Supplementary Materials for
Imaging Unstained Synthetic Polymer Crystals and Defects on Atomic Length-scales using Cryogenic Electron Microscopy

Xi Jiang[1], Douglas R. Greer[1,2], Joyjit Kundu[3,4], Colin Ophus[3], Andrew M. Minor[3,5], David Prendergast[3], Ronald N. Zuckermann[3], Nitash P. Balsara[1,2], Kenneth H. Downing[6]**

* Correspondence to: khdowning@lbl.gov, nbalsara@berkeley.edu

This PDF file includes:

Other Supplementary Materials for this manuscript includes the following:

Atomic models in PDB format

Image processing

Images of 2D crystals are generally not perfect due to dislocations in individual unit cells, distortions from stress and image distortion within the microscope, which cause high resolution diffraction spots to be smeared out [1]. In order to recover the high spatial frequency signal, a crystal unbending process was conducted on all micrographs. The motion-corrected and summed low-dose micrographs were imported into 2dx, an image processing package for 2D electron crystallography. Details of the unbending process in 2dx can be found elsewhere. Briefly, the position of each unit cell in the image is found by cross correlation with a small reference area, and a smooth function is defined for displacements from the ideal lattice. This function is then used to reinterpolate the image onto the regular lattice [1]. The defocus values, astigmatism and specimen tilt were determined using CTFFIND4 and corrected after unbending in 2dx [2]. Density maps were computed from structure factors extracted from the Fourier transforms of the unbent images.

We found that there were substantial variations of unit cell averaged images from different micrographs (Fig. S2), as well as higher symmetry than expected from plausible atomic models of the structure. In order to determine whether these variations arise from differences in crystal structure, we took advantage of the fact that the coordinates of the unit cells in the unbent micrographs are determined during crystal unbending by 2dx. Small boxes, just larger than one unit cell in one direction, were extracted from the micrographs at the locations of each unit cell where the correlation with the reference was above a threshold value, which was set to half the average value for each micrograph. Figure S3 shows typical results, where a substantial number of boxes around the crystal edges were excluded from analysis. The selected boxes were sorted into image classes using the Relion software package [3]. The number of classes requested can be set within a wide range, and we found that 35 classes gave the most informative results, shown in Fig. 2 and S4. Figure S5 shows Fourier transforms of each class average. This process clearly identified the presence of lower symmetry, consistent with the identified plane group, and a surprising number of highly populated image classes that apparently represent different crystal packing arrangements. Using information in the Relion-generated star file, the class of each box in the individual micrographs was plotted, as shown in Figs. 2 and S6. To display the information from each crystal more clearly, the boxed images from each micrograph were separately sorted by Relion into a smaller number of classes (typically six). In each case two of these resulting class averages were similar to the principal classes found in processing all micrographs together. Relion did not generally identify a class corresponding to the domain boundary. In order to visualize this boundary, the boxes along the domain edges were selected separately and averaged. Results are shown in Fig. 3.

In the Relion software package, there are several parameters that can be adjusted in order to obtain robust classification results. In this study, most of these parameters were set to the default values. Relion is designed primarily for work with isolated particles, and the signal-to-noise ratio is estimated from an area presumed not to include the particle. The presence of crystalline sample throughout the box could confound this estimate, which can be sensitive to both the size of the box and the size of a mask within the box used to isolate the particle of interest. Several parameters, including box size, mask size, number of classes and number of iterations were examined in order to determine their effect on the 2D classification results. The box size did not affect the 2D classification at all. However, with a large mask size (diameter close to the box size), the high symmetry artifacts were found in the final classes, particularly with a small number of classes (~15). Therefore, the mask size was fixed at 52 Å diameter (slightly larger

than a unit cell along one direction) with a box size of 87 Å for processing of ~200,000 boxes from 12 micrographs. However, when the 2D classification was carried out on one micrograph with a smaller number of classes (3-6), the mask size was allowed to be larger and gave good segregation of classes, such as shown in Figure S6. While a large number of classes (~40 – 50) was helpful in eliminating highly symmetrical artifacts, there were more empty and duplicate classes.

Molecular dynamics simulation

Based on the experimental observations, we prepared a number of ordered $pNdc_9$-$pNte_9$ peptoid configurations within a water-box as initial structures for molecular dynamics (MD) simulations. We considered all-cis and all-trans conformations of this polymer having a polar N-terminal block and a hydrophobic C-terminal block (each molecule has a backbone directionality) to build the monolayer nanosheet. Although a bilayer structure, where the backbone chains run roughly perpendicular to the crystal surface and hydrophobic blocks of the polymers are hidden from water, could be a candidate structure, initial modeling was restricted to monolayer conformations where adjacent molecules are arranged parallel in *a* direction and antiparallel in the *b* direction as shown in Fig. 3. We found that the interaction strength in the *b* direction is not as strong as that in the *a* direction (due to very small surface area of interaction), leading to the possibility of the types of defects or grain boundaries that are observed along the *b* direction.

