Trace_y: Software algorithms for structural analysis of individual helical filaments by three-dimensional contact point reconstruction atomic force microscopy

Wei-Feng Xue

School of Biosciences, Division of Natural Sciences, University of Kent, CT2 7NJ, Canterbury, UK; W.F.Xue@kent.ac.uk

Key words: atomic force microscopy / structural biology / amyloid / helix / filaments / image analysis

ABSTRACT

Atomic force microscopy (AFM) is a powerful and increasingly accessible technology that has a wide range of imaging applications. AFM is capable of producing detailed three-dimensional topographical height images with a high signal-to-noise ratio. This is a key capability of AFM, which enables the structural features of individual molecules to be studied without the need for ensemble averaging, and could offer structural analysis applications where heterogeneity of molecular populations, structural variations between individual molecules, or population distribution properties in general, hold important information. Here, a software tool called Trace_y designed to reconstruct the three-dimensional (3D) surface envelopes of individual helical filament structures from topographical AFM height images is presented. Workflow using Trace_y is demonstrated on the structural analysis of individual helical amyloid protein fibrils where the assembly mechanism of heterogeneous, complex and diverse fibril populations from a single amino-acid sequence due to structural polymorphism is not fully understood. The software and the workflow presented here allows the structural information encoded in topographical AFM height images to be extracted and understood as 3D contact point clouds. This approach will facilitate the use of AFM in structural biology, in integrated methodologies together with other structural analysis tools to understand the structures and behaviours at individual molecule level.

47 # INTRODUCTION

48

49 Atomic force microscopy (AFM) is a powerful multimodal technique with imaging
50 applications ranging from arrays of living cells to single molecules [1]. AFM is an increasingly
51 accessible method that has been used to study molecular dynamics [2] and molecular responses
52 to force [3]. It is also used for nanoscale imaging of biological molecules and structures in air or
53 in liquid (e.g. [4-6]). AFM topographical imaging operates by sensing the physical contact
54 between the probe, typically consisting of a sharp conical tip attached to a cantilever, and the
55 specimen surface, to generate a three-dimensional (3D) height map of the sample surface [7].
56 While the imaging resolution of AFM is limited by the physical size of the AFM probe tip,
57 molecular images acquired using AFM operating in non-contact mode with chemically
58 functionalised probe tips is capable of reaching atomic resolution on samples of small planar
59 molecules [8]. Imaging resolution using un-modified probe tips can also be enhanced to sub-
60 nanometres through recent image analysis advances [9]. For structural biology applications,
61 AFM offers several capabilities complementary to techniques such as cryogenic electron
62 microscopy (cryo-EM) and X-ray crystallography. These include the ability of imaging in
63 ambient conditions and in liquid environment, allowing imaging of biological samples such as
64 live cells in a close to native environment [10-12]. A key, sometimes overlooked, capability of
65 AFM is that topographical images acquired by AFM can achieve a high signal-to-noise ratio,
66 which enables structural features of individual molecules to be studied without the need for
67 ensemble averaging [13,14]. This means that relevant structural information can be gained
68 potentially from a single observation of a molecule on a single image. Therefore, this capability
69 could offer new exciting opportunities in elucidating rare molecular structures in complex,
70 heterogeneous samples and offer new information in the cases where the heterogeneity of

71 molecular populations or the structural variations between individual molecules hold important
72 biological relevance.
73
74 Amyloid fibrils represent a group of filamentous protein structures where the heterogeneity of
75 the complex and mixed amyloid populations as well as the structural variations between the
76 individual fibrils or even within individual fibrils hold relevant information regarding their
77 biological impact [15]. Amyloid fibrils are typically ordered helical nano-structures around 10
78 nm wide and between few nm to several microns long. They are assembled from monomeric
79 proteins or peptide precursors. Some amyloid structures are associated with devastating
80 neurodegenerative disorders including Alzheimer's and Parkinson's Diseases [16-20]. In contrast,
81 some amyloid fibrils have been identified to fulfil essential biological functions, providing
82 scaffolding, adhesion and protection for organisms from bacteria to fungi, and more recently
83 discovered to provide regulatory control of memory [21] and for viral infection [22]. The formation
84 of these functional amyloid fibrils is assumed to be tightly controlled. In contrast, the assembly
85 of pathologically associated amyloid, even from a single type of precursor, can result in highly
86 heterogeneous molecular mixtures, as well as a spectrum of different individual filament
87 structures called polymorphs. Thus, it is clear that amyloid structural polymorphism is common
88 for disease associated amyloid and this polymorphism phenomenon has been suggested to be
89 a property differentiating disease associated amyloid from functional ones [15]. AFM imaging
90 has previously been used to characterise the morphological features of individual fibrils in a
91 sample (e.g. [13,23-25]) and the size distributions of amyloid populations (e.g. [26-30]). Recent
92 advances in cryo-EM and solid-state nuclear magnetic resonance spectroscopy (ssNMR) have
93 provided atomic-detailed structures showing multiple polymorphic forms for fibrils formed of
94 identical precursors [15]. However, to resolve the extent of the polymorphic diversity and the
95 roles polymorphs play in biology, one must be able to determine the structures of individual

amyloid fibril particles, and monitor their individual interactions in a complex and heterogeneous population to sufficient detail so that the diverse amyloid assembly landscapes, their polymorphic extents and biological properties linked to the heterogeneous populations can be mapped.

Here, an image analysis software Trace_y (https://github.com/wfxue/Trace_y), featuring a collection of algorithms for 3D contact point reconstruction (CPR) [14] and structural analysis of individual helical filaments, is presented. The individual filament structural analysis workflow using Trace_y is demonstrated here on the analysis of helical amyloid protein fibrils formed from the 42 residues long Amyloid β 1-42 peptide ($A\beta_{42}$) associated with Alzheimer's disease, where heterogeneous, complex and diverse fibril populations can assemble from a single monomer amino-acid sequence in the same sample due to structural polymorphism. Individual filament level structural analysis approach through 3D CPR-AFM, recently developed to utilise the high signal-to-noise topographical imaging capability of AFM [13,14,31], has the potential to tackle the challenges presented by the structural analysis of complex and heterogeneous samples such as polymorphous amyloid fibrils [32]. The Trace_y software and the workflow for individual filament structural analysis presented here also demonstrate that the three-dimensional information encoded in AFM height images can be understood as 3D point clouds, which is a type of spatial information that could be integrated with other structural analysis methodologies [33].

