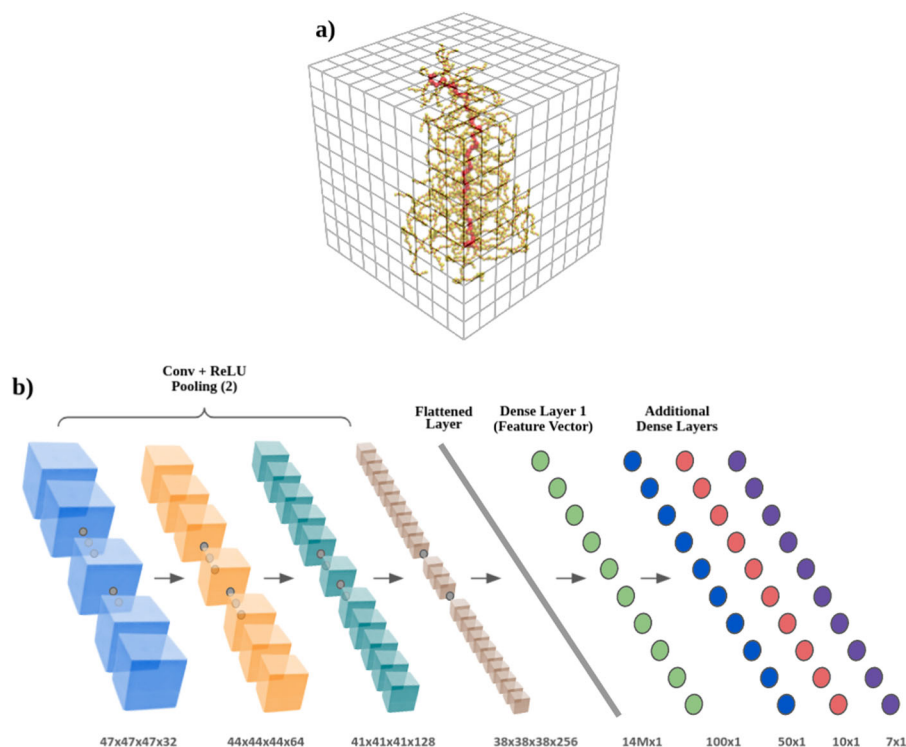
Fig. 4 CNN model development. a Schematic of the grid generation. **b** Architecture diagram for the developed CNN model.

used in this manuscript were considerably different from each other and the voxelization process utilized adequately captured the differences between these shapes, the CNN model was able to accurately classify the shapes with a wide variety of hyperparameter sets without overfitting. Results from this hyperparameter tuning process can be found in Supplementary Figs. 32–34. A dropout of 0.5 was applied after the third convolutional layer to regularize the model and avoid overfitting. Using L1 and L2 kernel regularization methods was also explored but they were found to reduce the validation accuracy of the models (Supplementary Fig. 35). This was followed by four dense layers, where the output of the first dense layer represented 100 features of the input. Softmax function was applied on the output of the last layer to obtain the probability scores for each class[119]. The Adam optimizer along with a categorical cross-entropy loss function was used[120]. A detailed architecture diagram of the 3D CNN model is provided in Fig. 4b.

The model was trained for 30 epochs to prevent the model from overfitting the training data and to simultaneously reach the optimal accuracy value for the validation dataset (Supplementary Fig. 36). The training and validation were carried out for data points at 300 K and 320 K separately to train different models at different temperatures. Visualization of a representative tensor slice after different convolutional layers is also presented in Supplementary Fig. 37. The testing dataset, which was independent from the training and validation datasets was then used to classify existing BBPs into their respective shapes. Successful classification of all shapes by the CNN indicated that the model was able to convert our one-hot encoded grid of input data points into translational equivariant responses known as feature maps[121]. These feature maps, in our case are feature vectors consisting of 100 numbers that cannot be directly interpreted into particular structural features of the BBP, but are formed entirely as a result of the same and are capable of classifying the structures based on their shape and microstructure. An independent feature vector is also generated for each tested data point and the structure is classified based on similarities between this feature vector with

other feature vectors for the different classes in question. This approach of converting known structures into feature vectors forms the basis of this paper as well as several CNN-based approaches used for image recognition and segmentation, image and video classification, natural language processing etc[122–125]. These 100 features, which reflect microstructural characteristics of the different BBP structures that may not be apparent to the more macroscopic analyses[105], were extracted from the first dense layer of our architecture for further analysis. This process, known as feature engineering––which is quite cumbersome using conventional methods, and usually requires extensive knowledge of the system to be studied––was thus carried out very easily using our deep learning model[106–108,126–128]. Finally, to identify and quantify the similarities between different BBP shapes, these different extracted feature vectors (FVs) were used to obtain important insights into the behavior of these BBPs below and above LCST. However, since the obtained features may vary with every model trained using different training data points, it was essential to first make these values statistically significant.

