Notice: This manuscript has been authored by UT-Battelle, LLC, under Contract No. DE-AC0500OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for the United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).
Exploring particle dynamics during self-organization processes via rotationally invariant latent representations

Sergei V. Kalinin,1,* Shuai Zhang,2,3 Mani Valleti,4
Harley Pyles,5,6 David Baker,5,6,7 James J. De Yoreo,2,3 and Maxim Ziatdinov1,8,†

1 Center for Nanophase Materials Sciences, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA
2 Materials Science and Engineering, University of Washington, Seattle, WA, USA
3 Physical Sciences Division, Pacific Northwest National Laboratory, Richland, WA, USA
4 Bredesen Center for Interdisciplinary Research, University of Tennessee, Knoxville, TN 37996, USA
5 Department of Biochemistry, University of Washington, Seattle, WA, USA
6 Institute for Protein Design, University of Washington, Seattle, WA, USA
7 Howard Hughes Medical Institute, University of Washington, Seattle, WA, USA
8 Computational Sciences and Engineering Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA

The dynamic of complex ordering systems with active rotational degrees of freedom exemplified by protein self-assembly is explored using a machine learning workflow that combines deep learning-based semantic segmentation and rotationally invariant variational autoencoder-based analysis of orientation and shape evolution. The latter allows for disentanglement of the particle orientation from other degrees of freedom and compensates for shifts. The disentangled representations in the latent space encode the rich spectrum of local transitions that can now be visualized and explored via continuous variables. The time dependence of ensemble averages allows insight into the time dynamics of the system, and in particular, illustrates the presence of the potential ordering transition. Finally, analysis of the latent variables along the single-particle trajectory allows tracing these parameters on a single particle level. The proposed approach is expected to be universally applicable for the description of the imaging data in optical, scanning probe, and electron microscopy seeking to understand the dynamics of complex systems where rotations are a significant part of the process.

* sergei2@ornl.gov
† ziatdinovma@ornl.gov
Emergence of ordered patterns in the systems of interacting particles is one of the foundational phenomena in chemistry, condensed matter physics, materials science, and biology, encompassing areas ranging from formation of atomic lattices to self-assembly of viruses, polymers, proteins and lipids in membranes to self-organization phenomena. Correspondingly, understanding of the evolution of such systems and the mechanisms guiding the emergence of the order remained on the forefront of physical research for over half a century. This effort includes both the theoretical and simulation analysis, exploration of real-world structures, and model systems such as colloidal crystals that allow for tunability of underpinning interactions.

One of the challenges in understanding the dynamics of pattern emergence in these systems is the nature of the descriptors of the systems, i.e. the compact representation of the local and global geometries. In cases where the long-range discrete translational symmetry emerges as is the case for crystalline lattices, these can be naturally described in the form of the lattice and primitive unit cells, and the deviations from ideal behaviors can be further analyzed in terms of the symmetry breaking phenomena. This in turn yields the concepts of order parameters, etc. However, the descriptions are considerably more complex in the absence of long-range translational symmetry, as is the case for liquids, liquid crystals, etc. In these cases, the local descriptors are often constructed based on local geometries and nearest-neighbor distributions, e.g. tetratic and hexatic order parameters, and the spatial organization and global order are quantified via corresponding correlation functions and structure factors. The advantage of this approach is that it provides the descriptors that can derived from macroscopic scattering experiments, opening the pathway for experimental studies of such systems as a function of global stimuli such as temperature or chemical potentials, external fields, and time.

The rapid emergence of real space imaging methods and corresponding model systems have provided the insight into the mechanisms and dynamics of the self-organization phenomena on the single particle levels. Several notable examples include optical microscopy studies of colloid crystals assembly, scanning probe microscopy studies of nanoparticle and molecular self-assembly, and environmental electron microscopy studies of nanoparticle dynamics. As such, observations of system evolution particle-by-particle and even the partial or full trajectory reconstructions have become common. Remarkably, the progress in the dynamic scanning transmission electron microscopy enabled these studies even on the atomic level, providing insight
into the mechanisms of temperature-, environment-, and beam induced phase transformations and chemical reactions.\textsuperscript{18,19}