RESULTS

Overview of the Trace_y workflow and its algorithms

121 The purpose of the workflow presented here is to recover the 3D structural information of
122 individually observed helical filaments encoded in topographical AFM height images.
123 **Figure 1** shows an overview of the workflow with a flowchart highlighting the key steps and
124 the algorithms in the workflow with the example application to an individual fibril formed from
125 the human $A\beta_{42}$ sequence [31]. The input for the workflow is an AFM height image where each
126 pixel in the image contains the information of its x/y coordinate in the image lattice as well as
127 the pixel value representing the recorded z-coordinate of AFM probe tip when intermittent
128 contact between the tip and the molecular surface of the sample was detected. Due to the
129 physics behind AFM, topographical height images have high vertical (z-axis) signal-to-noise
130 ratio, with the noise typically in the order of sub-Ångströms range on current commercial
131 instruments. One challenge of 3D structural reconstruction using AFM height image is that the
132 recorded x, y and z-coordinates of the tip are not the actual tip-sample contact points [14,34],
133 which are not known. This issue is commonly referred to as the tip-sample convolution artefact.
134 An important step in the workflow, therefore, is to deconvolute the image to recover the actual
135 tip-sample contact points. Once the coordinates for tip-sample contact points are recovered,
136 full 3D surface envelopes can be rendered for structures with helical symmetry, despite that
137 only the top surface of the sample is accessible to the AFM tip, because structural information
138 on the whole of a helical surface can be gained from a single direction of observation
139 perpendicular to its helical axis due to their periodic twist around the helical axis. This requires
140 the knowledge on the location of the helical axis, the handedness of the twist and the pitch of
141 the helical structure, all of which can be gained or estimated from the height images (**Figure**
142 **1c-e**). The output of the workflow is a 3D contact point cloud of the surface of the helical
143 filament, which can be further visualised as a 3D surface envelope model of the filament
144 structure. The algorithms performing the key steps of the Trace_y workflow are explained in

145 the sections below and exemplified on a fibril formed from recombinant human $A\beta_{42}$ (**Figure**
146 **1b-c**).
147
148 **Individual filament tracing and estimation of the helical axis**
149 A semi-automatic fibril tracing algorithm was developed to allow the user to select and to trace
150 segments of individual filaments observed on AFM height images. Since the central contour
151 line of a filament appearing in a height image typically corresponds to the projection of the
152 filament's helical axis on the x/y-plane, an accurate estimate of the central contour line of the
153 selected filament is needed. Therefore, a curved Gaussian wall [35] model (Equation 1, Methods)
154 was used and fit by least-squares method to regions of the height image (image squares with
155 edges twice the length of filament's apparent width) cropped along the coarse user-selected
156 contours of the filament (**Figure 2**). This approach allows the x/y-coordinates of points on the
157 central contour line of the filament to be determined with sub-pixel accuracy because the
158 coordinates are estimated using information from the entirety of the local features rather than
159 few individual pixels. The z-coordinate of the filament helical axis was subsequently estimated
160 as half of the average peak height (for example, the peaks values at locations marked with
161 triangles in **Figure 1d**). This simple estimate assumed a rigid fibril that rests on the support
162 surface without substantial deformation during deposition and by the peak probing force during
163 imaging.
164
165 **3D contact point reconstruction by image deconvolution**
166 The 3D coordinates of tip-sample contact points were next reconstructed by modelling rigid-
167 body interactions between the AFM probe tip and the sample filament using the image
168 deconvolution strategy as previously described [14]. First, the tip geometry was estimated. A
169 rounded cone tip model made by the rotation of the half cross-section described by Equation

170 (2) around the y-axis is used and the tip radius was estimated using a iterative approach where
171 image dilation caused by the contact between the tip model and the circular cross-section of a
172 featureless cylinder is matched to the experimentally observed dilation for the analysed
173 filament. This procedure is illustrated by **Figure 3a**, where least-squares regression analysis is
174 performed with a calculated convoluted cross-section (red lines and crosses) of a featureless
175 cylinder (black) fit to the observed convoluted cross-section of the filament averaged through
176 the length of the filament along its estimated helical axis (blue circles). This approach is able
177 to provide a useful initial tip radius estimate using the helical axis coordinates as input since
178 the cross-section of a helical structure with its cross-section rotating around its helical axis
179 approaches a circular shape when averaged over the length of its helical axis. The initial tip
180 radius estimate can be further optimised in a later optional step by minimising the difference
181 between the image data of the filament and a simulated image calculated using the surface
182 envelope model of the same filament. An important advantage of this sample feature-based
183 blind tip radius estimation approach is that additional dedicated tip characterisation step, which
184 can additionally blunt the tip during characterisation, can be avoided.
185
186 Once the tip radius is estimated, deconvolution of the image data by CPR [14] was subsequently
187 carried out to recover the 3D coordinates of tip-sample contact points (**Figure 3b-d**). This step
188 outputs a point cloud containing the estimated tip-sample surface contact points by moving the
189 3D x/y/z-coordinates of the recorded tip positions off the x/y pixel lattice of the image to the
190 appropriate sub-pixel locations estimated by CPR step. **Figure 3b** shows the AFM height
191 image of a segment of the example $A\beta_{42}$ filament where each pixel is shown as a point in 3D
192 space. All of the points have their x/y coordinates in the pixel lattice of the image, and the pixel
193 lattice along the x-axis is particularly visible from the viewing angle used in **Figure 3b**. **Figure**
194 **3c** demonstrates the tip-sample contact point cloud after image deconvolution by CPR where

195 the data points are no longer bound to the x/y pixel lattice and instead echo the actual sample
196 surface coordinates. The points that are localised at the top of the filament (red points in **Figure**
197 **3d**) were then selected to be used in subsequent steps to create a surface envelope model of the
198 individual filament. In summary, the CPR step reconstructed the 3D contact points between
199 the AFM tip and the filament, creating a point cloud that represents the top of the individual
200 filament being analysed.
201
202 **Visualisation of the filament cross-section and construction of 3D surface envelope model**
203 The point cloud data that represents the structure of the filament top obtained from the CPR
204 step was next used to create a 3D surface envelope model of the filament. The helical properties
205 of the individual filament being analysed were first determined by manual inspection, this is
206 shown for the example Aβ_{42} filament in **Figure 1c-e**. Since topological AFM height images
207 are not transmission images, only one side of the helical filament is observed. Hence, the helical
208 handedness of the filament twist is typically readily visible by the tilt direction of the striation
209 pattern (exemplified by the white lines in **Figure 1d**). For left-hand twisted filaments, the tilt
210 of the pattern goes from bottom left to top right whereas the tilt of the pattern goes from bottom
211 right to top left for right-hand twisted fibrils when the filaments image is aligned horizontally.
212 The handedness can also be deduced by 2D Fourier transform of the filament image
213 (**Figure 1e**). For vertically aligned filaments images, the main peaks in the 2D Fourier
214 transform power spectrum image follow the same bottom left to top right or bottom right to top
215 left tilt patterns relative to the image centre for left-handed or right-handed twists, respectively
216 (arrow heads in **Figure 1e** for the example Aβ_{42} filament). The helical pitch of the filament can
217 also be deduced from the z-height peak patterns of the central contour line of the filament
218 image, and the peaks in the 2D Fourier transform of the filament image. In this case, a direct
219 determination of the helical pitch by spotting repeating patterns in the z-height profile or