For each simulation, a periodic box consisting of 96 polymers (12 along *a*, 8 along *b*) was considered where we put explicit water on top and bottom of the nanosheet. All the simulations were performed with the simulation package NAMD [4] using the CHARMM-based [5] force field for peptoid backbones, MFTOID [6]. Each system is equilibrated for at least 160 ns in isothermal-isobaric (NPT) ensemble at 298 K and 1 atmospheric pressure where all three orthogonal box dimensions were allowed to fluctuate independently. Constant pressure is maintained using the Nosé-Hoover Langevin piston method [7, 8] with an oscillation period of 100 fs and decay time of 50 fs. All the bonds involving hydrogen atoms were kept rigid (using SHAKE [9]. Particle-mesh Ewald summation [10] was used to evaluate long-range electrostatic interactions with a real-space cut-off of 12 Å. Van der Waals interactions were calculated up to a distance of 12 Å, with a smoothing function applied from 10 Å to 12 Å to ensure vanishing energies and forces at the cutoff. All the systems were simulated using a leap-frog integration algorithm with a 2-fs time-step. We checked equilibration for each model by ensuring that the final configuration is independent of the initial preparations. The data were recorded once the steady state (energy and box dimensions fluctuate around the respective mean values) was reached. Using the equilibrated configurations obtained from MD simulations, we calculated histograms of atom spacings for pNdc-pNte. The spacings calculated from MD simulations of the cis-PA model are in good agreement with the cryo-EM results.

The major features observed in the EM images, including the V-shape arrangement of side chains and reflections near 4.7 Å along equator in the Fourier transform of the image are reproduced in simulated crystal model (Figure 3B). The distance between backbones is 4.7 Å, and there is strong contrast of this spacing along the vertex of the V. The experimental cryo-EM images suggest the presence of grain boundaries in the nanosheet. To check the stability of the structures with grain boundary, we created defects in the middle of nanosheet atomic model as shown in Fig. S8 and measured their energy penalty. We used the equilibrated (in NPT ensemble) configuration to estimate the internal energy. Energy cost per defect/mismatch along

b can vary from negligible (within numerical precision) up to 16 kcal/mol depending on whether the defect introduces a mismatch of tilt angle between two adjacent domains. When the backbone tilt in the two parallel chains is the same, the energy cost for the crystal defect is negligible.

References

1. Henderson, R.; Baldwin, J. M.; Downing, K. H.; Lepault, J.; Zemlin, F. Structure of Purple Membrane from Halobacterium-Halobium - Recording, Measurement and Evaluation of Electron-Micrographs at 3.5 a Resolution. *Ultramicroscopy* **1986,** 19 (2), 147-178 DOI: Doi 10.1016/0304-3991(86)90203-2.
2. Rohou, A.; Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *Journal of structural biology* **2015,** 192 (2), 216-21 DOI: 10.1016/j.jsb.2015.08.008.
3. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *Journal of structural biology* **2012,** 180 (3), 519-30 DOI: 10.1016/j.jsb.2012.09.006.
4. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J Comput Chem* **2005,** 26 (16), 1781-1802 DOI: 10.1002/jcc.20289.
5. Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: The Biomolecular Simulation Program. *J Comput Chem* **2009,** 30 (10), 1545-1614.
6. Mirijanian, D. T.; Mannige, R. V.; Zuckermann, R. N.; Whitelam, S. Development and Use of an Atomistic CHARMM-Based Forcefield for Peptoid Simulation. *J Comput Chem* **2014,** 35 (5), 360-370.
7. Martyna, G. J.; Tobias, D. J.; Klein, M. L. Constant-Pressure Molecular-Dynamics Algorithms. *J Chem Phys* **1994,** 101 (5), 4177-4189 DOI: Doi 10.1063/1.467468.
8. Feller, S. E.; Zhang, Y. H.; Pastor, R. W.; Brooks, B. R. Constant-Pressure Molecular-Dynamics Simulation - the Langevin Piston Method. *J Chem Phys* **1995,** 103 (11), 4613-4621 DOI: Doi 10.1063/1.470648.
9. Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes. *J Comput Phys* **1977,** 23 (3), 327-341 DOI: Doi 10.1016/0021-9991(77)90098-5.
10. Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. *J Chem Phys* **1993,** 98 (12), 10089-10092 DOI: Doi 10.1063/1.464397.