However, this proliferation of the experimental data has brought forth the challenge of associated local descriptors, as a necessary step towards comparing local mechanisms to the theoretical models and scattering studies. For spherical particles interacting through isotropic or angle-dependent force fields, such local descriptors can be derived from local bonding geometries. However, extension of this approach for the anisotropic particles, e.g. rod-like or having more complex geometries, leads to the rapid growth of the number of possible ad-hoc descriptors. Furthermore, comparison with the experiment becomes progressively more complex. These challenges are obvious even in theory, where the analysis of interaction and structure evolution in systems of non-spherical particles evolved much slower than for spherical ones. These difficulties are progressively magnified for the experimental systems, characterized by the presence of the particle size distributions, noise, etc.

Here we introduce the approach for the analysis of the structure evolution in the system of interacting anisotropic particles based on a combination of deep fully convolutional neural networks and variational autoencoders (VAEs) with rotational invariance. This methodology particularly relies on the recognized capability of the VAEs to reduce high-dimensional data sets to the low-dimensional continuous latent variables, and disentangle the representations, i.e. discover the significant trends in data.\textsuperscript{20-24} The example of such trends are e.g. writing styles in hand-written digit data bases or emotional states in human face databases. Introduction of the rotation angle in the image plane as one of the latent variables allows identification of the variants of the same shape at multiple orientations.\textsuperscript{25} In this manner, we introduce the parsimonious particle level descriptors for the system. The time dynamics of the global averages provide the insight into the system evolution, emergence of order, and phase transitions, whereas trajectory-level analysis yields insight into particle dynamics. Note that while this approach is illustrated here for a specific case of particle self-assembly visualized via liquid atomic force microscopy, it is in fact universal and can be applied for the analysis of structure evolutions across broad experimental and theoretical domains.

As a model system, we explore the self-assembly of a \textit{de novo} designed helical repeat protein DHR10-micaX(X=18) \textsuperscript{26}, where X is the number of repeat subunits, on muscovite mica (\textit{m}-mica) with 100 mM KCl (pH 7) was recorded by high-speed atomic force microscopy (HS-
Inspired by ice-binding proteins, DHR10-mica18 has a designed interface with 54 carboxylate residues geometrically matching the Potassium ($K^+$) sublattice on the plane of $m$-mica (001). With 100 mM KCl, 0.025 uM DHR10-mica18 self-assemble into numerous discrete domains, align along one of the three closed-packed $K^+$ lattice-directions. However, the close-packed matrix does not have a steady state. Instead the domains exhibit fluctuations in directions and sizes simultaneously. In the meantime, the relative position of each individual proteins within the domains remain dynamics, too (Figure 1). Hence, the big data of this in-situ assembly of DHR10-mica18 with 100 mM KCl on $m$-mica (hundreds of frames with tens of protein molecules in each frame), generated by HS-AFM, beyond the capability of manual approaches for statistical analysis, defining a descriptors of the self-assembly process.

**Figure 1.** The AFM studies of the dynamics of protein organization. Shown are raw AFM images (top row) and neural network reconstructions (bottom row) for frames (a,b) +131.6 s, (c,d) +263.2 s and (e,f) +394.7 s. The dots in (b,d,f) mark the positions of the centers of gravity of the particles. Image size is 200 nm and frame rate is 0.38 Hz.
To analyze particle positions and orientations from the AFM data, we first use the deep convolutional neural network (DCNN) approach to remove noise and identify particles centers of the mass. Briefly, the training set for DCNN is made from the sub-set of original images with clearly discernible particles as features and manually labeled categorical image (particle/not particle) as a target. The training set is augmented via rotations, Gaussian noise, and horizontal or vertical flips. The DCNN is based on the in-house developed dilnet neural network architecture, which uses multiple blocks with spatial pyramids of dilated convolutions and only one max-pooling operation (and the corresponding up-sampling operation) to preserve the maximum amount of information. To improve the network’s predictions we use the deep ensemble approach\textsuperscript{27,28}, where multiple networks with different “training trajectories” are trained in parallel and used for prediction. The ensemble mean prediction and the associated variance provide improved recognition/generalization and uncertainty estimates\textsuperscript{27}. Note that thus trained network ensemble yields the semantically segmented images, i.e. for each pixel in the raw image Fig. 1 (a,c,e) the decoded image Fig. 1 (b,d,f) yields the “probability” that it belongs to the particle.