220 estimating the magnitude of the tilt of the pattern in 2D Fourier transformed image may not be
221 straight forward as the differences expected for different cross-sectional symmetries may be
222 small. Therefore, a best guess of the cross-sectional symmetry and resulting helical pitch was
223 used for the initial 3D model building, and the validity of this initial guess was checked by
224 comparing simulated image and its 2D Fourier transform power spectrum with the image data
225 of the filament in a subsequent step. This is illustrated for the example $A\beta_{42}$ filament in
226 **Figure 1d** where the helical pitch was estimated to be the peak-peak distances of every-other
227 peak in the central line z-height profile of the filament (e.g. filled blue triangles in **Figure 1d**),
228 possibly due to a C_2 cross-sectional symmetry or a 2_1-screw axis symmetry. This guess for the
229 example filament was validated because the simulated image based on the 3D model
230 constructed using the guessed pitch values (**Figure 1f**) matched well with the image data, and
231 the location of the main peaks in the 2D Fourier transformed image matched with those in the
232 2D Fourier transformed power spectrum image of the data (arrow heads in **Figure 1e**).
233
234 Once the helical pitch and twist handedness is determined for the individual filament being
235 analysed, the point cloud data representing the structure of the filament top is moved from its
236 original AFM height image coordinate system to a helical coordinate system (**Figure 4a**) where
237 the helical axis of the filament is aligned with the y-axis and the x/z-axes (the cross-sectional
238 plane) rotates with the same rate along the helical axis as the helical filament being analysed.
239 In this case, the x/z-plane rotates with a θ rotation angle that is locally estimated, referenced
240 onto the same peaks used to calculate the local helical pitch values (filled blue triangles in
241 **Figure 1d**). For the example $A\beta_{42}$ filament, the θ rotation angle at any given y-coordinate along
242 the helical axis is shown in **Figure 4b** where the blue crosses are the local reference points at
243 full integral number of complete turns around the helical axis, and the red line represents the
244 interpolated θ values along the helical axis by cubic splines interpolation. The cross-sectional

envelope of the helical filament was then visualised on the helical x/z plane when the point cloud was registered onto the rotating x/z-frame. **Figure 1g** shows the visualisation of the cross-sectional envelope of the example Aβ_{42} filament as a density map of the contact points in the x/z-plane constructed using a non-parametric kernel density estimator [36]. The spatial resolution of this density map was subsequently estimated by a Monte Carlo (MC) approach where the points in the point cloud were randomly split into two equal sized sets. The density map of each set of points were separately constructed (an example pair is shown in Figure 4c) and their Fourier ring correlation were evaluated using the ½-bit information criterion [37]. For the example Aβ_{42} filament shown in **Figure 1g**, evaluated using 500 MC cycles, the resulting spatial resolution estimate was around 13.1Å (**Figure 4d**). Finally, the 3D surface envelope of the whole of the filament was constructed by fitting an average cross-sectional envelope constructed by a cubic spline curve in the x/z-plane to the point cloud data. A moving window approach to select only points local to each evaluated y-coordinate (within ±1 helical pitch) were used to construct cross-sectional envelopes that takes into account local structural variations as previously described [14]. The final 3D surface envelope of a 400 nm segment of the example Aβ_{42} filament is shown in **Figure 4e** and **Figure 4f**. The filament surface was coloured by its perpendicular distance to the helical axis in **Figure 4e** to highlight its helical properties and by its local twist angle per unit length in **Figure 4f** to highlight the structural variations in its twist along this single individual filament.

DISCUSSION

Atomic force microscopy has recently embarked on a renaissance in terms of its application in imaging of biological/soft materials and in molecular structural biology, driven by the recent

270 developments in high-speed AFM instrumentation [38] and several new emerging image analysis
271 algorithms [9,14,39-42]. Despite the limitations of the AFM physics due to the nanometre size of
272 the probe tip and that only the surface features of molecules of interest are probed, AFM is
273 capable of time-resolved imaging with ~100 ms temporal resolution in aqueous conditions,
274 allowing the dynamics of molecules to be studied under ambient conditions. AFM is also
275 capable of high signal-to-noise topology imaging with typically sub Ånström noise and ~1Å
276 resolution in the z height-axis. This feature allows for structural analysis of individual
277 molecules without cross-molecule averaging [13,14,31]. Therefore, this allows the structural
278 variations between individual molecules in a population or within the same molecular assembly
279 to be resolved. Together, these two notable features of AFM fill an important gap in the type
280 of molecular structural information that can be gained, and creat an unique niche for AFM in
281 structural biology where methods such as cryo-electron microscopy, nuclear magnetic
282 resonance spectroscopy and X-ray crystallography all provide ensemble averaged structural
283 information. Here, the high signal-to-noise topology imaging feature of AFM is exploited to
284 allow 3D structural studies of individual helical filaments.
285
286 The Trace_y software and the workflow based on 3D CPR-AFM presented here is a post-
287 experiment analysis approach. As such, it is widely applicable for the analysis of topographical
288 AFM height images. Importantly, the application of Trace_y on helical amyloid fibrils
289 demonstrated here show unequivocally that 3D structural reconstruction is possible using AFM
290 data as input. The CPR approach allows this by unlocking the 3D information encoded in the
291 images and show that this information can be understood and visualised as 3D contact point
292 clouds. Contact point clouds can be quantitatively analysed once aligned to an appropriate and
293 meaningful coordinate system and/or to each other, this is a process also referred to as point
294 cloud registration. Indeed, point cloud registration is a principal problem that features

295 prominently in pattern recognition and computer vision fields. In the case of Trace_y, the
296 current presented version of the 3D reconstruction workflow and the algorithms that underpin
297 the key steps of the workflow establishes many opportunities for further developments,
298 refinements and improvements at each and every step by the bio-AFM community. For
299 example, refinements to the helical axis estimate and subsequently the alignment of the cross-
300 sectional axes will improve the overall quality of the 3D point cloud in representing the surface
301 envelope of the helical structures of interest. Integrative approaches involving spatial 3D data
302 sources from cryo-EM, molecular PDB models from NMR or X-ray crystallography studies,
303 or predicted molecular models (e.g. by AlphaFold [43]) can further extend the use of the 3D point
304 cloud data and provide models by which the point clouds can be compared to or registered onto.
305 In addition, it may be possible to map chemical or nanomechanical information onto the 3D
306 surface envelopes (e.g. using analogous approach as demonstrated in **Figure 4e, f**)
307 reconstructed from multi-modal AFM-IR, AFM-Raman or nanomechanics AFM experiments.
308 Therefore, these opportunities may make it possible to distinguish between different
309 conformational or ligand binding states of biological molecules or complexes, facilitating the
310 analysis of protein structure and function to high details on a single, individual molecule level.
311
312

