Supporting Information

Visualization of Bacterial Microcompartment Facet Assembly Using High-Speed Atomic Force Microscopy

Markus Sutter[†‡#], Matthew Faulkner[§#], Clément Aussignargues[†], Bradley C. Paasch[†], Steve Barrett[⊥], Cheryl A. Kerfeld[†‡‖▽⊤*], and Lu-Ning Liu[§*]

[†] MSU-DOE Plant Research Laboratory, Michigan State University, East Lansing, Michigan, USA.

[‡] Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California, USA.

[§] Institute of Integrative Biology, University of Liverpool, Liverpool L69 7ZB, UK.

[⊥] Department of Physics, University of Liverpool, Liverpool, UK.

[‖] Department of Plant and Microbial Biology, UC Berkeley, Berkeley, California, USA.

[▽] Berkeley Synthetic Biology Institute, Berkeley, California, USA.

[⊤] Department of Biochemistry & Molecular Biology, Michigan State University, East Lansing, Michigan, USA

[#] These authors contributed equally to this work

*** Corresponding Author:**

Lu-Ning Liu, Email: Luning.Liu@liverpool.ac.uk or

Cheryl Kerfeld, Email: ckerfeld@lbl.gov

Materials and Methods

Cloning, Expression, and Protein Purification

E. coli codon optimized full-length HO BMC-H gene (locus tag Hoch_5815) (including stop codon) was synthesized by Genscript and cloned into a pET-11b vector at the BamH1 and Nde1 restriction sites of the multiple cloning site. Protein expression was induced in *E. coli* BL21 (DE3) cells by the addition of 0.45 mM IPTG at mid log phase (OD = 0.6-0.8) and proceeded for 4 hours at 37°C before cells were harvested and stored at -20°C. Frozen cells were resuspended in 30 mL lysis buffer (50 mM Tris-HCl, pH 7.8, 100 mM NaCl, 10 mM $MgCl_2$), 20 µL DNase (25 mg/mL), 20 µL lysozyme (50 mg/mL) per 1 L culture and incubated at room temperature with gentle agitation for 20 minutes followed by lysis using a French press. Upon lysis, soluble and insoluble cell material was separated by centrifugation (20 min at 20,000 x g). The pelleted fraction (containing most of the BMC-H protein) was repeatedly washed by resuspending in 30 mL wash buffer (50 mM Tris-HCl, pH 7.8, 100 mM NaCl, 10 mM $MgCl_2$, 3% (v/v) Triton X-100). A final wash was performed without detergent to remove residual Triton X-100.

Generation of HO BMC-H mutants

K28A and R78A mutants were generated using mutagenic primers and the protocol supplied with the QuikChange Lightning site-directed mutagenesis kit (Agilent Technologies).

Crystallization and Structure Determination

Purified BMC-H was adjusted to 15 mg/ml as measured by absorption at 280 nm (ε = 2,980 $M^{-1}cm^{-1}$) in 10 mM Tris-HCl, pH 7.8, 50 mM NaCl. Initial crystal screening was

conducted using a Phoenix liquid handling robot (Art Robbins Instruments) in sitting-drop vapor diffusion trays. Upon optimization, 6 µL well volume sitting-drop trays were set up with a 3:1 ratio of protein to crystallization buffer (1 M ammonium tartrate, pH 6.5, 4% (v/v) formamide) and equilibrated against a 500 µL reservoir at room temperature. Well drops were adjusted to 24% glycerol prior to harvesting the crystals for cryoprotection. Diffraction data were collected at 1 Å wavelength at 100K on beamline 5.0.2 of the Advanced Light Source (ALS) at the Lawrence Berkeley National Laboratory.

Data were integrated with XDS[1] and scaled with SCALA[2]. The structure was solved by molecular replacement in Phaser[3] with EutM from *E. coli* as the template (PDB ID: 3I6P). The structure was built using multiple cycles of phenix.refine[4] and manual rebuilding in Coot[5]. The data was found to be twinned and the twinning operator –k, -h, -l was applied during refinement with a twin fraction of 0.50. Final R_{work} was 17.0% and R_{free} 20.4% with no Ramachandran outliers. Data collection and refinement statistics are given in Table S1. The coordinates for the HO BMC-H structure were deposited in the Protein Data Bank under ID 5DJB.

Structural alignments were made against BMC-H homologs (monomers) deposited in the PDB using cealign in PyMOL[6]. Surface area measurements of hexamer-hexamer interface interactions were made with ProFunc[7] and the contributing areas of individual buried residues were calculated with PISA[8]. Models were visualized and figures were prepared in PyMOL (The PyMOL Molecular Graphics System, Version 1.7.4 Schrödinger, LLC).

AFM imaging

HO BMC-H protein samples were adsorbed to the mica substrate immersed in 40 µl buffer (50 mM Tris-HCl, pH 7.8, 100 mM NaCl, 10 mM $MgCl_2$). After 5 minutes the sample

was rinsed with and imaged in the same buffer. Series of images with the addition of extra protein were captured before and after the addition of 5 μg of protein into the 40 μl of buffer solution (50 mM Tris-HCl, pH 7.8, 100 mM NaCl, 10 mM $MgCl_2$) on the mica.

For conventional AFM imaging, a Multimode 8 AFM (Bruker), equipped with a 160-μm scanner (J-scanner) and oxide-sharpened Si_3N_4 cantilevers (k = 0.09 $N \cdot m^{-1}$, Olympus) was operated in contact mode in buffer (50 mM Tris-HCl, pH 7.8, 100 mM NaCl, 10 mM $MgCl_2$) at ambient temperature and pressure. Minimal loading forces of approximately 100 picoNewton were applied during AFM imaging, at scan frequencies of 3-4 Hz using optimized feedback parameters and a resolution of 512 by 512 pixels[9-12].

HS-AFM images were captured at between 30 and 60 Hz in buffer in AC mode using a NanoWizard ULTRA speed A (JPK) equipped with an ULTRA Speed 2.8 μm scanner and 'Ultra-Short Cantilever' USC-0.3 MHz probes (0.3 $N \cdot m^{-1}$, Nano World), or in tapping mode using a Dimension Fastscan AFM (Bruker) equipped with 30 μm Icon scanner and ScanAsyst fluid plus probes (0.7 $N \cdot m^{-1}$, Bruker).

Image processing and analysis

Image analysis was initially performed using JPKSPM Data Processing (JPK) and NanoScope Analysis (Bruker). Particle averaging and pair correlation analysis were performed using cross-correlation based Java routines for ImageJ[13]. HS-AFM imaging analysis was performed using ImageSXM (http://www.ImageSXM.org.uk). AFM images were automatically registered using a custom script to align the protein position in each frame. Using a custom macro with user checks, AFM images were processed with 8th order plane fitting, 3rd order flattening and median filtering to reduce noise and XY tilt. The processed images were

thresholded using a custom LUT and converted to a binary format distinguishing protein and non-protein. To analyze the protein dynamics, binary images were subtracted from the previous image in the series to show the differences between frames of AFM data. After excluding the noise, the differences (± standard deviation) were then measured using particle analysis.

