

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23

Quantification of amyloid fibril polymorphism by nano- morphometry reveals the individuality of filament assembly

Liam D. Aubrey ^{1†}, Ben J. F. Blakeman ^{1†}, Liisa Lutter ¹, Christopher J. Serpell ², Mick F. Tuite ¹, Louise C. Serpell ³, Wei-Feng Xue ^{1*}

¹ Kent Fungal Group, School of Biosciences, University of Kent, CT2 7NJ, Canterbury, UK

² School of Physical Sciences, University of Kent, CT2 7NH, Canterbury, UK

³ Sussex Neuroscience, School of Life Sciences, University of Sussex, BN1 6NN, Falmer, Brighton, UK

[†] Authors contributed equally to this study

* Correspondence to: W.F.Xue@kent.ac.uk; Tel +44-(0)1227 824821

Key words: amyloid / fibril structure / polymorphism / self-assembly / atomic force microscopy / image analysis

24

25 **ABSTRACT**

26

27 Amyloid fibrils are highly polymorphic structures formed by many different proteins. They
28 provide biological function but also abnormally accumulate in numerous human diseases. The
29 physicochemical principles of amyloid polymorphism are not understood due to lack of
30 structural insights at the single-fibril level. To identify and classify different fibril polymorphs
31 and to quantify the level of heterogeneity is essential to decipher the precise links between
32 amyloid structures and their functional and disease associated properties such as toxicity,
33 strains, propagation and spreading. Employing gentle, force-curve based AFM, we produce
34 detailed images, from which the 3D reconstruction of individual filaments in heterogeneous
35 amyloid samples is achieved. Distinctive fibril polymorphs are then classified by hierarchical
36 clustering, and sample heterogeneity is objectively quantified. These data demonstrate the
37 polymorphic nature of fibril populations, provide important information regarding the energy
38 landscape of amyloid self-assembly, and offer quantitative insights into the structural basis of
39 polymorphism in amyloid populations.

40

41

42

43 INTRODUCTION

44

45 Amyloid fibrils are well known for their association with protein misfolding diseases,
46 collectively known as Amyloidoses, as well as several neurodegenerative disorders. Each
47 disease is characterised by a specific protein that accumulates in the peripheral tissues, in
48 peripheral organs or in the brain. All amyloid fibrils have a core cross- β molecular architecture
49 composed of β strands that stack perpendicular to the long fibril axis to form protofilaments up
50 to several microns in length ¹. Multiple protofilaments can laterally associate to form twisted
51 fibrils with a hydrophobic core ^{2,3}. Alzheimer's A β is one of the best known of the amyloid
52 proteins, others include islet amyloid polypeptide in diabetes type 2 ⁴, β -2 microglobulin in
53 dialysis related amyloidosis ⁵⁻⁷, transthyretin in familial amyloidotic polyneuropathy ^{8,9}
54 amyloid A (AA) in acute phase amyloidosis ¹⁰, tau in Alzheimer's and other tauopathies, and
55 α -synuclein in Parkinson's disease ¹¹. There are also many known amyloid forming proteins
56 that are not associated with disease but with biological function, including structural
57 components of biofilms such as Curli expressed in *E. coli*, ^{12,13} and amyloid fibrils involved in
58 human skin pigmentation such as Pmel17 ^{14,15}.

59

60 Advances in high resolution structural determination techniques such as X-ray crystallography,
61 solid-state nuclear magnetic resonance (ssNMR) and cryo-transmission electron microscopy
62 (cryo-EM) have recently revolutionised our understanding of amyloid fibril structures ¹⁶⁻¹⁹. For
63 example, recent cryo-EM and ssNMR structures of A β ^{17,20,21} and tau ^{16,22,23} reveal parallel, in
64 register organisation of the proteins within the cross-beta core. These advances provided
65 molecular information on the 3D fold of the monomeric subunits within the core of distinct
66 fibril types in addition to information about the intermolecular interactions between amino acid

67 residues²⁴. The impact of mutations on the core fold, for example in familial forms of
68 Alzheimer's disease, can therefore be accessed²⁰. Thus, on an atomic scale, these techniques
69 have advanced our understanding of the core structures that make up individual amyloid fibril
70 types.

71

72 On a mesoscopic (micrometre to nanometre) scale, amyloid fibrils display a high degree of
73 polymorphism and amyloid populations are often highly heterogeneous. This has been
74 demonstrated in disease related amyloid fibrils in both *in vitro* and *ex vivo* samples, as well as
75 in samples of functional amyloid²⁵⁻²⁷. For example in a study involving A β , ssNMR data was
76 compared for fibril structures taken from various patients which showed that the structures of
77 the fibrils observed were highly varied from patient to patient in both A β ₁₋₄₀ and A β ₁₋₄₂²⁸.
78 Structural characterisation has revealed that many amyloidogenic proteins can form
79 polymorphic structures, revealing that, for example, AA can fold into different conformations
80 dependent on the species as well as the individual patients²⁹, while tau has been shown to form
81 different polymorphs in different tauopathies^{16,22,30}.

82

83 Structural polymorphism in amyloid fibrils, can be dependent on the number and orientation
84 of the protofilaments that are arranged within a fibril³¹, as well as the conformation of the
85 monomeric subunits that make up the core of the fibril structure which is in turn dependent on
86 the primary sequence³². Fibril structural polymorphism can be identified by the twists along
87 the longitudinal axis of the fibrils producing varied periodic cross-over distances, and by the
88 shape and size of the cross-sectional area of the fibril³³. Polymorphism is of interest due to the
89 possibility that different fibril polymorphs may have different physical properties, such as the
90 rate of fragmentation^{34,35} or the ability to act as a catalytic surface for secondary nucleation^{35,36}.
91 Different physical properties could in turn result in different biological activities, such as the

92 potential to propagate in a prion-like manner ³⁷, the ability to associate with cytotoxic active
93 species ³⁸ or the impermeability of a biofilm matrix ³⁹. Another example of an amyloid structure
94 function relationship is demonstrated by the difference in the structures of murine and human
95 AA which results in a lower efficiency for the induction of amyloidosis between species ²⁹.
96 Using electron microscopy (EM) and atomic force microscopy (AFM), it has been possible to
97 observe and differentiate between polymorphs in amyloid fibril samples ²⁷. However, since
98 fibril polymorphism has generally only been assessed in a qualitative manner with respect to
99 the overall heterogeneity of amyloid fibril data sets ²⁷, the structural, physicochemical and
100 mechanistic origin of polymorphism, the extent to which it exists and its effect on biological
101 activity is not known.

102

103 In order to identify, classify and quantify amyloid fibril polymorphism, as well as enumerate
104 the heterogeneity of amyloid samples, the structure of individual fibrils in a sample population
105 must first be resolved at sufficiently high-resolution so that individual twist patterns, twist
106 handedness and cross-sectional profiles are distinguishable at a single-fibril level without
107 cross-filament averaging. Here, using a force-curve based AFM imaging at low contact forces,
108 we are able to image, in sufficiently high detail, individual fibrils made from three different
109 peptides with sequences HYFNIF, VIYKI and RVFNIM. These peptides are short
110 amyloidogenic sequences originating within larger proteins (Human Bloom syndrome protein,
111 Drosophila Chorion protein and Human eIF-2, respectively ⁴⁰) identified via the Waltz
112 algorithm ⁴¹. Structural models of the amyloid protofilament cores from the Waltz peptide
113 assemblies have been presented previously (**Figure 1**) which were validated using X-Ray Fibre
114 Diffraction (XRFD) data, but these assembly reactions have also been previously observed to
115 produce a large range of fibril polymorphs upon assembly ⁴⁰. Thus, they provide excellent
116 model systems for investigating the structural basis of polymorphism within amyloid

117 populations. Here, using the AFM image data sets collected, each individual fibril observed in
118 AFM images were reconstructed as a distinct 3D model, and the structural parameters of
119 individual fibrils in the sample populations were measured and compared. Finally, we
120 employed an agglomerative hierarchical clustering method to classify fibril polymorphs by
121 measuring the structural differences observed between individual fibrils. The results show that
122 variation in the degree of structural polymorphism and the heterogeneity of amyloid fibril
123 samples are highly sequence specific. More importantly, the data and analysis here demonstrate
124 that that each individual amyloid fibril has characteristics that are different to those of the
125 population average.

126

127

128 **RESULTS**

129

130 **Gentle force-curve based AFM imaging identifies distinct fibril polymorphs to high detail**

131

132 Considerable heterogeneity has been displayed in several amyloid fibril samples previously
133 ^{26,27}, including in samples made from Waltz peptide amyloid assembly reactions ⁴⁰. To achieve
134 the high level of detail required for quantitative structural analysis for individual amyloid fibrils
135 without cross-particle averaging, we imaged the three Waltz peptide assembly samples,
136 HYFNIF, RVFNIM and VIYKI, using a force-curve based AFM imaging method. AFM is a
137 high signal-to-noise method that has been used previously to image amyloid fibrils at high
138 resolution ⁴²⁻⁴⁵. Here, force-curve based imaging (Peak-force tapping mode, Bruker) was
139 employed rather than traditional tapping mode imaging so that the force applied to the sample
140 was kept consistent and minimal ensuring that the specimens were not deformed and that the
141 surface details were tracked faithfully. Typical images of the three Waltz peptide assembly

142 samples collected using force-curve based imaging are shown in **Figure 2**. Closer inspection
143 of these images shows that fine details such as the length and handedness of repeating patterns
144 in the fibrils can be readily observed without any further processing of the raw image data.
145 Qualitatively, it is clear that there is a high degree of heterogeneity in these samples since many
146 fibrils of different heights and twist patterns, including some with different handedness can be
147 observed from within the same image.

