Supporting Information

Epidermal cell surface structure and chitin-protein co-assembly determine fiber architecture in the Locust cuticle

Authors: Sanja Sviben, Oliver Spaeker, Mathieu Bennet, Marie Albéric, Jan-Henning Dirks, Bernard Moussian, Peter Fratzl, Luca Bertinetti and Yael Politi

Methods

FIB/SEM Image analysis. Image processing and segmentation of data cubes were performed as follows: first, images belonging to each stack were registered taking advantage of the Fourier shift theorem. We calculated the displacement vector of each image to the previous one (taking the first image of the stack as a reference), $\boldsymbol{d}_i$, using the phaseCorrelate function of OpenCV (v 2.4.7). Then, the shift of each image with respect to the first one, $\boldsymbol{D}_i$, was calculated as $\boldsymbol{D}_i = \sum_i \boldsymbol{d}_i$. In many cases, curtaining artifacts were visible in the stacks, giving rise to vertical contrast modulation (stripes). To correct for this, we followed the approach described in a recent paper [1], where the depth of the wavelet transform, the wavelet family and the width of the Gaussian filter were visually optimized to obtain a corrected stack with the least contrast and information loss and the largest stripe removal. Finally, to improve the signal to noise ratio, the image stacks were denoised using the skimage.restoration non-local denoising filters of the scikit-image python library (v. 0.14.0). Again, the filtering algorithms and parameters were chosen by visually optimizing the corrected images to obtain the largest noise removal with the lowest information loss.

Segmentation of structures of interest was performed semi automatically using a combination of Amira 3D (FEI, USA) and a custom python code implementing state of the art 2D and 3D machine learning methods based on fully convolutional networks using Keras and GPU accelerated

Tensorflow. For the 2D segmentation between 20 and 30 representative 256 x 256 regions of the image stack were sparsely annotated using Amira 3D. The images and ground truth labels were augmented using the Keras ImageDataGenerator class by allowing a rotation and shearing. The augmented images were used to train a U-net like network which consists, for the contracting path, of the repeated application of 3 x 3 convolutions (padding = same) followed by a ReLU activation and 2 x 2 max pooling for downsampling. For the expanding path, every step consists of a 2 x 2 upconvolution concatenated with the corresponding layer of the contracting path and the repeated application of 3 x 3 convolutions followed by a ReLU activation. A dropout of 0.25 after each max pooling and upconvolution operations was applied. A total of 4 max pooling and upconvolutions were used (depth of the network = 4). On the final layer, a sigmoidal activation was employed.

The input images and the corresponding ground truth labels were used to train the network on a workstation with a single Nvidia Quadro M5000 GPU with 8GB of memory with the adam optimization algorithm, a learning rate of 10^{-4} and binary cross entropy as a loss function. Once the loss dropped below 1%, the training was stopped and the network was used to infer the labels on the whole image stack. Although the network would predict the microvilli with a high accuracy, many of the images would contain a small number of predicted labels not coinciding with the objects of interests. These were manually deleted from the image stack using Amira 3D.

For the 3D segmentation, a similar workflow was used: first, 2 representative sub volumes (512 x 512 x 512) px^3 of the image stack were selected and densely manually annotated using Amira 3D. From these 2 volumes, we extracted randomly 30 smaller volumes (92 x 92 x 92) px^3. Each of these 30 volumes was randomly transformed 3 times to generate 3 ‘augmented’ volumes using one of the following transformations: identity, reflection with respect to the x or y axis, 10% zoom in or out and 10% dilation or compression on a random axis (x, y or z). The generated augmented volumes were randomly cut to a final size of (64 x 64 x 64) px^3 not containing boundary artifacts associated with the transformation (like for zooming out and compressing). The final 90 volumes (64 x 64 x 64) px^3 were used to train a 3D network with an architecture slightly modified with respect to the one used for the

2D segmentation. The basic building blocks of the network remained unchanged, but we used convolution kernels of (5 x 5 x 5) px^3, a network's depth of 2 instead of 4 and the layers in the contracting and expanding path were not concatenated.

The network was trained with the same hardware and software implementation as in two dimensions. In this case, the training was stopped once the loss reached a value of 0.05 and the network used to infer the whole volume. Because of the limitation of the GPU memory size (8GB), the inference was run by tiling the whole volume in tiles of (128 x 128 x 128) px^3 overlapping for half of their size. For each tile, only the inferred labels in the central (64 x 64 x 64) px^3 were used to generate the final predictions on the original image stack. Because the training sub-volumes were taken from the image stack the inference was run on, the quality of the predictions in those regions was superior to the others. Nevertheless, large portions of the stack neighboring that were selected for the training exhibited also accurate predictions. Also, in this case, the predicted labels were checked with Amira 3D and the wrongly predicted labels were manually deleted. The Drishti volume exploration and representation tool (v.2.6.5) was used for volume rendering [2]. Orientation maps and distributions were generated with Orientation J plugin of Fiji (http://bigwww.epfl.ch/demo/orientation/). Three-dimensional Fourier transforms of the stack of the labelled microvilli were calculated using scipy (v 1.3.0). Azimuthal averages of the 2D Fourier transforms were calculated using a custom script in python.

