Continuum dynamics and statistical correction of compositional heterogeneity in multivalent IDP oligomers resolved by single-particle EM

Barmak Mostofian[1,†], Russell McFarland[2,†], Aidan Estelle[3], Jesse Howe[3], Elisar Barbar[3,*], Steve L. Reichow[2,*], Daniel M. Zuckerman[1,*]

[1] Department of Biomedical Engineering, Oregon Health and Science University, Portland OR 97239, U.S.A.

[2] Department of Chemistry, Portland State University, Portland OR 97201, U.S.A.

[3] Department of Biochemistry and Biophysics, Oregon State University, Corvallis, Oregon 97331

† Indicates equal contribution

* Correspondence: D.M.Z. (zuckermd@ohsu.edu); S.L.R (reichow@pdx.edu); E.B. (elisar.barbar@oregonstate.edu)

Classification: Biophysics and Computational Biology

Keywords: Intrinsically disordered proteins (IDP), multivalency, LC8, electron microscopy (EM), single-molecule

This PDF file includes:

Main Text

Figures 1 to 6

Supplemental Figures and Tables

Abstract

Multivalent intrinsically disordered protein (IDP) complexes are prevalent in biology and control diverse cellular functions, including tuning levels of transcription, coordinating cell-signaling events, and regulating the assembly and disassembly of complex macromolecular architectures. These systems pose a significant challenge to structural investigation, due to the continuum dynamics imparted by the IDP and compositional heterogeneity resulting from characteristic low-affinity interactions. Traditional single-particle electron microscopy (EM) is a powerful tool for visualizing IDP complexes. However, the IDPs themselves are typically "invisible" by EM, undermining methods of image analysis and structural interpretation. To overcome these challenges, we developed a pipeline for automated analysis of common 'beads-on-a-string' type of assemblies, composed of IDPs bound at multivalent sites to the ubiquitous ~20 kDa cross-linking hub protein LC8. This approach quantifies conformational and compositional heterogeneity on a single-particle basis, and statistically corrects spurious observations arising from random proximity of bound and unbound LC8. After careful validation of the methodology, the approach was applied to the nuclear pore IDP Nup159 and the transcription factor ASCIZ. The analysis unveiled significant compositional and conformational diversity in both systems that could not be obtained from traditional single particle EM class-averaging strategies, and shed new light on how these architectural properties contribute to their physiological roles in supramolecular assembly and transcriptional regulation. Ultimately, we expect that this approach may be adopted to many other intrinsically disordered systems that have evaded traditional methods of structural characterization.

Significance Statement

Intrinsically disordered proteins (IDPs) or protein regions (IDRs) represent >30% of the human proteome, but mechanistically remain some of the most poorly understood classes of proteins in biology. This dearth in understanding stems from these very same intrinsic and dynamic properties, which make them difficult targets for quantitative and structural characterization. Here, we present an automated approach for extracting quantitative descriptions of conformational and compositional heterogeneity present in a common 'beads-on-a-string' type of multivalent IDP system from single-particle images in electron micrographs. This promising approach may be adopted to many other intrinsically disordered systems that have evaded traditional ensemble methods of characterization.

MAIN TEXT

INTRODUCTION

The role of intrinsically disordered proteins (IDPs) in organizing multivalent recruitment of regulatory proteins has been established in a wide range of systems, from metabolic enzymes, signal transduction scaffolds, kinases and gene regulation (1, 2). This range of functionality is made possible by the unique degree of conformational plasticity exhibited by IDP platforms that can be exploited for the recruitment of multiple binding partners with temporally regulated assembly. With the additional potential for tight control by post-translational modifications, IDP systems provide ideal substrates for their roles in cellular regulation (1, 3, 4). Despite the prevalence of IDPs (constituting as much as 1/3 of the human proteome; (5)) and their critical roles in cellular regulation, these systems remain some of the most mechanistically enigmatic and poorly understood components of molecular biology. This dearth in understanding stems from the very same intrinsic and dynamic properties that make multivalent IDP systems difficult targets for quantitative and structural characterization. The continuous and highly diverse conformational heterogeneity, in combination with often transient and/or multivalent binding properties that enable rapid and responsive regulatory roles, are notoriously difficult to characterize, as these features are often lost by traditional ensemble methods of structural characterization.

Our focus is on multivalent IDP strands which can form a duplex ladder-like assembly, reversibly cross-linked by the LC8 hub protein (DYNLL1) that form the 'rungs' of the ladder-like assembly (Fig. 1). Although the highly stable homodimer LC8 was originally characterized in complex with dynein (6, 7), a much broader role has now been well established with over 100 IDP binding partners in the cell (8, 9), impacting nucleopore assembly (10, 11), regulation of mitochondrial apoptosis (12), signal transduction (13), gene regulation (14-16) and many other processes (8, 17, 18). The complex binding and heterogeneity of LC8/IDP systems appears to be at the heart of its diverse functional roles: compositional heterogeneity is responsive to post-translational modifications and local LC8 concentration - *e.g.,* affecting spindle positioning in mitosis (19) and modular sensing for transcriptional activity (20), while large-scale conformational heterogeneity provides plasticity that is required for dynamic molecular machines, such as the dynein motor complex (21).

Despite these established features of LC8/IDP systems, crucial structural and mechanistic knowledge gaps remain due to the inherent dynamical properties and transient formation of multiple oligomeric states that are key to their cellular function (22, 23). Notably, quantification of structural and compositional heterogeneity is lacking. Indeed, the dynamic nature of the disordered LC8/IDP complexes render structural determination by crystallography intractable. Aggregation, limited solubility, and conformational heterogeneity add to the challenges for characterization by NMR (24, 25). Previous analysis of the