148

149 Around 90 well-separated fibrils for each of the Waltz peptide assemblies were chosen from
150 the images. Their contours along the filament centres were individually traced and digitally
151 straightened. **Figure 3** shows eight typical 500 nm long fibril segment examples from each of
152 the three data sets (almost all traced fibrils were longer than 500 nm, only a portion of the traces
153 are shown in **Figure 3**). All three peptide precursors result in a set of visibly distinguishable,
154 unique fibrils. All of the traced fibrils are displayed in **Supplementary Figure SI 1**.

155

156 **3D models of each individual amyloid fibril in the heterogeneous samples can be**
157 **reconstructed**

158

159 An informative method for visualising and comparing polymorphic filament structures in high
160 detail is to reconstruct 3D models of the filaments from the AFM image data. In this case,
161 because AFM is a high signal-to-noise imaging method, we were able to reconstruct 3D
162 structural models with sufficient detail without averaging across multiple filaments, thus
163 producing individual models for each of the traced fibrils. 3D reconstruction also allows for
164 visualising and comparing filament structures without the influence of the varying degree of
165 tip-sample convolution on the AFM images due to varying cantilever tip dimensions across
166 different images (evident in the varied apparent width of the fibrils within the images in **Figure**

167 **3).** **Figure 4** shows examples of the 3D fibril models from the three data sets. All 266 models
168 are shown in **Supplementary Figure SI 2**. In addition, the 3D models reveal the shape and
169 size of the cross-section of each of the fibrils and how the cross-section rotates along the length
170 of the fibril (**Supplementary Figure SI 3**). In this respect, the VIYKI data set stands out as all
171 of the fibril models show near-circular cross-sections. This is particularly striking when
172 compared to the HYFNIF and RVFNIM data sets in which there are numerous fibril types with
173 ellipsoidal cross-sections as well as a few models which display asymmetric features. In
174 summary, all three datasets display fibrils with varying degree of structural individualities, and
175 a range of polymorph classes that can be distinguished qualitatively.

176

177 **Analysis of height, twist periodicity and cross-sectional area enables the quantification of** 178 **heterogeneity due to fibril polymorphism**

179

180 Identifying the fibril polymorphs at the level of individual fibrils led to the qualitative
181 observation that all three samples display a high degree of heterogeneity. In order to quantify,
182 enumerate and compare the overall heterogeneity of the samples, we performed nano-
183 morphometric measurements on individual filaments in our datasets. For each of the ~90 fibrils
184 from each of the three datasets, six different structural parameters were measured
185 (**Supplementary Table 1, Figure 5**): maximum height (h_{max}), minimum height (h_{min}), average
186 height (h_{mean}), handedness, periodic frequency, and average cross-sectional area (csa). The
187 periodic frequency refers to the frequency of the most common repeating pattern observed in
188 the height profile of a straightened filament identified by Fast Fourier Transform (FFT) of the
189 height profile, corresponding to the twist pattern of the filaments. The frequency of the
190 periodicity per nm was used rather than the periodicity as this would allow any twisted “ribbon”
191 like fibrils with very long twist periodicity to be analysed and visualised in the same way as

192 fibrils with a very short periodicity. The handedness of the twist was determined separately by
193 manual inspection of the fibril images and 2D Fourier transform (2D power spectral density
194 map) of the fibril images. The periodic frequency of fibrils with left-handed twists were
195 assigned negative frequency values while the periodic frequency of right-handed twisted fibrils
196 retained positive values to give directional periodic frequency (*dpf*) values that enabled the
197 visualisation of the distribution between left and right handedness in each data set. The standard
198 deviation (SD) of the 5-dimensional (h_{mean} , h_{max} , h_{min} , *dpf* and *csa*) Euclidian distance to the
199 global mean for each dataset was subsequently calculated as a quantitative measure of sample
200 heterogeneity (Equation 2 in Materials and Methods). For the three peptide sequences studied
201 here, VIYKI peptide assembly showed the most structural heterogeneity with a SD of 2.57,
202 RVFNIM had a SD of 1.98 and HYFNIF showed the least structural heterogeneity with a SD
203 of 1.40.

204

205 In order to display the whole fibril population of morphometric measurements, the average
206 height was plotted against the directional periodic frequency as a contour map which is shown
207 in **Figure 6**. The contours represent the density of the data in specific regions (seen as two-
208 dimensional projections in **Figure 6**). The information gained from the contour map is
209 analogous to a plot of an energy landscape for the assembly reactions where deeper energy
210 wells represent more likely structures. As shown in **Figure 6**, both HYFNIF and VIYKI peptide
211 assembly reactions appear to be dominated by the formation of left-handed twisted fibrils.
212 Interestingly, in contrast, RVFNIM peptide assembly displays no overall dominant preference
213 for twist handedness. There is, therefore, a sequence specific energetic favourability for
214 forming predominantly left-handed or right-handed fibrils. In contrast to RVFNIM and VIYKI,
215 HYFNIF fibril polymorphs were mostly populated in one narrow region of the contour map
216 with a few outliers, suggesting that there was a deeper preference for one class of assembly

217 compared to the other two sequences. Fibrils made from the VIYKI peptide however, covered
218 a large range of polymorphs with different average heights and periodic frequencies,
219 populating a large area in the contour diagram in **Figure 6**, suggesting that there were multiple
220 classes of structure that were almost equally likely to occur, and therefore similar in terms of
221 the free energy associated with protofilament assembly. RVFNIM fibrils displayed an
222 intermediate level of heterogeneity, with some preference for specific classes of both left-hand
223 and right-hand twisted polymorphs. On a contour plot of the minimum height vs the maximum
224 height, shown in the **Supplementary Figure SI 4**, VIYKI fibrils showed a tendency to have
225 more similar values for minimum and maximum heights suggesting a preference for cross-
226 sections with a particularly rounded shape whereas HYFNIF and RVFNIM assembled into
227 fibrils that may have a larger maximum height than minimum height. This was also observed
228 when the average cross-sectional areas of the models were plotted against the average height
229 of the fibrils where the cross-sectional areas of VIYKI fibrils are distributed close to the line
230 expected for circular cross-sections (cross-sectional area proportional to width squared). These
231 observations support the qualitative assessment of the models in which the VIYKI fibrils
232 overall appear to be more cylindrical. In summary, the three different assembly reactions,
233 despite all resulting in amyloid fibrils with cross-beta core, display different assembly free
234 energy landscapes, which are influenced by their side chain composition. This results in
235 sequence specific preferences in the polymorphs that they form (e.g. in terms of width, twist
236 and handedness) is due to variations in the packing of the side-chains within the protofilaments,
237 the packing of the protofilaments in the fibrils or a combination of both, and results in
238 differences in the heterogeneity of the fibril populations.

239

240 **Amyloid fibril polymorphs can be classified by hierarchical clustering**

241

242 Classifying fibril polymorphs within a sample population can provide visualisation and
243 organisation of the single-fibril level structural data that allows for further analysis of the
244 activities and behaviours of types of structural polymorphs. Here, natural divisions in the data
245 set of amyloid fibrils can be found by identifying which fibrils have shared or similar features
246 and which fibrils do not, as determined by the nano-morphometric measurements.
247 Subsequently, agglomerative or ‘bottom-up’ hierarchical clustering was performed to resolve the
248 natural divisions within each dataset and therefore objectively classify the fibrils. The
249 standardised Euclidean distance between each possible pair of fibrils in 5-dimensions (average
250 height, maximum height, minimum height, directional periodic frequency and cross-sectional
251 area), was calculated and linked together iteratively to generate a linked tree representing the
252 structural and morphological relationships between individual fibrils. The morphometric data
253 can then be represented in a dendrogram with the clusters on the x-axis and the 5-dimensional
254 standardised Euclidean distance from one cluster to the next on the y-axis. In the
255 **Supplementary Figure SI 5**, full dendrograms displaying the entire trees of the three datasets
256 are presented. In order to separate the fibrils into ‘classes’, the data must be analysed at
257 appropriate distance cut-offs. The distance cut-off determines the maximum distance within
258 any clusters (or class). Therefore, the number of clusters at a given distance cut-off is indicative
259 of the heterogeneity of the data set. A greater number of clusters suggest that the data is more
260 spread out and therefore more heterogeneous. The dendrograms shown in **Figure 7a** have been
261 collapsed based on a distance cut-off of 1 (standardised Euclidean distance) show how the
262 datasets were subsequently classified in **Figure 7b and 7c**.

263

264 In total, at a distance cut-off of 1 in the standardised 5-dimensional Euclidean space, 13
265 separate clusters were generated by HYFNIF polymer assembly, 22 clusters were generated by
266 RVFNIM polymer assembly and 19 clusters were generated by VIYKI polymer assembly.

267 These clusters can be seen as classes of fibrils, where member fibrils of the same class show
268 similar structure in terms of their morphometric appearances. Further analysis of the data at
269 different distance cut-offs, shown in **Supplementary Figure SI 6**, corroborate that HYFNIF
270 fibrils display considerably less heterogeneity than RVFNIM and VIYKI. Interestingly, when
271 compared to VIYKI, the RVFNIM data appears to be spread more evenly resulting in larger
272 number of clusters compared to VIYKI despite showing smaller distance SD value. This
273 suggests that VIYKI forms fibril classes that are very different but structurally and
274 energetically similar within each class, whereas the RVFNIM fibrils are overall more similar
275 across fibril classes compared to VIYKI fibrils. This may indicate that the energy landscape
276 for the assembly reaction is flatter and more rugged for RVFNIM so that the structural
277 differences between possible RVFNIM fibril polymorphs are on a more continuous scale
278 compared with VIYKI fibrils with distinct but distant polymorphs.