Supporting Table

Table S1. Data collection and refinement statistics

	HO BMC-H
Data collection	
Space group	P3
Cell dimensions	
a, *b*, *c* (Å)	69.2, 69.2, 120.4
α, β, γ (°)	90, 90, 120
Resolution (Å)	33.3–1.80 (1.86–1.80)*
R_{merge}	0.150 (0.828)
$I/\sigma I$	10.3 (2.2)
Completeness (%)	99.4 (94.6)
Redundancy	5.5 (5.0)
Refinement	
Resolution (Å)	33.3–1.80
No. reflections	118444
R_{work}/R_{free} (%)	17.0/20.4
No. atoms	
Protein	6443
Ligand/ion	n/a
Water	991
B-factors	16.1
Protein	15.6
Ligand/ion	n/a
Water	18.8
R.m.s deviations	
Bond lengths (Å)	0.003
Bond angles (º)	0.73

1 crystal per dataset

*Highest resolution shell is shown in parenthesis.

Supporting Figures

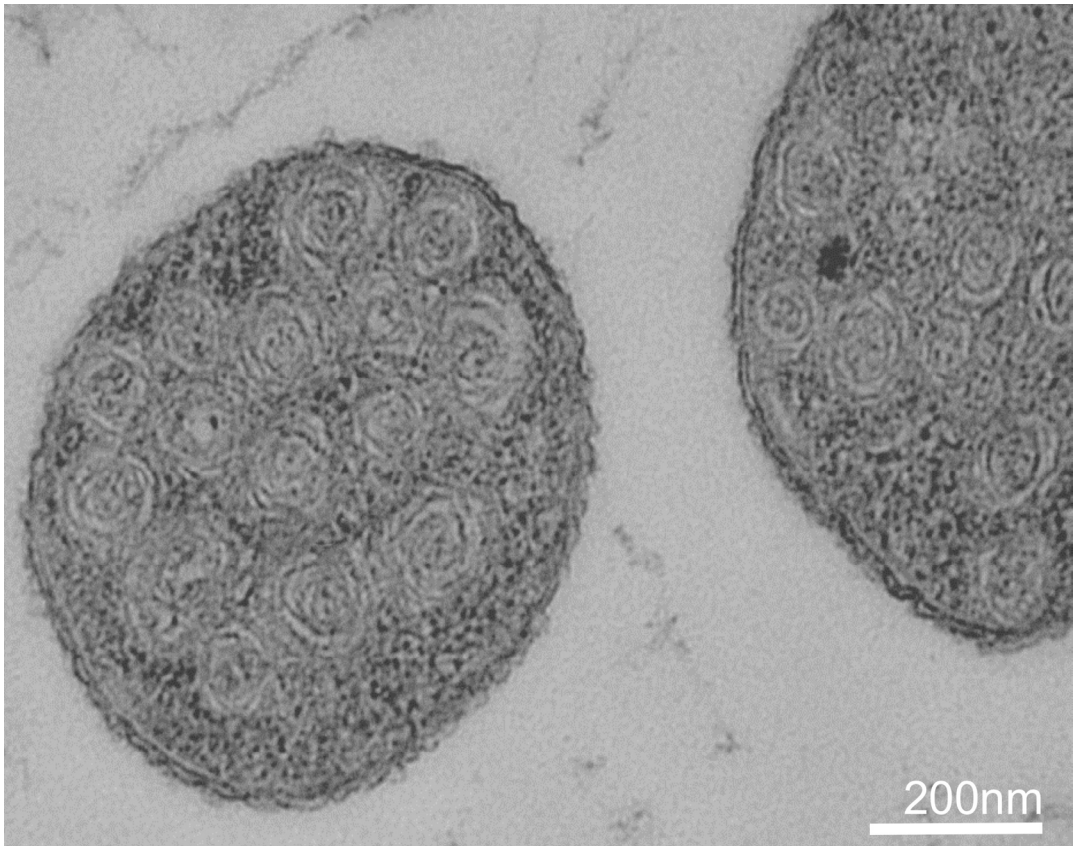

Figure S1. Thin section electron micrograph of *E. coli* overexpressing the HO BMC-H proteins resulting in "swiss rolls" formed by sheets of hexamers.

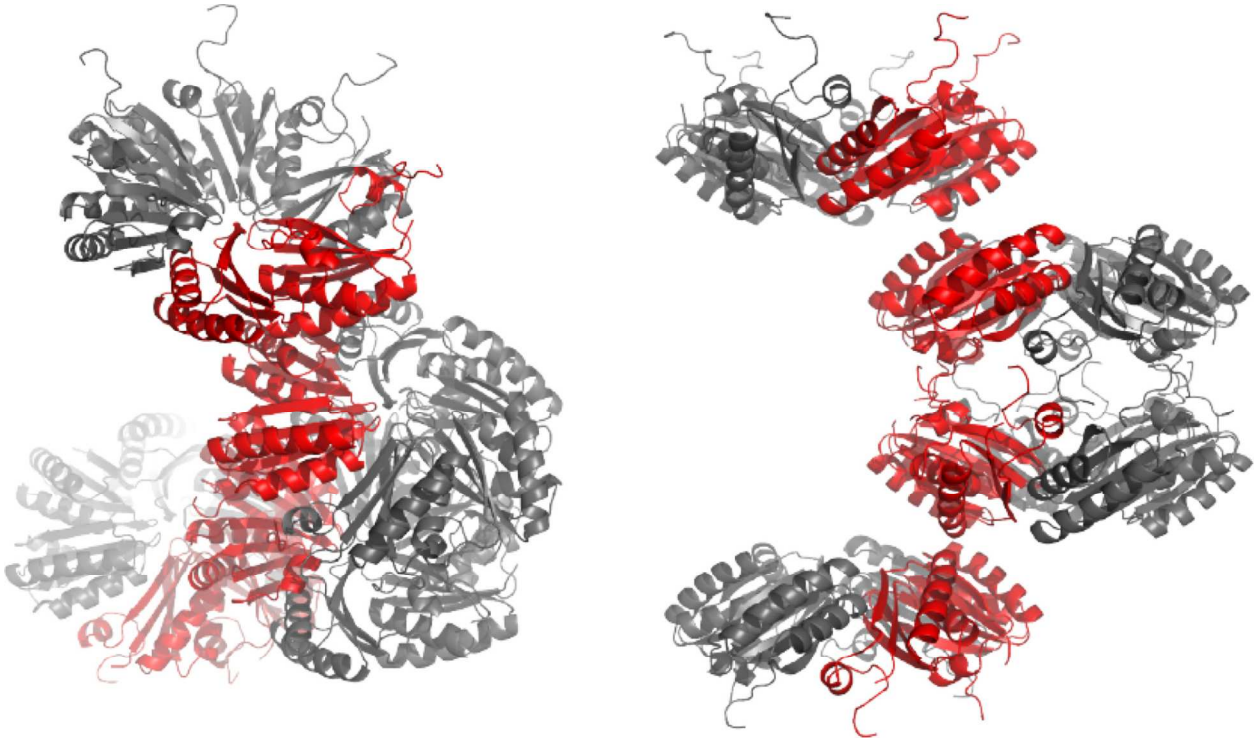

Figure S2. Asymmetric unit and corresponding hexamers. Cartoon representation of HO BMC-H in the asymmetric unit of the crystal (red) and generation of the biological hexamers by applying the space group symmetry (symmetry related chains in grey). Overhead view on the left and side view on the right.