SAXS/XRD and XRF mapping. X-ray fluorescence and scattering experiments on *L. migratoria* tibia sections were performed at the microXAS – X05LA beamline at the Swiss Light Source (SLS) synchrotron radiation facility (Paul Scherrer Institute, Villigen, Switzerland). The X-ray beam was defined to 15.2 keV (0.816 Å) using a Si(111) monochromator and focused to (1 x 1) μm^2 using KB mirrors. XRD data was obtained using a 2D Dectris Eiger 9M detector ((2070 x 2167) px^2) behind the sample, while a single element Si(Li) detector (Ketek, Munich, Germany) was used for XRF measurements located perpendicular to the beam main axis. Samples were mounted at a sample-to-detector distance of around 250 mm on a rotating x, y, z-stage. XRD and XRF signals were recorded

simultaneously with an integration time of 1 s. The scanning step size in y direction was 1 µm, in x direction 0.5 µm. Quartz powder was used to calibrate the sample to detector distance.

XRD data from 24 h day sample was obtained at the µSpot beamline, station of the l-Spot beam line at the synchrotron BESSY II (Helmholtz Center, Berlin, Germany). The X-ray energy of 15 keV was defined by a multilayered monochromator. The incident X-ray beam was focused on the sample by a toroidal mirror, and the final beam size was defined by a pinhole of 10 µm diameter placed in front of the sample. WAXS data were collected using a large-area 2D detector (MarMosaic 225, Mar USA Evanston, USA), situated approximately 300 mm behind the sample.

Calibration, integration and data analysis of the 2D diffraction patterns were performed using the software DPDAK [3]. DPDAK was also used for background removal and peak-fitting the XRD patterns. For each peak, several fits were performed where the fitting parameters (q position, intensity and peak width) were first varied and then fixed. We used Lorentzian peak-shape for fitting in q space (neglecting micro-strain contribution) and Gaussian peak-shape for fitting in azimuthal space. OriginPro 2015 software was used for averaging of 1D diffraction patterns. 2D heatmaps were plotted using the matplotlib library in Spyder 3.3.2 from the Anaconda package[4].

Supplementary figures

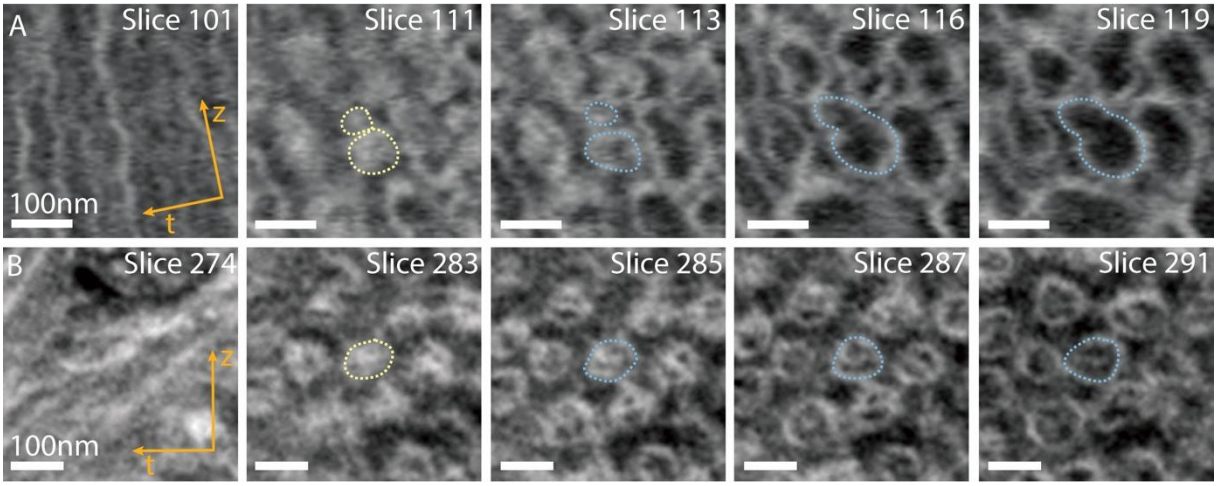

Figure S1. **FIB-SEM slices of cryo-fixed locust tibiae from fifth instar animals sampled 2 days after ecdysis showing microvilli structures**. (A) *Night* (B) *Day*. The plaques are indicated with a light yellow dotted line and the microvilli base with a light blue dotted line. The slices are presented clockwise from the cuticle to the cells, i.e. along *(r)* direction.