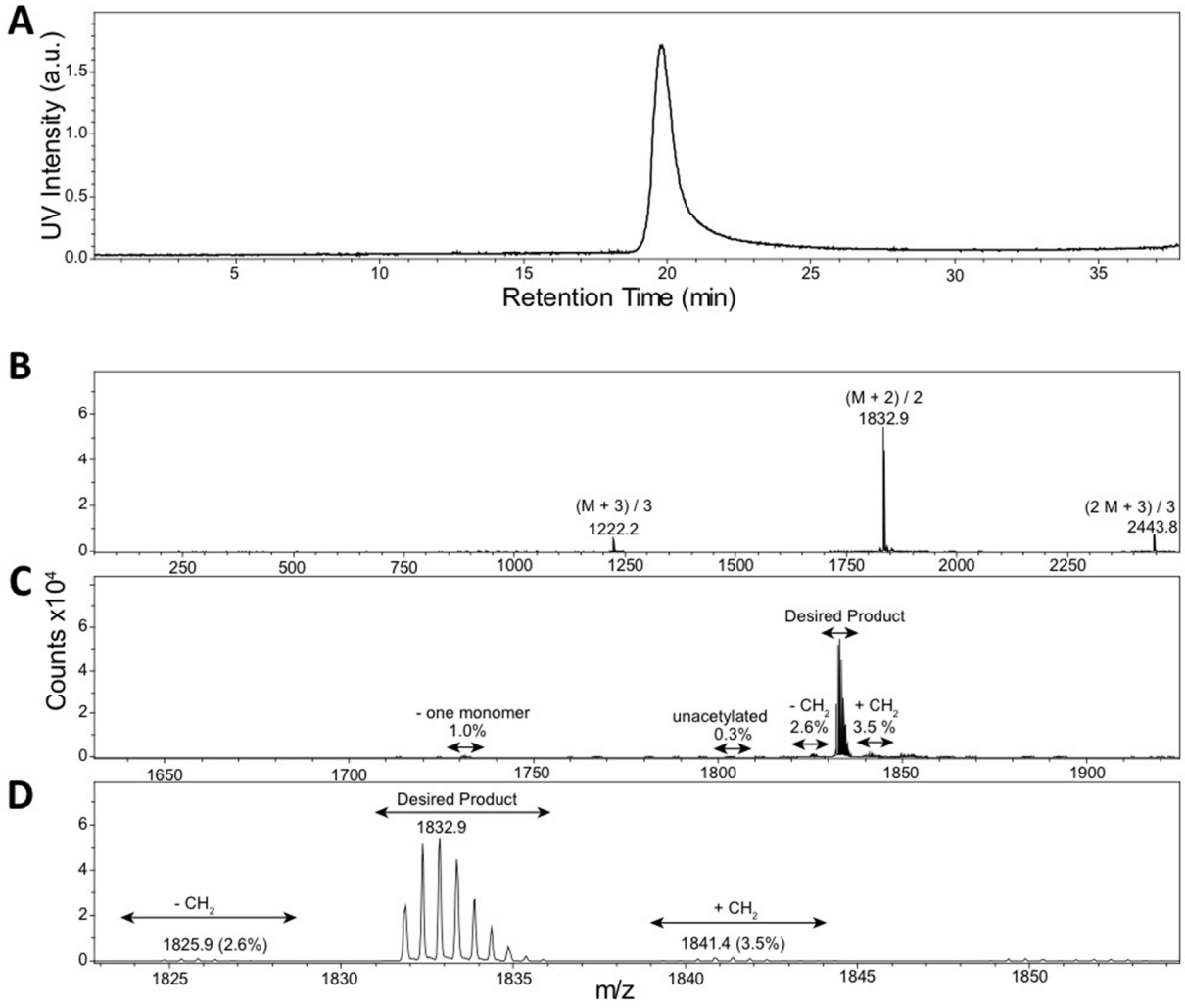

Fig. S1 Analytical HPLC-MS analysis of pNdc$_9$-pNte$_9$. A, UV chromatogram at 215 nm of the purified compound. B, The electrospray mass spectrum from the product peak (retention time of 18.9 - 22.7 min) over an m/z range of 0 – 2500 amu. Peaks for the doubly and triply charged parent ion are seen, as well as a dimer peak. C, The desired product peak range is identified, as well as ranges of synthetic byproducts and their relative mass spectrum peak intensities. The most abundant byproduct is the addition of a single -CH_2- group in one of the decyl sidechains, likely resulting from the presence of a small amount of undecylamine in the decylamine submonomer. D, Increased m/z detail of the desired product peak range.