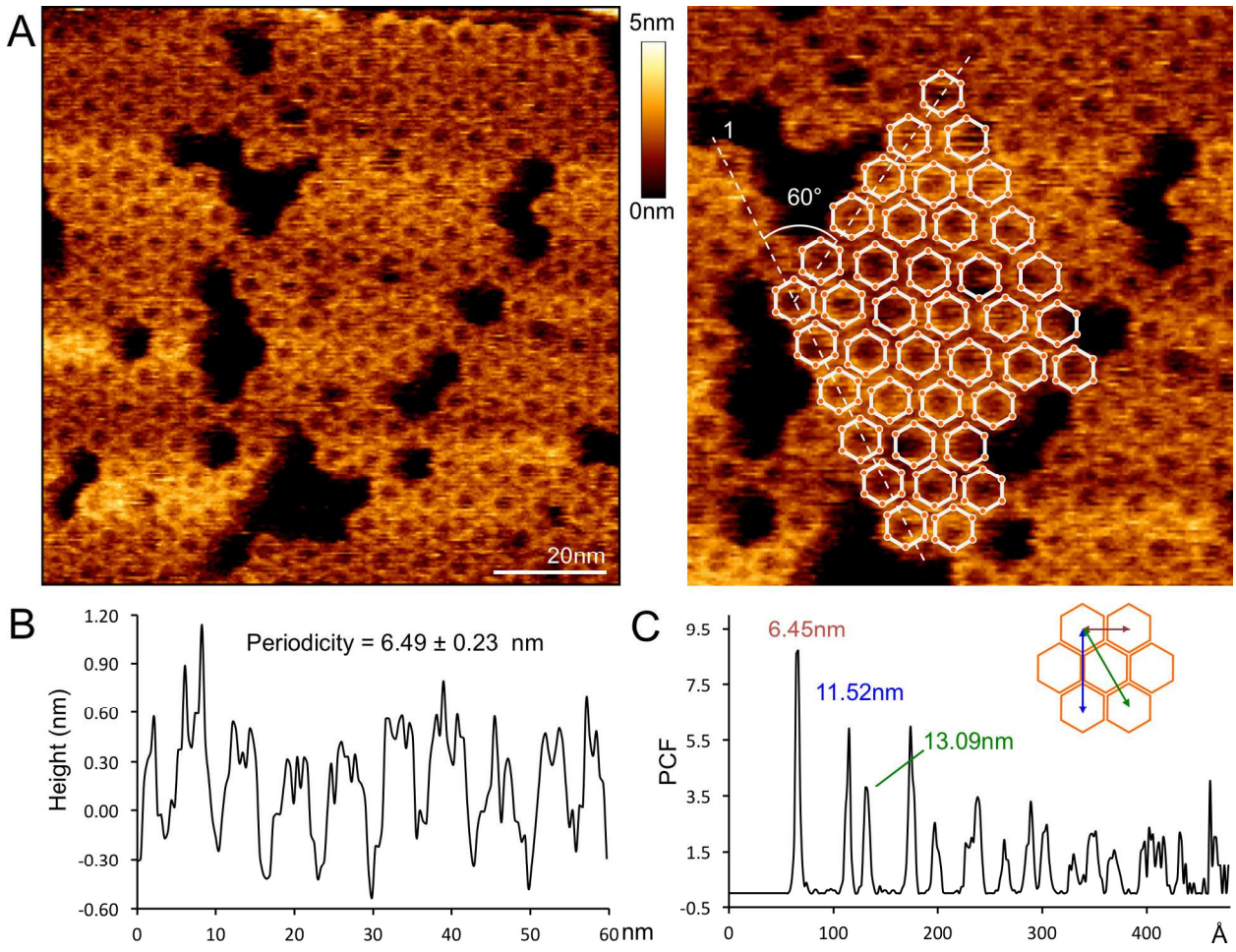

Figure S3. Organization of HO hexamers in the self-assembled patches. (A) High-resolution AFM image showing the structure of individual hexamers and the "honeycomb" packing pattern of hexagons (60° angle) in the self-assembled layers. (B) Height profile along Line 1 in panel (A) indicates the periodicity of hexamer packing (6.49 ± 0.23 nm, n = 10) in the patch. (C) Pair correlation function analysis of the hexamer organization in protein sheets. The closest interacting distances are 6.45 nm, 11.52 nm and 13.09 nm respectively, representing the typical assembly of hexamers as illustrated in the structural model (insert, red, blue and green arrows).

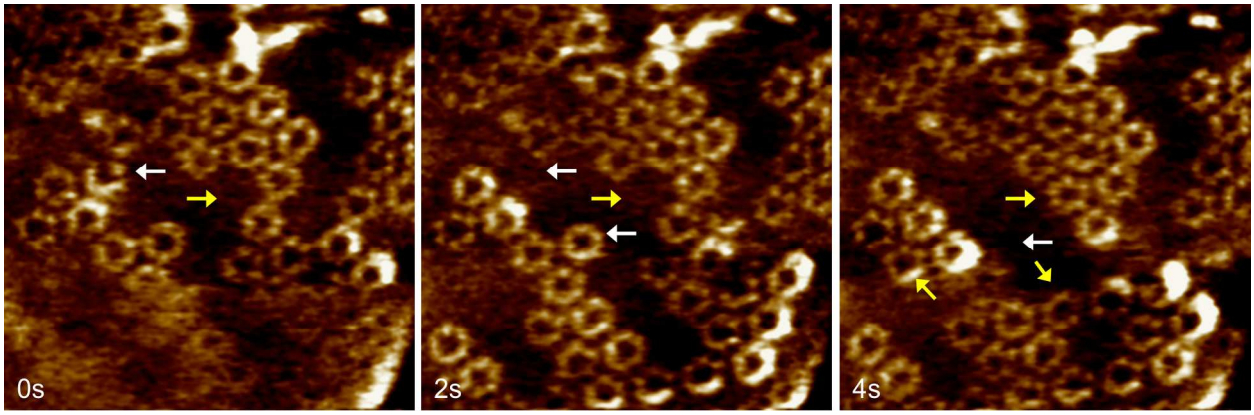

Figure S4. High-speed AFM images of the self-assembled hexamer sheets. Three aligned AFM images (50 × 50 nm) were captured at higher speed (2 s/frame). The assembly and disassembly of HO hexamers in the patches are shown in yellow and white arrows, respectively. Scale bar: 10 nm. See Movie S2.