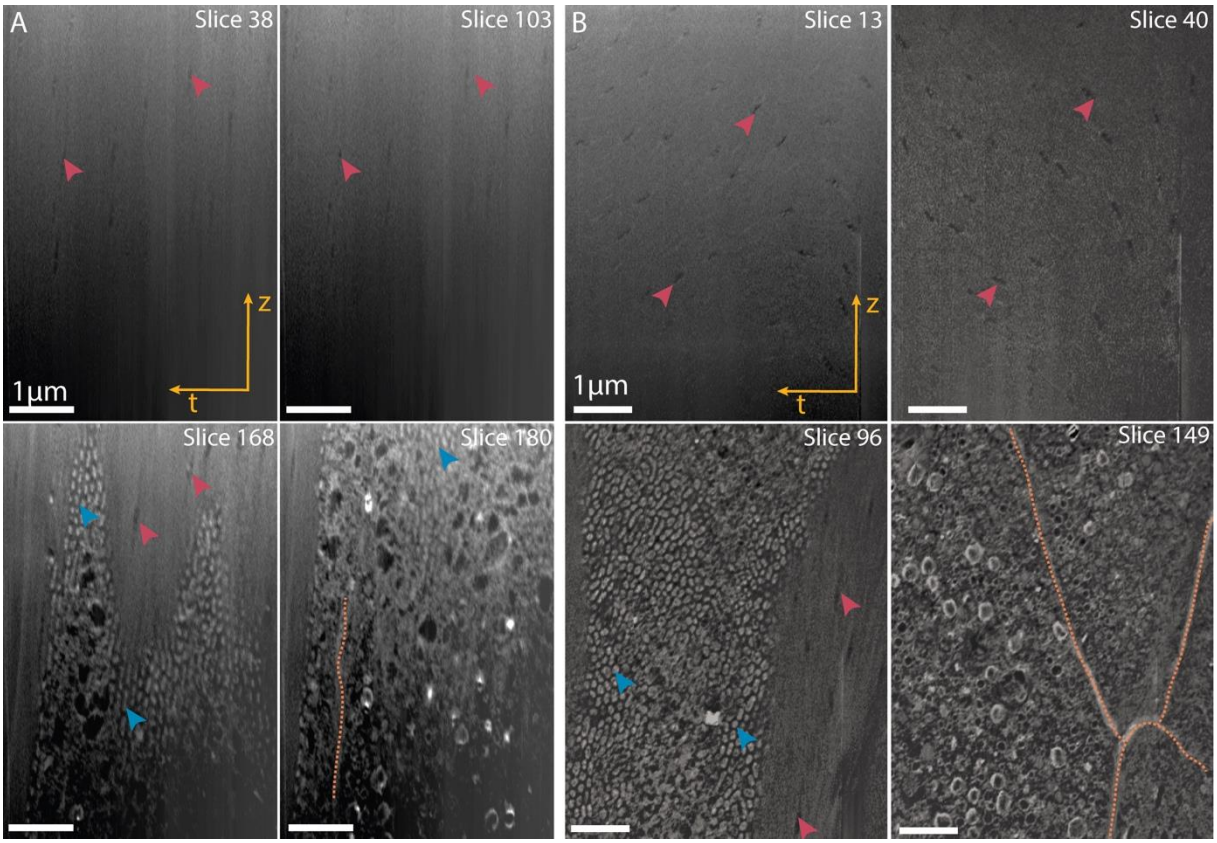

Figure S2. FIB/SEM slices of chemically fixed locust tibiae from fifth instar animals sampled 2 days after ecdysis. (A) *Night* sample. (B) *Day* sample. Marked: Pore canals in the cuticle (magenta arrowheads), cell borders (dotted orange lines), apical cell protrusions (blue arrowheads). The slices are presented clockwise from the cuticle to the cells, i.e. along *(r)* direction.

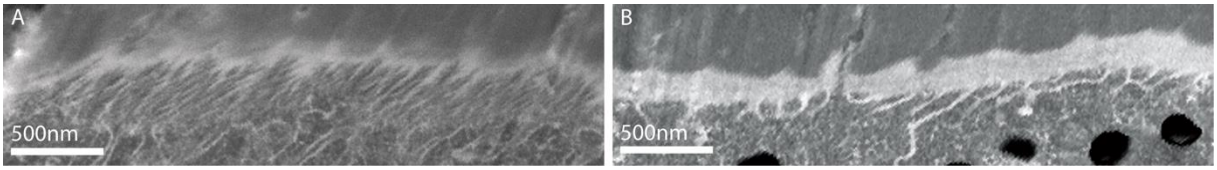

Figure S3. **TED micrographs of the assembly zone in a fifth instar animal 2 days after ecdysis**. Comparison of the assembly zone in areas where microvilli are present (A) or absent (B). TED micrographs obtained from a 100 nm thick section, electron dense structures are bright, as in FIB/SEM micrographs. Arrowhead in (A) points to the smallest/thinnest resolved fiber.