With this information in hand, the positions of the center of mass of individual particles can be readily mapped. Previously, we used the ellipsoid fit to find the characteristic particle sizes and rotation angles\textsuperscript{29}. Based on these top-down particle level descriptors, the system evolution in terms of relevant distribution functions and their spatial correlations can be analyzed. In certain cases when the same particle can be traced across multiple frames, the evolution of these parameters along the trajectory can be traced. Note that while these characteristics are readily available in the simulation studies, discovery of these in the experimental data represents a considerable challenge and analysis is naturally limited to the objects that can be discerned by the deep learning network and, for trajectories, the displacements of which is sufficiently small to allow for reconstruction. That said, the subsequent discussion and analysis can be equally applied for both decoded experimental and simulation data.

To get insight into particle dynamic and structure evolution during the self-organization process, here we use a rotationally invariant extension\textsuperscript{25} of the variational autoencoder (rVAE). Generally, autoencoders (AE) refer to the class of the neural networks that compress the data set to a small number of latent features, and then expand back to original data set. The training aims to minimize information loss between the initial and reconstructed images via usual backpropagation. This process tends to select the relevant features in the data set and reject the
noise, giving rise to applications for denoising, etc. At the same time, the latent features allow for efficient encoding of the original data set.

Variational autoencoders (VAEs) expand on this concept by substituting the bottleneck latent layer by the latent space, from which the variables for decoding are drawn from a prescribed prior distribution. In this manner, VAEs represent a hybrid of the AE approach for creating generalized encoding and decoding functions, and Bayesian priors for feature selection. Since the Bayesian layer is non-differentiable, the training of the VAEs is based on sampling latent vector using the reparameterization trick introduced by Kingma and Welling.\textsuperscript{20,21} The unique aspect of VAEs is that they tend to structure the latent space in such a way that the decoded data will have clear variations along the latent directions. This behavior, referred to as disentangled representations,\textsuperscript{30,31} allows for determination of styles of handwriting of fashion, style transfer, etc. More generally, VAEs allow projection of high-dimensional and potentially discrete spaces onto low-dimensional differentiable manifold, potentially allowing for mapping equations of motions, enabling Bayesian optimization, etc.

Here, we adapt the variational autoencoder to include the rotation and offsets in x- and y-directions as three of the latent variables, in addition to classical latent variables.\textsuperscript{25} In this manner, the rotations of the particles in the image plane are separated as one latent variable, and non-idealities in determination of particle center of mass are captured via offsets. The remaining latent variables provide the information on particle shape, structure of the nearest neighborhood, etc. depending on the size of the sampling window (size of subimage cropped around each detected particle). The encoder and decoder of rVAE are chosen to be simple fully-connected (“dense”) neural networks.

The semantically segmented DCNN output is used to create an input into the rVAE. The use of the raw data led to relatively smooth decoded features that allow for partial orientation mapping only and was not actively pursued. The latent angles (in this specific case) give rise to clear multimodal distributions corresponding to 6 possible orientations of the particles on the surface, whereas offset distributions are reasonably narrow and sharp. These criteria were used to identify optimal training and sampling window parameters. Here, we have chosen the windows slightly above the particle length (see Fig. 2d-f), to fully capture the particle shape and the lateral interactions but at the same time avoid excessive details that necessitate high dimensionalities of latent space to describe.
Figure 2. (a,b,c) Several sub-images of the raw data and (d,e,f) corresponding DCNN output. The latter is used for rVAE training. (g) The learned latent space of the rVAE projected onto the image space showing the evolution of the decoded images as a function of latent variables, $d(L_1, L_2)$.