106 LC8/Nup159 system by single particle EM have been successful and visualizing the fully assembled
107 oligomer, where five LC8 dimers appear uniformly arranged into a ladder-like assembly in two-dimensional
108 class averages (26). However, the underlying complexity of conformational and configurational states
109 present in this system were not characterized in this study.
110
111 Significant developments have been made in the field of high-resolution single-particle CryoEM image
112 analysis for the characterization of conformational heterogeneity (*e.g.,* (27-33)). While extremely powerful,
113 these approaches are most effective at characterizing large complexes (typically >100 kDa) needed to
114 generate the necessary contrast in CryoEM images for accurate 3D alignment and are most effective at
115 resolving discrete states of conformational heterogeneity. However, the small size of the LC8 dimer (~20
116 kDa), coupled with the broad continuum of conformational states in LC8/IDP complexes make these
117 "beads-on-a-string" systems intractable to current high-resolution methods in CryoEM. Recently, we
118 leveraged the high-contrast (and low-resolution) method of negative stain EM (NSEM) to directly visualize
119 the highly heterogeneous multivalent LC8/IDP assemblies formed by the transcription factor ASCIZ, which
120 regulates LC8's own cellular concentration (22). However, this workflow was tremendously labor-intensive
121 and subject to manual interpretation. Furthermore, quantifying the continuum of conformational states that
122 abrogated the validity of traditional 2D classification results could not be readily assessed by such manual
123 methods.
124
125 Here, we present an automated approach for single-particle *distribution* analysis, in contrast to standard
126 single-particle class averaging that can suppress heterogeneity, for quantifying the conformational and
127 compositional states of LC8/IDP 'beads-on-a-string' that are resolved in NSEM micrographs. To overcome
128 potential artifacts arising due to low-affinity interaction that are characteristic of LC8/IDP systems (K_d ~1 –
129 10 μM), we apply a statistical correction process to estimate the effects caused by random proximity of
130 free LC8 particles. The methodology was developed and validated using an artificial LC8/IDP system
131 designed with 4 equivalent LC8 binding sites (termed *syn*-4mer), and then applied to two biological IDP
132 systems, the nuclear pore protein Nup159 and the transcription factor ASCIZ (Fig. 1B). This approach
133 recapitulated previous results for ASCIZ obtained by intensive manual methods, while further correcting
134 over-estimation of small oligomeric species due to random proximity. For Nup159, we demonstrate a high-
135 degree of conformational and configurational heterogeneity that had been obscured by previous ensemble
136 methods of EM image analysis (26), and predict the presence of off-register type assemblies (Fig. 1C).
137 Ultimately, we expect that this method may be generalized to obtain quantitative measurements on
138 structure/assembly of many other 'beads-on-a-string' type multi-valent IDP systems.
139
140 **RESULTS**
141 **Traditional characterization of a synthetic 4-site IDP construct by single-particle EM**

142 To develop and validate our automated image analysis pipeline, we designed a synthetic 4-site LC8/IDP
143 system (termed *syn*-4mer) (Fig. 1A,B). The QT-binding motif used in this IDP construct is based on a
144 peptide sequence from the protein CHICA (34), selected for the reasonable binding affinity (K_d ~ 0.4 μM
145 for the single site peptide). Each QT-motif is separated by a model flexible linker design (Methods). A tight
146 LC8/*syn*-4mer interaction was validated by isothermal titration calorimetry (K_d of ~40 nM) and analytical
147 ultra-centrifugation (AUC) (Supplemental Fig. 1). For initial structural characterization, the purified complex
148 of LC8/*syn*-4mer was negatively stained and visualized by traditional single-particle EM image analysis
149 (Fig. 2A). As expected under the dilute conditions required for NSEM (*i.e.,* below the Kd of 40 nM), a
150 mixture of free LC8 dimers and assembled LC8/*syn*-4mer complexes were readily observed. Free LC8
151 dimers appear as small punctate densities (~5 nm diameter), while assembled oligomers appear as chains
152 of 2 – 4 LC8's separated by a characteristic spacing (4.7 ± 0.43 nm) dictated by the designed *syn*-4mer
153 IDP.
154
155 A total of ~4150 LC8/*syn*-4mer oligomers and free LC8 particles were extracted from 34 micrographs, and
156 subjected to traditional reference-free 2D classification procedures (Fig. 2B,C) (35). The most well resolved
157 classes depict free LC8 and various LC8/*syn*-4mer assemblies, consisting of two, three or four LC8 dimers
158 arranged in a nearly linear fashion (Fig. 2B). In addition to these 'ideal' classes, other conformational states
159 are resolved, depicting arched and/or corrugated assemblies that are most recognizable in complexes with
160 four bound LC8 dimers (Fig. 2C). Such variability is consistent with the range of conformational states
161 expected to be accessible by the flexible linkers connecting neighboring LC8 binding sites. However, the
162 degree of conformational heterogeneity resolved in 2D class averages appears to represent only a fraction
163 of conformational states presented in the raw EM micrographs. The limitations of this ensemble approach
164 is further apparent in several of the resulting 2D class averages, where LC8 densities often appear weak
165 or blurred due to the underlying heterogeneity present in the images contributing to the ensemble
166 representations (asterisks in Fig. 2C). Similar artifacts were present in our previously reported 2D class
167 averages of LC8/ASCIZ system (22), and other IDP assemblies (36-38). Such artifacts are characteristic
168 of EM 2D class averages where connected proteins, or domains, exhibit uncoupled and/or continuum
169 dynamic behavior.
170
171 This analysis demonstrates that while the traditional approach of 2D class averaging significantly improves
172 the overall signal-to-noise present in the raw images, the resulting average representations depict only a
173 fraction of the underlying structural heterogeneity that can be observed at the single-particle level for such
174 highly disordered beads-on-a-string type assemblies. Indeed, a majority of conformers that are part of the
175 continuum of states are not represented by these results. While additional insights into the underlying
176 conformational heterogeneity may be obtained by characterizing a much larger image dataset, such brute-
177 force approaches would still be challenged by the continuum dynamics that are characteristic of IDPs.

178

179 **Automated single-particle distribution analysis resolves the continuum of conformational states in**
180 **the LC8/*syn*-4mer system**

181 The limitations of class averaging methods described above inspired the development of an automated
182 image analysis pipeline that provides oligomer species populations and conformational distributions as
183 assessed at the single-particle level. Our single-particle distribution analysis builds on two pillars: first, a
184 straightforward, interpretable scoring function based on geometric and signal intensity criteria that is
185 trained on a small set of manually selected oligomers; and second, a novel self-consistency analysis
186 capable of correcting naïvely assigned oligomer populations based on the possibility of random proximity
187 of oligomers and free LC8 particles.

188

189 To facilitate this approach, we treated each LC8 dimer (*a.k.a.* bead) independently, that is by not assuming
190 the assembly state prior to analysis (Fig. 2D, white circles). The obtained coordinates of LC8 dimers were
191 then subjected to single-linkage clustering. A scoring function was then applied to all possible oligomers
192 within a cluster, with priority given to the largest possible oligomer that scored above a defined threshold.
193 The score threshold was set to a low (permissive) value based on calibrated bead-to-bead distances and
194 angles obtained from a small training dataset of hand-selected oligomers (Supplemental Fig. 4). Finally, a
195 distance filtering step is applied to avoid assignments within crowded regions of the micrograph, by setting
196 a minimum distance of 9 nm between assigned oligomers and other neighboring LC8 particles
197 (Supplemental Fig. 5).