Comparison and quantification of BBP shapes using CNN

This was achieved by training 100 different CNN models, each with the same hyperparameters. Each model was fed a unique, randomly selected set of training data points from the complete data set. For each model trained, FVs for every structure from the testing data were extracted, with those for the same shape averaged to obtain one representative FV. The cosine similarity between these mean FVs for different shapes was calculated and stored as a similarity matrix of dimensions 7×7. These similarity matrices were used to obtain their mean and standard deviation matrices after every 10 models to observe their evolution as the number of CNN models increased. It was observed that after about 50 models, the overall values of these matrices converged to a near constant value, thus indicating that all possible sets of unique FVs for each shape had been accounted for in our final matrices (Supplementary Fig. 38). These matrices for the mean similarity values and standard deviations are presented in Fig. 5.
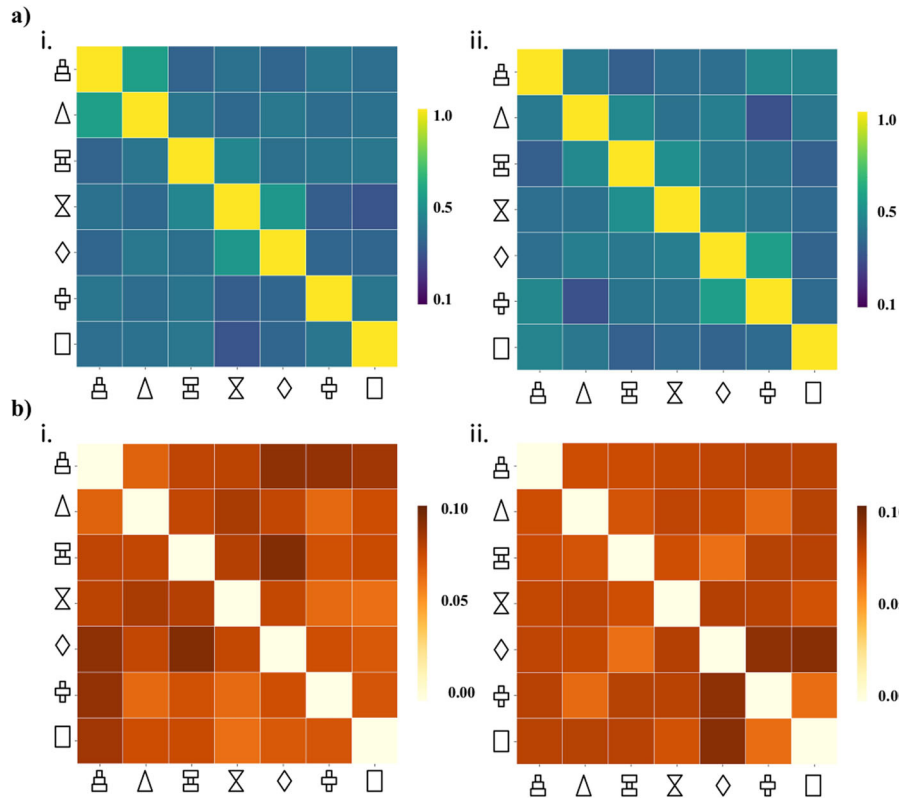
Fig. 5 Quantification of similarities and differences in BBP shapes. Shape-wise comparison of the (**a**) mean cosine similarity and (**b**) mean similarity standard deviation between the extracted feature vectors for each shape at (i) 300 K and (ii) 320 K. Legend for the similarity and standard deviation values is presented alongside each plot. Shape legend same as Fig. 3.

It was also observed that more BBP pairs showed higher similarity values at 320 K as compared to 300 K. This indicates that below LCST, when side chains are in a stretched, coil-like state, the BBP shapes are different from each other as identified by the 3D CNN model. Above LCST, however, these differences in shapes were much lower than those below LCST. This could be because at 320 K a coil-to-globule transition of PNIPAM side chains may have altered their shapes to contain more similar microscopic features than at 300 K.

The diagonal values approached the value of 1, indicating that the CNN model was able to accurately capture and classify BBPs based on their shapes. The traditional analysis (e.g., SANS, free volume, RDF, asphericity, and R_g) also suggests that the internal structure of the BBPs changed in response to their local environment and change in the temperature due to the LCST behavior of PNIPAM. Indeed, when CNN compared analogous BBPs with two different shapes it suggested that dumbbell-hourglass and plus-diamond pairs have higher similarities at 320 K and cake-cone pairs have higher similarities at 300 K. Thus, these results generally suggest that the combination of the one-hot encoded grid and BBP coordinates can capture two important aspects of the BBPs: (i) their shape and (ii) partially internal structure. These data, being different from our 3-feature comparison of the different shapes (Fig. 2) indicate that the CNN model was able to find granular similarities between shapes of BBPs, which were not captured by geometry based analysis such as R_g and SANS. The mean cosine similarity values also indicated that BBP pairs like cake-dumbbell, cake-hourglass, diamond-cone showed similar similarity values at 300 K and 320 K. However, some BBP pairs displayed a considerable change in their similarity values at the two temperatures. For instance, the similarity between cake-cone, plus-cone, and hourglass-diamond BBP pairs was higher at 300 K than 320 K, but for the pairs cone-

diamond, cake-plus, and cone-dumbbell, the opposite behavior was observed. These results can be attributed to the LCST behavior of these BBPs and the arrangement of side chains around the backbone. It was also observed that the similarity of the worm-like BBP with most other shapes was lower at 320 K than at 300 K. The standard deviation matrix showed that similarity values at 300 K generally displayed more fluctuations than at 320 K. This indicates that the magnitude of similarities between structures at 300 K is not deterministic enough as compared to at 320 K. A detailed account of the similarity values between different BBP pairs is further provided in Supplementary Note 10 of the SI.