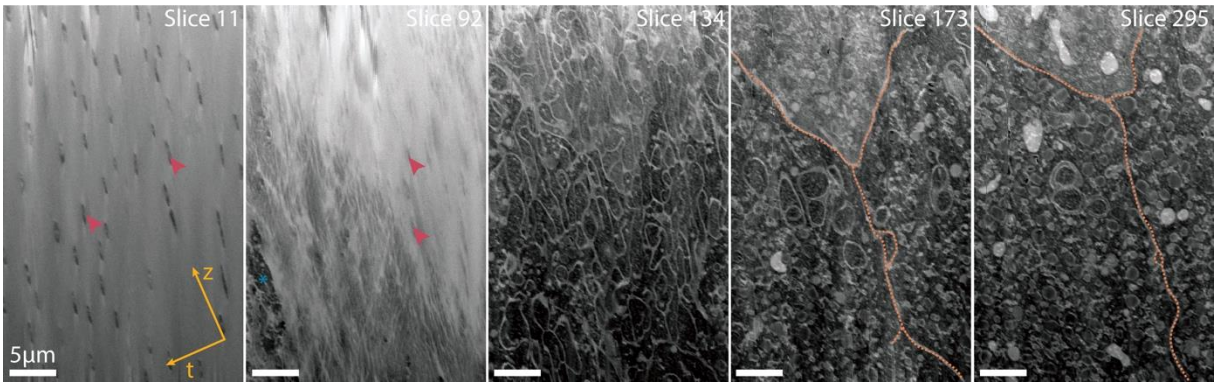

Figure S4. **FIB/SEM slice series from regions in a *Day* sample from a fifth instar animal 2 days after ecdysis lacking microvilli**. Magenta arrows point to pore-canals. Dashed orange lines indicate lateral cell membranes. Reconstructed volumes are presented in Figure 6 in the main text.

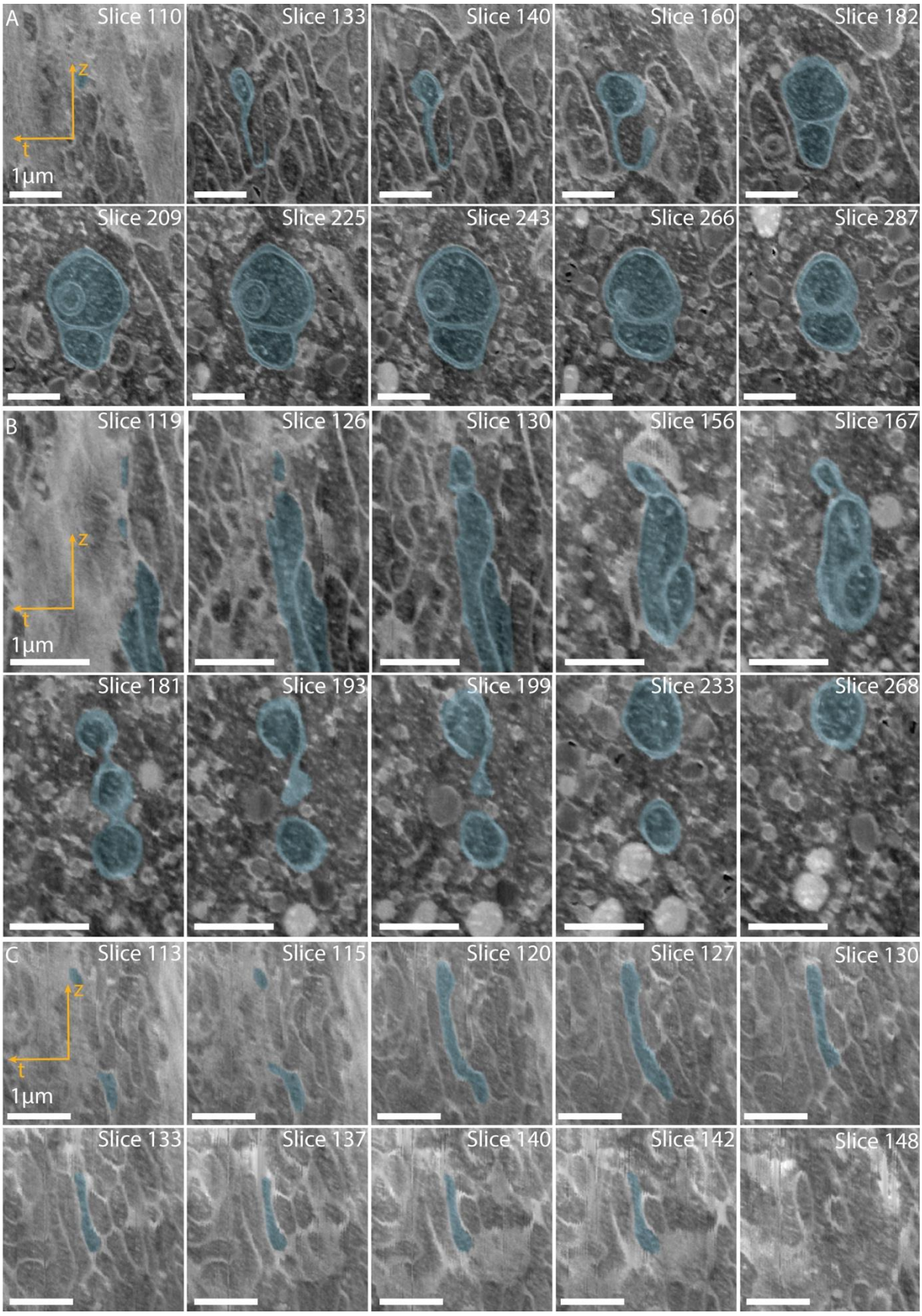
A
Slice 110
Slice 133
Slice 140
Slice 160
Slice 182
z
t
1µm
Slice 209
Slice 225
Slice 243
Slice 266
Slice 287
B
Slice 119
Slice 126
Slice 130
Slice 156
Slice 167
z
t
1µm
Slice 181
Slice 193
Slice 199
Slice 233
Slice 268
C
Slice 113
Slice 115
Slice 120
Slice 127
Slice 130
z
t
1µm
Slice 133
Slice 137
Slice 140
Slice 142
Slice 148