279

280

281 **DISCUSSION**

282

283 High-resolution images of amyloid fibril samples assembled from three amyloidogenic peptide
284 sequences using a gentle, force-based AFM approach allowed for the detailed quantitative
285 structural identification of polymorphs at a single fibril level. Individual 3D reconstructed fibril
286 models were made for every traced fibril in the data set which included ~90 fibrils from each
287 peptide assembly and 266 fibrils in total. This demonstrates the utility of AFM as an imaging
288 method for the structural analysis of amyloid fibril populations at the single fibril level, and is
289 highly complementary to cryo-EM methodologies. Morphometry is the process of measuring
290 differences in the structures of objects and is often performed to measure differences in
291 anatomy such as physiological differences in the brain^{46,47}. Here, we perform quantitative nano-

292 morphometric measurements on individual fibrils in heterogeneous amyloid samples. Using
293 our AFM data, we were able to make quantitative measurements of different structural
294 parameters which allowed us to distinguish between different polymorphs of amyloid fibril
295 structure. Indeed, the data shows that each fibril in the amyloid samples is unique. The
296 advances reported here opens up the future possibility of analysing and comparing the
297 structures present in entire populations of amyloid fibril species on a single fibril level, thereby
298 opening up the possibility of linking population level and single fibril level properties. This
299 connection may be key in order to decipher the relationships between specific polymorphs or
300 structural properties and phenotypic behaviours or biological consequences of amyloid.

301

302 The overall heterogeneity of the samples made from each different Waltz peptide precursor
303 was quantified and enumerated. These results also provide an indication of the energy
304 landscape associated with the filament assembly reaction of each peptide. In the cases
305 compared here, it was revealed that VIYKI assembly resulted in an overall more heterogeneous
306 population than the other two peptides and that VIYKI and HYFNIF fibril assemblies are
307 dominated by the formation of left-handed fibrils, whereas RVFNIM assemblies shows no
308 overall preference to handedness as similar sized populations of both left-hand and right-hand
309 twisted fibrils were formed. It has been thought that left-handed predominance in amyloid
310 fibrils is a result of the natural curvature of the β -strands that make up the monomeric subunits
311 which comes from the backbone of the polypeptide chains from L-amino acids ⁴⁸. The results
312 here supports the idea that preference for left-handedness is not absolute and the preference for
313 handedness could be modulated by side-chain sequence ^{49,50}. In this case, the side-chains in
314 RVFNIM assemblies may counteract the natural curvature of the β -strand due to the influence
315 of specific side chains interactions in some polymorphic arrangements. Overall, the specific
316 energetic explanation for the difference in behaviour between the three peptide assemblies

317 remains unknown. However, the data here demonstrate the considerable influence of amino
318 acid sequence in filament assembly reactions and opens up the possibility of systematic
319 sequence-structure analysis of cross- β assemblies. Furthermore, utilising an agglomerative
320 hierarchical clustering approach, we were able to objectively assign individual fibrils to
321 different classes of fibril polymorph. By analysing the dendrograms that allowed visualisation
322 of the similarities and dissimilarities between individual fibrils, it was possible to identify the
323 natural divisions in the datasets and existence of clusters that existed in the data. This
324 information was then used to classify individual fibrils to a cluster. The number of clearly
325 defined clusters also corroborated the overall heterogeneity of the data sets. The distance
326 between the fibril classes may also reflect the possible plasticity of the types of polymorph
327 observed in each data set. For example, VIYKI fibrils are likely belong to a predominant type
328 of polymorph, which is very different to any other type of possible VIYKI fibril polymorph.
329 As a consequence, such a population if presented in biological context may not be able to shift
330 to a different polymorph easily under pressure from environmental changes. The most common
331 types of HYFNIF fibril, however, are very similar and so HYFNIF fibril populations may be
332 able to shift their morphology much more easily when conditions change. Thus, this work
333 demonstrates a general utility to quantify polymorphism in amyloid fibril data sets by AFM on
334 a single fibril level as a possible means to quantify and comparing fitness of individual amyloid
335 or prion polymorphs under any given set of conditions.

336

337 Structural polymorphism could account for some of the biological activities associated with
338 amyloid observed *in vivo*. Individual fibril polymorphs will have different physical properties.
339 For example, thicker fibril polymorphs with a circular cross section will have a different second
340 moment of inertia compared to thin polymorphs with a more oval shaped cross section, which
341 could result in differential stability to fibril fragmentation. Thicker fibrils with large surface

342 areas per length may provide more sites for secondary nucleation events than thinner fibrils
343 with small surface areas. Thus, fibril polymorphs displaying different properties would likely
344 have different biological activities, even formed from identical precursors and present within
345 the same population. In this case, individual fibrils within a population may also have higher
346 rates of fragmentation that could result in a greater likelihood for propagation ^{51,52} or
347 cytotoxicity ³⁴, or higher rates of secondary nucleation that could result in more cytotoxic active
348 species ³⁸. Various proteins and peptides with different amino acid sequences are
349 amyloidogenic ⁵³. Yet, despite all of them being capable of forming amyloid fibrils *in vivo*,
350 they display different biological activities ⁵⁴⁻⁵⁶. Therefore, the variation in heterogeneity caused
351 by polymorphism, as well as individual polymorphs could account for the differences in
352 biological activity between different amyloidogenic precursors.

353

354 Prion and prion-like amyloid fibrils are able to propagate between cells ^{57,58}. Segregation of
355 specific fibril polymorphs could result in the cell-specific propagation of polymorphs ⁵⁹ likely
356 manifest as strains in mammals or variants in yeast. This could result in different biological
357 activities in different cells infested with the same amyloid sample. Examples of this include
358 the identification of numerous strains of Tau, each of which causes a different pathology, in
359 different brain regions and propagates at different rates ⁶⁰ and the identification of two α -
360 synuclein strains which show the same type of strain dependent phenomena ⁵⁸. This is of
361 particular importance as some fibril polymorphs may react differently to potential inhibitors,
362 which could result in strains that become 'resistant' to therapeutics over time. The strains
363 phenomenon seen in prions and prion-like amyloid, in which structural characteristics
364 presumably determine the biological activity of the fibrils has been demonstrated in human
365 prions ⁶¹. This behaviour has also been implicated in the Alzheimer's related A β peptides ⁶². It
366 is, therefore, important to investigate amyloid polymorphism in heterogeneous populations

367 from a single fibril perspective, as this can account for differences in specific constructive or
368 destructive biological activities associated with amyloid structures since each individual
369 amyloid fibril particle has characteristics that may be vastly different to those of the population
370 average.

371

372

373 **MATERIALS AND METHODS**

374

375 **Waltz Peptide synthesis**

376 The Waltz peptides with the sequences HYFNIF, RVFNIM and VIYKI were purchased from
377 JPT peptide technologies, or the Biomolecular analysis facility at the University of Kent. The
378 peptides were synthesised with N-terminal acetylation and C-terminal amidation. Multistage
379 solid phase synthesis using Fmoc protection chemistry generated a lyophilised powder with >
380 95% purity as measured by HPLC.

381

382 ***In vitro* polymerisation**

383 To prepare amyloid fibril samples formed from the three Waltz peptides, 1 mg of the respective
384 lyophilised powder was dissolved in 100 μ l of filter sterilized milli-Q water to a final
385 concentration of 10 mg/ml. The solution was incubated at room temperature for 1 week prior
386 to imaging. This allowed for the ‘maturation’ of the fibrils from all three Waltz peptide samples
387 to occur within a standardised amount of time.

388

389 **AFM sample preparation and imaging**

390 Each Waltz peptide assembly sample was diluted to 0.05 mg/ml in a dilute solution of HCl (pH
391 2.0, made up using filter sterilised milli-Q water). Immediately after dilution, 20 μ l samples

392 were deposited onto freshly cleaved mica surfaces (Agar scientific, F7013) and incubated for
393 10 minutes. Following incubation, the sample was washed with 1 ml of filter sterilised milli-Q
394 water and then dried using a stream of nitrogen gas. Fibrils were imaged using a Multimode
395 AFM with a Nanoscope V (Bruker) controller operating under peak force tapping mode using
396 ScanAsyst probes (silicon nitride triangular tip with tip height = 2.5-2.8 μm , nominal tip radius
397 = 2 nm, nominal spring constant 0.4 N/m, Bruker). Each collected image had a scan size of 6
398 x 6 μm and 2048 x 2048 pixels or 12 x 12 μm and 4096 x 4096 pixels. A scan rate of 0.305 Hz
399 was used with a noise threshold of 0.5 nm and the Z limit was reduced to 1.5 μm . The peak
400 force set point was set automatically, typically to ~ 675 pN during image acquisition.
401 Nanoscope analysis software (Version 1.5, Bruker) were used to process the image data by
402 flattening the height topology data to remove tilt and scanner bow.

403

404 **Structural data extraction**

405 Fibrils were traced⁶³ and digitally straightened⁶⁴ using an in-house application and the height
406 profile for each fibril was extracted from the centre contour line of the straightened fibrils. The
407 periodicity of the fibrils was then determined using fast-Fourier transform of the height profile
408 of each fibril. Some fibrils do not display any clear periodicity and have a height profile that
409 appears to show an erratic pattern so frequency of the highest peak in the frequency domain
410 per nm was used. The final datasets consist of a varying multitude of images comprising ~ 90
411 individually traced fibrils per Waltz sequence. The same pixel density is maintained for all
412 images within the dataset.

413

414 **3D modelling of fibril structures**

415 Straightened fibril traces were corrected for the tip-convolution effect using an algorithm based
416 on geometric modelling of the tip-fibril contact points (Lutter et al., 2019a, in preparation).