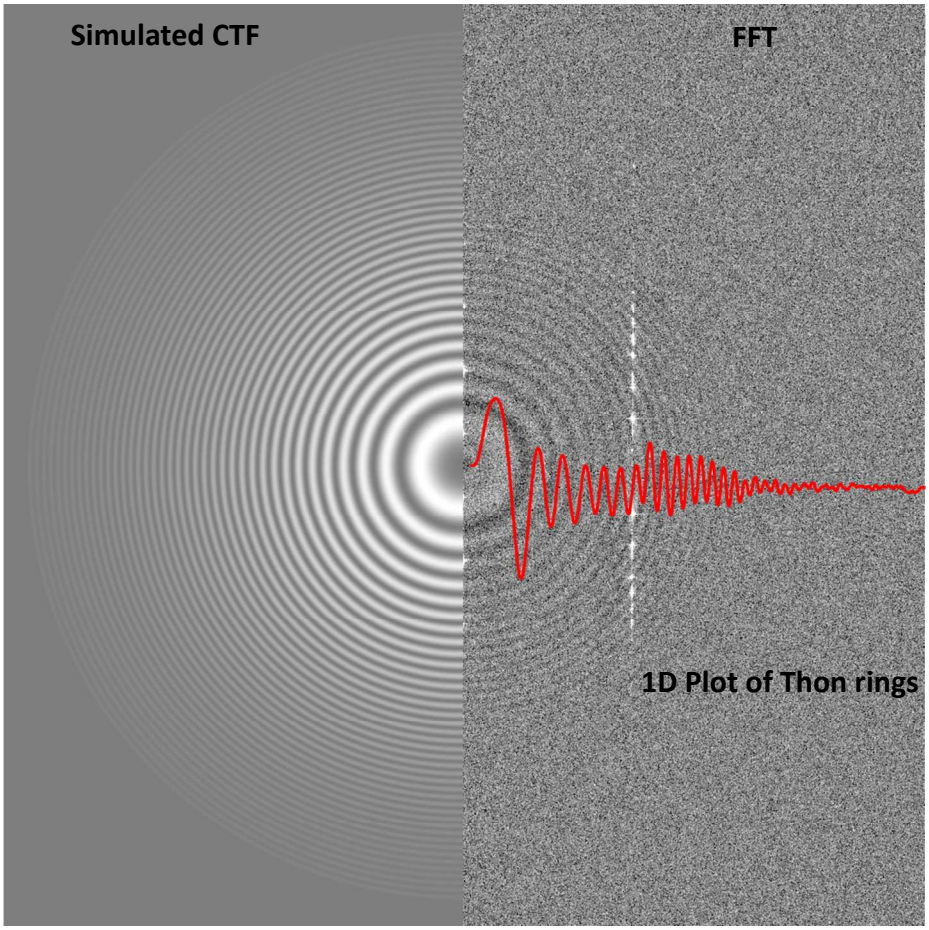

Fig. S2 CTF estimation of the cryo-EM micrograph in Figure 1E. Simulated CTF (left) and 1D plot of Thon rings (right), red line represents the rotation averaged Thon rings after subtracting background, that generated using Gctf Thon rings are detectable close to Nyquist limitation (1.8 Å). Reflections cannot be observed in 1D plot of Thon rings.

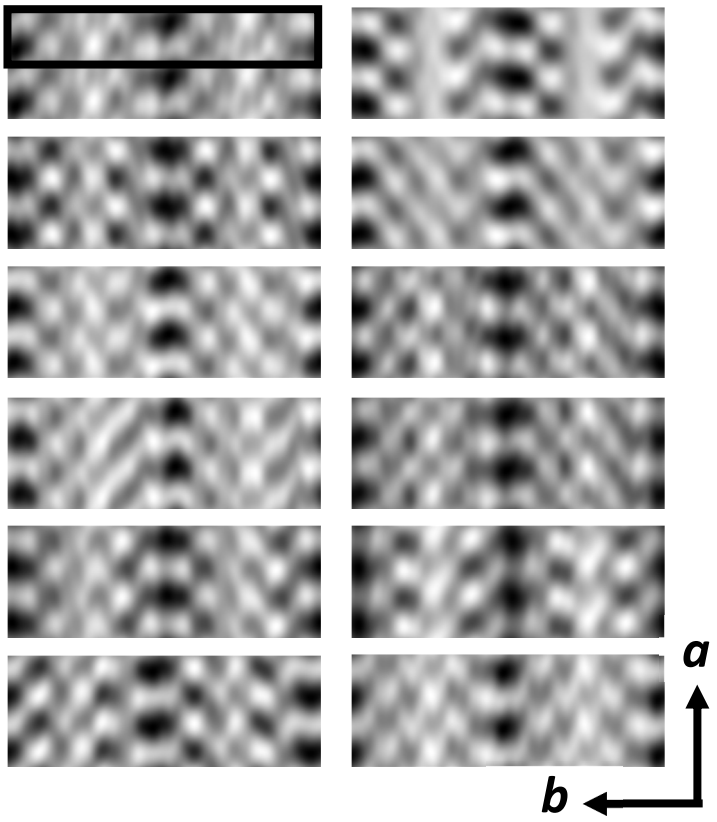

Fig. S3 Averaged images of unit cells obtained by crystallographic averaging over the full area of 12 low-dose cryo-EM micrographs of nominally untilted crystals. This procedure does not account for the presence of anti-parallel adjacent molecules. The black rectangular box indicates the size of unit cell with *a*=4.7Å, *b*=50Å.