Here, the rVAE’s training was performed using Adam optimizer\textsuperscript{32} for 3000 epochs with a mini-batch size of 200 for the data set of the ~13,200 sub-images, as limited by the number of the decoded particles in the data set. The typical sub-images with raw experimental data and corresponding decoded images are shown in Figure 2. rVAE analysis converts each decoded sub-image into the angle, offsets, and latent variables that now describe the state of each particle. The system behavior can then be analyzed via statistical analysis of the time dependence of relevant distributions within each frame as a function of frame number, correlation function analysis, or trajectory analysis.

A convenient way to represent the rVAE operation is though the analysis of latent space representations as shown in Fig. 2 (g). The encoding of the sub-images transforms each of the sub-images into 3+2 latent variables. The distribution of the latent variables determines the size of the latent variable space bound by minimal and maximal values of $L_1$ and $L_2$. We can further introduce the uniform rectangular grid of points in the latent space and decoded these values to yield the
particles geometry. This latent space representation is shown in Fig. 2 (g). Note that the decoded features have a clear physical meaning, representing the single and multiple particles assemblies.

The characteristic aspect of VAEs is their potential for disentangling the data representations, where each latent variable describes the certain trait in data set. This examination of this behavior is visible in Fig. 2 (g), where on transition from left to right (i.e. for constant $L_2$) the particles become smaller, whereas on vertical the number of particles in decoded sub-image increases. The projected latent space for the vanilla VAE can be found in the Supplemental Material.

Figure 3. (a,b,c) Magnified cross-sections along the $L_1$ axis at fixed values of $L_2$ and (d,e) along the $L_2$ axis along the fixed values of $L_1$.

To get further insight into these behaviors, shown in Fig. 3 are the expanded cross-sections from Figure 2 (g) obtained for higher sampling density in latent space. Here, the Fig. 3 (a) shows the gradual evolution of a particle shape, eventually converting to non-physical mixed contrast. Similar variation between physical and unphysical shapes is observed in (d). At the same time, Fig. 3 (b,c,e,f) illustrate the evolution of density of particles across the row.
**Figure 4.** Distributions of latent variables across the full experiment. (a) Angle distribution clearly illustrates the presence of six dominant orientations due to the interactions between the nanoparticles and substrate. Note that here particles are not assumed to be symmetric perpendicular to the long axis. (b, c) Distributions of the $x$- and $y$-offsets. These distributions are relatively featureless but indicate the convergence of the rVAE. (d,e,f) Joint distributions between the angle and latent variables $L_1$, $L_2$. Shown are the points corresponding to each particle and superimposed is kernel density estimate.

The latent representation analysis allows to get further insight into the global system dynamics via the analysis of the latent parameter distributions and their time dynamics. Shown in Figure 4 is the global (i.e. averaged over the full data set) distributions for the latent variables. Here, the angle distributions clearly show 6 peaks, corresponding to preferential orientation of the protein particles due to anisotropic interactions with the substrate. The offset distributions are relatively featureless and generally confined within one pixel, indicative of successful particle finding (in cases when rVAE fails to converge, much broader distributions are observed). Finally, the joint distributions in latent variable space are shown in Figs. 4 (d-f). Here, we visualize both the individual data points and the superimposed kernel density estimates. Note that the $L_1-\theta$ distribution is almost marginalizable, with clear 6-fold maxima associated with angle distribution. At the same time, the $L_2$ distribution is nontrivial, illustrating the presence of the multimodal $L_2$ distributions for each angle. Comparison with Fig. 2 (g) illustrates that these distributions differ
by the laterals pacing between the particles. This behavior is further illustrated in Fig 4 (f), where two primary maxima corresponding to ~1 and -1 for L2 are clearly seen.

**Figure 5.** Time dynamics of (a) latent angle, (b) first and (c) second latent variables. NMF analysis of latent angle distribution. (d,e) time dependence of the first and second NMF component and (f,g) corresponding component shapes. Note that features with lost contrast were excluded from NMF analysis. Clear transitions at frame ~170 is visible (a,d).

To get insight into the dynamics of the system, we analyze the time dynamics of the latent variable distribution. To accomplish it, we calculate the 1D kernel density estimates (KDE) for the distribution of the corresponding latent variables, as shown in Fig. 5. The time dynamics of the angle clearly indicates the presence of six rotational variants in the early times, with the transition to only 4 dominant variants on later stages.