417 Briefly, the variation in the tip radius from their nominal value was first determined. Tip radii
418 can be estimated by imaging standards such as gold nanoparticles with known dimensions ⁶⁶.
419 It is also important to consider that the tip can become blunter with scanning action and the tip
420 radius can widen over time. Therefore, the tip radius was estimated for each individual fibril
421 on an image from the extent of convolution seen in data by assuming the twisted amyloid fibrils
422 have ideal corkscrew symmetry with circular average cross-section of the fibril perpendicular
423 to its axis of rotation. The algorithm then corrects for the lateral dilation of nano-structures
424 resulting from the finite dimensions of the AFM probe, without the loss of structural
425 information, by resampling of the fibril images to tip-sample contact points. This results in
426 recovering subpixel resolution of lateral sampling. Each pixel value in the straightened fibril
427 data is then corrected in their x, y and z coordinates. Filament helical symmetry was estimated
428 by building 3D models with various symmetries from the data, back calculating a dilated AFM
429 image and comparing the angle of the fibril twist pattern with that of the straightened fibril
430 trace in the simulated images and in the 2D Fourier transform of the simulated images. Then
431 for construction of the 3D models, the degree of twist per pixel along the y-axis was found by
432 dividing 360° with the product of fibril periodicity and its symmetry number (e.g. 1 2 or 3 etc.).
433 The 3D models were made assuming a helical symmetry using a moving-window approach, in
434 which a window, centred at a pixel n , contained the pixels $n - x$ to $n + x$ where x is the axial
435 length covered by 180° twist. The central pixel n is not rotated while neighbouring pixels on
436 both sides along the y-axis are rotated by a rotation angle, which is the product of the twist
437 angle and the distance from n in pixels. Rotation angle values are negative in one direction
438 from n and positive in the other direction, with the specific direction depending on the
439 handedness of the fibril, determined by manual inspection of the straightened fibril image and
440 its 2D Fourier transform image.
441

442 **Nano-morphometric measurements on individual fibrils**

443 Fibril image datasets for RVFNIM, HYFNIF and VIKYKI were analysed with 92, 89 and 85
444 individually traced and characterised fibrils, respectively. Five morphometric parameters were
445 extracted from the image data for each of the fibrils and the 3D models of the individual fibrils
446 as shown in **Figure 5** The maximum, minimum and average heights were measured on the
447 central ridge of each fibril. The periodic frequencies were obtained by Fourier-transform of the
448 z-coordinates along the central ridge of each fibril. The average cross-sectional areas were
449 obtained by averaging of the numerical polar integration along the fibril axis of the
450 reconstructed fibril 3D models.

451

452 **Hierarchical Clustering**

453 The standardised Euclidean distance $d(x, y)$ in 5-dimensional space representing the five
454 morphometric parameters average height, maximum height, minimum height, directional
455 periodic frequency and average cross-sectional area (each data point consists of these five
456 values, representing a single fibril segment) was calculated for every possible pair of data
457 points (e.g. fibrils x and y) using equation (1) where V is the 5-by-5 diagonal matrix whose j th
458 diagonal element is the variance (standard deviation squared) for each of the five morphometric
459 parameters. In Equation (1), x and y are 1 by 5 row vectors representing each of the data points,
460 with the elements in the vectors representing each of the 5 morphometric parameters for each
461 individual fibril. Here, the diagonal elements of the V matrix contain the list of the variance
462 for each parameter individually across all 3 data sets.

463 *Equation (1)*

$$464 \quad d(x, y) = \sqrt{(x - y)V^{-1}(x - y)'}$$

465 As expressed in Equations (1), therefore, the Euclidian distances are standardised by dividing
466 the distance in each of the morphometric parameter with their global standard deviation. Hence,

467 the standardised Euclidean distance, d , between two fibril data-points is a unitless measure of
468 the similarity (or dissimilarity) between the two. The distance standard deviation SD of the
469 morphometric data as a measure of the heterogeneity of the fibril populations was defined as
470 the standard deviation of $d(x, \bar{x})$, which is the standardised Euclidean distance between each
471 of the data points and the mean of all the data points, as shown in equation (2).

472 *Equation (2)*

$$473 \quad SD = \sqrt{\frac{\sum_{i=1}^n (d(x_i, \bar{x}))^2}{n - 1}}$$

474 In Equation (2), n is the total number of data points. Agglomerative hierarchical clustering was
475 performed using the average linkage function shown in equation 3, which hierarchically links
476 clusters using the average distance between all pairs of data points in any two clusters.

477 *Equation (3)*

$$478 \quad d(r, s) = \frac{1}{n_r n_s} \sum_{i=1}^{n_r} \sum_{j=1}^{n_s} d(x_{ri}, x_{sj})$$

479 In Equation (3), $d(r, s)$ is the average distance between cluster r with n_r data points and cluster
480 s with n_s data points. Explained briefly, the shortest distance between any two data points
481 within a data set is found. Those two data points are then considered to be a cluster, and the
482 average coordinate of the two data points in the 5-dimensional space is used as the coordinate
483 of the cluster. This is then repeated until all of the data is linked under one cluster.

484

485

486 **AUTHOR CONTRIBUTIONS**

487 L.D.A. and B.J.F.B. designed the research, conducted the experiments, and analyzed the data.

488 L.L. wrote the analytical software tools. C.J.S. and M.F.T. designed the research and analyzed

489 the data. L.C.S. designed the research, provided assembly reagents and methods, and analyzed

490 the data. W.F.X. designed the research, wrote the analytical software tools, analyzed the data,
491 and managed the research. The manuscript was written through contributions of all authors.

492

493

494 **ACKNOWLEDGEMENTS**

495 We thank the members of the Xue group, the LCS group, the CJS group and the Kent Fungal
496 Group for helpful comments throughout the preparation of this manuscript. We also thank Ian
497 Brown and Kevin Howland for technical support. This work was supported by funding from
498 the University of Kent (B.J.F.B), Biotechnology and Biological Sciences Research Council
499 (BBSRC), UK grant BB/S003312/1, as well as Engineering and Physical Sciences Research
500 Council (EPSRC), UK DTP grant (EP/R513246/1 for L.L).

501

502

503 **REFERENCES**

- 504 1. Sunde, M. *et al.* Common Core Structure of Amyloid Fibrils by Synchrotron X-ray
505 Diffraction. *11* (1997).
- 506 2. Jiménez, J. L. *et al.* The protofilament structure of insulin amyloid fibrils. *Proc. Natl. Acad.*
507 *Sci. U. S. A.* **99**, 9196–9201 (2002).
- 508 3. Serpell, L. C. *et al.* The protofilament substructure of amyloid fibrils. *J. Mol. Biol.* **300**,
509 1033–1039 (2000).
- 510 4. Raleigh, D., Zhang, X., Hastoy, B. & Clark, A. The β -cell assassin: IAPP cytotoxicity. *J.*
511 *Mol. Endocrinol.* **59**, R121–R140 (2017).
- 512 5. Iadanza, M. G. *et al.* The structure of a β 2 -microglobulin fibril suggests a molecular basis
513 for its amyloid polymorphism. *Nat. Commun.* **9**, 1–10 (2018).

- 514 6. Iwata, K. *et al.* 3D structure of amyloid protofilaments of β 2-microglobulin fragment
515 probed by solid-state NMR. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 18119–18124 (2006).
- 516 7. Stoppini, M. & Bellotti, V. Systemic Amyloidosis: Lessons from β 2-Microglobulin. *J. Biol.*
517 *Chem.* **290**, 9951–9958 (2015).
- 518 8. Teng, M. H. *et al.* Amyloid and nonfibrillar deposits in mice transgenic for wild-type
519 human transthyretin: a possible model for senile systemic amyloidosis. *Lab. Investig. J.*
520 *Tech. Methods Pathol.* **81**, 385–396 (2001).
- 521 9. Zhao, L., Buxbaum, J. N. & Reixach, N. Age-related oxidative modifications of
522 transthyretin modulate its amyloidogenicity. *Biochemistry* **52**, 1913–1926 (2013).
- 523 10. Rokita, H., Shirahama, T., Cohen, A. S. & Sipe, J. D. Serum amyloid A gene expression
524 and AA amyloid formation in A/J and SJL/J mice. *Br. J. Exp. Pathol.* **70**, 327–335 (1989).
- 525 11. Goedert, M. Alpha-synuclein and neurodegenerative diseases. *Nat. Rev. Neurosci.* **2**, 492–
526 501 (2001).
- 527 12. Deshmukh, M., Evans, M. L. & Chapman, M. R. Amyloid by Design: Intrinsic Regulation
528 of Microbial Amyloid Assembly. *J. Mol. Biol.* **430**, 3631–3641 (2018).
- 529 13. Erskine, E., MacPhee, C. E. & Stanley-Wall, N. R. Functional Amyloid and Other Protein
530 Fibers in the Biofilm Matrix. *J. Mol. Biol.* **430**, 3642–3656 (2018).
- 531 14. Jackson, M. P. & Hewitt, E. W. Why are Functional Amyloids Non-Toxic in Humans?
532 *Biomolecules* **7**, 71 (2017).
- 533 15. Watt, B., Niel, G. van, Raposo, G. & Marks, M. S. PMEL: a pigment cell-specific model
534 for functional amyloid formation. *Pigment Cell Melanoma Res.* **26**, 300–315 (2013).
- 535 16. Fitzpatrick, A. W. P. *et al.* Cryo-EM structures of tau filaments from Alzheimer’s disease.
536 *Nature* **547**, 185–190 (2017).
- 537 17. Gremer, L. *et al.* Fibril structure of amyloid-beta(1-42) by cryo-electron microscopy.
538 *Science* **358**, 116+ (2017).