198

199 This approach was applied to a test dataset of ~17k isolated LC8 particles retrieved from only 5
200 micrographs (Fig. 2D-G). The output of this automated analysis provides a quantitative geometric
201 description of the conformational state of each oligomer, defined by the center-to-center distance
202 separating neighboring LC8 dimers (*d*), and the bend angle (θ) defined by three neighboring LC8 'beads'
203 (Fig. 2E-G). The resulting coordinates were plotted to visualize a representative ensemble of
204 conformational states present in each oligomer class (Fig. 2F). The distribution of separation distances (*d*)
205 and bend angles (θ) for each class was very similar, and consistent with the symmetrical design of the *syn*-
206 4mer IDP (Fig. 2G). In each class, the average separation distance (*d*) was equal to ~4.8 nm (± 0.5 nm),
207 while the average bend angle (θ) was ~29° (± 20°).

208

209 The LC8-to-LC8 separation distances are consistent with those obtained in 2D class averages and with
210 the length of the synthetic IDP, designed with 15 residues separating each QT recognition motif (Fig. 1B).
211 A fully extended polypeptide of 15 residues would be expected to extend ~5.3 nm (*i.e.,* ~3.5 Å per residue),
212 while a completely random polypeptide chain would be expected to follow a random-walk distribution,
213 resulting in an average separation distance of ~1.4 nm (3.5 Å * $\sqrt{N_{residues}}$ (39)). Thus, the center-to-center

214 distance distribution obtained for the LC8/*syn*-4mer complexes suggests that the IDP adopts a primarily
215 extended state, with only partial random character. Such characteristics are consistent with atomic models
216 of LC8/IDP complexes where 10 amino acids of the QT recognition motif adopt an extended conformation
217 when bound to LC8, but with a 5 residue linker between recognition motifs remaining flexible (Fig. 1A and
218 Fig. 2H,I). This short flexible linkage is sufficient to facilitate the high degree of bend angles that lead to
219 the continuum of conformational states presented in Fig. 2F,I, and minimum end-to-end distances between
220 terminal LC8s that reach ~10 nm for the fully assembled 4-mer (Supplemental Fig. 4).
221
222 In comparison to the results obtained by traditional 2D class averaging methods, the single particle
223 distribution approach applied here harvested a much greater degree of conformational states (distribution
224 of bend angles), providing a much closer reflection of the heterogeneity observed in the raw micrographs.
225 The conformational flexibility of this IDP complex is dramatized by a series of movies concatenating
226 snapshots of oligomer images extracted from the micrographs (Supplemental Movies 1 and 2). This range
227 of conformational motion would be difficult if not impossible to extract from class-averaged data. At the
228 same time, geometric descriptions of the formed assemblies may be readily analyzed for quantitative
229 measures and/or comparison between systems to assess the effects of LC8 assembly onto a variety of
230 biological IDP scaffolds.
231
232 **Statistical correction for unbound LC8 particles refines species population profiles: 4-site system**
233 An additional strength of our single-particle workflow is the ability to extract quantitative counts of identified
234 species populations, which presents both new challenges and opportunities. Of particular interest is the
235 question of whether off-register (or daisy-chain) type assembly occurs in LC8/IDP systems (Fig. 1C), which
236 may be inferred by the presence of species with five or more assembled LC8 dimers in the *syn*-4mer
237 dataset. However, oligomer assignments based purely on visible criteria such as geometry and signal
238 intensity (labeled 'initial' assignments in Fig. 3), must be considered naïve because they cannot account
239 for the possibility of random association among oligomer and/or free LC8 species. For example, a bonafide
240 LC8/*syn*-4mer with four LC8's randomly deposited on the EM grid in close proximity to a free LC8 dimer
241 may be naïvely interpreted as evidence for off-register assembly because the IDP itself is not directly
242 resolved by NSEM. Therefore, the invisibility of IDPs stringing together LC8 particles requires additional
243 analysis of particle positions, extending ideas of correlation and Ripley's K function analysis (40).
244
245 To provide an estimate for the actual number of the underlying oligomers, the experimental process of
246 randomly depositing single LC8 particles onto the EM grid was iteratively simulated and the degree of
247 artifactual oligomer creation was evaluated in order to obtain a self-consistent estimate of the true
248 underlying oligomer populations (Fig. 3A). In every stage of the iterative process, synthetic micrographs
249 are generated by randomly positioning the population of free LC8 particles while the positions of initially

assigned oligomers remained fixed, and the resulting synthetic micrograph is re-classified and scored. This rescattering procedure inevitably alters the outcome of the classification process, as oligomers may be lengthened or created by random proximity of free LC8. By comparing these results to initial assignments, the putative list of oligomers can be iteratively refined, as detailed in Methods and Supplemental Fig. 6. Self-consistency is assessed by agreement of the pre-corrected populations from the synthetic micrographs (fluctuating gray lines in Fig. 3B) and the initial populations from the experimental micrograph (dashed black lines).

The results of the statistical correction analysis are presented in Fig. 3B,C. Populations of 3-mer and 4-mers are only mildly perturbed as compared to the naïve predictions; however, the population of 2-mers decreases dramatically, indicating a significant fraction of the original assignments reflected random proximity of free LC8 under these experimental conditions. Of greater interest, the analysis suggests that the population of 5-mers (*i.e.,* evidence of off-register/daisy-chain assembly) that were originally assigned are likely artifacts of random association between assembled 4-mers and free LC8 particles. In other words, we find it unlikely that off-register binding occurs in the 4-site system at the concentrations used for EM, fitting with the solution-state analysis by AUC that shows a homogeneous population assembly formed at much higher concentrations (Supplemental Fig. 1).

Validation of the single-particle distribution analysis routine

The oligomers initially assigned by our scoring algorithm were validated by comparison to manually assigned oligomers in two phases: *(i)* assessment of whether the automated pipeline recovered oligomers assigned manually by the microscopist, and *(ii)* microscopist assessment of the quality of additional automated assignments not originally selected by the microscopist (Supplemental Fig. 3). Our assessment explicitly acknowledges that we lack “ground truth” oligomer assignments because it is impossible to unambiguously distinguish background from noise, or to distinguish visually between random proximity and true oligomerization. The latter point, in fact, motivated the statistical correction analysis (described in the previous section). However, because our correction procedure is based on statistical inferences and not on observable structural features, the results of this approach could not be validated in a similar way.