The relevant aspects of this behavior can be further analyzed using the suitable dimensionality reduction method. Given that KDE are positively defined, we use the non-negative matrix factorization (NMF),

\[
KDE(c, t) = \sum_{i=1}^{N} l_{i}(Lc)w_{i}(t)
\]

where \(l_{i}(Lc)\) are the NMF weights and \(w_{i}(t)\) are the endmembers that determine characteristic time behaviors. The number of components \(N\) is set at the beginning of the analysis and can be
chosen based on the quality of decomposition, anticipated physics of the systems, etc. Here, after experimentation, the $N = 2$ was found to be sufficient to represent the observed dynamics.

The time dependence of the 1$^{\text{st}}$ and 2$^{\text{nd}}$ NMF component are shown in Figure 5 (d,e) respectively. Note the sharp change at frame $\sim 140$ associated with the disappearance of 2 out of 6 orientational variants, i.e. spontaneous symmetry lowering in the system.

![Figure 5](image)

**Figure 6.** Latent encoding in particle assemblies. (a) Latent angle and (b) first and (c) second latent variables. Note that for ordered phase both parallel (ferroelectric-like) and antiparallel (antiferroelectric-like) arrangements of the particles can be observed. (d, e) Example of a trajectory of a single particle encoded by angle (d) and the first latent variable (e).

Finally, the latent representations allow the exploration of the dynamics on a single particle level. Shown in Figure 6 is the raw images at several times with the color markers indicative of the corresponding latent variables. In some case, the particle trajectories can be reconstructed by tracking individual particles from one frame to another via the nearest neighbor search. Here the search radius was set to 21 px, which corresponds to 8.2 nm. In this case, the evolution of the latent variable along the trajectory can be explored, as shown in Fig. 6. Similarly, the particle dynamics
can be explored in latent space, representing the changes in particle geometry and nearest neighborhood during the evolution.

To summarize, the particle dynamics using protein self-assembly was explored using machine learning workflow combining DCNN based semantic segmentation and rotationally invariant VAE analysis of orientation shad shape dynamics. The chosen VAE architecture allows disentangling the particle orientation from other degrees of freedom and compensates for shifts. The disentangled representations in the latent space encode the rich spectrum of local transitions that can now be visualized and explore via continuous variables. The time dependence of ensemble averages allows insight into the time dynamics of the system, and in particular illustrates the presence of the potential ordering transition. Finally, analysis of the latent variables along the single particle trajectory allows mapping evolution of particle shape and nearest environment.

The proposed approach is expected to be universally applicable for the description of the imaging data in optical, scanning probe, and electron microscopy seeking to understand dynamics of complex systems where rotations is a significant part of the process. We note that while here both the DCNN and rVAE were applied to a single class, it can be expanded to multiclass features in a straightforward way. Furthermore, this approach can be used for exploration of computationally generated datasets, including the evolution of the electronic density and lattice displacements desiring diffusion and reactions in atomistic modelling, dynamics of macromolecules, etc.

Acknowledgment
This research including AFM (S.Z., J.J.D.,), data analytics (S.V.K.) and protein design and synthesis (H.P., D.B.) are supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences, as part of the Energy Frontier Research Centers program: CSSAS–The Center for the Science of Synthesis Across Scales–under Award Number DE-SC0019288, located at University of Washington. The machine learning was performed and partially supported (M.Z.) at the Oak Ridge National Laboratory’s Center for Nanophase Materials Sciences (CNMS), a U.S. Department of Energy, Office of Science User Facility. High-speed AFM experiments were performed at the Department of Energy's Pacific Northwest National Laboratory (PNNL). PNNL
Materials and methods:

Particle synthesis
DHR10-mica18 proteins were expressed in *E. coli*, purified with nickel NTA affinity and size exclusion chromatography, and dialyzed into 20mM Tris buffer (pH=8). For more details see reference [7].