Figure S5. **Segmentation of FIB/SEM data.** FIB/SEM slice series from regions in a *Day* sample from a fifth instar animal sampled 2 days after ecdysis presented in 3D volume rendering in Figure 6. A, B and C correspond to features depicted in Figure 6E, F and G in the main text.

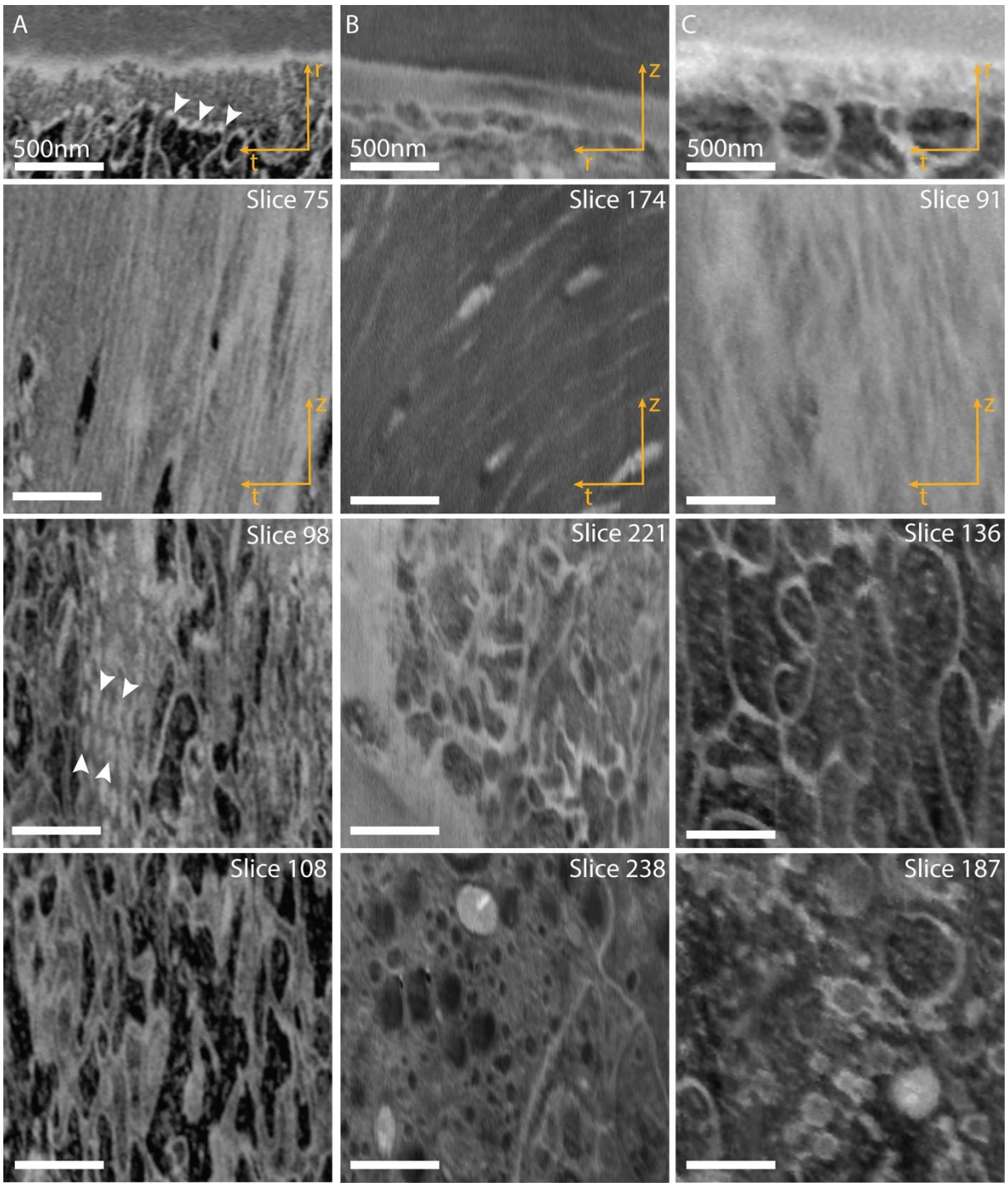

Figure S6. FIB/SEM data of epidermal cell apical surfaces in adult specimen (A) and fifth instar (B, C) sampled 2 days after ecdysis. Data acquired in regions in *Day* (A, C) and *Night* (B) samples where

microvilli are partially (A) or fully absent (B, C). Upper panel shows the (*rt)* or *(rz)* planes, followed by slice series along the (*r*) direction displaying the (*tz)* planes. Plaque-like structures in (A) are marked by white arrowheads.