- 539 18. Guerrero-Ferreira, R. *et al.* Cryo-EM structure of alpha-synuclein fibrils. *eLife* **7**, e36402
540 (2018).
- 541 19. Eisenberg, D. S. & Sawaya, M. R. Structural Studies of Amyloid Proteins at the Molecular
542 Level. *Annu. Rev. Biochem.* **86**, 69–95 (2017).
- 543 20. Wälti, M. A. *et al.* Atomic-resolution structure of a disease-relevant A β (1–42) amyloid
544 fibril. *Proc. Natl. Acad. Sci.* **113**, E4976–E4984 (2016).
- 545 21. Tycko, R. Molecular Structure of Aggregated Amyloid- β : Insights from Solid-State
546 Nuclear Magnetic Resonance. *Cold Spring Harb. Perspect. Med.* **6**, (2016).
- 547 22. Falcon, B. *et al.* Novel tau filament fold in chronic traumatic encephalopathy encloses
548 hydrophobic molecules. *Nature* **568**, 420–423 (2019).
- 549 23. Falcon, B. *et al.* Tau filaments from multiple cases of sporadic and inherited Alzheimer’s
550 disease adopt a common fold. *Acta Neuropathol. (Berl.)* **136**, 699–708 (2018).
- 551 24. Nelson, R. *et al.* Structure of the cross- β spine of amyloid-like fibrils. *Nature* **435**, 773–
552 778 (2005).
- 553 25. Riek, R. & Eisenberg, D. S. The activities of amyloids from a structural perspective. *Nature*
554 **539**, 227–235 (2016).
- 555 26. Adamcik, J. & Mezzenga, R. Amyloid Polymorphism in the Protein Folding and
556 Aggregation Energy Landscape. *Angew. Chem. Int. Ed.* **57**, 8370–8382 (2018).
- 557 27. Fändrich, M. *et al.* Amyloid fibril polymorphism: a challenge for molecular imaging and
558 therapy. *J. Intern. Med.* **283**, 218–237 (2018).
- 559 28. Qiang, W., Yau, W.-M., Lu, J.-X., Collinge, J. & Tycko, R. Structural variation in amyloid-
560 β fibrils from Alzheimer’s disease clinical subtypes. *Nature* **541**, 217–221 (2017).
- 561 29. Liberta, F. *et al.* Cryo-EM fibril structures from systemic AA amyloidosis reveal the
562 species complementarity of pathological amyloids. *Nat. Commun.* **10**, 1–10 (2019).

- 563 30. Falcon, B. *et al.* Structures of filaments from Pick's disease reveal a novel tau protein fold.
564 *Nature* **561**, 137–140 (2018).
- 565 31. Fändrich, M., Meinhardt, J. & Grigorieff, N. Structural polymorphism of Alzheimer A β
566 and other amyloid fibrils. *Prion* **3**, 89–93 (2009).
- 567 32. Lutter, L., Serpell, C. J., Tuite, M. F. & Xue, W.-F. The molecular lifecycle of amyloid -
568 Mechanism of assembly, mesoscopic organisation, polymorphism, suprastructures, and
569 biological consequences. *Biochim. Biophys. Acta Proteins Proteomics* **1867**, 140257
570 (2019).
- 571 33. Meinhardt, J., Sachse, C., Hortschansky, P., Grigorieff, N. & Fändrich, M. A β (1-40)
572 fibril polymorphism implies diverse interaction patterns in amyloid fibrils. *J. Mol. Biol.*
573 **386**, 869–877 (2009).
- 574 34. Xue, W.-F. *et al.* Fibril Fragmentation Enhances Amyloid Cytotoxicity. *J. Biol. Chem.* **284**,
575 34272–34282 (2009).
- 576 35. Arosio, P., Knowles, T. P. J. & Linse, S. On the lag phase in amyloid fibril formation. *Phys.*
577 *Chem. Chem. Phys.* **17**, 7606–7618 (2015).
- 578 36. Cohen, S. I. A. *et al.* Proliferation of amyloid-42 aggregates occurs through a secondary
579 nucleation mechanism. *Proc. Natl. Acad. Sci.* **110**, 9758–9763 (2013).
- 580 37. Jucker, M. & Walker, L. C. Self-propagation of pathogenic protein aggregates in
581 neurodegenerative diseases. *Nature* **501**, 45–51 (2013).
- 582 38. Cohen, S. I. A. *et al.* A molecular chaperone breaks the catalytic cycle that generates toxic
583 A β oligomers. *Nat. Struct. Mol. Biol.* **22**, 207–213 (2015).
- 584 39. Vidakovic, L., Singh, P. K., Hartmann, R., Nadell, C. D. & Drescher, K. Dynamic biofilm
585 architecture confers individual and collective mechanisms of viral protection. *Nat.*
586 *Microbiol.* **3**, 26 (2018).

- 587 40. Morris, K. L. *et al.* Exploring the sequence–structure relationship for amyloid peptides.
588 *Biochem. J.* **450**, 275–283 (2013).
- 589 41. Maurer-Stroh, S. *et al.* Exploring the sequence determinants of amyloid structure using
590 position-specific scoring matrices. *Nat. Methods* **7**, 237–242 (2010).
- 591 42. Adamcik, J., Castelletto, V., Bolisetty, S., Hamley, I. W. & Mezzenga, R. Direct
592 Observation of Time-Resolved Polymorphic States in the Self-Assembly of End-Capped
593 Heptapeptides. *Angew. Chem. Int. Ed.* **50**, 5495–5498 (2011).
- 594 43. Goldsbury, C., Frey, P., Olivieri, V., Aebi, U. & Müller, S. A. Multiple assembly pathways
595 underlie amyloid-beta fibril polymorphisms. *J. Mol. Biol.* **352**, 282–298 (2005).
- 596 44. Lara, C., Handschin, S. & Mezzenga, R. Towards lysozyme nanotube and 3D hybrid self-
597 assembly. *Nanoscale* **5**, 7197–7201 (2013).
- 598 45. Reynolds, N. P. *et al.* Competition between crystal and fibril formation in molecular
599 mutations of amyloidogenic peptides. *Nat. Commun.* **8**, 1338 (2017).
- 600 46. Noble, K. G. *et al.* Family income, parental education and brain structure in children and
601 adolescents. *Nat. Neurosci.* **18**, 773–778 (2015).
- 602 47. Lerch, J. P. *et al.* Studying neuroanatomy using MRI. *Nat. Neurosci.* **20**, 314–326 (2017).
- 603 48. Harper, J. D., Lieber, C. M. & Lansbury, P. T. Atomic force microscopic imaging of seeded
604 fibril formation and fibril branching by the Alzheimer’s disease amyloid- β protein. *Chem.*
605 *Biol.* **4**, 951–959 (1997).
- 606 49. Dzwolak, W. Chirality and Chiroptical Properties of Amyloid Fibrils. *Chirality* **26**, 580–
607 587 (2014).
- 608 50. Usov, I., Adamcik, J. & Mezzenga, R. Polymorphism Complexity and Handedness
609 Inversion in Serum Albumin Amyloid Fibrils. *ACS Nano* **7**, 10465–10474 (2013).
- 610 51. Kundel, F. *et al.* Measurement of Tau Filament Fragmentation Provides Insights into Prion-
611 like Spreading. *ACS Chem. Neurosci.* **9**, 1276–1282 (2018).

- 612 52. Marchante, R. *et al.* The physical dimensions of amyloid aggregates control their infective
613 potential as prion particles. *eLife* **6**, e27109 (2017).
- 614 53. Takács, K., Varga, B. & Grolmusz, V. PDB_Amyloid: an extended live amyloid structure
615 list from the PDB. *FEBS Open Bio* **9**, 185–190 (2019).
- 616 54. Chiti, F. & Dobson, C. M. Protein Misfolding, Amyloid Formation, and Human Disease:
617 A Summary of Progress Over the Last Decade. *Annu. Rev. Biochem.* **86**, 27–68 (2017).
- 618 55. Hobley, L., Harkins, C., MacPhee, C. E. & Stanley-Wall, N. R. Giving structure to the
619 biofilm matrix: an overview of individual strategies and emerging common themes. *FEMS*
620 *Microbiol. Rev.* **39**, 649–669 (2015).
- 621 56. Wickner, R. B. *et al.* Yeast Prions Compared to Functional Prions and Amyloids. *J. Mol.*
622 *Biol.* **430**, 3707–3719 (2018).
- 623 57. Goedert, M., Eisenberg, D. S. & Crowther, R. A. Propagation of Tau Aggregates and
624 Neurodegeneration. *Annu. Rev. Neurosci.* **40**, 189–210 (2017).
- 625 58. Bousset, L. *et al.* Structural and functional characterization of two alpha-synuclein strains.
626 *Nat. Commun.* **4**, 2575 (2013).
- 627 59. Bateman, D. A. & Wickner, R. B. The [PSI⁺] Prion Exists as a Dynamic Cloud of Variants.
628 *PLOS Genet.* **9**, e1003257 (2013).
- 629 60. Kaufman, S. K. *et al.* Tau Prion Strains Dictate Patterns of Cell Pathology, Progression
630 Rate, and Regional Vulnerability In Vivo. *Neuron* **92**, 796–812 (2016).
- 631 61. Safar, J. G. *et al.* Structural Determinants of Phenotypic Diversity and Replication Rate of
632 Human Prions. *PLOS Pathog.* **11**, e1004832 (2015).
- 633 62. Cohen, M., Appleby, B. & Safar, J. G. Distinct prion-like strains of amyloid beta implicated
634 in phenotypic diversity of Alzheimer’s disease. *Prion* **10**, 9–17 (2016).

- 635 63. Xue, W.-F., Homans, S. W. & Radford, S. E. Amyloid fibril length distribution quantified
636 by atomic force microscopy single-particle image analysis. *Protein Eng. Des. Sel.* **22**, 489–
637 496 (2009).
- 638 64. Egelman, E. H. An algorithm for straightening images of curved filamentous structures.
639 *Ultramicroscopy* **19**, 367–373 (1986).
- 640 65. Lutter, L., Serpell, C. J., Tuite, M. F., Serpell, L. & Xue, W.-F. Deconvolution and 3D
641 reconstruction of nano-filaments from atomic force microscopy topographs - Manuscript
642 in preparation. (2019).
- 643 66. Hill, S. E., Robinson, J., Matthews, G. & Muschol, M. Amyloid Protofibrils of Lysozyme
644 Nucleate and Grow Via Oligomer Fusion. *Biophys. J.* **96**, 3781–3790 (2009).

645

646

647 **Figure legends**

648

649 **Figure 1. Amyloid core models of protofilaments formed from the three Waltz peptides.**

650 Predicted models of amyloid protofilaments made from each of the 3 Waltz peptides, HYFNIF
651 (a), RVFNIM (b) and VIYKI (c). These models were generated and validated by comparing
652 simulated and experimental XRFD data as detailed in *Morris et al*⁴⁰.