High-speed atomic force microscopy
DHR10-mica18 protein stock solution was diluted to 0.025 μM with 20 mM Tris buffer (pH=7) having 100 mM KCl. 20 μl diluted protein solution was dropped onto freshly cleaved muscovite-mica (001) (SPI Supplies) and characterized by Cypher Video-Rate AFM (Asylum Research) in liquid amplitude-modulation mode. The probe USC-F1.2-K0.15 (NanoWorld) was used. The imaging force was adjusted to minimize any interruption. Tris-HCl buffer (pH 7, 1 M), and KCl were bought from Sigma-Aldrich. Nuclease-free water was bought from Ambion.

Data analysis
The DCNN and rVAE were implemented via AtomAI package\textsuperscript{34}. To train a deep DCNN ensemble we first trained a baseline model for \( N \) epochs until test loss reached a plateau. We then trained 12 individual ensemble models for \( n \) epoch (\( n \ll N \)) starting each time with the weights of the baseline model and performing training data shuffling with different random seed. The Adam optimizer\textsuperscript{32} with a learning rate of 0.001 was used for optimizing weights of all the ensemble models. The encoder and decoder of rVAE were 2-layer perceptrons with 128 neurons in each layer activated by \( \text{tanh}(\cdot) \) function whose weights were optimized using Adam optimizer with a learning rate of 0.0001.
References

1. Sharma, V., Park, K. & Srinivasarao, M. Colloidal dispersion of gold nanorods: Historical background, optical properties, seed-mediated synthesis, shape separation and self-assembly. *Mater. Sci. Eng. R-Rep.* **65**, 1-38, doi:10.1016/j.mser.2009.02.002 (2009).

2. Boles, M. A., Engel, M. & Talapin, D. V. Self-Assembly of Colloidal Nanocrystals: From Intricate Structures to Functional Materials. *Chem. Rev.* **116**, 11220-11289, doi:10.1021/acs.chemrev.6b00196 (2016).

3. Wang, T. *et al.* Self-Assembled Colloidal Superparticles from Nanorods. *Science* **338**, 358-363, doi:10.1126/science.1224221 (2012).

4. He, J. B. *et al.* Self-Assembly of Tobacco Mosaic Virus at Oil/Water Interfaces. *Langmuir* **25**, 4979-4987, doi:10.1021/la803533n (2009).

5. Marsh, J. A. & Teichmann, S. A. Structure, Dynamics, Assembly, and Evolution of Protein Complexes. *Annual Review of Biochemistry* **84**, 551-575, doi:10.1146/annurev-biochem-060614-034142 (2015).

6. Luo, Q., Hou, C., Bai, Y., Wang, R. & Liu, J. Protein Assembly: Versatile Approaches to Construct Highly Ordered Nanostructures. *Chemical Reviews* **116**, 13571-13632, doi:10.1021/acs.chemrev.6b00228 (2016).

7. Baule, A., Mari, R., Bo, L., Portal, L. & Makse, H. A. Mean-field theory of random close packings of axisymmetric particles. *Nat. Commun.* **4**, 11, doi:10.1038/ncomms3194 (2013).

8. Mahynski, N. A., Pretti, E., Shen, V. K. & Mittal, J. Using symmetry to elucidate the importance of stoichiometry in colloidal crystal assembly. *Nat. Commun.* **10**, 11, doi:10.1038/s41467-019-10031-4 (2019).

9. Li, B., Zhou, D. & Han, Y. L. Assembly and phase transitions of colloidal crystals. *Nature Reviews Materials* **1**, doi:10.1038/natrevmats.2015.11 (2016).

10. Whitelam, S. Control of Pathways and Yields of Protein Crystallization through the Interplay of Non-specific and Specific Attractions. *Physical Review Letters* **105**, 088102, doi:10.1103/PhysRevLett.105.088102 (2010).

11. Mannige, R. V. *et al.* Peptoid nanosheets exhibit a new secondary-structure motif. *Nature* **526**, 415-420, doi:10.1038/nature15363 (2015).

12. Wittkowski, R. *et al.* Scalar $\phi(4)$ field theory for active-particle phase separation. *Nat. Commun.* **5**, 9, doi:10.1038/ncomms5351 (2014).