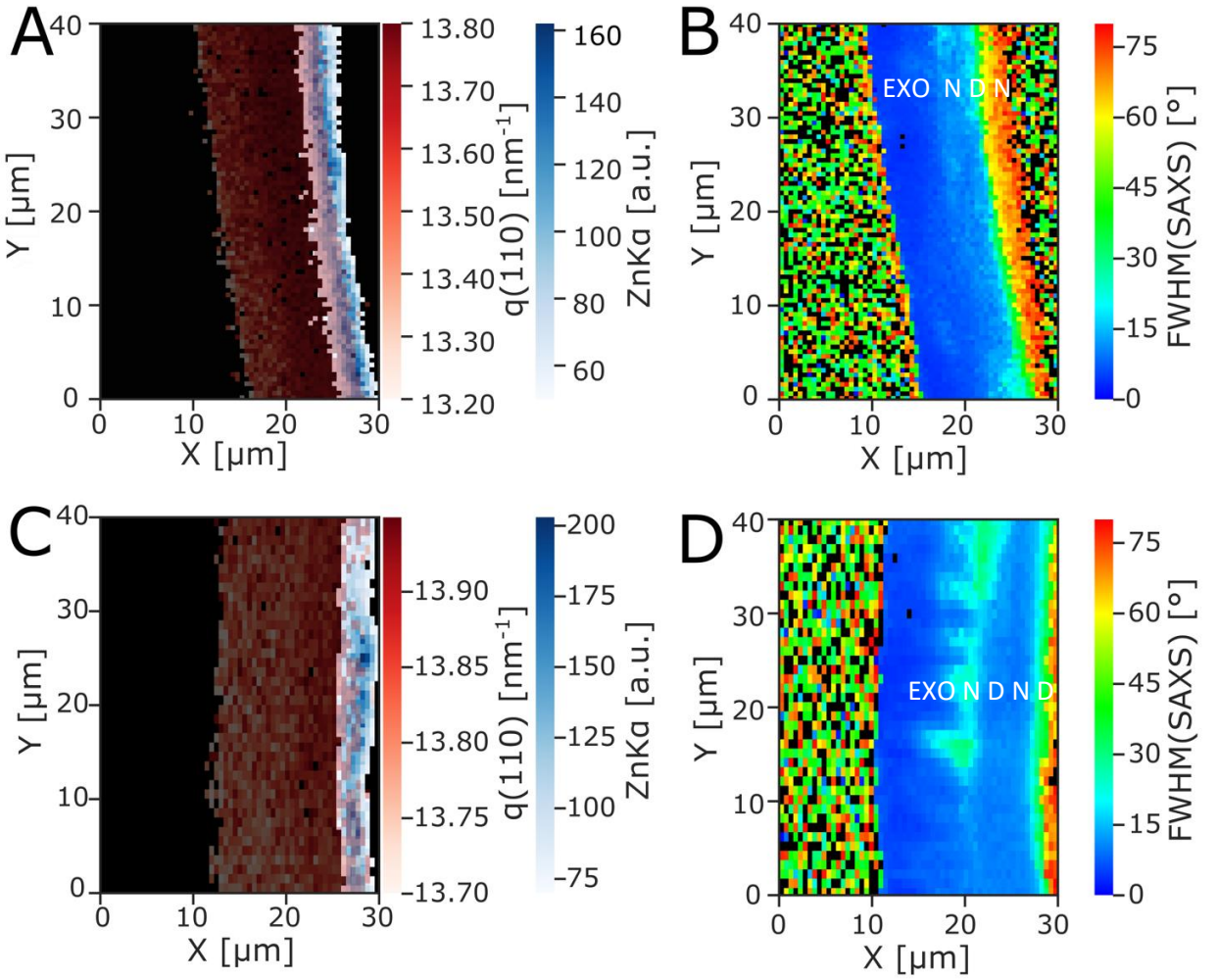

Figure S7. Scanning X-ray scattering and fluorescence results of locust tibia cross sections from adults three days after ecdysis. (A, C) Overlay of a map plotting the scattering vector (q) position of the (110) chitin reflection, and the Zn Kα X-ray fluorescence signal originating from the epidermal cells. The overlap between the two signals marks the assembly zone in (A) *Night* and (C) *Day* samples. (B, D) Heatmaps of the SAXS peak width, representing the degree of fiber alignment in (B) *Night* and (D) *Day* samples. The increased width of the SAXS peak in the assembly zone indicates decreased

fiber alignment and lower fiber compactization. Exo marks the region of exocuticle, D marks *Day* regions and N marks *Night* regions.