653

654 **Figure 2. High resolution AFM imaging of Waltz peptide assemblies.**

655 Fibrils generated by each peptide were deposited onto freshly cleaved mica and imaged using peak force tapping
656 mode AFM. Each row shows representative image data from each assembly reaction, HYFNIF
657 (a), RVFNIM (b) and VIYKI (c). The original 6 μm by 6 μm images are shown in the left
658 column and each subsequent column shows a 2-fold increase in magnification indicated by the
659 white boxes. Different structural polymorphs are readily visible at high magnification including

660 different twist patterns and height profiles. The colour scale represents the height range from
661 0 to 12.5 nm and the scale bar represents 1 μm in all images.

662

663 **Figure 3. Comparison of distinct polymorphs from straightened fibril image data.** Around
664 90 fibrils from each data set were traced and 8 examples from each data set, HYFNIF (a),
665 RVFNIM (b) and VIYKI (c), are displayed. The traced fibrils displayed here were straightened
666 and cropped to 500 nm segments, and no further processing occurred. Qualitatively, each fibril
667 can be distinguished from all of the other fibrils in each data set. Different twist patterns and
668 average heights were readily visible. All of the fibril images analysed (around 90 fibrils for
669 each data set) are displayed in the **Supplementary Figure SI 1**.

670

671 **Figure 4. 3D models of Waltz peptide assemblies.** A selection of typical reconstructed fibril
672 3D models is shown. Each fibril in the three data sets was reconstructed as a 3D models using
673 information and 3D coordinates extracted directly from the AFM data. The average cross-
674 sectional area and the helical symmetry was determined from the generation of each 3D model.
675 Models for 8 fibrils from each assembly reaction, HYFNIF (a), RVFNIM (b) and VIYKI (c)
676 are displayed here. The models displayed here represent the same fibrils displayed in the image
677 data in **Figure 3**. All of the models are shown with identical scale and the colour scale represent
678 the local radius to the screw axis for visualisation. All of the models (around 90 fibrils for each
679 data set) are displayed in the **Supplementary Figure SI 2**.

680

681 **Figure 5: Nano-morphometry on individual fibrils results in quantitative structural**
682 **parameters.** (a) Two example images of straightened fibrils, including a left-hand twisted
683 fibril (left) and a right hand twisted fibril (right), are shown at the top of the figure. The
684 corresponding height profiles across the centre line of each straightened fibril are shown below

685 them. The minimum, maximum and average heights can be determined directly from the height
686 profiles. (b) FFT of the height profiles are shown, with peaks represent the periodicity
687 describing the repeating units in the height profile. The average length covered by the repeating
688 unit, representing the periodicity of the fibrils, was extracted from this analysis. Because the
689 two example fibrils have different twist handedness, the directional periodic frequency (*dpf*)
690 was assigned as a negative value for left-hand and a positive value for right-hand twisted fibrils.
691 (c) A schematic diagram illustrating the quantitative structural parameters obtained from each
692 individual fibril. The solid black line at the bottom representing the mica surface during AFM
693 imaging. The parameters measured include the average height (h_{mean}), the minimum height
694 (h_{min}), the maximum height (h_{max}), the directional periodic frequency (*dpf*), and the cross-
695 sectional area (*csa*).

696

697 **Figure 6. Comparing the heterogeneity of the polymorphic Waltz peptide assemblies.** (a)
698 The average height of the fibrils plotted against the number of repeating units per nm, with
699 negative and positive values to distinguish handedness (directional periodic frequency, *dpf*).
700 (b) The average cross-sectional area of the fibrils plotted against the average height. The data
701 is represented as a smoothed 2D histogram and visualised as a contour map, where the
702 colouring represents the density of the data-points. The HYFNIF peptide assembly reaction
703 favours one region, whereas the RVFNIM and VIYKI data was distributed across multiple
704 regions. HYFNIF and VIYKI were also predominantly left-handed whereas the RVFNIM
705 fibrils were almost evenly split between left and right-handed. See the **Supplementary Figure**
706 **SI 4** for visualisation of the data in other pairs of structural parameters.

707

708 **Figure 7 Natural separations in the clustered data define fibril polymorph classes.** The
709 standardised Euclidean distance was measured in the 5-dimensional space between every

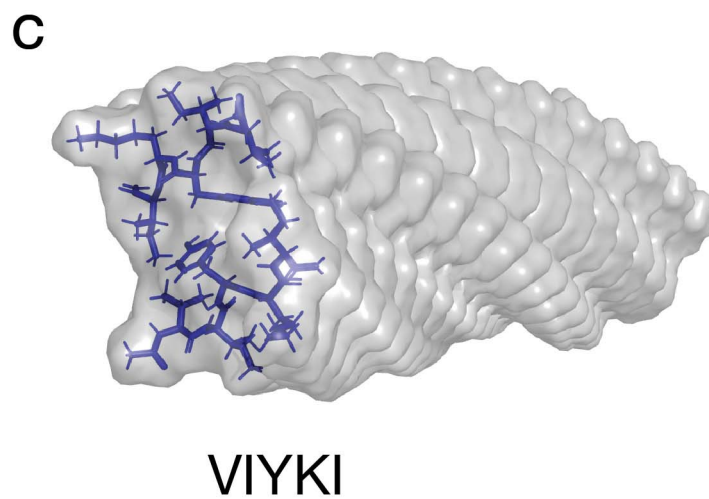
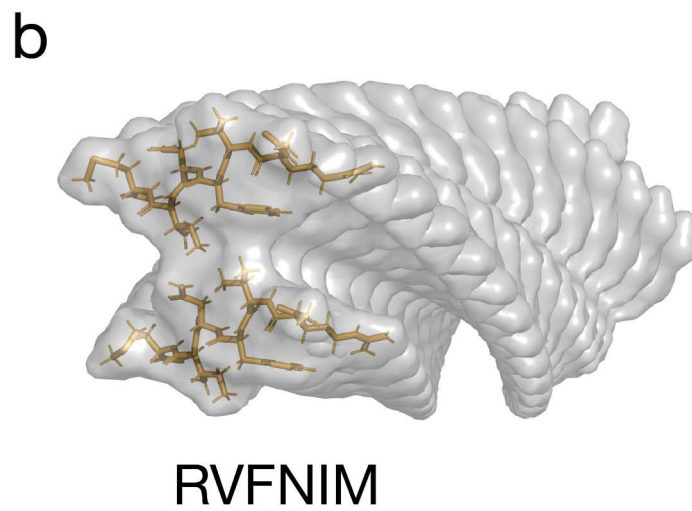
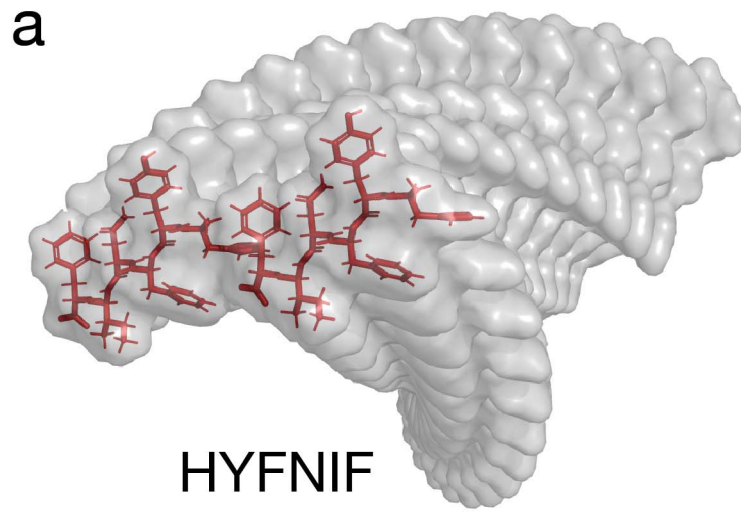
710 possible pair of fibrils in each data set and the data was separated into clusters. (a) The top row
711 shows dendrograms in which the x-axis represents the clusters that were generated, and the y-
712 axis is the standardised Euclidean distance between each cluster. The four data clusters with
713 largest numbers of fibril members are indicated as circles with size corresponding to cluster
714 size, and typical fibril structural models for each of these clusters are shown below the
715 dendrograms. In the dendrograms, the overall distance required to cluster each entire data set
716 reflects the overall heterogeneity of the data. (b) Scatter plot of clusters shown as spheres for
717 directional periodic frequency vs average height where the data points are coloured dependent
718 on the clustering at the cut-off level shown by the red line on the dendrograms above (1
719 standardised Euclidean distance). (c) 3D scatter plot of clusters, with directional periodic
720 frequency vs minimum height vs maximum height with the same colouring of the clusters and
721 projections of the various 2D plots projected onto the back walls of the plot. At this cut-off
722 level, there were 13 clusters in the HYFNIF data set, 22 clusters in RVFNIM and 19 clusters
723 in VIYKI with some clusters containing proportionally more of the data than others, here
724 visualised as sphere size. Further visualisations of the hierarchical clustering analysis are
725 shown in the **Supplementary Figure SI 5 and 6**.