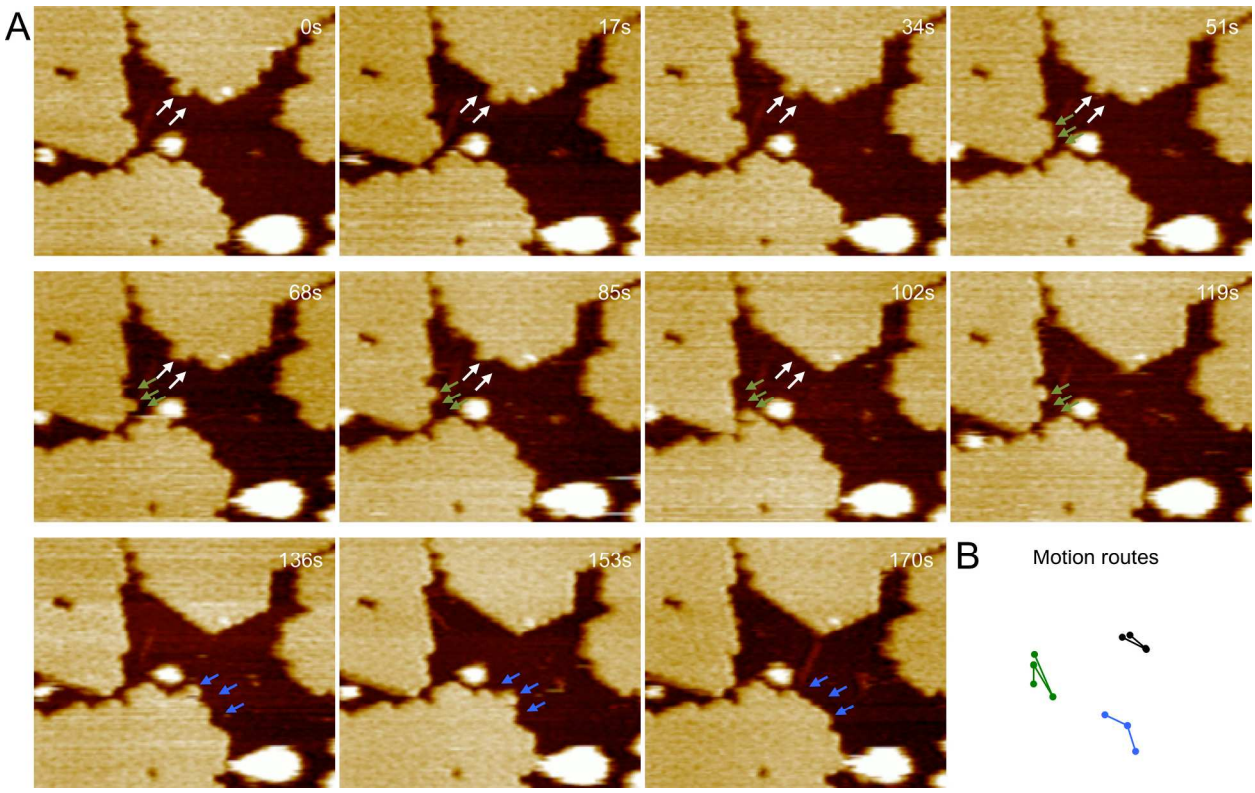

Figure S5. Dynamic associations among hexamers at the non-crystalline patch edges. (A) Eleven aligned AFM frames captured at 17 s/frame. Three individual dynamics events of proteins were visualized (white, green and blue arrows). Scale bar: 20 nm. See Movie S3. (B) The motion routes of the three dynamics events in (A) were shown in black, green and blue respectively, by tracking the hexamers between frames.

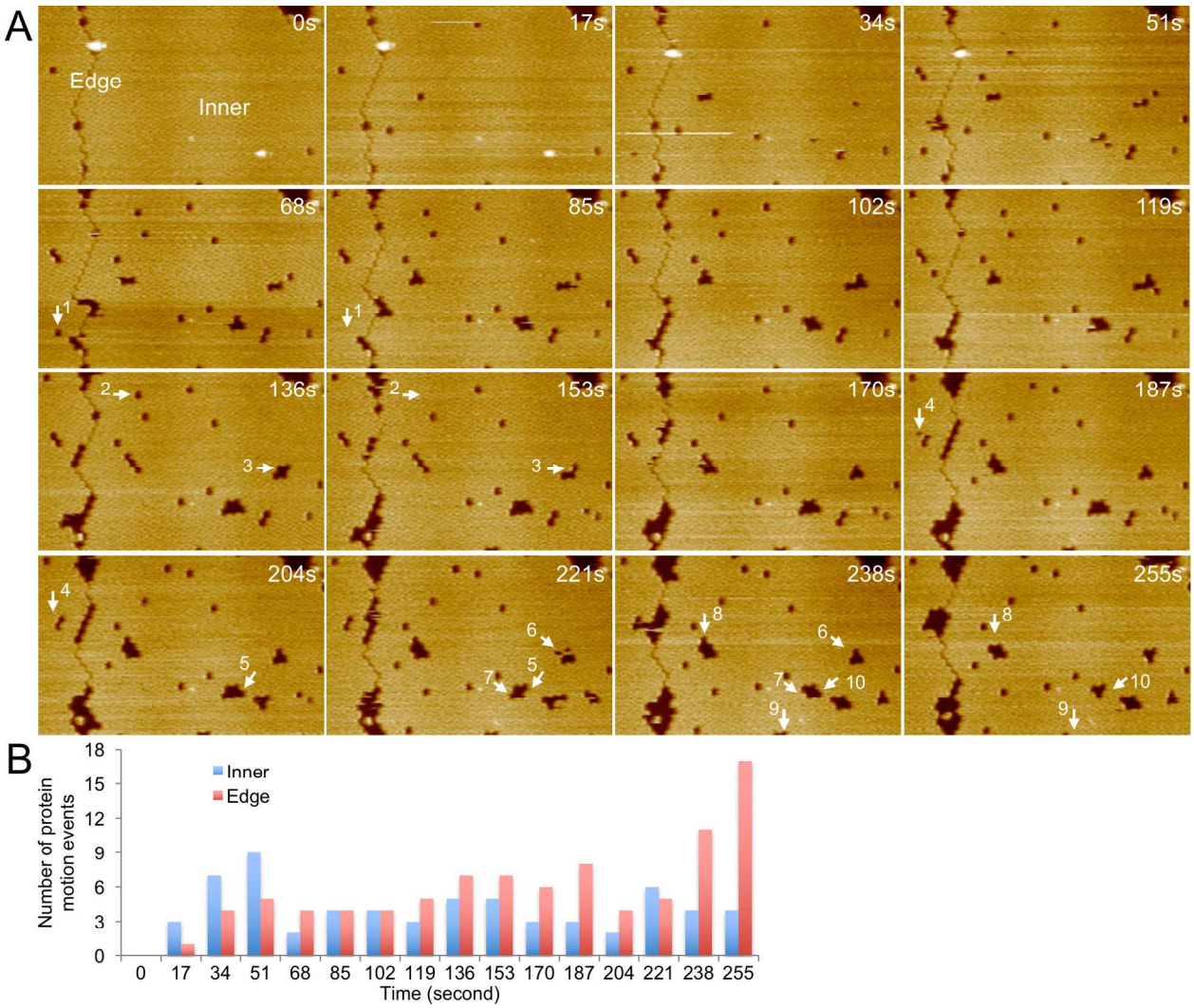

Figure S6. The motions of HO hexamers within the protein patch. (A) Sixteen aligned AFM frames captured at 17 s/frame display both protein assembly and disassembly events. Arrows represent the assembly events in which diffusing hexamers were trapped by other hexamers within the protein sheet. Scale bar: 20 nm. See Movie S4. (B) The number of protein motion events that occurred at both the patch interior and edge per high-speed AFM frame as a function of time.