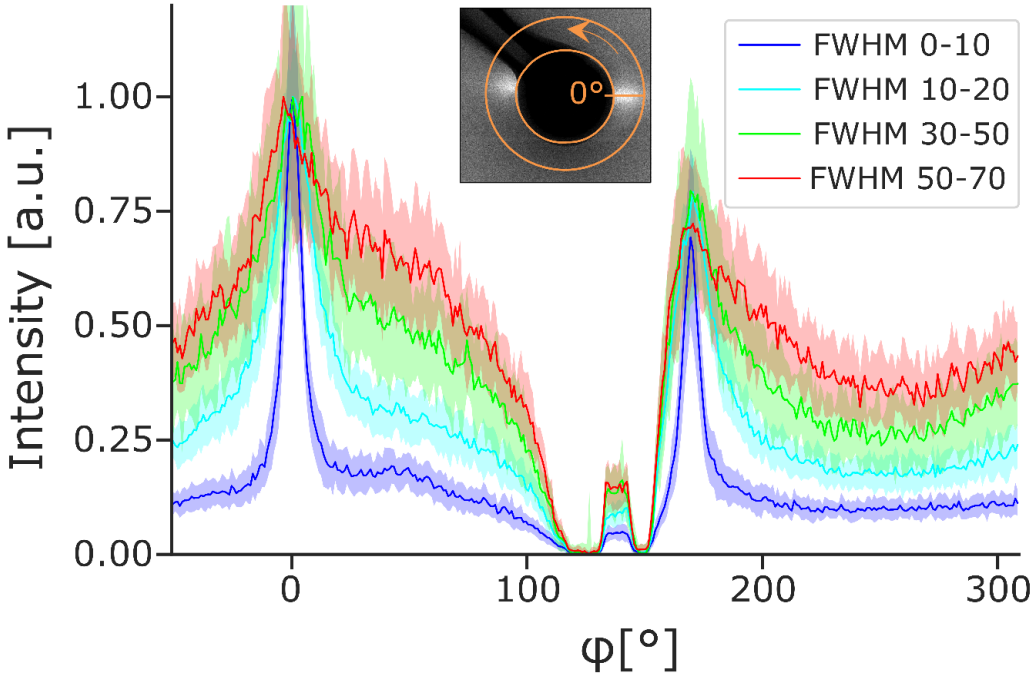

Figure S8. **SAXS peak azimuthal width**. Radial integration of the SAXS patterns obtained from various regions in the cuticle and assembly zone on a locust tibia cross section from an adult three days after ecdysis. The widths of the peaks (FWHM) are displayed in Figure S7 B and D. Inset shows an example of a 2D SAXS pattern from the cuticle region. The orange arrow marks the integration direction, and the orange contour shows the radial and azimuthal range used for the integration. Single profiles are averaged from multiple spectra in the given range with the standard deviation displayed in light color around it. The spectra were generated from the sample shown in Figure S7 A+B.