The first validation analysis revealed that the automated pipeline recovered ~90% or more of manually assigned 2-mer, 3-mer, and 4-mer oligomers (Supplemental Fig. 4). A lower proportion (~75%) of manually assigned 5-mers were recovered, but as shown (Fig. 3), it is likely that the naïvely assigned 5-mers are the result of random association of smaller oligomers with free LC8s.

In addition, our automated analysis revealed the presence of oligomers beyond those identified by the microscopist. The second phase of validation revealed a range of phenomena that could be attributed to

286 the discrepancy between microscopist-identified complexes and the automated procedure (Fig. 4). Most
287 notably, among the discovered complexes (*i.e.,* those not originally identified by the microscopist), 40% or
288 more were found by the microscopist to be acceptable upon inspection (Fig. 4A,B), with the remainder
289 judged to be invalid (Fig. 4C,D). The invalid complexes, in turn, were approximately evenly split into two
290 cases: in the first group, some of the assigned LC8 density(s) were judged to be too ambiguous to permit
291 confident identification of an oligomer (*i.e.,* containing poorly resolved or weak LC8 particle density) (Fig.
292 4C). The second group of invalid assignments was characterized by overall unconvincing picks of single
293 LC8 particles (*i.e.,* deemed to be falsely picked particles that were not clearly distinguishable from
294 background noise) (Fig. 4D).
295
296 Hence, it appears that most automated oligomer assignments that were deemed by the microscopists to
297 be erroneous stemmed from unreliable particle picks, rather than from intrinsic problems with the oligomer
298 identification algorithm. Although significant care was taken to optimize the automated particle picking
299 parameters used in this study, spurious background picks were unavoidable. Indeed, the identification of
300 such small ~20 kDa particles in NSEM micrographs is often ambiguous via manual inspection as well, and
301 the automated approach we employed compared favorably to a variety of established particle picking tools
302 (see Methods). Importantly, the population of spurious background picks leading to erroneous assignments
303 is relatively small and may be further filtered by selecting a more stringent scoring threshold; the particle
304 picker is a fully modular component of our pipeline.
305
306 **Single-particle distribution analysis of LC8/Nup159**
307 With the protocol for single-particle distribution analysis validated using the LC8/*syn*-4mer system, we went
308 on to examine more heterogeneous and biologically relevant LC8/IDP assemblies. The first case was the
309 nuclear pore protein Nup159, an IDP which binds up to 5 LC8 dimers in a duplex fashion (Fig. 1) (11, 26).
310 In contrast to the LC8/*syn*-4mer system, LC8/Nup159 displays a high-degree of compositional
311 heterogeneity even under saturating conditions, as assessed by AUC (Supplemental Fig. 1). For our
312 comparative analysis by NSEM, we performed both traditional 2D class-averaging and the single-particle
313 distribution analysis protocol on a dataset obtained from 30 micrographs (Fig. 5 and Supplemental Fig. 7).
314 Scoring and distance thresholds were defined as described in Methods and Supplemental Fig. 8.
315
316 Traditional 2D classification analysis of ~5875 identified oligomers resolved a range of species,
317 representing free LC8 dimers to fully assembled LC8/Nup159 complexes with 5 bound LC8 dimers (Fig.
318 5A and Supplemental Fig. 7). The fully-assembled complex appears similar to what was previously
319 described by Stelter et al (26), with five LC8 dimers arranged in an ordered and nearly linear fashion, with
320 each LC8 dimer separated by 4.5 ± 0.66 nm. However, in addition to these well-resolved oligomeric
321 assemblies, many of the other 2D class averages demonstrate the presence of underlying conformational

322 heterogeneity, resulting in the appearance of blurred LC8 densities and arranged in a non-linear fashion
323 (Fig. 5A, asterisk and Supplemental Fig. 7).
324
325 The results of our automated single-particle distribution analysis once again revealed the extent of
326 significant conformational heterogeneity that is apparent in the raw micrograph images (Fig. 5B and
327 Supplemental Fig. 7). The quantified separation distances between LC8 particles were similar among the
328 classified oligomer states, ranging from 5.0 ± 1.0 to 5.28 ± 1.0. Again, such distances are consistent with
329 the length of linkers separating the 10-residue long QT-motifs (2 – 15 residue linkers, see Fig. 1) and the
330 expected percentage of disordered versus structured IDP character that is induced by LC8 binding. The
331 longer linker region within the Nup159 IDP (separating motifs 2 and 3), as compared to the *syn*-4mer,
332 appear to result in a broader distribution of bend angles, which range from 39° ± 26 to 44° ± 27 in the
333 Nup159/IDP assemblies (Fig. 5B). The result of this flexibility is an apparent continuous ensemble of
334 conformational states supported by the IDP scaffold, and in the distribution of end-to-end distances
335 measured between terminal LC8 dimers (Supplemental Fig. 8).
336
337 The population counts of oligomeric assemblies decay significantly with increasing size, with the 2-mer
338 state being the most populated (Fig. 5C). Such behavior is expected for the moderate binding affinity
339 between LC8 and the Nup159 IDP (K_d ~3 μM) and the sample dilution to nanomolar level (41). In addition
340 to the expected stoichiometries of 2 – 5-mers, the initial automated assignment of oligomeric states also
341 identified a significant population of 6-mers (*i.e.,* Nup159 bound to 6 LC8 dimers) (Fig. 5B), again indicating
342 the possible existence of off-register/daisy-chain type assemblies even at these low concentrations (see
343 Fig. 1). Remarkably, application of our statistical correction protocol does not rule out the existence of the
344 off-register 6-mer species in this system (Fig. 5C and Supplemental Fig. 8). The significance of this
345 intriguing finding is further discussed below. On the other hand, similar to the LC8/*syn*-4mer system, the
346 corrected population of 2-mers is significantly reduced from the initial population based only on visible
347 features. Indeed, the formation of randomly proximal 2-mers is expected whenever there is a substantial
348 population of free LC8 particles. In contrast to the *syn*-4mer dataset, however, the relative population of 2-
349 mers is not significantly altered following our statistical correction protocol (Fig. 5C).
350
351 **Single-particle distribution analysis of LC8/*d*ASCIZ**
352 To further assess the effectiveness of our heterogeneity analysis, we characterized the transcription factor
353 ASCIZ, which regulates synthesis of the LC8 protein to which it binds in a multivalent fashion. Drosophila
354 ASCIZ (*d*ASCIZ) has seven LC8 binding sites (Fig. 1). We have shown that the LC8/*d*ASCIZ assembly
355 exhibits significant compositional and conformational heterogeneity by NSEM, NMR and analytical ultra-
356 centrifugation (22) and by native mass spectrometry (23). Manual analysis of the NSEM data was used to
357 obtain a quantitative assessment of oligomer populations, and while conformational heterogeneity could