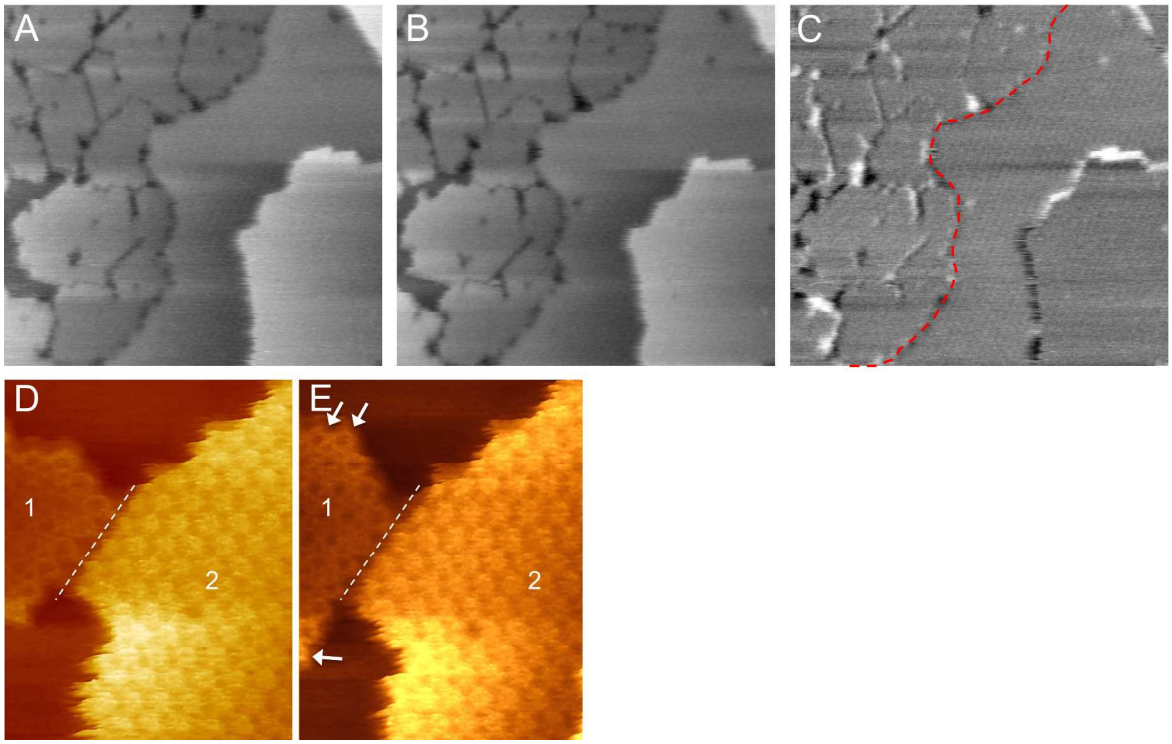

Figure S7. The assembly of protein patches depends on the protein orientation. (A) AFM topograph of protein patches in the same region at 0 min. (B) AFM topograph of protein patches in the same region at 3 min. (C) Difference AFM image by comparing the two AFM topographs (A) and (B). Black and white areas in the difference AFM image indicate the regions where protein dynamics were detected. The red line represents the interface between sheets with different orientations. The results revealed that the dynamic events were only observed at interfaces between sheets with the same orientations, rather than at the interface between sheets with different orientations. It suggested that the same orientation is essential to ensure the protein sheet development. Scale bar: 100 nm. (D, E) The interface of protein patches (dashed lines) with distinct orientations did not merge during three minutes of AFM imaging. This phenomenon was observed in multiple distinct locations (n = 16). Areas 1 and 2 represent the protein sheets with concave and convex faces exposed to the AFM probe, respectively. Arrows indicate the protein dynamics events captured. Scale bar: 10 nm.

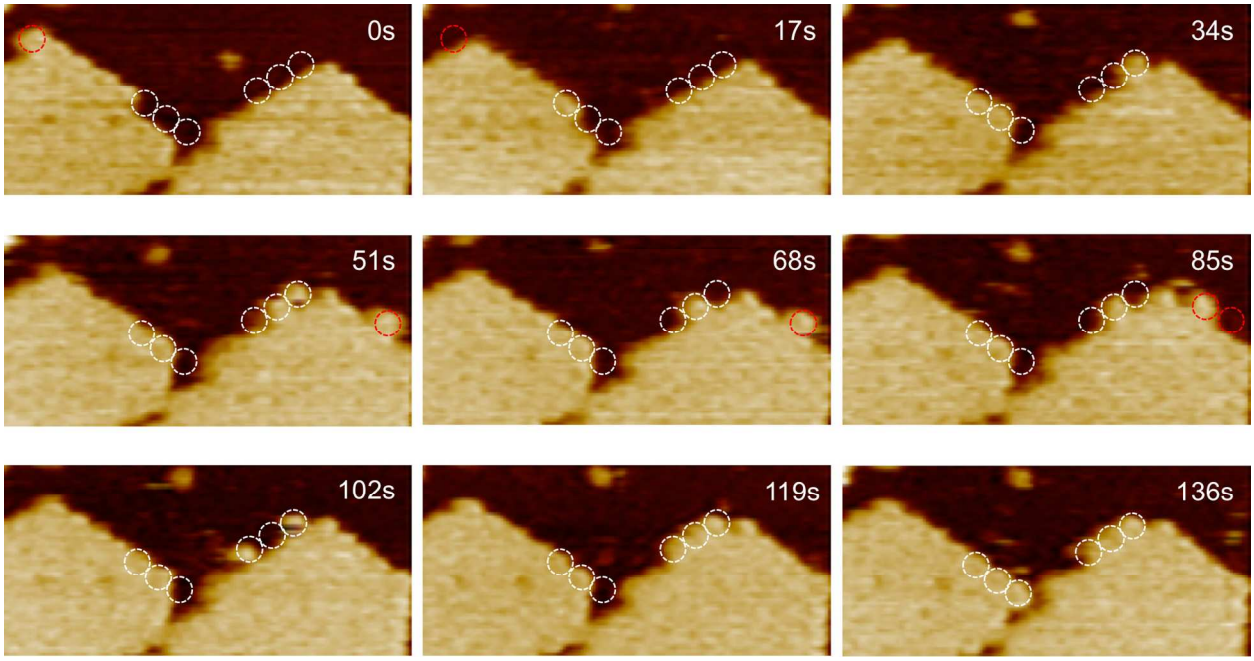

Figure S8. Formation of new edges of hexamer sheets by the assembly of individual hexamers. Nine aligned AFM frames were captured at 17 s/frame. The association of hexamers to the sheet edges and their disassociation from the edges are shown in white and red circles, respectively. Scale bar: 20 nm. See Movie S5.