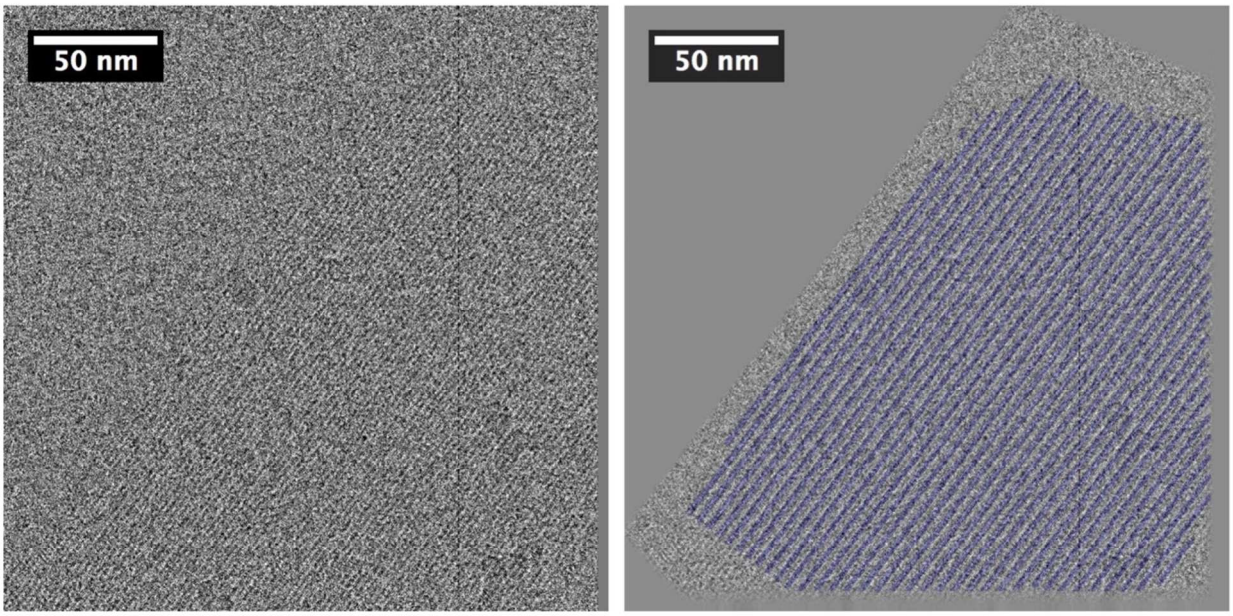

Fig. S4 Micrograph and crystal extraction. Original low-dose cryo-EM micrograph (left) and positions of the unit cells identified in the course of lattice unbending (marked with blue crosses superimposed on unbent crystal micrograph at right) which were also used as centers of extracted boxes. The threshold used for selecting unit cells causes exclusion of some around the periphery of the crystal.