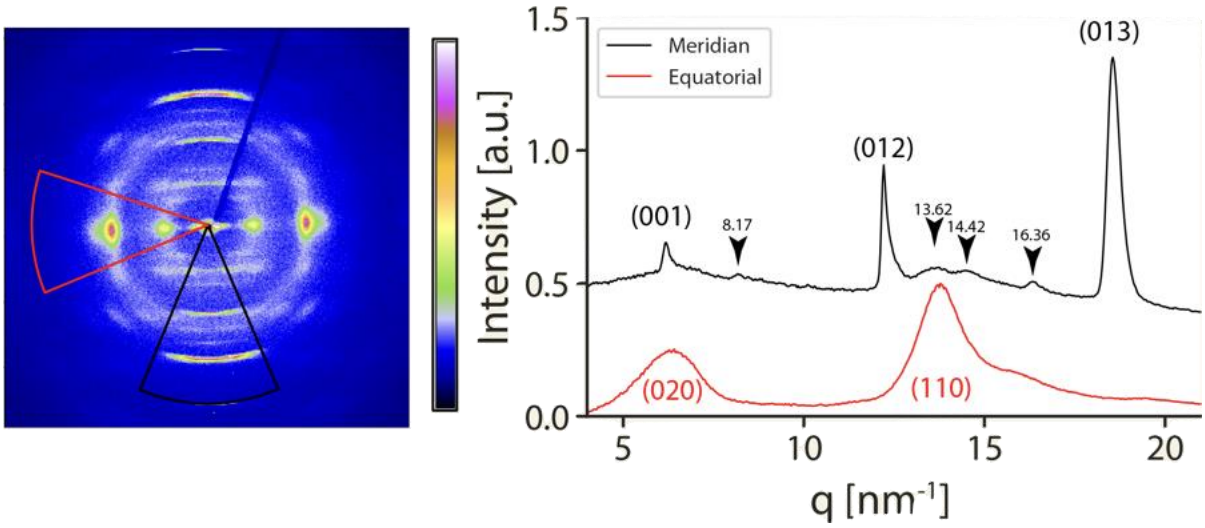

Figure S9. **WAXS data of adult locust tibia grown in 24 h light cycle for 14 days**. 2D XRD pattern and azimuthally integrated 1D XRD profiles in the equatorial direction (red line, red cake on 2D pattern) and meridian (black line, black cake on 2D pattern), obtained from the endocuticle of a locust reared in 24 hours day conditions. Black arrows point to protein diffraction peaks at q = 8.17 nm^{-1}, 13.62 nm^{-1}, 14.42 nm^{-1} and 16.36 nm^{-1} corresponding to d spacing of 0.77 nm, 0.46 nm, 0.44 nm and 0.38 nm.

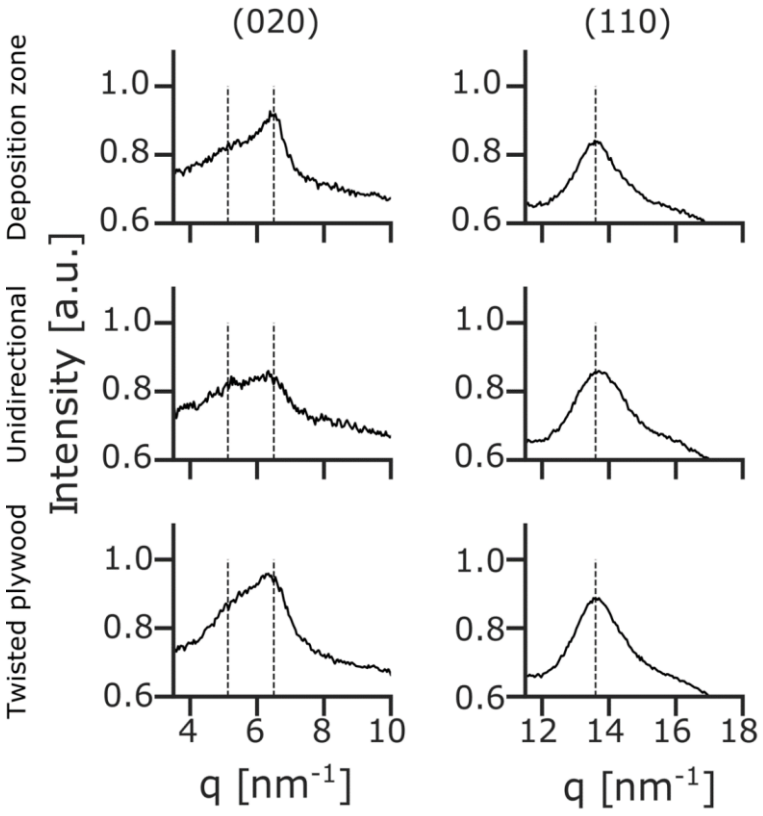

Figure S10. Averaged azimuthally integrated X-ray diffraction data. Data showing the (020) and (110) reflections from the assembly zone (upper panel), the unidirectional *Day* regions (middle panel) and the helicoidal *Night* regions in the cuticle. The spectra were generated from the sample shown in Figure 7.

References

1. Münch B, Trtik P, Marone F, Stampanoni M. 2009 Stripe and ring artifact removal with combined wavelet - Fourier filtering. *Opt Express* **17**, 8567–859.

2. Limaye A. 2012 Drishti: a volume exploration and presentation tool. In *Proc. SPIE 8506, Developments in X-Ray Tomography VIII, 85060X*, (doi:10.1117/12.935640).)

3. Benecke G *et al.* 2014 DPDAK - A customizable software for fast reduction and analysis of large X-ray scattering data sets: applications of the new DPDAK package to small angle X-ray scattering and grazing-incidence small angle X-ray scattering. *J. Appl. Cryst.* **47**, 1797-1803,.

4. Anaconda. 2014 Anaconda Software Distribution. *Comput. Softw.*