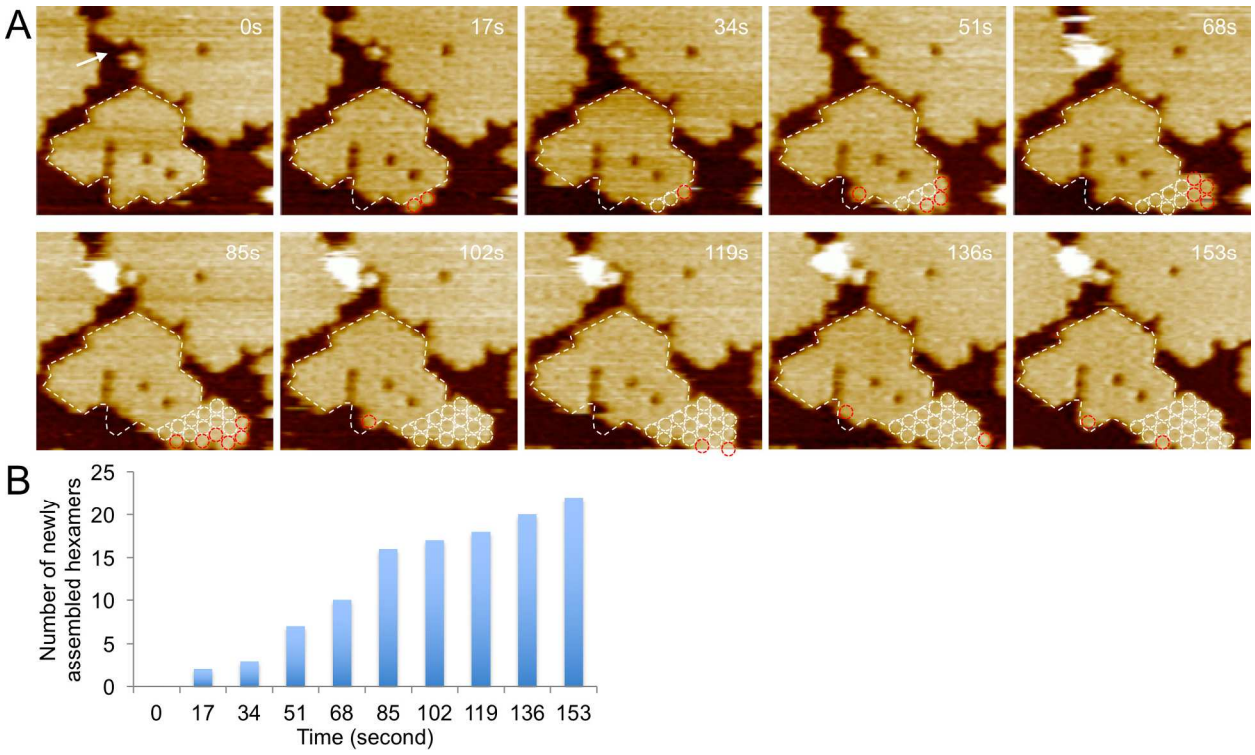

Figure S9. Progression of shell patch formation by the assembly of individual hexamers. (A) Ten aligned AFM frames captured at 17 s/frame. The dashed line indicates the original protein patch at the beginning of high-speed AFM imaging. Red circles in the frames indicate the newly assembled proteins, compared to previous frame, whereas white circles show the settled proteins during the assembly process. Scale bar: 20 nm. See Movie S6. (B) The number of self-assembled hexamers observed in the patch per high-speed AFM frame as a function of time.

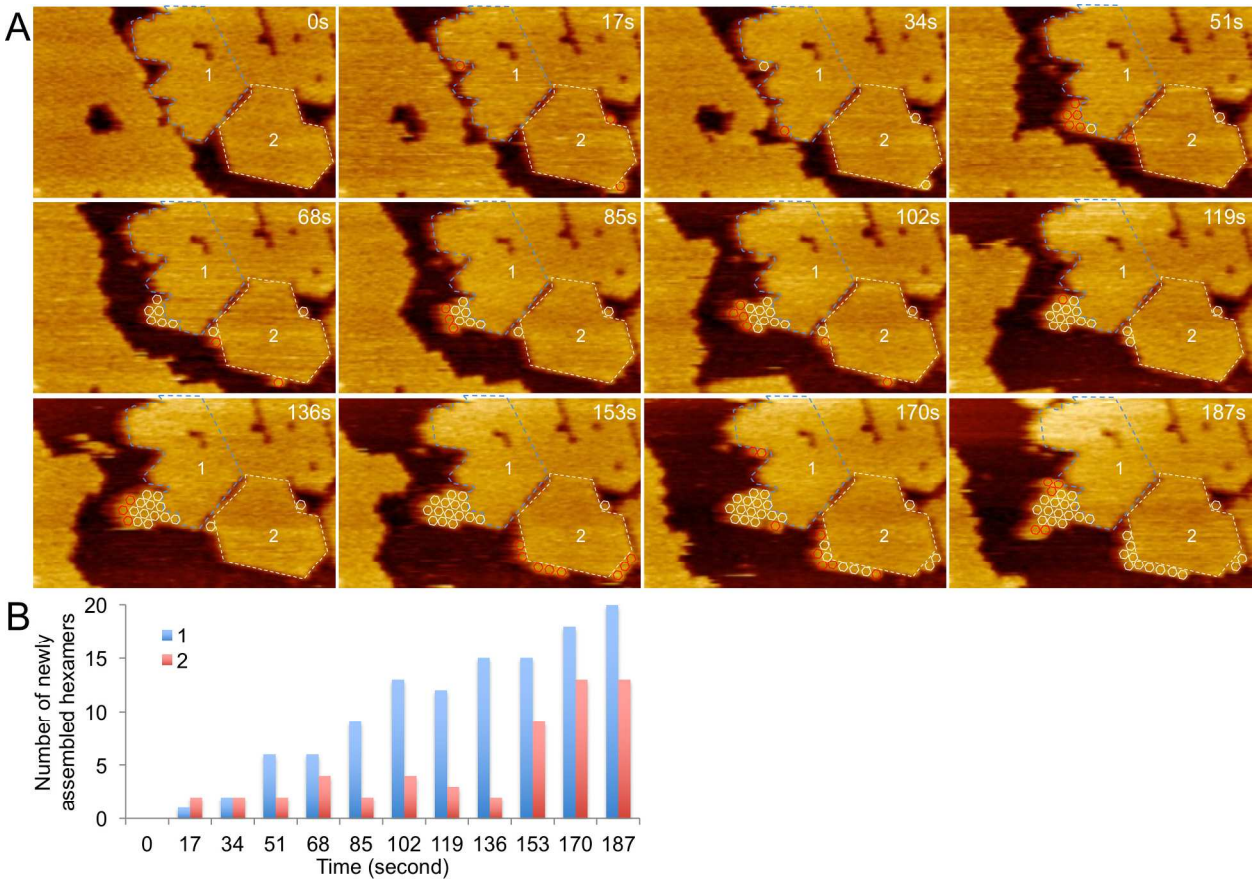

Figure S10. Assembly and disassembly dynamics of hexamer sheets. (A) Twelve aligned AFM frames captured at 17 s/frame. Dashed lines 1 and 2 indicate two protein patches that are growing, whereas the large patch on the left side in the frames shows disassociation events. Red circles in the frames indicate the newly assembled proteins, compared to previous frame, whereas white circles show the settled proteins during the assembly. Scale bar: 20 nm. See Movie S7. (B) The numbers of newly assembled hexamers in patch 1 and 2 per high-speed AFM frame as a function of time.

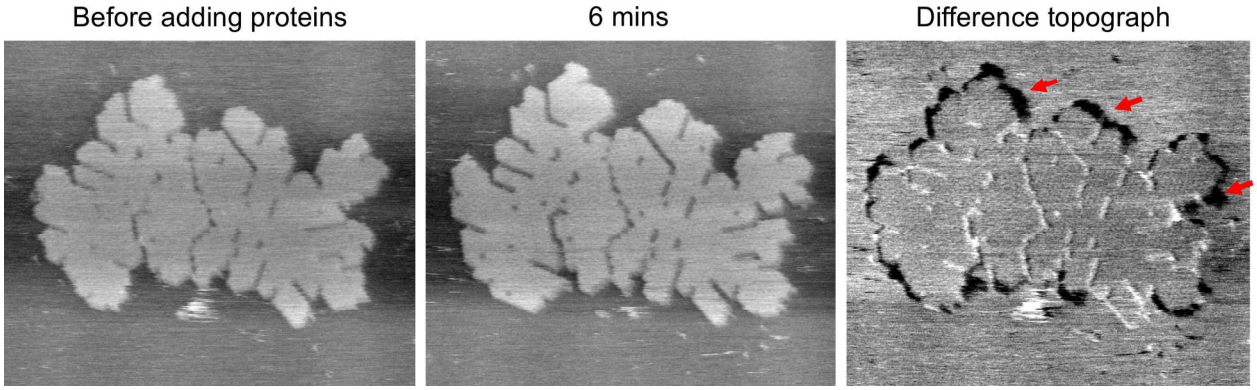

Figure S11. The synthesis of protein patches triggered by the addition of extra hexamers into the AFM imaging buffer. Left, AFM topograph of the large protein patches before adding extra hexamers. Middle, AFM topograph of the same protein patches captured at 6 min after adding extra hexamers. After the addition of proteins, a series of AFM images were captured at several locations for 6 minutes, all showing a similar growth in patch size (n = 9). Right, difference AFM image by comparing panel left and middle. Black areas (shown in arrows) indicate the newly developed patch areas. Scale bar: 20 nm.