In conclusion, the current study showed that the shapes of BBPs greatly influenced their overall conformations, dictating the compactness and bending in these structures. BBPs with the same theoretical molecular weights were shown to behave differently when the side chains were grafted onto the backbone in different patterns––with the structure with higher PNIPAM-PNIPAM interactions resulting in a smaller, more compact structure. The length of PNIPAM side chains and their immediate environments were also found to affect their coil-to-globule transition as well as their solvation dynamics, further influencing the overall structure of the BBP. Finally, a new CNN-based analysis was developed to quantify the degree of similarity between different BBP structures below and above the LCST of PNIPAM. This model, by itself, identified the best 100 features to define each BBP structure, without the need for human intervention. These features were obtained for 100 different CNN models to generate a statistically significant quantification of the similarity between different shapes. The CNN-based model was able to identify several structural and shape-based features of BBPs, not captured by conventional methods to provide unique, new insights into the similarities between these BBPs. In the near

future, we plan to use this CNN model to study architectures of Glycomaterials including their BBPs.

METHODS

Structure generation and computational details

Six shapes of BBPs were designed by gradually and abruptly changing the length of PNIPAM side chains between 7-mers to 30-mers using an in-house Python code. All generated bottlebrushes consisted of a hydrophobic backbone of 72 beads and PNIPAM side chains and a constant grafting density of 100% (one side chain on every backbone bead; 72 side chains). Individual BBPs (~4,250 beads) were solvated in a box of length 250 Å in x, y, and z directions with explicit 1-site CG water beads. The box dimensions were chosen to prevent interactions between the BBP and its periodic images while the number of water beads (~ 200,000) was calculated to account for the experimental density of water. Solvation was achieved by using the Packmol package developed by Martinez et al.[129] Coarse-grained (CG) molecular dynamics (MD) simulations were employed to study these systems at two temperatures––300 K (below LCST) and 320 K (above LCST) for 500 ns in NPT ensemble (1 bar pressure). Simulations were carried out using the NAMD package developed by the Theoretical and Computational Biophysics Group at the University of Illinois at Urbana-Champaign[130]. Simulation time-step of 10 fs was used to perform all the simulations. The cut-off for evaluating the nonbonded interactions was set to 12 Å. Temperature and pressure control during the CG MD simulations was achieved by Langevin thermostat and barostat, respectively. The first 50 ns of the simulation were set aside for equilibration of the system and 50–500 ns was considered as the production run. Detailed information can be found in Supplementary Notes 1, 2 and 3 of the SI.

DATA AVAILABILITY

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

CODE AVAILABILITY

The calculation and analysis codes used in this paper including the convolutional neural network (CNN) code are available from the corresponding author upon reasonable request.