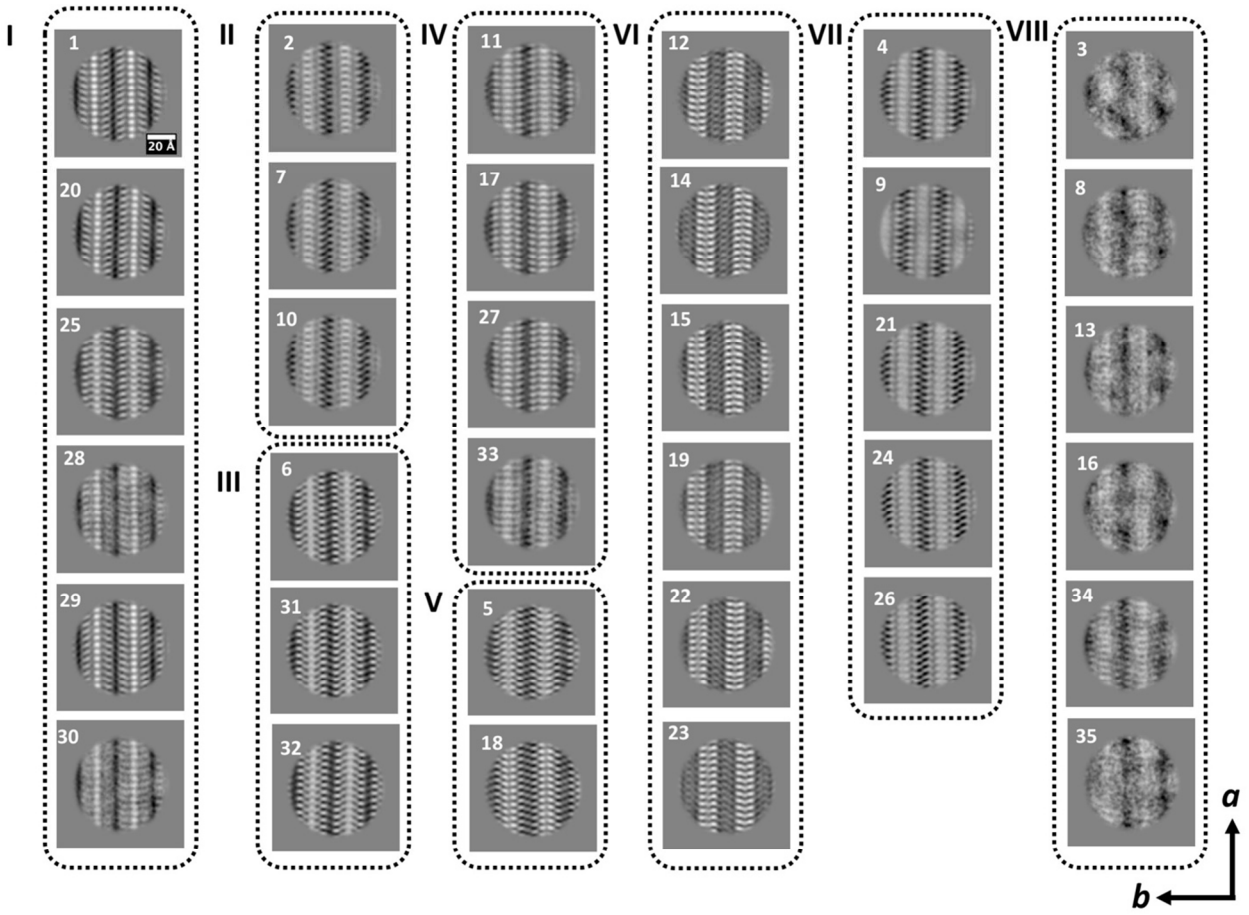

Fig. S5 Grouping class averages. The 35 class averages, derived from ~200,000 boxes from 12 micrographs, can be grouped into 8 distinct motifs. Each motif contains class averages that are mirror images, and there are a number of apparent duplicates.