313 # MATERIALS AND METHODS

314

315 **AFM specimen preparation**

316 The 3D contact-point reconstruction AFM (CPR-AFM) workflow to analyse the 3D surface
317 envelopes of individual helical filaments using Trace_y is demonstrated on amyloid fibrils
318 formed from the 42 residues long Amyloid β 1-42 peptide ($A\beta_{42}$). Helical filaments formed
319 from $A\beta_{42}$ is associated with neurodegenerative diseases such as Alzheimer's disease (AD), in

320 which extracellular deposits known as plaques containing $A\beta_{42}$ amyloid fibrils are found in
321 brains of patients [44]. Here, fibrils were formed *in vitro* from recombinant $A\beta_{42}$ [45] in 20mM
322 HEPES buffer under quiescent conditions at pH 7.4 and 37°C [31]. The fibrils were deposited
323 onto freshly cleaved mica discs as previously described [31].
324
325 **Topographical AFM image data acquisition**
326 $A\beta_{42}$ amyloid fibrils were imaged by AFM using a Multimode 8 SPM with a Nanoscope V
327 controller (Bruker) operating under peak-force tapping mode as detailed previously [13]. Briefly,
328 ScanAsyst silicon nitride probes (Bruker) with nominal tip radius of 2 nm and nominal spring
329 constant of 0.4 N/m were used. Topographical height images with a scan size of 4 x 4 μm and
330 2048 x 2048 pixels or 8 x 8 μm and 4096 x 4096 pixels were collected maintaining a pixel
331 density of 0.512 pixels/nm. A scan rate of 0.305 Hz was used for the 4 x 4 μm and 0.2 Hz for
332 the 8 x 8 μm scans. The fibrils were probed gently using ~400 pN of peak-force, with a noise
333 threshold of 0.5 nm and the Z limit was of 1.5 μm were used.
334
335 **Initial image processing**
336 Nanoscope analysis software supplied with the Multimode 8 SPM instrument (updated to
337 version 1.5, Bruker) were used for the initial processing of the z-height image data. Line
338 flattening were performed to remove tilt and scanner bow by fitting a 3rd-order polynomial
339 baseline to each scan line of the z-height channel image. For background flattening, the
340 thresholding option in Nanoscope analysis was used and the thresholding level was manually
341 adjusted prior to the polynomial fit. The height channel of the data was then imported into
342 Trace_y for the subsequent steps of the workflow.
343
344 **Algorithm implementation and image analysis**

The algorithms in the Trace_y workflow were implemented in Matlab (2022a, The MathWorks, Natick, MA) and include graphical user-interface for various steps in the workflow that requires user inputs. Equation (1) describes the Gaussian wall model [35] used to determine a point's coordinates $\hat{x}$ and $\hat{y}$ along the central filament contour.

$$z = h \cdot \exp\left[-\frac{1}{2\sigma^2}\left((x-\hat{x})\sin\varphi + (y-\hat{y})\cos\varphi + c\left((x-\hat{x})\cos\varphi + (y-\hat{y})\sin\varphi\right)^2\right)^2\right]$$

Eq. (1)

In Equation (1), h is the height of the Gaussian wall, σ is the standard deviation of the Gaussian cross-section, φ is the angle of the filament on the x/y-plane and c is the curvature parameter. The 3D reconstruction of tip-sample contact points were performed by image deconvolution as previously described [14]. For the deconvolution procedure, a rounded cone tip model made by rotation of the half cross-section described by Equation (2) around the y-axis is used.

$$y = \begin{cases} -\sqrt{r^2 - x^2} + \dfrac{r}{\sin\alpha}, & 0 \le x < r\cos\alpha \\ x\cot\alpha, & x \ge r\cos\alpha \end{cases}$$

Eq. (2)

In Equation (2), α is the half-angle of the tip cone and was set to 18° here (estimated from the tip geometry information provided by the manufacturer). The r parameter is the tip-radius. Software package with stand-alone executable of Trace_y was prepared using Matlab Compiler (2022a) and can be downloaded from the GitHub (https://github.com/wfxue/Trace_y).

ACKNOWLEDGMENTS

Help with the software development and insightful discussions by Liisa Lutter, and acquisition of the Aβ$_{42}$ amyloid image data used here as examples by Liam Aubrey are gratefully acknowledged. Testing and helpful comments provided by the members of the Xue and the Louise Serpell research groups are also gratefully acknowledged. The AFM image examples used is acquired on an instrument supported by funding from the Biotechnology and Biological Sciences Research Council (BBSRC), UK grants BB/J008001/1, BB/M02427X/1 and BB/S003312/1

REFERENCES

1 Dufrene, Y. F., Ando, T., Garcia, R., Alsteens, D., Martinez-Martin, D., Engel, A., Gerber, C. & Muller, D. J. Imaging modes of atomic force microscopy for application in molecular and cell biology. *Nat Nanotechnol* **12**, 295-307, (2017). https://doi.org/10.1038/nnano.2017.45

2 Parsons, E. S., Stanley, G. J., Pyne, A. L. B., Hodel, A. W., Nievergelt, A. P., Menny, A., Yon, A. R., Rowley, A., Richter, R. P., Fantner, G. E., Bubeck, D. & Hoogenboom, B. W. Single-molecule kinetics of pore assembly by the membrane attack complex. *Nat Commun* **10**, 2066, (2019). https://doi.org/10.1038/s41467-019-10058-7

3 Chen, Y., Radford, S. E. & Brockwell, D. J. Force-induced remodelling of proteins and their complexes. *Curr Opin Struct Biol* **30**, 89-99, (2015). https://doi.org/10.1016/j.sbi.2015.02.001

4 Ares, P., Fuentes-Perez, M. E., Herrero-Galan, E., Valpuesta, J. M., Gil, A., Gomez-Herrero, J. & Moreno-Herrero, F. High resolution atomic force microscopy of double-stranded RNA. *Nanoscale* **8**, 11818-11826, (2016). https://doi.org/10.1039/c5nr07445b

5 Moreno-Herrero, F., Perez, M., Baro, A. M. & Avila, J. Characterization by atomic force microscopy of Alzheimer paired helical filaments under physiological conditions. *Biophys J* **86**, 517-525, (2004). https://doi.org/10.1016/S0006-3495(04)74130-2

6 Nowakowski, R., Luckham, P. & Winlove, P. Imaging erythrocytes under physiological conditions by atomic force microscopy. *Biochim Biophys Acta* **1514**, 170-176, (2001). https://doi.org/10.1016/s0005-2736(01)00365-0

7 Cappella, B. & Dietler, G. Force-distance curves by atomic force microscopy. *Surf Sci Rep* **34**, 1-104, (1999). https://doi.org/10.1016/S0167-5729(99)00003-5

8 Gross, L., Mohn, F., Moll, N., Liljeroth, P. & Meyer, G. The chemical structure of a molecule resolved by atomic force microscopy. *Science* **325**, 1110-1114, (2009). https://doi.org/10.1126/science.1176210