REFERENCES

1. Bornstein, M. H. Colour-name versus shape-name learning in young children*. *J. Child Lang.* **12**, 387–393 (1985).
2. Biederman, I. Recognition-by-components: a theory of human image understanding. *Psychol. Rev.* **94**, 115–147 (1987).
3. Landau, B., Smith, L. B. & Jones, S. S. The importance of shape in early lexical learning. *Cogn. Dev.* **3**, 299–321 (1988).
4. Diez-Silva, M., Dao, M., Han, J., Lim, C.-T. & Suresh, S. Shape and biomechanical characteristics of human red blood cells in health and disease. *MRS Bull.* **35**, 382–388 (2010).
5. Zhao, G. et al. Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature* **497**, 643–646 (2013).
6. Ge, P. et al. Cryo-EM model of the bullet-shaped vesicular stomatitis virus. *Science* **327**, 689–693 (2010).
7. Ensign, L. M., Cone, R. & Hanes, J. Oral drug delivery with polymeric nanoparticles: the gastrointestinal mucus barriers. *Adv. Drug Deliv. Rev.* **64**, 557–570 (2012).
8. Banerjee, A., Qi, J., Gogoi, R., Wong, J. & Mitragotri, S. Role of nanoparticle size, shape and surface chemistry in oral drug delivery. *J. Control. Release* **238**, 176–185 (2016).
9. Nangia, S. & Sureshkumar, R. Effects of nanoparticle charge and shape anisotropy on translocation through cell membranes. *Langmuir* **28**, 17666–17671 (2012).
10. Wang, W., Gaus, K., Tilley, R. D. & Justin Gooding, J. The impact of nanoparticle shape on cellular internalisation and transport: what do the different analysis methods tell us? *Mater. Horiz.* **6**, 1538–1547 (2019).
11. Xie, X., Liao, J., Shao, X., Li, Q. & Lin, Y. The effect of shape on cellular uptake of gold nanoparticles in the forms of stars, rods, and triangles. *Sci. Rep.* **7**, 3827 (2017).
12. Vasudevan, S. & Fullerton-Shirey, S. K. Effect of nanoparticle shape on the electrical and thermal properties of solid polymer electrolytes. *J. Phys. Chem. C.* **123**, 10720–10726 (2019).
13. Han, L. et al. The effects of nanoparticle shape on electrical conductivity of Ag nanomaterials. *J. Mater. Sci.: Mater. Electron.* **25**, 3870–3877 (2014).
14. Elias, M. M. et al. Effect of nanoparticle shape on the heat transfer and thermodynamic performance of a shell and tube heat exchanger. *Int. Commun. Heat. Mass Transf.* **44**, 93–99 (2013).
15. Roni, M. R. H., Morshed, A. M., Tikadar, A., Paul, T. C., Khan, J. A. Nanoparticles shape effect on thermal conductivity of nanofluids: a molecular dynamics study. *Proc. ASME. Int. Mech. Eng. Congress. Expo.* **8**, 1–9 Salt Lake City, Utah, USA. November 11–14, 2019. V008T09A080. ASME.
16. Arno, M. C. et al. Exploiting the role of nanoparticle shape in enhancing hydrogel adhesive and mechanical properties. *Nat. Commun.* **11**, 1420 (2020).
17. Chow, T. S. The effect of particle shape on the mechanical properties of filled polymers. *J. Mater. Sci.* **15**, 1873–1888 (1980).
18. Darr, J. A., Zhang, J., Makwana, N. M. & Weng, X. Continuous hydrothermal synthesis of inorganic nanoparticles: applications and future directions. *Chem. Rev.* **117**, 11125–11238 (2017).
19. Radzinski, S. C. et al. Tapered bottlebrush polymers: cone-shaped nanostructures by sequential addition of macromonomers. *ACS Macro Lett.* **6**, 1175–1179 (2017).
20. Walsh, D. J., Dutta, S., Sing, C. E. & Guironnet, D. Engineering of molecular geometry in bottlebrush polymers. *Macromolecules* **52**, 4847–4857 (2019).
21. Shani, L. et al. DNA-assembled superconducting 3D nanoscale architectures. *Nat. Commun.* **11**, 5697 (2020).
22. Wang, Y. et al. A combined experimental and computational approach reveals how aromatic peptide amphiphiles self-assemble to form ion-conducting nanohelices. *Mater. Chem. Front.* **4**, 3022–3031 (2020).
23. Deshmukh, S. A., Solomon, L. A., Kamath, G., Fry, H. C. & Sankaranarayanan, S. K. R. S. Water ordering controls the dynamic equilibrium of micelle-fibre formation in self-assembly of peptide amphiphiles. *Nat. Commun.* **7**, 12367 (2016).
24. Li, Z. et al. Bottlebrush polymers: from controlled synthesis, self-assembly, properties to applications. *Prog. Polym. Sci.* **116**, 101387 (2021).
25. Foster, J. C., Varlas, S., Couturaud, B., Coe, Z. & O'Reilly, R. K. Getting into shape: reflections on a new generation of cylindrical nanostructures' self-assembly using polymer building blocks. *J. Am. Chem. Soc.* **141**, 2742–2753 (2019).
26. Chremos, A. & Douglas, J. F. A comparative study of thermodynamic, conformational, and structural properties of bottlebrush with star and ring polymer melts. *J. Chem. Phys.* **149**, 044904 (2018).
27. Zhao, B. Shape-changing bottlebrush polymers. *J. Phys. Chem. B* **125**, 6373–6389 (2021).
28. Kim, E. J. et al. Molecular weight dependent morphological transitions of bottlebrush block copolymer particles: experiments and simulations. *ACS Nano* **15**, 5513–5522 (2021).
29. Bejagam, K. K., Singh, S. K., Ahn, R. & Deshmukh, S. A. Unraveling the conformations of backbone and side chains in thermosensitive bottlebrush polymers. *Macromolecules* **52**, 9398–9408 (2019).
30. Walsh, D. J. & Guironnet, D. Macromolecules with programmable shape, size, and chemistry. *Proc. Natl Acad. Sci. U.S.A.* **116**, 1538–1542 (2019).
31. Theodorakis, P. E. & Fytas, N. G. Molecular dynamics simulations of bottle-brush polymers with a flexible backbone under theta and good solvent conditions. *Am. J. Condens. Matter Phys.* **2**, 101–108 (2012).
32. Azzaroni, O., Brown, A. A. & Huck, W. T. S. UCST wetting transitions of polyzwitterionic brushes driven by self-association. *Angew. Chem. Int. Ed.* **118**, 1802–1806 (2006).
33. Xu, Y. et al. pH and salt responsive poly(N,N-dimethylaminoethyl methacrylate) cylindrical brushes and their quaternized derivatives. *Polym. J.* **49**, 3957–3964 (2008).
34. Henn, D. M., Holmes, J. A., Kent, E. W. & Zhao, B. Worm-to-sphere shape transition of thermoresponsive linear molecular bottlebrushes in moderately concentrated aqueous solution. *J. Phys. Chem. B* **122**, 7015–7025 (2018).
35. Buckwalter, J. A., Kuettner, K. E. & Thonar, E. J. Age-related changes in articular cartilage proteoglycans: electron microscopic studies. *J. Orthop. Res.* **3**, 251–257 (1985).
36. Deshmukh, S. A., Sankaranarayanan, S. K. R. S., Suthar, K. & Mancini, D. C. Role of solvation dynamics and local ordering of water in inducing conformational transitions in poly(N-isopropylacrylamide) oligomers through the LCST. *J. Phys. Chem. B* **116**, 2651–2663 (2012).
37. Podewitz, M. et al. Coil-Globule Transition Thermodynamics of Poly(-isopropylacrylamide). *J. Phys. Chem. B* **123**, 8838–8847 (2019).