358 be deduced from the raw micrograph images, a quantitative procedure to characterize these states was
359 not practical.
360
361 To facilitate comparison of results to the *syn*-4mer and Nup159 systems described here, we re-analyzed
362 the LC8/*d*ASCIZ NSEM dataset using the same workflow, as described in Methods (Fig. 6 and
363 Supplemental Figs. 9 and 10). Traditional 2D classification analysis appeared to resolve only a subset of
364 oligomeric states (Fig. 6A and Supplemental Fig. 9), while fully assembled 7-mers were not identified under
365 the dilute conditions required for NSEM (Fig. 6C). These results are consistent with our previous analysis,
366 and the characterized negative cooperativity that is present in this system (22). Furthermore, the same
367 artifacts described for the *syn*-4mer and Nup159 systems resulting from the underlying conformational
368 heterogeneity of this system are readily identified in the results of 2D class averaging (Fig. 6A, asterisks
369 and Supplemental Figs. 8), and as previously described (22).
370
371 Our single-particle distribution analysis once again provides a much more detailed and quantitative picture
372 of the underlying conformational and compositional heterogeneity present in the LC8/*d*ASCIZ system. Due
373 to complications associated with the extreme level of heterogeneity in this sample, our prior manual
374 analysis was limited to < 350 total oligomers (extracted from ~300 micrographs) (22), whereas the present
375 automated pipeline yielded more than an order of magnitude more oligomers from the same set of
376 micrographs (Fig. 6C). Geometric distributions of LC8 bound to *d*ASCIZ portray a system with significant
377 structural variability (Fig. 6B). Remarkably, despite the presence of a variety of linker lengths separating
378 the 10-residue QT-motifs (ranging from 1 – 28 flexible residues), separation distances between bound
379 LC8's are similar to the *syn*-4mer and Nup159 assemblies, ranging from 5.8 ± 1.4 to 6.1 ± 1.7 nm (Fig.
380 6B). The end-to-end distance of terminal LC8 dimers becomes widely distributed with increasing valency,
381 and reaches a minimum of ~10 nm for assembled 6-mers (Supplemental Fig. 10). This continuum of states
382 is facilitated by the accessible bend angles, which are slightly more narrowly distributed as compared to
383 Nup159, and more similar to the *syn*-4mer assemblies, ranging from 30° ± 23 to 36° ± 26 (Fig. 6B). The
384 mechanistic explanation for this behavior is not yet clear, but might reflect some intrinsic behavior or
385 communication between LC8 binding sites that is responsible for the characterized negative cooperativity
386 displayed by *d*ASCIZ.
387
388 Population profiles of the assembled oligomeric states determined by manual assessment, by the initial
389 automated process, and after statistical correction are in rough agreement (Fig. 6C). Notably, neither the
390 manual nor the statistically corrected counts tally any fully saturated 7-mers, let alone larger potentially off-
391 register species. In all cases, the 2-mer population is the most abundant species, representing almost 80%
392 of the population by our automated workflow, while 3-mers and 4-mers make up the majority of other
393 species detected. This assessment is consistent with results for 2D class averages, where well-resolved

classes corresponded to only the 2-mer and 3mer populations (Fig. 6A), lending further validity to the results of the presented automated approach and statistical correction process.

DISCUSSION

The continuum dynamics (conformational fluctuations) of IDP systems makes them generally difficult to structurally characterize with precision, and the emerging class of semi-ordered beads-on-a-string systems, such as the LC8-organized systems studied here, compound the challenges. The LC8 dimers 'beads' are too small for current high-resolution CryoEM methods, forcing the use of NSEM in this work where the IDPs themselves are not directly detectable. Adding to these challenges, multivalent LC8/IDP systems exhibit substantial compositional heterogeneity, originating from their relatively low binding affinities. Class-averaging methods common in EM analysis are inappropriate for such systems because they suppress heterogeneity by construction. We have therefore developed a single-particle distribution analysis pipeline. Such tools are expected to become increasingly valuable given that the multivalent systems studied here appear to exploit weak, reversible, multivalent binding in support of diverse nano-architectural and sensory roles, with the full scope of functions still being revealed (42-45).

The idea to extract single-particle information directly from individual electron micrograph images is not new (22, 36-38), but here we exploit geometric characteristics of polymeric systems to create an intuitive and reliable automated approach applicable to the growing, important class of beads-on-string systems (46-48). We show that the "functional form" of polymeric conformational properties in terms of bead-bead distances and three-bead angles, enabled adequate training based on only a few dozen manually picked oligomers. The automated approach then generates thousands of candidate structures that represented the full breath of conformational states observed in the single-particle population, which can be analyzed and/or filtered as needed. The statistical correction process allowed the microscopists to gain a more faithful visualization of the underlying compositional heterogeneity that may be obscured by large populations of unbound ligand/proteins. In this way, we were able to address possible off-register binding (Fig. 1), an emerging phenomenon (23). For example, for Nup159, it is possible that off-register binding is required to bridge the IDP dimers into the 8-fold symmetry of the nuclear pore complex and stabilize the higher order assembly (49) – providing an intriguing basis for future investigation. However, for other LC8/IDP systems, the potential for such types of off-register assembly may require suppression for LC8 to orchestrate its physiological roles.

The pipeline presented here can be improved in future studies. Better optimized single-particle picking algorithms would enhance oligomer assignment quality most directly, though we note that preliminary testing of some modern particle picking platforms (35, 50, 51) yielded sub-optimal results, presumably because of the small size and close proximity of LC8 particles. Technical aspects of the scoring and

statistical correction procedures might also be improved, such as giving preference to pruning low-scoring oligomers/particles as part of the statistical correction procedure. The oligomer scoring function could also include the possibility of LC8 “beads” appearing in a non-sequential fashion, *e.g.,* accounting for cases where an LC8 dimer is missing in the middle of an oligomer. At the same time, it may be possible in some systems to correlate the extracted values of bead-to-bead distances with the known lengths of the IDP linkers, allowing identification of occupied binding sites.