9 Heath, G. R., Kots, E., Robertson, J. L., Lansky, S., Khelashvili, G., Weinstein, H. & Scheuring, S. Localization atomic force microscopy. *Nature* **594**, 385-390, (2021). https://doi.org/10.1038/s41586-021-03551-x

10 Drake, B., Prater, C. B., Weisenhorn, A. L., Gould, S. A., Albrecht, T. R., Quate, C. F., Cannell, D. S., Hansma, H. G. & Hansma, P. K. Imaging crystals, polymers, and processes in water with the atomic force microscope. *Science* **243**, 1586-1589, (1989). https://doi.org/10.1126/science.2928794

11 Pasquina-Lemonche, L., Burns, J., Turner, R. D., Kumar, S., Tank, R., Mullin, N., Wilson, J. S., Chakrabarti, B., Bullough, P. A., Foster, S. J. & Hobbs, J. K. The architecture of the Gram-positive bacterial cell wall. *Nature* **582**, 294-297, (2020). https://doi.org/10.1038/s41586-020-2236-6

12 Radmacher, M., Tillamnn, R. W., Fritz, M. & Gaub, H. E. From molecules to cells: imaging soft samples with the atomic force microscope. *Science* **257**, 1900-1905, (1992). https://doi.org/10.1126/science.1411505

13 Aubrey, L. D., Blakeman, B. J. F., Lutter, L., Serpell, C. J., Tuite, M. F., Serpell, L. C. & Xue, W. F. Quantification of amyloid fibril polymorphism by nano-morphometry reveals the individuality of filament assembly. *Commun Chem* **3**, 125, (2020). https://doi.org/10.1038/s42004-020-00372-3

14 Lutter, L., Serpell, C. J., Tuite, M. F., Serpell, L. C. & Xue, W. F. Three-dimensional reconstruction of individual helical nano-filament structures from atomic force microscopy topographs. *Biomol Concepts* **11**, 102-115, (2020). https://doi.org/10.1515/bmc-2020-0009

15 Lutter, L., Aubrey, L. D. & Xue, W. F. On the Structural Diversity and Individuality of Polymorphic Amyloid Protein Assemblies. *J Mol Biol* **433**, 167124, (2021). https://doi.org/10.1016/j.jmb.2021.167124

16 Knowles, T. P., Vendruscolo, M. & Dobson, C. M. The amyloid state and its association with protein misfolding diseases. *Nature reviews. Molecular cell biology* **15**, 384-396, (2014). https://doi.org/10.1038/nrm3810

17 Eisenberg, D. & Jucker, M. The amyloid state of proteins in human diseases. *Cell* **148**, 1188-1203, (2012). https://doi.org/10.1016/j.cell.2012.02.022

18 Ke, P. C., Zhou, R., Serpell, L. C., Riek, R., Knowles, T. P. J., Lashuel, H. A., Gazit, E., Hamley, I. W., Davis, T. P., Fandrich, M., Otzen, D. E., Chapman, M. R., Dobson, C. M., Eisenberg, D. S. & Mezzenga, R. Half a century of amyloids: past, present and future. *Chem Soc Rev* **49**, 5473-5509, (2020). https://doi.org/10.1039/c9cs00199a

19 Scheres, S. H., Zhang, W., Falcon, B. & Goedert, M. Cryo-EM structures of tau filaments. *Curr Opin Struct Biol* **64**, 17-25, (2020). https://doi.org/10.1016/j.sbi.2020.05.011

20 Lutter, L., Serpell, C. J., Tuite, M. F. & Xue, W. F. The molecular lifecycle of amyloid - Mechanism of assembly, mesoscopic organisation, polymorphism, suprastructures, and biological consequences. *Biochim Biophys Acta Proteins Proteom* **1867**, 140257, (2019). https://doi.org/10.1016/j.bbapap.2019.07.010

21 Hervas, R., Rau, M. J., Park, Y., Zhang, W., Murzin, A. G., Fitzpatrick, J. A. J., Scheres, S. H. W. & Si, K. Cryo-EM structure of a neuronal functional amyloid implicated in memory persistence in Drosophila. *Science* **367**, 1230-1234, (2020). https://doi.org/10.1126/science.aba3526

22 Leger, P., Nachman, E., Richter, K., Tamietti, C., Koch, J., Burk, R., Kummer, S., Xin, Q., Stanifer, M., Bouloy, M., Boulant, S., Krausslich, H. G., Montagutelli, X., Flamand, M., Nussbaum-Krammer, C. & Lozach, P. Y. NSs amyloid formation is associated with the virulence of Rift Valley fever virus in mice. *Nat Commun* **11**, 3281, (2020). https://doi.org/10.1038/s41467-020-17101-y

23 Adamcik, J. & Mezzenga, R. Amyloid Polymorphism in the Protein Folding and Aggregation Energy Landscape. *Angew Chem Int Ed Engl* **57**, 8370-8382, (2018). https://doi.org/10.1002/anie.201713416

24 Ruggeri, F. S., Flagmeier, P., Kumita, J. R., Meisl, G., Chirgadze, D. Y., Bongiovanni, M. N., Knowles, T. P. J. & Dobson, C. M. The Influence of Pathogenic Mutations in alpha-Synuclein on Biophysical and Structural Characteristics of Amyloid Fibrils. *ACS Nano* **14**, 5213-5222, (2020). https://doi.org/10.1021/acsnano.9b09676

25 Koloteva-Levine, N., Aubrey, L. D., Marchante, R., Purton, T. J., Hiscock, J. R., Tuite, M. F. & Xue, W. F. Amyloid particles facilitate surface-catalyzed cross-seeding by acting as promiscuous nanoparticles. *Proc Natl Acad Sci U S A* **118**, e2104148118, (2021). https://doi.org/10.1073/pnas.2104148118

26 Marchante, R., Beal, D. M., Koloteva-Levine, N., Purton, T. J., Tuite, M. F. & Xue, W. F. The physical dimensions of amyloid aggregates control their infective potential as prion particles. *Elife* **6**, e27109, (2017). https://doi.org/10.7554/eLife.27109

27 Sanami, S., Purton, T. J., Smith, D. P., Tuite, M. F. & Xue, W. F. Comparative Analysis of the Relative Fragmentation Stabilities of Polymorphic Alpha-Synuclein Amyloid Fibrils. *Biomolecules* **12**, 630, (2022). https://doi.org/10.3390/biom12050630

28 Beal, D. M., Tournus, M., Marchante, R., Purton, T. J., Smith, D. P., Tuite, M. F., Doumic, M. & Xue, W. F. The Division of Amyloid Fibrils: Systematic Comparison of Fibril Fragmentation Stability by Linking Theory with Experiments. *iScience* **23**, 101512, (2020). https://doi.org/10.1016/j.isci.2020.101512