38. Li, X. et al. Thermoresponsive PNIPAAM bottlebrush polymers with tailored side-chain length and end-group structure. *Soft Matter* **10**, 2008–2015 (2014).

39. Radzinski, S. C., Foster, J. C., Lewis, S. E., French, E. V. & Matson, J. B. Factors affecting bottlebrush polymer synthesis by the transfer-to method using reversible addition–fragmentation chain transfer (RAFT) polymerization. *Polym. Chem.* **8**, 1636–1643 (2017).

40. Dalsin, S. J. et al. Bottlebrush block polymers: quantitative theory and experiments. *ACS Nano* **9**, 12233–12245 (2015).

41. Chatterjee, D. & Vilgis, T. A. Scaling laws of bottle-brush polymers in dilute solutions. *Macromol. Theory Simul.* **25**, 518–523 (2016).

42. Joshi, S. Y. & Deshmukh, S. A. A review of advancements in coarse-grained molecular dynamics simulations. *Mol. Sim.* **47**, 1–18 (2020).

43. Mohammadi, E., Joshi, S. Y. & Deshmukh, S. A. A review of computational studies of bottlebrush polymers. *Comput. Mater. Sci.* **199**, 110720 (2021).

44. Decherchi, S., Berteotti, A., Bottegoni, G., Rocchia, W. & Cavalli, A. The ligand binding mechanism to purine nucleoside phosphorylase elucidated via molecular dynamics and machine learning. *Nat. Commun.* **6**, 6155 (2015).

45. Riniker, S. Molecular dynamics fingerprints (MDFP): machine learning from MD data to predict free-energy differences. *J. Chem. Inf. Model.* **57**, 726–741 (2017).

46. Karamzadeh, R. et al. Machine learning and network analysis of molecular dynamics trajectories reveal two chains of red/ox-specific residue interactions in human protein disulfide isomerase. *Sci. Rep.* **7**, 3666 (2017).

47. Kadupitiya, J. C. S., Sun, F., Fox, G. & Jadhao, V. Machine learning surrogates for molecular dynamics simulations of soft materials. *J. Comput. Sci.* **42**, 101107 (2020).

48. Wang, J. & Ferguson, A. L. Nonlinear machine learning in simulations of soft and biological materials. *Mol. Sim.* **44**, 1090–1107 (2018).

49. Ferguson, A. L. Machine learning and data science in soft materials engineering. *J. Phys. Condens. Matter* **30**, 043002 (2018).

50. So, S., Mun, J. & Rho, J. Simultaneous inverse design of materials and structures via deep learning: demonstration of dipole resonance engineering using core-shell nanoparticles. *ACS Appl. Mater. Interfaces* **11**, 24264–24268 (2019).

51. Coli, G. M., Boattini, E., Filion, L. & Dijkstra, M. Inverse design of soft materials via a deep-learning-based evolutionary strategy. *Sci. Adv.* **8**, eabj6731 (2022).

52. Swanson, K., Trivedi, S., Lequieu, J., Swanson, K. & Kondor, R. Deep learning for automated classification and characterization of amorphous materials. *Soft Matter* **16**, 435–446 (2020).

53. Bejagam, K. K., An, Y., Singh, S. & Deshmukh, S. A. Machine-learning enabled new insights into the coil-to-globule transition of thermosensitive polymers using a coarse-grained model. *J. Phys. Chem. Lett.* **9**, 6480–6488 (2018).

54. An, Y., Bejagam, K. K. & Deshmukh, S. A. Development of new transferable coarse-grained models of hydrocarbons. *J. Phys. Chem. B* **122**, 7143–7153 (2018).

55. An, Y., Bejagam, K. K. & Deshmukh, S. A. Development of transferable non-bonded interactions between coarse-grained hydrocarbon and water models. *J. Phys. Chem. B* **123**, 909–921 (2019).