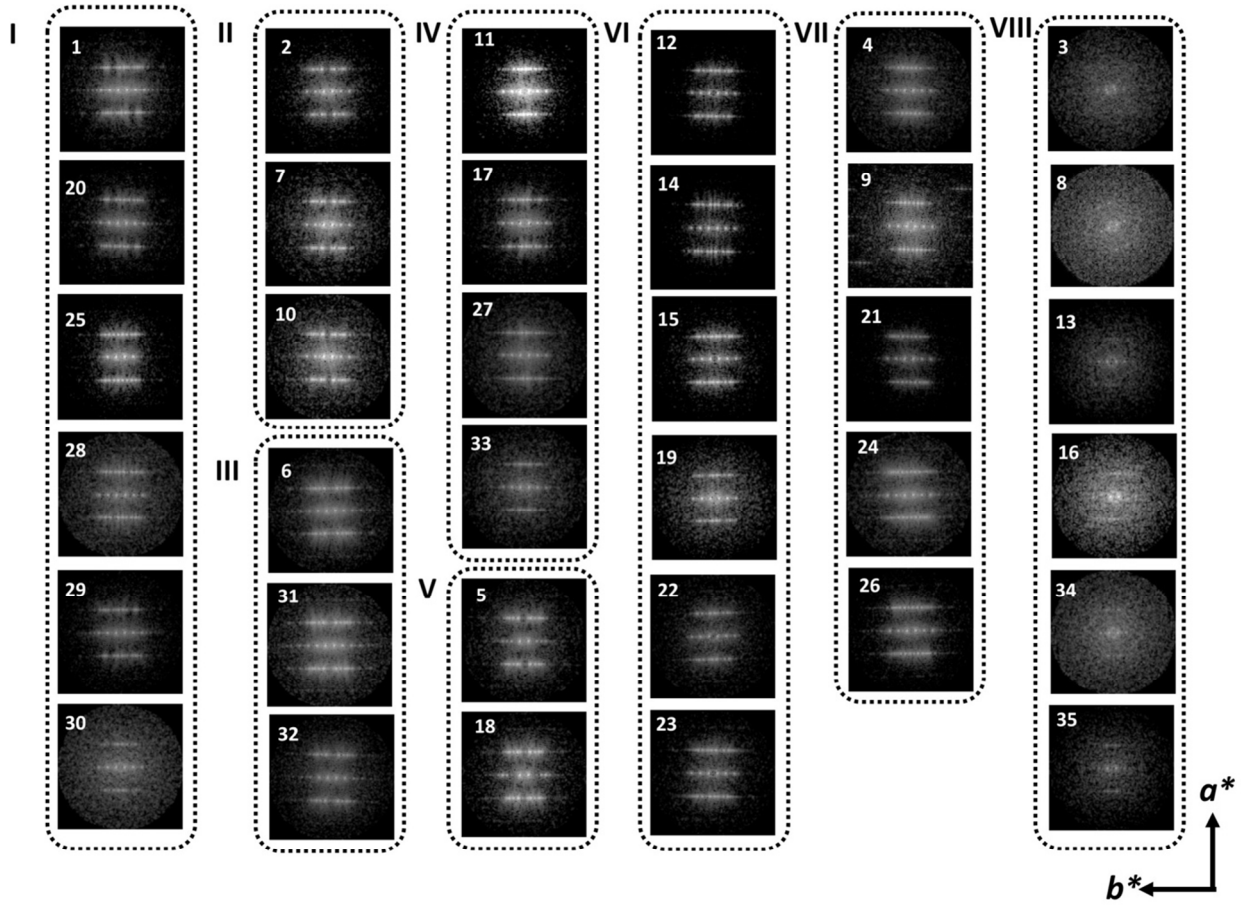

Fig. S6 Fourier transforms of the class averages.

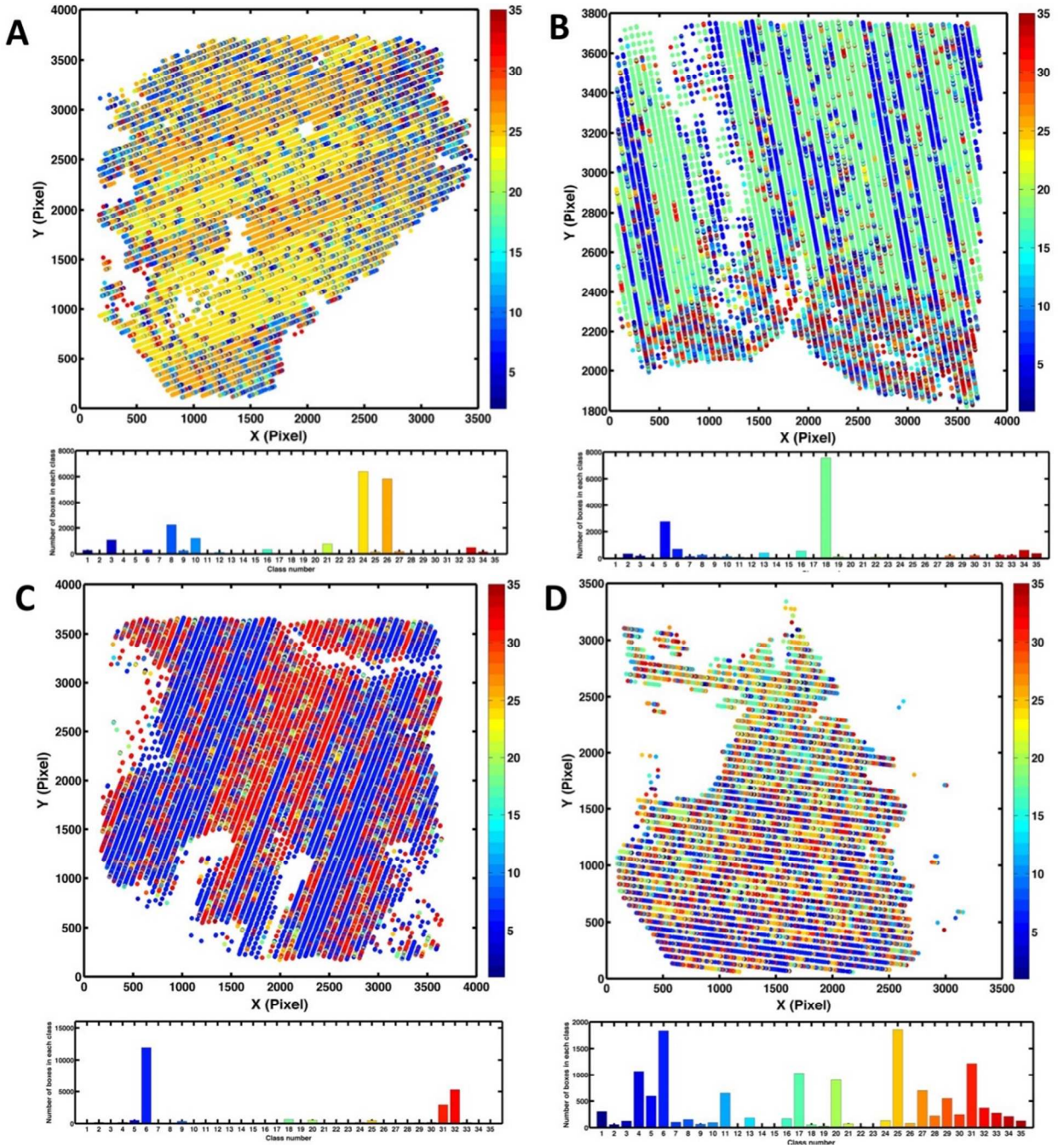

Fig. S7 Mapping distribution of class average members in various micrographs. A, Distribution map shows two major classes, class 24 and class 26, in micrograph. B, Distribution map shows two major classes, class 5 and class 18, in micrograph. C, Distribution map shows two major classes, class 6 and class 32, in micrograph. Class 28 and class 30 are essentially identical to class 29. D, Distribution map shows no dominant class in micrograph; several classes can be observed.