Importantly, the methods developed here are readily extensible to other heterogeneous and dynamic multivalent IDP systems. It should be straightforward to generate synthetic micrographs for self-consistent analysis within the statistical correction procedure demonstrated here. Likewise, determining and scoring geometric and intensity criteria should also present few obstacles for other systems.

441 **METHODS**

442 **Design of a synthetic 4-site LC8-binding protein**

443 For algorithm development and training, we designed a novel LC8-binding peptide (termed *syn*-4mer)
444 using a series of 4 repeats of the amino acid sequence RKAIDAATQTE, taken from the tight-binding LC8
445 motif of the protein CHICA (Uniprot Q9H4H8), which has a 0.4 μM affinity to LC8, making it one of the
446 tightest-known LC8-binding motifs (34). The motif is spaced by uniform disordered linker sequences,
447 totaling 3 linkers, and flanking GSYGS sequences were added to the N- and C- termini of the constructs
448 to allow for quantification by absorbance at 280 nm. The final sequence is:
449 GSYGSR**KAIDAATQTE**PKETR**KAIDAATQTE**PKETR**KAIDAATQTE**PKETR**KAIDAATQTE**GSYGS. In
450 bold is the 10 amino acid segment that packs as a beta strand when bound to LC8.

451

452 **Protein expression and purification**

453 A gene sequence for the LC8-binding *syn*-4mer peptide was purchased as a block (integrated DNA
454 technologies, Coralville, Iowa) and cloned into a pET24d expression vector with an N-terminal Hisx6 affinity
455 tag and a tobacco etch virus protease cleavable site. The protein was expressed in ZYM-5052 (52) auto-
456 induction media at 37° C for 24 hr. Cells were harvested, lysed by sonication and purified in denaturing
457 buffers containing 6 M urea on TALON resin. The 4-mer was dialyzed into non-denaturing buffer (25 mM
458 tris pH 7.5, 150 mM NaCl) and further purified by gel filtration on a Superdex 75 Hi-load column (GE
459 Health), in the same buffer. Domains of yeast Nup159 (residues 1096 – 1178) and drosophila ASCIZ
460 (*d*ASCIZ, residues 241 – 388), full length LC8 of Saccharomyces cerevisiae and LC8 of Drosophila
461 melanogaster were all expressed and purified as previously described (22, 34). All proteins were stored at
462 4° C and used within one week of purification.

463

464 **SEC-MALS**

465 Size-exclusion chromatography (SEC) coupled to a multiangle light scattering (MALS) instrument was
466 performed using an analytical SEC column of Superdex S200 resin (GE Healthcare) on an AKTA-FPLC
467 (GE Healthcare), then routed through a DAWN multiple-angle light scattering and Optilab refractive index
468 system (Wyatt Technology). The column was equilibrated to a buffer of 25 mM tris (pH 7.5), 150 mM NaCl,
469 and 5 mM BME, then injected with 100 μL of LC8/*syn*-4mer complex in the same buffer at an estimated 2
470 μM particle concentration (16 μM LC8 + 4 μM *syn*-4mer, assuming 2:8 binding stoichiometry). We
471 estimated the molar mass using the ASTRA software package, with a Zimm scattering model.

472

473 **Isothermal Titration Calorimetry**

474 Isothermal titration calorimetry was carried out at 25°C using a VP-ITC microcalorimeter (Microcal) in a
475 buffer of 25 mM tris (pH 7.5), 150 mM Nacl and 5 mM BME. A cell containing 9 μM syn-4mer was titrated

476 with a solution of 300 μM LC8, across 32 injections of 8 μL. Peaks were integrated and fit to a single-site
477 binding model in Origin 7.0.
478
479 **Analytical Ultracentrifugation**
480 Samples of the *syn*-4mer peptide in complex with LC8 were prepared for sedimentation velocity analytical
481 ultracentrifugation (SV-AUC) by mixing excess (8:1) LC8 with *syn*-4mer, then purifying the complex by gel
482 filtration on a Superdex 200 column in a buffer of 25 mM tris (pH 7.5), 150 mM NaCl, and 5 mM β-
483 mercaptoethanol. The estimated concentration of the syn-4mer/LC8 complex applied to SV-AUC was at a
484 4:1 ratio of syn-4mer (13.8 μM) and LC8 (55μM). The SV-AUC titration of LC8 into Nup159 was performed
485 by mixing Nup159 (12.5 μM) and LC8 at LC8:Nup159 ratios of 0.5:1 to 8:1 in a buffer of 50 mM sodium
486 phosphate (pH 7.5), 50 mM NaCl, 5 mM TCEP and 1 mM sodium azide. SV-AUC was performed on a
487 Beckman Coulter Optima XL-A ultracentrifuge, equipped with optics for absorbance. Complexes were
488 loaded into two-channel sectored centerpieces with a 12-mm path length and centrifuged at 42,000 rpm
489 and 20° C. We collected 300 scans at 280 nm with no interscan delay, and fit data to a c(S) distribution
490 using SEDFIT (53). Buffer density was calculated using Sednterp (54).
491
492 **LC8-IDP complex preparation for EM**
493 LC8/syn-4mer complexes were prepared for electron microscopy studies by mixing excess (8:1) of the
494 purified LC8 with the syn-4mer peptide and purifying the complexes by size-exclusion chromatography
495 (SEC; Superdex 200, in a buffer of 25 mM tris pH 7.5, 150 mM NaCl and 5 mM BME). The Nup159 complex
496 was formed by mixing equimolar amounts of LC8 and Nup159, without further purification. Negative stain
497 EM grids were prepared by diluting the LC8 complexes to a final particle concentration of 16 nM (presumed
498 to be fully bound complexes) in SEC buffer. A 3 μl drop of sample was applied to a glow-discharged
499 continuous carbon coated EM specimen grid (400 mesh Cu grid, Ted Pella, Redding, CA). Excess protein
500 was removed by blotting with filter paper and washing the grid two times with dilution buffer. The specimen
501 was then stained with freshly prepared 0.75% (wt vol^{-1}) uranyl formate (SPI-Chem).
502
503 **Electron microscopy**
504 Negatively stained specimens were imaged on a 120 kV TEM (iCorr, FEI) at a nominal magnification of
505 49,000x at the specimen level. Digital micrographs were recorded on a 2K × 2K CCD camera (FEI Eagle)
506 with a calibrated pixel size of 4.37 Å $pixel^{-1}$ and targeted a defocus of 1.5 – 2 μm. For the *syn*-4mer/LC8
507 specimen, a dataset of 34 micrographs was collected and picked in an automated fashion to select the
508 center of ~4 – 5 nm densities, corresponding to individual LC8 dimers, using DoG Picker (55). We note
509 that a variety of alternative automated particle picking tools were assayed for this workflow (35, 50, 51),
510 which included traditional blob-pickers, template-based methods, as well as neural net particle picking