56. Bejagam, K. K., Singh, S., An, Y., Berry, C. & Deshmukh, S. A. PSO-assisted development of new transferable coarse-grained water models. *J. Phys. Chem. B* **122**, 1958–1971 (2018).

57. Brassey, C. A., O'Mahoney, T. G., Chamberlain, A. T. & Sellers, W. I. A volumetric technique for fossil body mass estimation applied to Australopithecus afarensis. *J. Hum. Evol.* **115**, 47–64 (2018).

58. Brassey, C. A. & Gardiner, J. D. An advanced shape-fitting algorithm applied to quadrupedal mammals: improving volumetric mass estimates. *R. Soc. Open Sci.* **2**, 150302 (2015).

59. Zawada, K. J. A., Dornelas, M. & Madin, J. S. Quantifying coral morphology. *Coral Reefs* **38**, 1281–1292 (2019).

60. Kapoor, U. & Jayaraman, A. Self-assembly of allomelanin dimers and the impact of poly(ethylene glycol) on the assembly: a molecular dynamics simulation study. *J. Phys. Chem. B* **124**, 2702–2714 (2020).

61. Gartner, T. E. III & Jayaraman, A. Modeling and simulations of polymers: a roadmap. *Macromolecules* **52**, 755–786 (2019).

62. Joshi, S. Y., Singh, S., Bejagam, K. K. & Deshmukh, S. A. Dehydration of polymer chains initiates graphene folding in water. *Carbon N. Y.* **180**, 244–253 (2021).

63. Malod-Dognin, N. Protein structure comparison: from contact map overlap maximisation to distance-based alignment search tool. PhD thesis, University of Rennes (2010).

64. Emerson, I. A. & Amala, A. Protein contact maps: a binary depiction of protein 3D structures. *Phys. A* **465**, 782–791 (2017).

65. Bhattacharya, S. & Bhattacharya, D. Evaluating the significance of contact maps in low-homology protein modeling using contact-assisted threading. *Sci. Rep.* **10**, 2908 (2020).

66. Yethiraj, A. A Monte Carlo simulation study of branched polymers. *J. Chem. Phys.* **125**, 204901 (2006).

67. Rodgers, J. L. & Alan Nicewander, W. Thirteen ways to look at the correlation coefficient. *Am. Stat.* **42**, 59 (1988).

68. Svergun, D. I. et al. Protein hydration in solution: experimental observation by x-ray and neutron scattering. *Proc. Natl Acad. Sci. U.S.A.* **95**, 2267–2272 (1998).

69. Franke, D. et al. ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J. Appl. Crystallogr.* **50**, 1212–1225 (2017).

70. Pesek, S. L., Li, X., Hammouda, B., Hong, K. & Verduzco, R. Small-angle neutron scattering analysis of bottlebrush polymers prepared via grafting-through polymerization. *Macromolecules* **46**, 6998–7005 (2013).

71. Kikhney, A. G. & Svergun, D. I. A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS Lett.* **589**, 2570–2577 (2015).

72. Trulove, P. C. et al. The structure and dynamics of silk and cellulose dissolved in ionic liquids. *ECS Trans.* **16**, 111–117 (2019).

73. Rambo, R. P. & Tainer, J. A. Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. *Biopolymers* **95**, 559–571 (2011).

74. Karplus, M. & Kushick, J. N. Method for estimating the configurational entropy of macromolecules. *Macromolecules* **14**, 325–332 (1981).

75. Levy, R. M., Karplus, M., Kushick, J. & Perahia, D. Evaluation of the configurational entropy for proteins: application to molecular dynamics simulations of an α-helix. *Macromolecules* **17**, 1370–1374 (1984).

76. Karplus, M., Ichiye, T. & Pettitt, B. M. Configurational entropy of native proteins. *Biophys. J.* **52**, 1083–1085 (1987).

77. Chang, C.-E., Chen, W. & Gilson, M. K. Evaluating the accuracy of the quasi-harmonic approximation. *J. Chem. Theory Comput.* **1**, 1017–1028 (2005).

78. Schlitter, J. Estimation of absolute and relative entropies of macromolecules using the covariance matrix. *Chem. Phys. Lett.* **215**, 617–621 (1993).

79. Roe, D. R. & Cheatham, T. E. 3rd PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* **9**, 3084–3095 (2013).

80. Watts, C. R., Gregory, A. J., Frisbie, C. P. & Lovas, S. Structural properties of amyloid β(1-40) dimer explored by replica exchange molecular dynamics simulations. *Proteins* **85**, 1024–1045 (2017).

81. Drake, J. A. & Pettitt, B. M. Thermodynamics of conformational transitions in a disordered protein backbone model. *Biophys. J.* **114**, 2799–2810 (2018).

82. Deshmukh, S. A., Sankaranarayanan, S. K. R. & Mancini, D. C. Vibrational spectra of proximal water in a thermo-sensitive polymer undergoing conformational transition across the lower critical solution temperature. *J. Phys. Chem. B* **116**, 5501–5515 (2012).