13. Ando, T., Uchihashi, T. & Scheuring, S. Filming Biomolecular Processes by High-Speed Atomic Force Microscopy. *Chemical Reviews* **114**, 3120-3188, doi:10.1021/acs.chemrev.4b00387 (2014).

14. Rajendran, A., Endo, M. & Sugiyama, H. State-of-the-Art High-Speed Atomic Force Microscopy for Investigation of Single-Molecular Dynamics of Proteins. *Chemical Reviews* **114**, 1493-1520, doi:10.1021/acs.chemrev.1c00253x (2014).

15. Chen, J. *et al.* Building two-dimensional materials one row at a time: Avoiding the nucleation barrier. *Science* **362**, 1135, doi:10.1126/science.aau4146 (2018).

16. Ou, Z., Wang, Z., Luo, B., Luijten, E. & Chen, Q. Kinetic pathways of crystallization at the nanoscale. *Nature Materials* **19**, 450-455, doi:10.1038/s41563-019-0514-1 (2020).

17. Tan, S. F., Chee, S. W., Lin, G. & Mirsaidov, U. Direct Observation of Interactions between Nanoparticles and Nanoparticle Self-Assembly in Solution. *Accounts of Chemical Research* **50**, 1303-1312, doi:10.1021/acs.accounts.7b00063 (2017).

18. Song, M. *et al.* Oriented attachment induces fivefold twins by forming and decomposing high-energy grain boundaries. *Science* **367**, 40, doi:10.1126/science.aax6511 (2020).
19 De wael, A. et al. Measuring Dynamic Structural Changes of Nanoparticles at the Atomic Scale Using Scanning Transmission Electron Microscopy. *Physical Review Letters* **124**, 106105, doi:10.1103/PhysRevLett.124.106105 (2020).

20 Kingma, D. P. & Welling, M. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114* (2013).

21 Kingma, D. P. & Welling, M. An Introduction to Variational Autoencoders. *Foundations and Trends® in Machine Learning* **12**, 307-392, doi:10.1561/2200000056 (2019).

22 Doersch, C. Tutorial on variational autoencoders. *arXiv preprint arXiv:.05908* (2016).

23 Yamada, T., Hosoe, M., Kato, K. & Yamamoto, K. in *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*. 1019-1024.

24 Linh Tran, D., Walecki, R., Eleftheriadis, S., Schuller, B. & Pantic, M. in *Proceedings of the IEEE International Conference on Computer Vision*. 3190-3199.

25 Bepler, T., Zhong, E., Kelley, K., Brignole, E. & Berger, B. in *Advances in Neural Information Processing Systems*. 15409-15419.

26 Pyles, H., Zhang, S., De Yoreo, J. J. & Baker, D. Controlling protein assembly on inorganic crystals through designed protein interfaces. *Nature* **571**, 251-256, doi:10.1038/s41586-019-1361-6 (2019).

27 Lakshminarayanan, B., Pritzel, A. & Blundell, C. in *Proceedings of the 31st International Conference on Neural Information Processing Systems* 6405–6416 (Curran Associates Inc., Long Beach, California, USA, 2017).

28 Fort, S., Hu, H. & Lakshminarayanan, B. Deep Ensembles: A Loss Landscape Perspective. *arXiv preprint arXiv:.02757* (2019).

29 Ziatdinov, M. et al. Quantifying the dynamics of protein self-organization using deep learning analysis of atomic force microscopy data. *arXiv preprint arXiv:.03532* (2020).

30 Bengio, Y., Courville, A. & Vincent, P. Representation Learning: A Review and New Perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **35**, 1798-1828, doi:10.1109/TPAMI.2013.50 (2013).

31 Higgins, I. et al. Towards a definition of disentangled representations. *arXiv preprint arXiv:.02230* (2018).

32 Kingma, D. P. & Ba, J. Adam: A Method for Stochastic Optimization. *arXiv preprint arXiv:1412.6980* (2015).

33 Lee, D. D. & Seung, H. S. Learning the parts of objects by non-negative matrix factorization. *Nature* **401**, 788-791, doi:10.1038/44565 (1999).

34 Ziatdinov, M. AtomAI: deep and machine learning for atom-resolved data. *GitHub repository, https://git.io/JfUxl* (2020).