511 algorithms. Following this initial screen, DoG Picker was selected due to the ease-of-use and performance
512 as compared to these alternative methods.
513
514 DoG Picker settings were optimized for radius equal to 8 pixels and optimal thresholds ranging from 4.0 –
515 4.4, resulting in ~2000 – 3700 particle picks per micrograph with minimal contribution from background,
516 assessed manually. From this dataset, 4 micrographs were set aside for training that contained a total of
517 14,306 LC8 particles. A separate validation set of 5 micrographs was prepared similarly using DoG Picker,
518 yielding 17,245 particles. A dataset of 104 micrographs of the Nup159 construct were collected under
519 identical conditions, that yielded a total of 246,328 particles by automated selection using DoG Picker. For
520 the *d*ASCIZ construct, our previously collected dataset of 305 micrographs was re-processed and used for
521 automated analysis (22), and yielding 557,134 LC8 particles by DoG Picker.
522
523 For use in method development and validation studies, the *syn*-4mer training dataset was curated by the
524 microscopist familiar with the LC8-IDP structure to manually classify a representative set of LC8 oligomers
525 as 2-mers, 3-mers, 4-mers, etc. To minimize ambiguity, the microscopist selected complexes that were
526 well separated from neighboring particles on the micrograph. This procedure resulted in a curated set of
527 54 oligomers of varying valency (containing 216 LC8 particles in total) that were used for calibration of our
528 automated analysis workflow.
529
530 For further comparative analysis, additional single-particle datasets were obtained by manual selection
531 from the recorded micrographs for traditional 2D classification and averaging in EMAN (35). Obtained
532 image stacks contained 4151 putative oligomers extracted with a box size of 96 for the LC8/*syn*-4mer
533 dataset, 5875 oligomers and a box size of 96 for the LC8/Nup159 dataset, and 2434 complexes with a box
534 size of 160 for LC8/*d*AZCIS dataset.
535
536 **Automated identification and population counting of oligomers**
537 The automated pipeline for identifying beads-on-a-string LC8/IDP oligomers employed three stages: (i)
538 clustering (ii) oligomer identification based on a scoring function (Fig. 2E,F), and (iii) distance-filtering to
539 disregard crowded regions of micrographs (Supplemental Fig. 5). Single-linkage clustering (56) of all LC8
540 coordinates from the auto-picked micrographs was first performed. In this clustering method, data points
541 separated by less than a given distance are grouped together to distinguishing sets of particles that cannot
542 form an oligomer based on their inter-particle coordinates. The linkage distance was set to a value that is
543 larger separation of neighboring LC8 binding sites on the IDP, as derived by the distribution of separation
544 distances obtained in the curated training set (see Supplemental Figs. 3, 7, 9). In particular, the clustering
545 threshold was set to 7 nm for the 4-site IDP, 8 nm for the Nup159 system, and 15 nm for the ASCIZ system.
546

547 A scoring algorithm was developed to classify the heterogeneous oligomer populations. The scoring
548 algorithm is informed by the particle intensity (I), *i.e.,* the average pixel value within a picked particle as
549 reported by the DoG picker (55) and the oligomer geometry, *i.e.,* particle-to-particle separation distance,
550 (d) and angle (Θ) defined by three adjoining particles. The oligomer obtained from the hand curated training
551 data were used to calibrate these features from their distributions in the training data. The distributions of
552 these three metrics (Supplemental Fig. 4) provided parameters to score new oligomers, as described in
553 equation (1).
554
555 To treat all three metrics I, d, and Θ on an equal basis, we used their training data cumulative distribution
556 functions (CDFs) within their 0.5 to 99.5 percentile region. To give a higher score to small distances and
557 angles between two or three given particles, we used 1-CDF as the probability scores P_d and P_θ for these
558 two metrics, whereas the CDF was used as the probability score P_I for the intensity I of a given particle.
559 The total score for any *n*-mer is the normalized sum over all of its sequential intensity, distance, and angle
560 log-probability scores:

561 $$Score = \frac{1}{3n-3}\left[\sum_{i}^{n}\log\left(P_I(i)\right) + \sum_{j=1}^{n-1}\log\left(P_d(j,j+1)\right) + \sum_{k=1}^{n-2}\log\left(P_\theta(k,k+1,k+2)\right)\right] \quad (1)$$

562
563 Here P = CDF or 1-CDF as noted above.
564
565 To obtain oligomer assignments from the clustered particles, our program considers every possible
566 combination of particle sequences (or oligomeric states) within a cluster and scores them independently.
567 The potential oligomers were then ranked by their length and their total score, thus giving precedence to
568 longer assemblies (regardless of score) over shorter ones, and the highest-ranked non-overlapping
569 oligomers in each cluster were saved. Finally, a score threshold value was applied to discard oligomers
570 with a low total score, for instance such oligomers that consist of several low-intensity particles and with
571 geometry that is unfavorable compared to the training set statistics. Based on Supplemental Fig. 4, a small
572 (*i.e.,* more permissive) threshold value of 0.05 was selected for the analysis of LC8/*syn*-4mer and
573 LC8/Nup159 and, based on Supplemental Fig. 10, a threshold value of 0.3 was applied to the analysis of
574 LC8/*d*ASCIZ, which contains a significantly longer IDP.
575
576 In order to prevent assignments in crowded, ambiguous regions of the micrograph, oligomer assignments
577 were filtered by a distancing criterion, counting only those oligomers that were separated at least by a
578 specified distance from any other LC8 particles. For consistency with manual evaluation, an initial set of
579 micrographs with automatically assigned oligomers at different distance thresholds were examined by a
580 microscopist to identify an optimal value for this filtering distance. The threshold was set to 9 nm for the
581 *syn*-4mer and Nup159 systems and to 14 nm for the *d*ASCIZ system. The corresponding fractional

populations of all three systems are not significantly different at slightly larger or smaller filtering distance thresholds (+/- 2 nm) as shown in Supplemental Fig. 5. To facilitate direct comparison of oligomeric state populations from the corresponding manually and automatically assigned micrographs, the same threshold was applied to the manual dataset of *syn*-4mer and *d*ASCIZ systems.