83. Ahmed, Z., Gooding, E. A., Pimenov, K. V., Wang, L. & Asher, S. A. UV resonance raman determination of molecular mechanism of poly(N-isopropylacrylamide) Volume phase transition. *J. Phys. Chem. B* **113**, 4248–4256 (2009).

84. Sose, A. T. et al. Modelling drug adsorption in metal–organic frameworks: the role of solvent. *RSC Adv.* **11**, 17064–17071 (2021).

85. Singh, R. D., Mittal, A. & Bhatia, R. K. 3D convolutional neural network for object recognition: a review. *Multimed. Tools. Appl.* **78**, 15951–15995 (2019).

86. Maturana, D., Scherer, S. VoxNet: A 3D Convolutional Neural Network for real-time object recognition. *IEEE Int.* 922–928. https://ieeexplore.ieee.org/document/7353481 (2015).

87. Garcia-Garcia, A. et al. PointNet: A 3D Convolutional Neural Network for real-time object class recognition. *Proc. Int. Jt. Conf. Neural Netw. (IJCNN)* 1578–1584. https://arxiv.org/abs/1612.00593 (2016).

88. Lu, Y., Guo, Q., Lu, G. A Geometric Convolutional Neural Network for 3D Object Detection. *IEEE Glob. Conf. Signal Inf. Process. Proc.* 1–5. https://ieeexplore.ieee.org/document/8969077 (2019).

89. Pang, G., Neumann, U. 3D point cloud object detection with multi-view convolutional neural network. *Int. Conf. Pattern Recognit.* 585–590. https://ieeexplore.ieee.org/document/7899697 (2016).

90. Singh, S. P. et al. 3D Deep Learning on Medical Images: A Review. *Sensors* **20**, https://www.mdpi.com/1424-8220/20/18/5097 (2020).

91. Razzak, M. I., Naz, S., Zaib, A. Deep learning for medical image processing: overview, challenges and the future. *Lect. Notes Comput. Vis. Biomech.* 323–350. https://link.springer.com/chapter/10.1007/978-3-319-65981-7_12 (2018).

92. Li, H. M. Deep Learning for Image Denoising. *Int. J. Signal Process. Image Process. Pattern Recognit.* **7**, 171–180 (2014).

93. Voulodimos, A., Doulamis, N., Doulamis, A. & Protopapadakis, E. Deep learning for computer vision: a brief review. *Comput. Intell. Neurosci.* **2018**, 7068349 (2018).

94. Ioannidou, A., Chatzilari, E., Nikolopoulos, S. & Kompatsiaris, I. Deep learning advances in computer vision with 3D data. *ACM Comput. Surv.* **50**, 1–38 (2017).

95. Young, T., Hazarika, D., Poria, S. & Cambria, E. Recent trends in deep learning based natural language processing. *IEEE Comput. Intell. Mag.* **13**, 55–75 (2018).

96. Li, H. Deep learning for natural language processing: advantages and challenges. *Natl Sci. Rev.* **5**, 24–26 (2018).

97. Yamashita, R., Nishio, M., Do, R. K. G. & Togashi, K. Convolutional neural networks: an overview and application in radiology. *Insights Imaging* **9**, 611–629 (2018).

98. Ryan, K., Lengyel, J. & Shatruk, M. Crystal structure prediction via deep learning. *J. Am. Chem. Soc.* **140**, 10158–10168 (2018).

99. Danaee, P., Ghaeini, R. & Hendrix, D. A. A deep learning approach for cancer detection and relevant gene identification. *Pac. Symp. Biocomput.* **22**, 219–229 (2017).

100. Uziela, K., Menéndez Hurtado, D., Shu, N., Wallner, B. & Elofsson, A. ProQ3D: improved model quality assessments using deep learning. *Bioinformatics* **33**, 1578–1580 (2017).

101. Ching, T. et al. Opportunities and obstacles for deep learning in biology and medicine. *J. R. Soc. Interface* **15**, https://royalsocietypublishing.org/doi/10.1098/rsif.2017.0387 (2018).

102. Soukup, D., Huber-Mörk, R. Convolutional neural networks for steel surface defect detection from photometric stereo images. *Advances in Visual Computing* 668–677 (Springer International Publishing, 2014).

103. Tao, X., Zhang, D., Ma, W., Liu, X. & Xu, D. Automatic metallic surface defect detection and recognition with convolutional neural networks. *NATO Adv. Sci. Inst. Ser. E Appl. Sci.* **8**, 1575 (2018).

104. Yang, C., Kim, Y., Ryu, S. & Gu, G. X. Prediction of composite microstructure stress-strain curves using convolutional neural networks. *Mater. Des.* **189**, 108509 (2020).

105. Rao, C. & Liu, Y. Three-dimensional convolutional neural network (3D-CNN) for heterogeneous material homogenization. *Comput. Mater. Sci.* **184**, 109850 (2020).