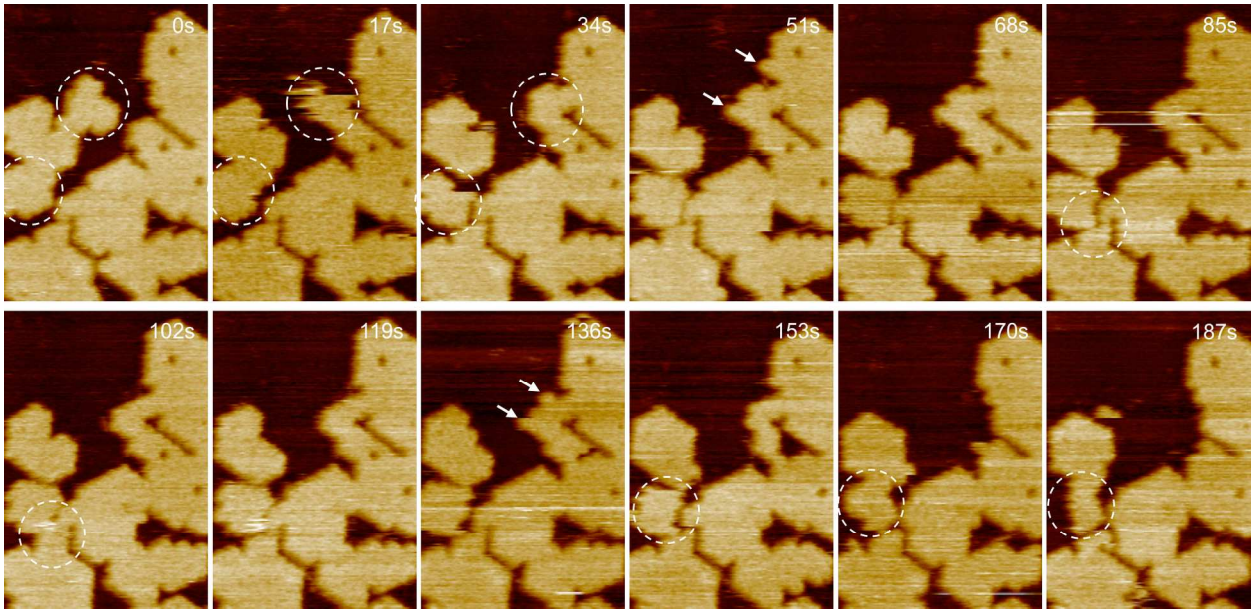

Figure S12. Development of shell patch by the merging with other patches. Twelve aligned AFM frames were captured at 17 s/frame. Circles indicate the merging and dynamic large protein patches during high-speed AFM imaging. Arrows show the dynamics individual proteins during high-speed AFM imaging. Scale bar: 20 nm. See Movie S8.

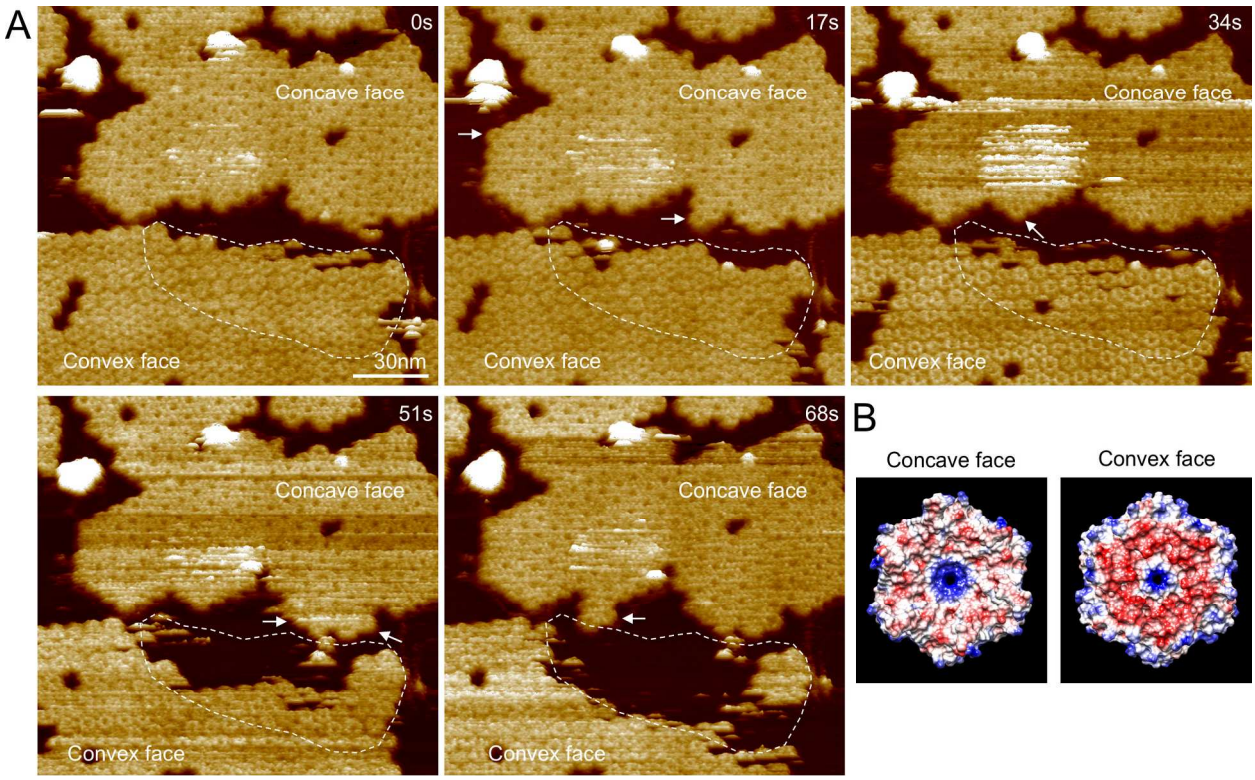

Figure S13. Different dynamic features of hexamer self-assemblies when concave and convex faces are attached to the mica substrate. (A) Five aligned AFM frames captured at 17 s/frame, showing that the two patches with distinct orientations present different dynamic features. The protein patch with convex face up displays greater protein dynamics (dashed circle), whereas the other sheet with concave face up is relatively stable (arrows indicate a few single-protein motion events). (B) The electrostatic properties of the hexamer, showing that the concave face is relatively less uniformly polar than the convex face.