726

727

728

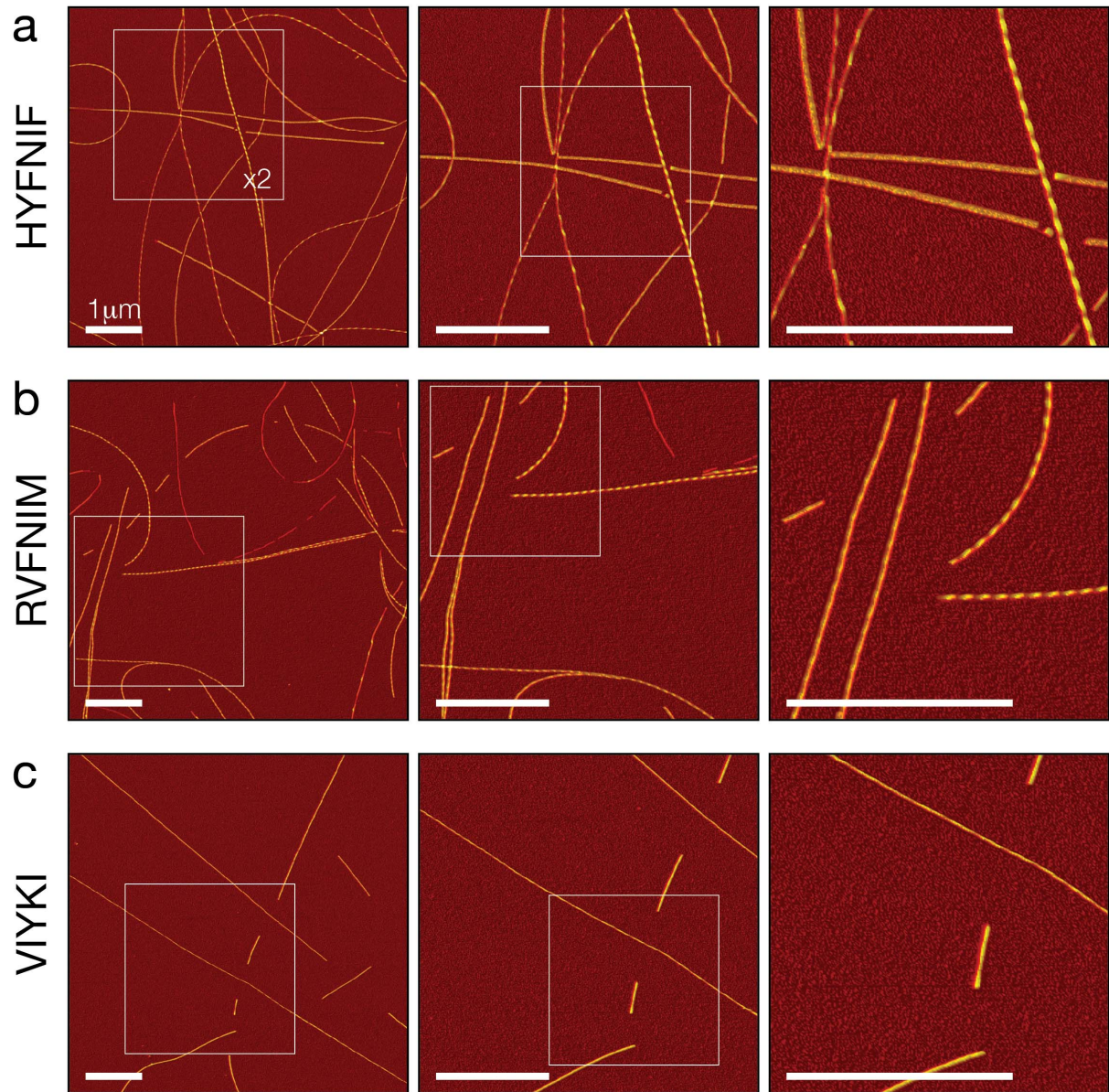
729



730

731 **Figure 1.**

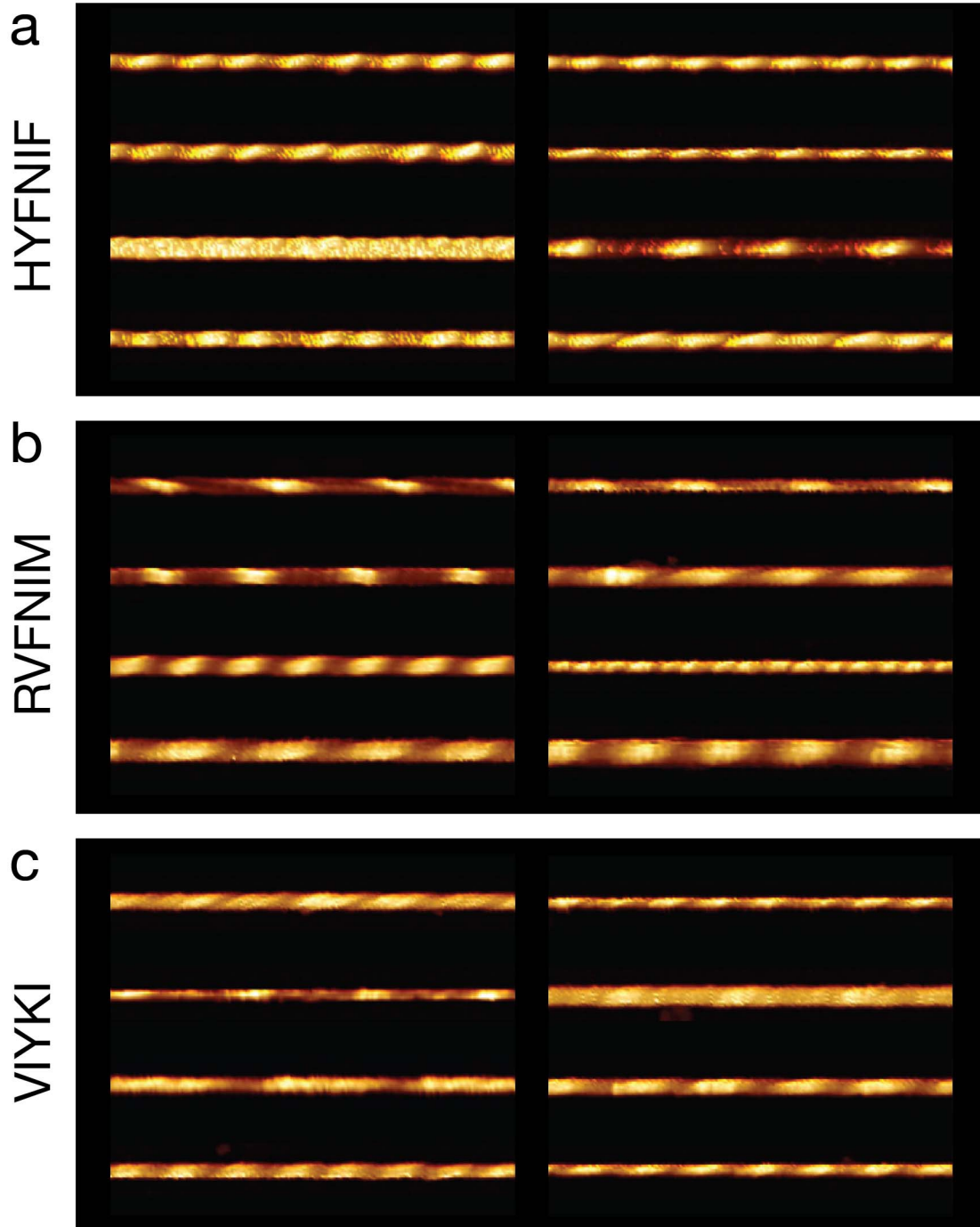
732



733

734 **Figure 2.**

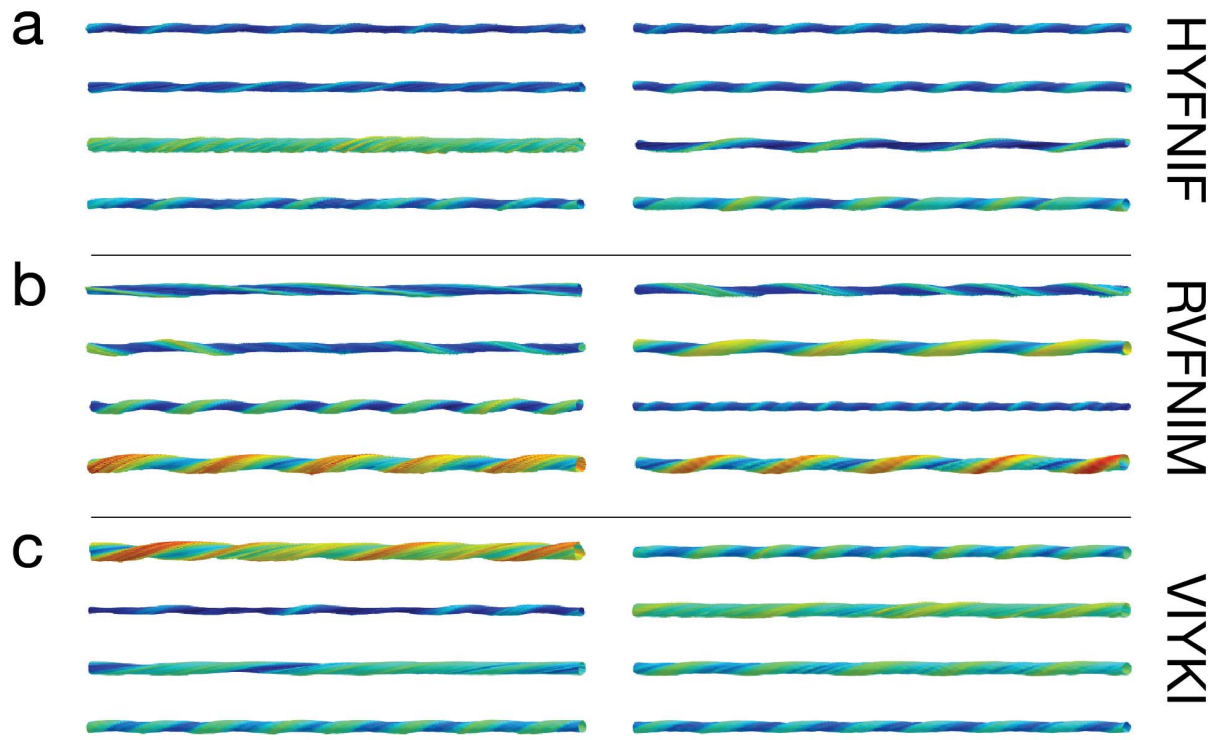
735



736

737 **Figure 3.**

738