106. Le, N. Q. K., Yapp, E. K. Y., Ou, Y.-Y. & Yeh, H.-Y. iMotor-CNN: Identifying molecular functions of cytoskeleton motor proteins using 2D convolutional neural network via Chou's 5-step rule. *Anal. Biochem.* **575**, 17–26 (2019).

107. Shao, Y. & Chou, K.-C. PLoc_deep-mVirus: a CNN model for predicting sub-cellular localization of virus proteins by deep learning. *Nat. Sci. (Irvine)* **12**, 388–399 (2020).

108. Shao, Y.-T. & Chou, K.-C. PLoc_deep-mAnimal: a novel deep CNN-BLSTM network to predict subcellular localization of animal proteins. *Nat. Sci. (Irvine)* **12**, 281–291 (2020).

109. Tayara, H., Tahir, M. & Chong, K. T. iSS-CNN: Identifying splicing sites using convolution neural network. *Chemometrics Intell. Lab. Syst.* **188**, 63–69 (2019).

110. van Ginneken, B. Fifty years of computer analysis in chest imaging: rule-based, machine learning, deep learning. *Radiol. Phys. Technol.* **10**, 23–32 (2017).

111. Ge, L., Liang, H., Yuan, J., Thalmann, D. 3D convolutional neural networks for efficient and robust hand pose estimation from single depth images. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* 5679–5688. https://ieeexplore.ieee.org/document/8100085 (2017).

112. Hachaj, T. & Ogiela, M. R. Rule-based approach to recognizing human body poses and gestures in real time. *Multimed. Syst.* **20**, 81–99 (2014).

113. Chen, X., Sun, Y., Zhang, Q. & Liu, F. Two-stage grasp strategy combining CNN-based classification and adaptive detection on a flexible hand. *Appl. Soft Comput.* **97**, 106729 (2020).

114. Wang, R. et al. Real-time soft body 3D proprioception via deep vision-based sensing. *IEEE Robot. Autom. Lett.* **5**, 3382–3389 (2020).

115. Gulli, A., Pal, S. *Deep Learning with Keras*. (Packt Publishing Ltd, 2017).

116. Shukla, N. *Machine Learning with TensorFlow*. (Manning Publications, 2018).

117. Chen, Y. & Qu, R. Study on infringement identification of art works based on CNN image recognition technology. *J. Phys. Conf. Ser.* **1802**, 032084 (2021).

118. Heaton, J., Goodfellow, I., Bengio, Y. & Courville, A. Deep learning. *Genet. Program. Evol. Mach.* **19**, 305–307 (2018).

119. Goodfellow, I., Bengio, Y., Courville, A. *Deep Learning*. (MIT Press, 2016).

120. Kingma, D. P., Ba, J. Adam: A Method for Stochastic Optimization. *arXiv:1412.6980* (2014).

121. Zhang, W., Itoh, K., Tanida, J. & Ichioka, Y. Parallel distributed processing model with local space-invariant interconnections and its optical architecture. *Appl. Opt.* **29**, 4790–4797 (1990).

122. Li, Q. et al. Medical image classification with convolutional neural network. *Int. Conf. Control Autom. Robot. Vis.* 844–848 (IEEE, 2014).

123. Qin, J., Pan, W., Xiang, X., Tan, Y. & Hou, G. A biological image classification method based on improved CNN. *Ecol. Inform.* **58**, 101093 (2020).

124. Liao, S., Wang, J., Yu, R., Sato, K. & Cheng, Z. CNN for situations understanding based on sentiment analysis of twitter data. *Procedia Comput. Sci.* **111**, 376–381 (2017).

125. Sharma, P., Berwal, Y. P. S. & Ghai, W. Performance analysis of deep learning CNN models for disease detection in plants using image segmentation. *Inf. Process. Agric.* **7**, 566–574 (2020).

126. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).

127. Zheng, A., Casari, A. Feature engineering for machine learning: principles and techniques for data scientists. ('O'Reilly Media, Inc.', 2018).

128. Aghdam, H. H., Heravi, E. J. Convolutional neural networks. *Guide to Convolutional Neural Networks* 85–130 (Springer, 2017).

129. Martínez, L., Andrade, R., Birgin, E. G. & Martínez, J. M. PACKMOL: a package for building initial configurations for molecular dynamics simulations. *J. Comput. Chem.* **30**, 2157–2164 (2009).

130. Phillips, J. C. et al. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 (2005).

ACKNOWLEDGEMENTS

AUTHOR CONTRIBUTIONS

S.Y.J.: Performed simulations, analyzed simulation trajectories, writing - original draft, writing - review & editing, visualization, developed code to generate the simulated structures, developed and implemented CNN-based analysis, feature engineering studies. S.S.: Developed CNN-based analysis methodology, feature engineering studies. S.A.D.: Conceptualized the idea, provided required guidance, funding acquisition, writing - review & editing, feature engineering studies.

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/s41524-022-00725-7.

Correspondence and requests for materials should be addressed to Sanket A. Deshmukh.

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.