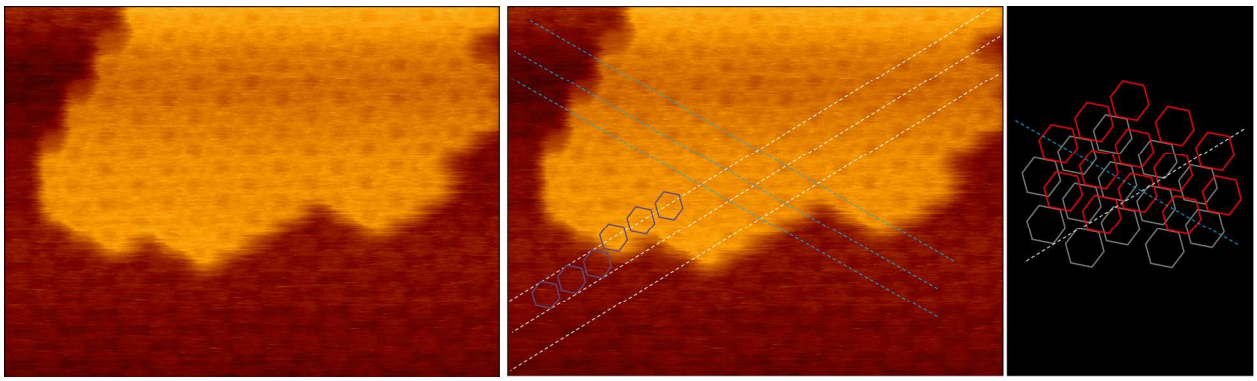

Figure S14. Stacking pattern of K28A hexamer double sheets characterized by AFM imaging. White lines show the directions where the protein centers in the upper layer are facing the interfaces of hexamers in the bottom layer, whereas blue lines indicating the centers of proteins in both layers are in identical line. Hexamers in upper layer are shown in red, those in the bottom layer are shown in grey. The data indicates that there is no direct pore superpositioning between the double shell sheets. Scale bar: 20 nm.

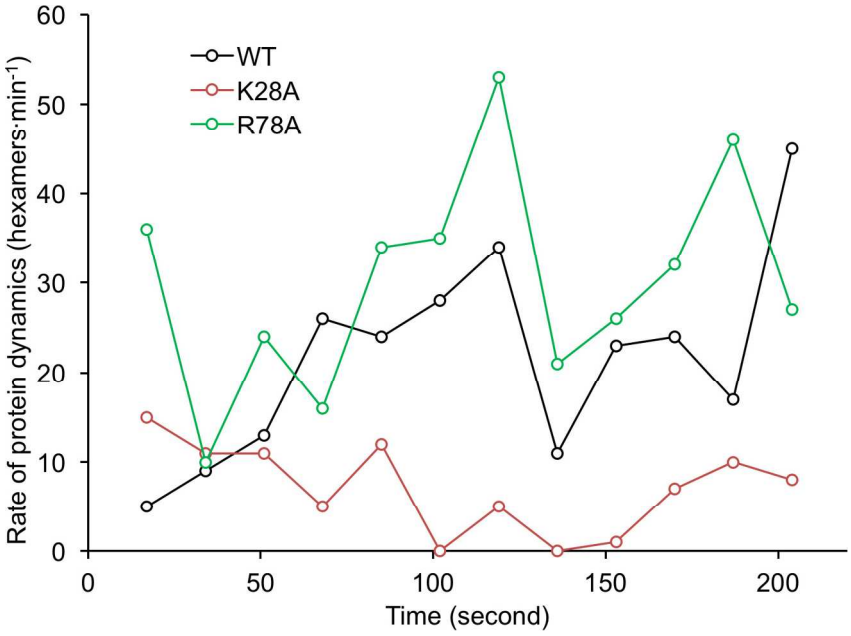

Figure S15. The rates of protein dynamics in the WT, K28A and R78A hexamers per HS-AFM frame as a function of time. See Figure 4 where the data were normalized to correct for differences in frame capture time, the scan area which have diverse ratios of protein to mica and the scan size, relative to WT dynamic events.

Supporting Movies

Movie S1. Dynamic assembly of the hexamer sheet. The data was taken at 17 s/frame and shown at 1 s/frame. See Figure 3.

Movie S2. Real-time AFM imaging of the dynamics of the hexamer sheet. The data was taken at 2 s/frame and shown at 2 s/frame. See Figure S4.

Movie S3. Dynamic associations among hexamers at the non-crystalline patch edges. The data was taken at 17 s/frame and shown at 0.5 s/frame. See Figure S5.

Movie S4. The motions of hexamers within the protein patch. The data was taken at 17 s/frame and shown at 0.5 s/frame. See Figure S6.

Movie S5. Formation of new edges of protein sheets by the assembly of individual hexamers. The data was taken at 17 s/frame and shown at 0.5 s/frame. See Figure S8.

Movie S6. The formation process of the hexamer patch by the assembly of individual hexamers. The data was taken at 17 s/frame and shown at 0.5 s/frame. See Figure S9.

Movie S7. Assembly and disassembly dynamics of hexamer patches. The data was taken at 17 s/frame and shown at 0.5 s/frame. See Figure S10.

Movie S8. Development of the hexamer patch by the merging with other patches. The data was taken at 17 s/frame and shown at 0.5 s/frame. See Figure S12.

Movie S9. Real-time dynamics of hexamer assemblies visualized using higher-speed AFM imaging (1 s/frame). The data was shown at 1 s/frame. The scanning area is 100 × 100 nm.

References:

(1) Kabsch, W. *Acta Crystallogr. Sect. D-Biol. Crystallogr.* **2010,** *66*, 125-132.

(2) Bailey, S. *Acta Crystallogr. Sect. D-Biol. Crystallogr.* **1994,** *50*, 760-763.

(3) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. *J. Appl. Crystallogr.* **2007,** *40* (Pt 4), 658-674.

(4) Afonine, P. V.; Grosse-Kunstleve, R. W.; Echols, N.; Headd, J. J.; Moriarty, N. W.; Mustyakimov, M.; Terwilliger, T. C.; Urzhumtsev, A.; Zwart, P. H.; Adams, P. D. *Acta Crystallogr. Sect. D-Biol. Crystallogr.* **2012,** *68*, 352-367.

(5) Emsley, P.; Cowtan, K. *Acta Crystallogr. Sect. D-Biol. Crystallogr.* **2004,** *60*, 2126-2132.

(6) Shindyalov, I. N.; Bourne, P. E. *Protein Eng.* **1998,** *11* (9), 739-747.

(7) Laskowski, R. A.; Watson, J. D.; Thornton, J. M. *Nucleic Acids Res.* **2005,** *33*, W89-W93.

(8) Krissinel, E.; Henrick, K. *J. Mol. Biol.* **2007,** *372* (3), 774-797.

(9) Scheuring, S.; Nevo, R.; Liu, L. N.; Mangenot, S.; Charuvi, D.; Boudier, T.; Prima, V.; Hubert, P.; Sturgis, J. N.; Reich, Z. *Biochim. Biophys. Acta* **2014,** *1837* (8), 1263-1270.

(10) Liu, L. N.; Sturgis, J. N.; Scheuring, S. *J. Struct. Biol.* **2011,** *173* (1), 138-45.

(11) Liu, L. N.; Duquesne, K.; Oesterhelt, F.; Sturgis, J. N.; Scheuring, S. *Proc. Natl. Acad. Sci. U. S. A.* **2011,** *108* (23), 9455-9.

(12) Liu, L. N.; Duquesne, K.; Sturgis, J. N.; Scheuring, S. *J. Mol. Biol.* **2009,** *393* (1), 27-35.

(13) Fechner, P.; Boudier, T.; Mangenot, S.; Jaroslawski, S.; Sturgis, J. N.; Scheuring, S. *Biophys. J.* **2009,** *96* (9), 3822-31.