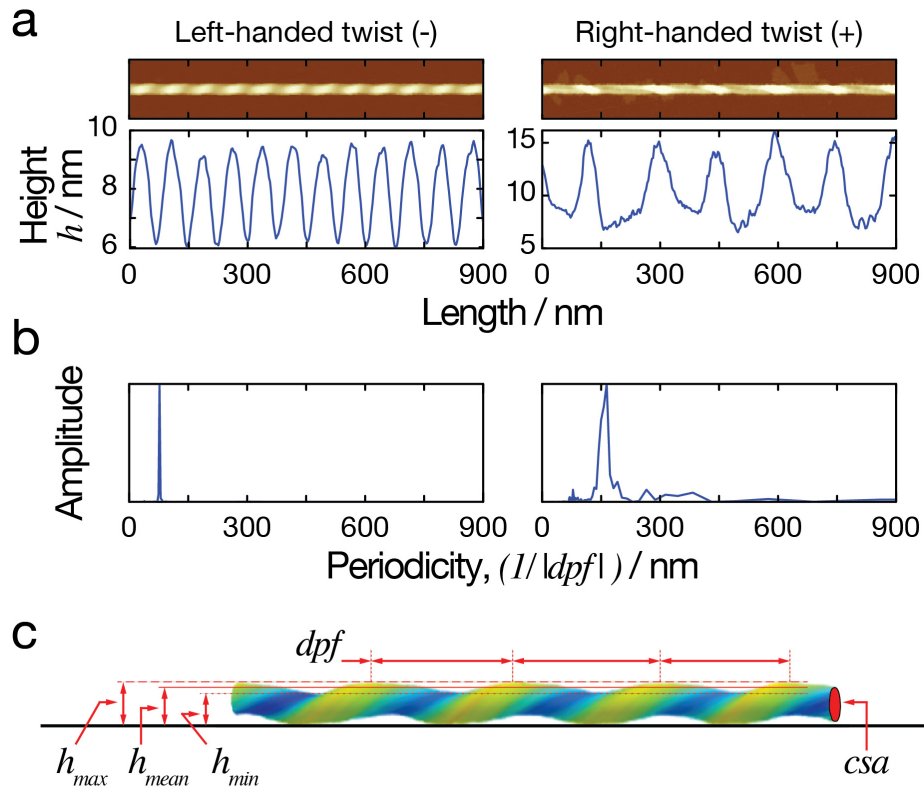
739

740 **Figure 4.**

741

742

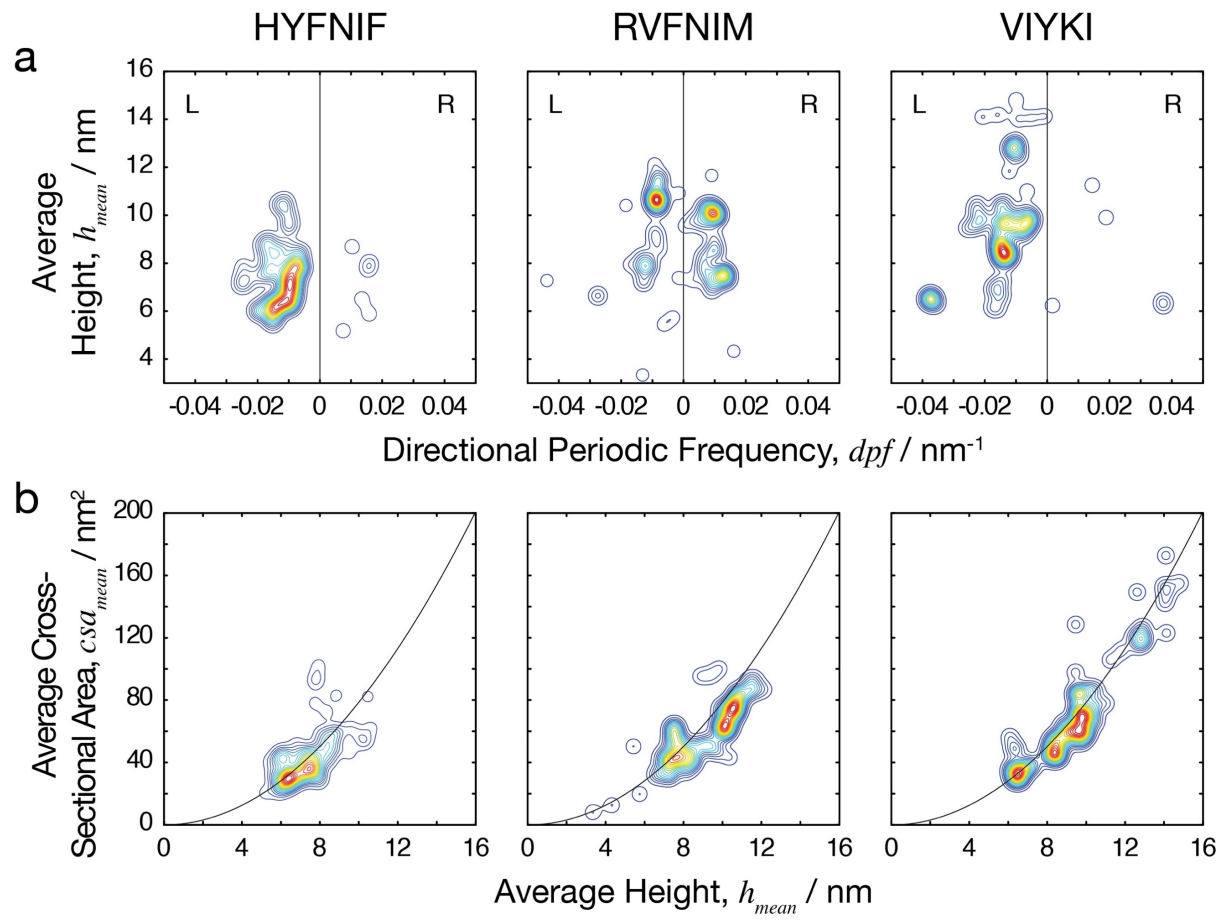
743



744

745 **Figure 5.**

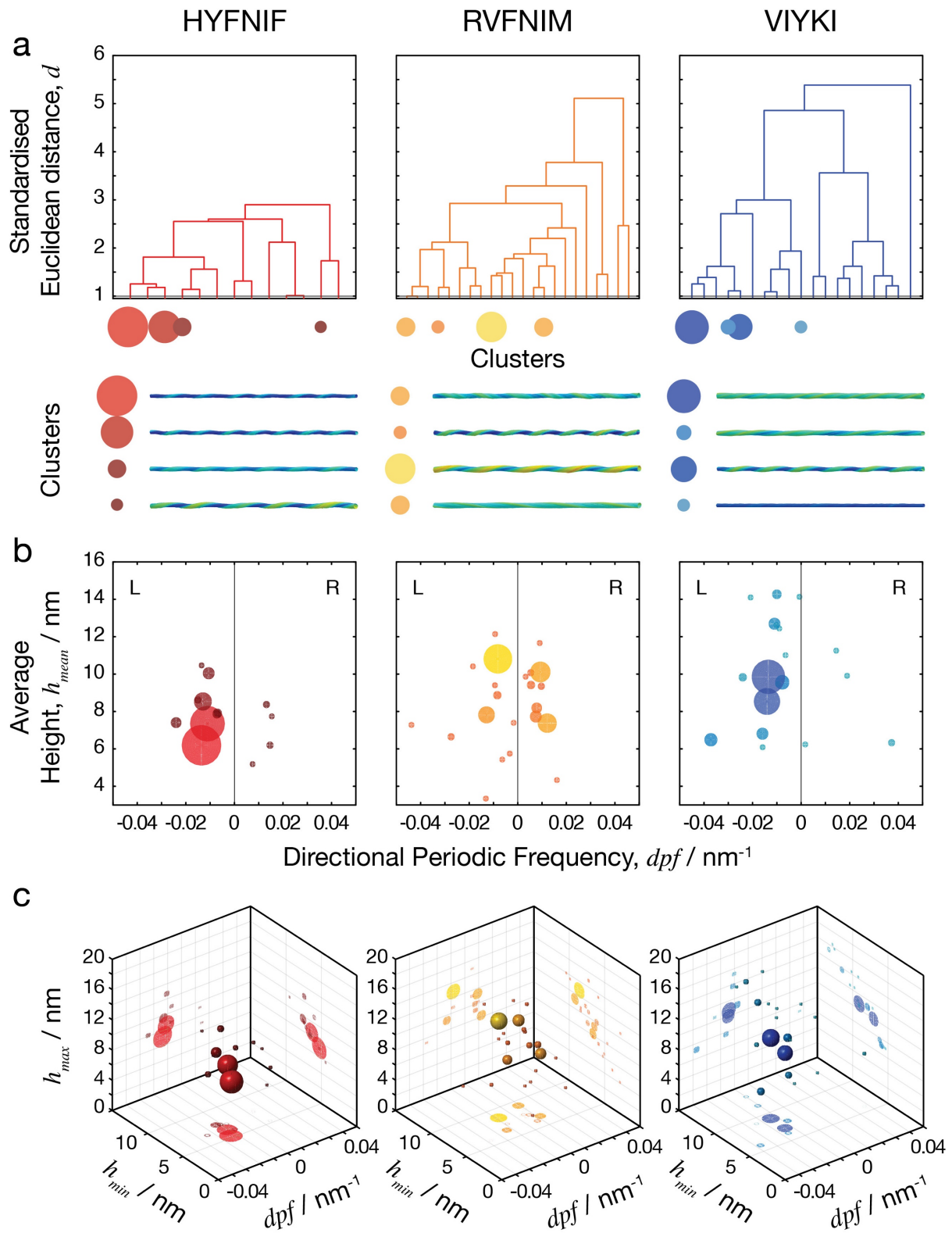
746



747

748 **Figure 6.**

749



750

751 **Figure 7.**

752

753