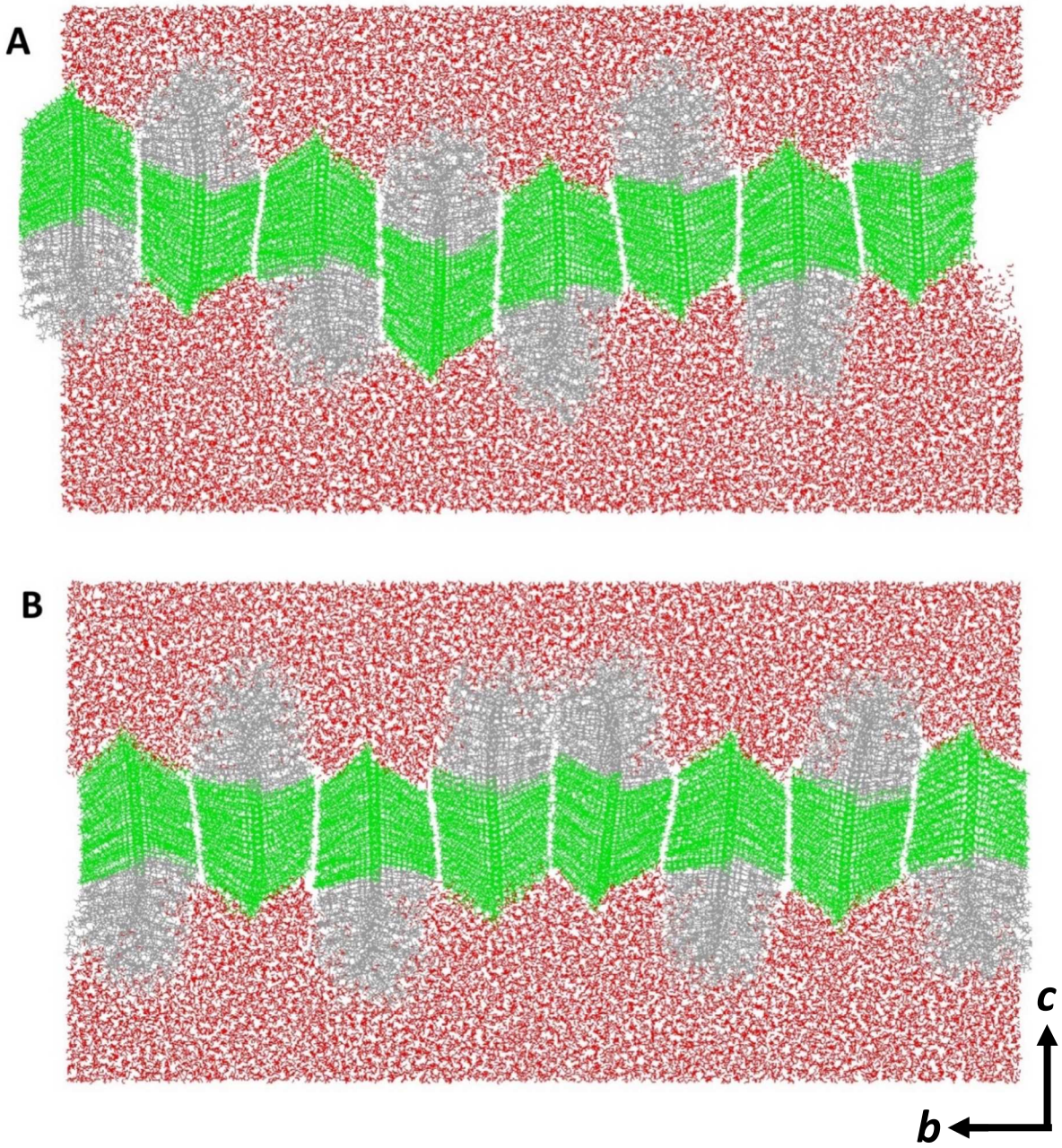

Fig. S8 (In-plane (front) view of relaxed atomic models. Atomic models were relaxed in water box for 200 ns. Red components represent water molecules, green represent pNdc crystals and silver represents non-crystalline pNte blocks. A, without grain boundary. B, with grain boundary in the middle of model. There was no measurable energy penalty for the presence of this defect within the numerical fluctuation. Coordinate files for these models are available as part of the Supplementary Materials.

Other materials

Coordinate files for atomic models of relaxed crystals without and with grain boundaries:

These files are stored in Protein Data Bank (PDB) format which is a standard for files containing atomic coordinates. They can be viewed by using UCSF Chimera or VMD or PyMol.

nanosheet_relaxed_without_grainboundary.pdb
nanosheet_relaxed_with_grainboundary.pdb