29 Xue, W. F. & Radford, S. E. An imaging and systems modeling approach to fibril breakage enables prediction of amyloid behavior. *Biophys J* **105**, 2811-2819, (2013). https://doi.org/10.1016/j.bpj.2013.10.034

30 Xue, W. F., Hellewell, A. L., Gosal, W. S., Homans, S. W., Hewitt, E. W. & Radford, S. E. Fibril fragmentation enhances amyloid cytotoxicity. *J Biol Chem* **284**, 34272-34282, (2009). https://doi.org/10.1074/jbc.M109.049809

31 Aubrey, Liam D, Lutter, Liisa, Fennell, Kate, Purton, Tracey J, Ward, Natasha, Serpell, Louise C. & Xue, Wei-Feng. Structural reconstruction of individual filaments in Abeta42 fibril populations assembled in vitro reveal rare species that resemble ex vivo amyloid polymorphs from human brains. *bioRxiv*, 2023.07.14.549001, (2023). https://doi.org/10.1101/2023.07.14.549001

32 Konstantoulea, K., Guerreiro, P., Ramakers, M., Louros, N., Aubrey, L. D., Houben, B., Michiels, E., De Vleeschouwer, M., Lampi, Y., Ribeiro, L. F., de Wit, J., Xue, W. F., Schymkowitz, J. & Rousseau, F. Heterotypic Amyloid beta interactions facilitate amyloid assembly and modify amyloid structure. *EMBO J* **41**, e108591, (2022). https://doi.org/10.15252/embj.2021108591

33 Lutter, L., Al-Hilaly, Y. K., Serpell, C. J., Tuite, M. F., Wischik, C. M., Serpell, L. C. & Xue, W. F. Structural Identification of Individual Helical Amyloid Filaments by Integration of Cryo-Electron Microscopy-Derived Maps in Comparative Morphometric Atomic Force Microscopy Image Analysis. *J Mol Biol* **434**, 167466, (2022). https://doi.org/10.1016/j.jmb.2022.167466

34 Villarrubia, J. S. Algorithms for Scanned Probe Microscope Image Simulation, Surface Reconstruction, and Tip Estimation. *J Res Natl Inst Stand Technol* **102**, 425-454, (1997). https://doi.org/10.6028/jres.102.030

35 Ruhnow, F., Zwicker, D. & Diez, S. Tracking single particles and elongated filaments with nanometer precision. *Biophys J* **100**, 2820-2828, (2011). https://doi.org/10.1016/j.bpj.2011.04.023

36 Botev, Z. I., Grotowski, J. F. & Kroese, D. P. Kernel Density Estimation Via Diffusion. *Ann Stat* **38**, 2916-2957, (2010). https://doi.org/10.1214/10-Aos799

37 van Heel, M. & Schatz, M. Fourier shell correlation threshold criteria. *J Struct Biol* **151**, 250-262, (2005). https://doi.org/10.1016/j.jsb.2005.05.009

38 Flechsig, H. & Ando, T. Protein dynamics by the combination of high-speed AFM and computational modeling. *Curr Opin Struct Biol* **80**, 102591, (2023). https://doi.org/10.1016/j.sbi.2023.102591

39 Amyot, R. & Flechsig, H. BioAFMviewer: An interactive interface for simulated AFM scanning of biomolecular structures and dynamics. *PLoS Comput Biol* **16**, e1008444, (2020). https://doi.org/10.1371/journal.pcbi.1008444

40 Beton, J. G., Moorehead, R., Helfmann, L., Gray, R., Hoogenboom, B. W., Joseph, A. P., Topf, M. & Pyne, A. L. B. TopoStats - A program for automated tracing of biomolecules from AFM images. *Methods* **193**, 68-79, (2021). https://doi.org/10.1016/j.ymeth.2021.01.008

41 Necas, D. & Klapetek, P. Gwyddion: an open-source software for SPM data analysis. *Cent Eur J Phys* **10**, 181-188, (2012). https://doi.org/10.2478/s11534-011-0096-2

42 Usov, I. & Mezzenga, R. FiberApp: An Open-Source Software for Tracking and Analyzing Polymers, Filaments, Biomacromolecules, and Fibrous Objects. *Macromolecules* **48**, 1269-1280, (2015). https://doi.org/10.1021/ma502264c

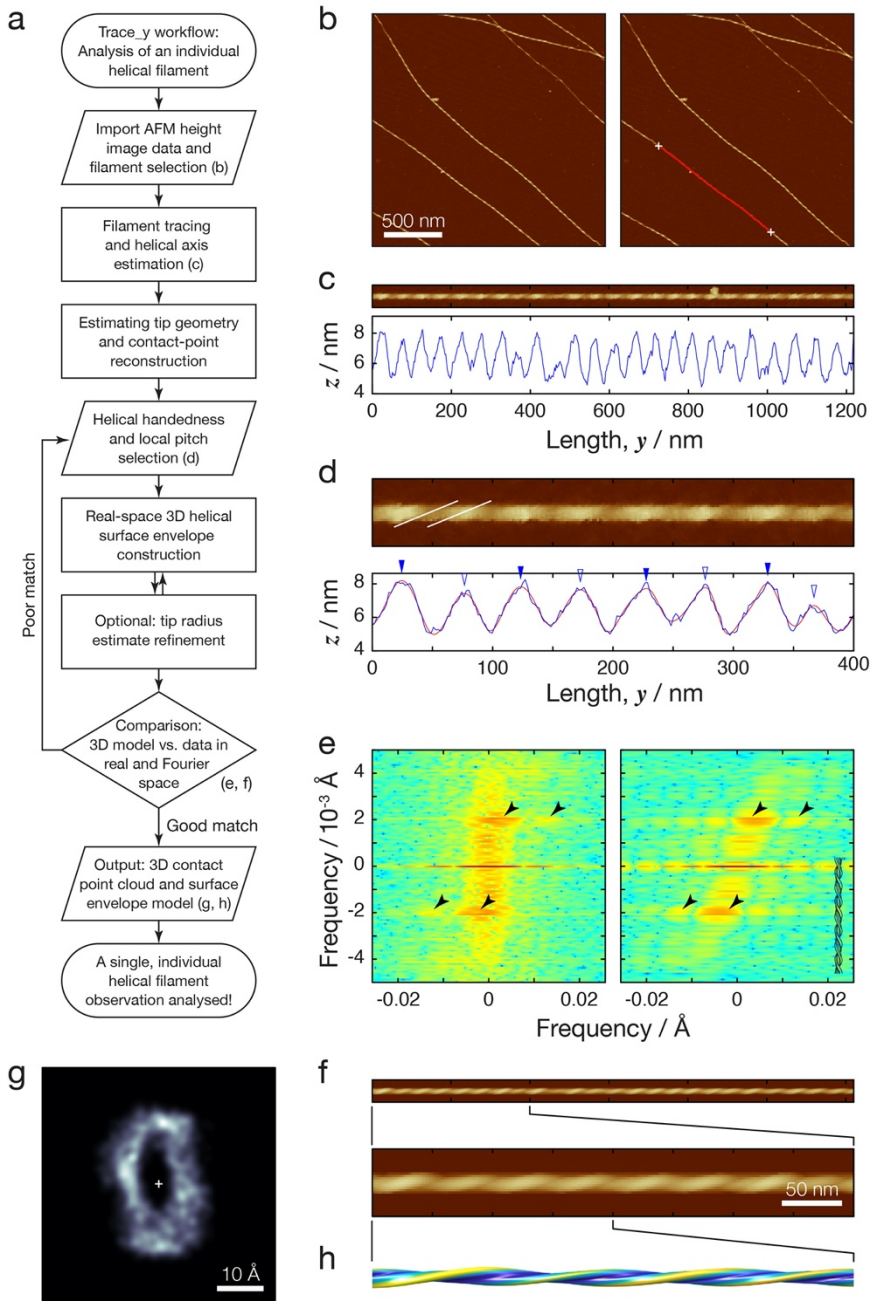
43 Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zidek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P. & Hassabis, D. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589, (2021). https://doi.org/10.1038/s41586-021-03819-2

44 Yang, Y., Arseni, D., Zhang, W., Huang, M., Lovestam, S., Schweighauser, M., Kotecha, A., Murzin, A. G., Peak-Chew, S. Y., Macdonald, J., Lavenir, I., Garringer, H. J., Gelpi, E., Newell, K. L., Kovacs, G. G., Vidal, R., Ghetti, B., Ryskeldi-Falcon, B., Scheres, S. H. W. & Goedert, M. Cryo-EM structures of amyloid-beta 42 filaments from human brains. *Science* **375**, 167-172, (2022). https://doi.org/10.1126/science.abm7285

45 O'Malley, T. T., Linse, S. & Walsh, D. M. Production and Use of Recombinant Abeta for Aggregation Studies. *Methods Mol Biol* **1777**, 307-320, (2018). https://doi.org/10.1007/978-1-4939-7811-3_19

540 **FIGURES**

541



542

543

544 **Figure 1. The Trace_y workflow demonstrated on an individual Aβ$_{42}$ amyloid fibril.** (a)
545 Flowchart overview of the Trace_y workflow for analysing the structure of an individual
546 helical filament. (b) Topographical AFM height images of helical twisted Aβ$_{42}$ fibrils. The
547 right image is identical to the left image, except with an example fibril and its central contour

line highlighted (red line). The white crosses indicate user defined start and end point of the traced fibril segment. The scale bar represent 500 nm in both images. (c) Digitally straightened fibril image showing the fibril highlighted in (b) and its central line height profile. (d) A zoomed image showing the first 400 nm segment the same fibril shown in (c). Its central line height profile (blue line) is shown together with a digitally smoothed line (red line) with which the peak positions were determined (blue triangles). The filled blue triangles indicate reference locations with one helical pitch distance in between each filled blue triangles. (e) Power spectrum of 2D Fourier transform of the fibril image data is shown to the left compared with that calculated from a simulated topograph image shown to the right. The black arrows indicate identical locations in both power spectra and show major peak locations. (f) Topographic AFM height image simulated from the 3D fibril envelope model (h) of the same fibril in (b) together with a zoomed in image showing the first 400 nm segment. (g) The cross-sectional contact-point density map of the same highlighted fibril from (b) Scale bar indicate distance of 10Å.

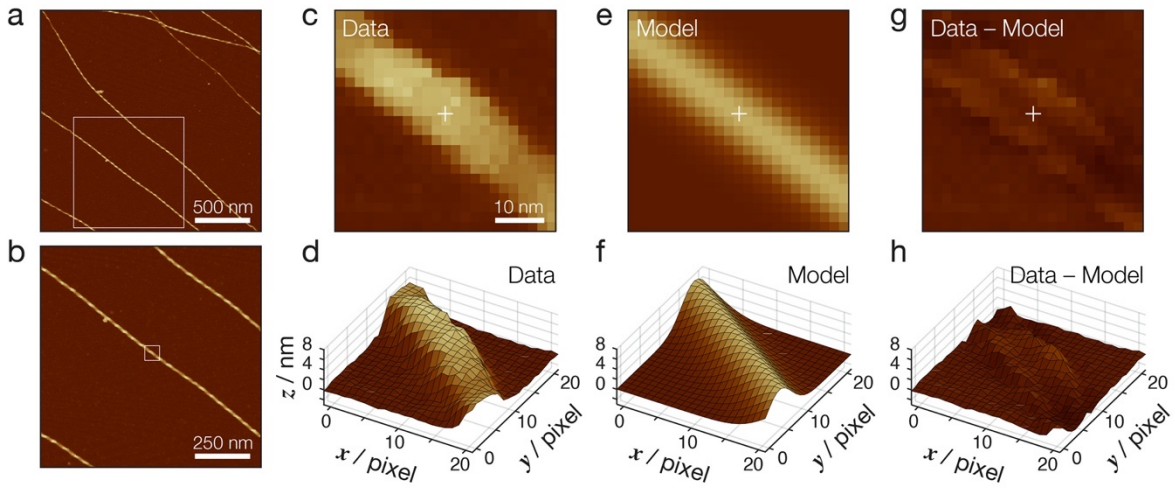


Figure 2. Tracing of individual filaments and estimation of helical axis coordinates by regression analysis using a curved Gaussian wall model. (a) Topographical AFM height images of helical twisted $A\beta_{42}$ fibrils. The same 2x2 μm image as **Figure 1b** is shown. The white box indicate 1x1 μm magnified area shown in (b) and scale bar indicating 500 nm. (b) 1x1 μm magnified area indicated by the white box in (a). The white box in (b) indicate 22x22 pixel area shown in (c)-(f), and scale bar indicating 250 nm. (c) and (d) 22x22 pixel area indicated in the image data in (b) shown as a 2D height image and as a 3D surface, respectively. (e) and (f) Best fit curved Gaussian wall model of the same area indicated in the image data in (b) shown as a 2D height image and as a 3D surface, respectively. (g) and (h) Pixel level residuals between the image data and the best fit model shown as a 2D height image and as a 3D surface, respectively. In (c)-(h), white cross indicate an estimated point of helical axis x/y coordinates. Scale bar indicate 10 nm in (c), (e) and (g).

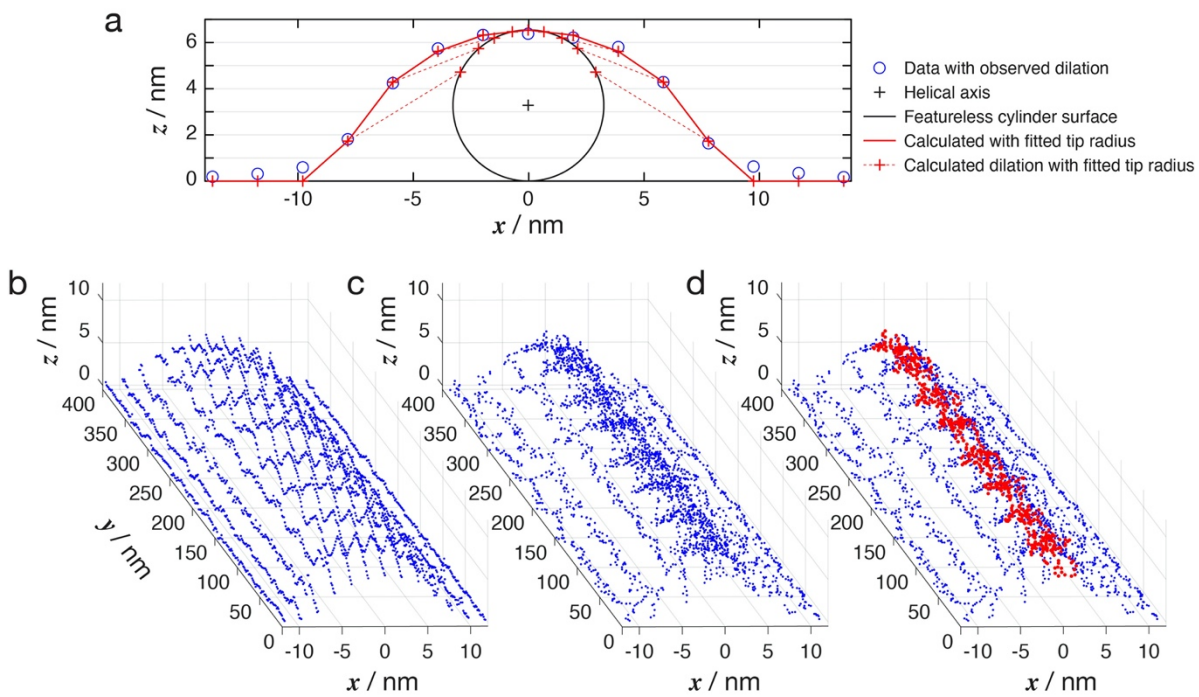


Figure 3. Cantilever tip radius estimation and contact point reconstruction result in a 3D point cloud that represents the tip-filament surface contact points. (a) Initial estimation of the tip radius by least-squares fit of calculated image dilation (red crosses, red solid line and red dotted lines) from a featureless cylinder (black line representing the cross-section of the cylinder) to observed average dilation along the fibril (blue circles). (b) The fibril image in **Figure 1d** shown as points in 3D space representing the pixels' x/y/z coordinates. (c) The point cloud output of CPR showing the x/y/z coordinates of the points in (b) moved to estimated tip-filament surface contact points. (d) Contact points on the top of the helical filament subsequently used for construction of 3D surface envelop is show as red points.

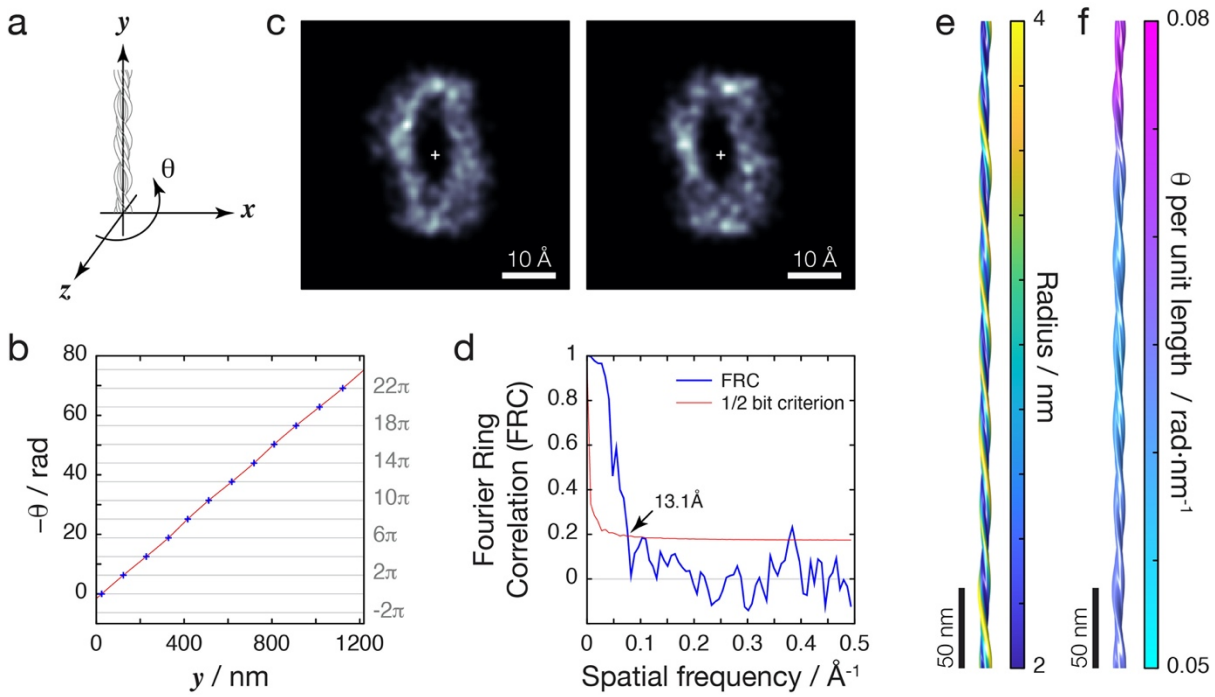


Figure 4. Construction of filament 3D surface envelope by point cloud registration in a helical coordinate system. (a) Schematic of the helical coordinate system used to align contact points in a point cloud representing a helical filament. The x/z plane of this coordinate system is set to rotate around the helical axis aligned to the y-axis with the same twist angles θ as observed for the filament. (b) The local twist angle θ (blue cross) of the example filament shown in **Figure 1c** measured using its height profile (e.g. filled blue triangles in **Figure 1d**). Red line represents the θ function along the whole filament constructed by spline interpolation with the measurement points (blue cross). (c) The cross-sectional contact-point density map of the same example filament. An example density map pair generated by randomly selected half-sets of the contact point cloud used for FRC measurement is shown. (d) The spatial resolution of the contact-point density map shown in **Figure 1g** estimated by FRC (blue line) and the ½ bit information criterion [37] (red line with the FRC crossing point indicated by the arrow). (e) 3D surface envelope model of the example filament (the first 400 nm segment as shown in **Figure 1d**) with surface colour indicating distance to helical axis. (f) The same 3D surface

envelope model of the example filament as (e) but with colour indicating helical twist mapped to its surface to highlight local helical structural features.