Correcting oligomer populations through self-consistent statistical re-scoring

Ultimately, the accuracy of oligomer prediction depends not only on the correct assignment of oligomers but also on the identification of artifactual structures that should not be counted – *i.e.,* spurious oligomers resulting from random proximity of free LC8 particles not bound to any IDP. Random proximity would extend actual *n*-mers to be wrongly counted as (*n+1*)-mers or longer. Recall that the IDPs "strings" themselves are not visible in the micrographs. To provide an estimate for the actual number of the underlying real oligomers, the experimental process of random placement of single LC8 particles was iteratively simulated and the degree of artifactual oligomer creation was evaluated in order to obtain a self-consistent estimate of the underlying populations.

The iterative correction procedure is initialized by randomly relocating all free LC8 particles, *i.e.,* those that were not assigned to be part of a putative oligomer during the initial identification and scoring process. This population of free particles are then positioned randomly and independently, but with a minimum distance of 2 nm from any other particle present on the micrograph, which roughly corresponds to the minimum distance of LC8 particles observed experimentally. This procedure produces a synthetic micrograph that includes all predicted oligomers from the original experimental micrograph, but with the single LC8s rearranged. Applying the same scoring and counting rules described above to this synthetic micrograph leads to different oligomer assignments due to the random relocation of free LC8 particles that can lead to the appearance of both new oligomer creations (randomly placed LC8s that meet our scoring criteria of an oligomer) and putative oligomer extensions (randomly placed LC8s that are now located near the terminus of a previously assigned *n*-mer). Another effect of this process is that oligomers that were previously counted in the experimental micrograph may now be "disqualified" because of the distancing criterion that is applied (to avoid assignments in crowded regions). At the same time, other oligomers which did not meet the distancing criteria in the original assignment process, because they were "blocked" by a nearby free LC8 particle(s), may now be "released" and counted.

The process just described is iterated until self-consistency is obtained between population counts from the synthetic micrographs, as compared to the originally assigned micrograph. Specifically, at each iteration *i*, the naive count of each *n*-mer oligomer species (abbreviated by *n*) in the synthetic micrograph is compared with that in the experimental micrograph to obtain the difference $\Delta_i(n)$. If at iteration *k*, the cumulative sum of these differences over all previous iterations was a positive integer, i.e., $\sum_{i=0}^{k} \Delta_i(n) > 0$,

suggesting that the number of directly counted *n*-mers in the given synthetic micrograph exceeded those in the experimental micrograph, then that many putative *n*-mers were selected at random and pruned. Pruning of oligomers was performed by stripping one of their terminal particles and adding it to the set of free particles at iteration (*k+1*), thereby also reducing the number of putative *n*-mers and increasing that of putative (*n-1*)-mers. This operation was performed at every iteration in a cascading fashion from longer oligomers to shorter oligomers and the 2-mers were pruned by splitting and adding both particles to the set of free LC8s. If $\sum_{i=0}^{k} \Delta_i(n) \leq 0$, then no pruning and updating of putative oligomer counts was performed. This iterative process was conducted until the counts of all oligomer species in the synthetic micrograph matched those in the experimental micrograph. At that point, the updated population of putative oligomers can be considered corrected with respect to artifacts arising from the large number of free LC8 particles. If continued, the populations fluctuate among a set of values consistent with the originally assigned dataset.

Convergence was reached within 100 iterations for the *syn*-4mer system and within 200 iterations for the Nup159 and *d*ASCIZ systems (Fig. 3 and Supplemental Figs. 7,9). For each system, the naïve and corrected oligomer populations were obtained as the arithmetic mean over the last 50 iterations and the total population counts over all analyzed micrographs were converted into fractional populations. Error bars are derived as the square-root of the sum of all per-micrograph variances computed from the last 50 iterations, representing the effective standard deviation – *i.e.,* the scale of variation of the obtained mean values from the correction procedure. A flowchart describing the statistical correction procedure is provided in Supplemental Fig. 6.

Code and data availability

All codes are available at https://github.com/ZuckermanLab/EM_OligomerAnalysis. Electron microscopy images and particle coordinate files are available at http://doi.org/10.5281/zenodo.4726027. Expression vectors are available upon request to E.B.

Acknowledgments and Funding Sources

We thank Patrick Reardon for his assistance with collecting the AUC data, and the staff at the OHSU Multiscale Microscopy Core for assistance and training. The research was funded by the National Science Foundation (MCB 1715823 to D.M.Z.) and (MCB 1617019 to E.B.) and the National Institutes of Health (R35GM124779 and R01EY030987 to S.L.R.).

Author Contributions: All authors contributed to the preparation of the manuscript. B.M. and R.M contributed equally. B.M. developed the single-particle distribution and statistical analysis programs. R.M. conducted the electron microscopy experiments and analysis. A.E. and J.H. performed the biochemistry and biophysical characterizations of LC8/IDP complexes. E.B. supervised the biochemical and biophysical studies. S.L.R. supervised the electron microscopy studies. D.M.Z. supervised the theoretical modeling and statistical analysis. E.B., S.L.R. and D.M.Z. provided overall oversight to the design and execution of the work.

Competing Interest Statement: The authors have no competing interests.

663 **REFERENCES**

664 1. P. E. Wright, H. J. Dyson, Intrinsically disordered proteins in cellular signalling and regulation.
665 *Nature reviews. Molecular cell biology* **16**, 18-29 (2015).

666 2. C. J. Oldfield, A. K. Dunker, Intrinsically disordered proteins and intrinsically disordered protein
667 regions. *Annu Rev Biochem* **83**, 553-584 (2014).

668 3. T. H. Kim *et al.*, Phospho-dependent phase separation of FMRP and CAPRIN1 recapitulates
669 regulation of translation and deadenylation. *Science* **365**, 825-829 (2019).

670 4. A. E. Conicella *et al.*, An intrinsically disordered motif regulates the interaction between the p47
671 adaptor and the p97 AAA+ ATPase. *Proc Natl Acad Sci U S A* **117**, 26226-26236 (2020).

672 5. J. J. Ward, J. S. Sodhi, L. J. McGuffin, B. F. Buxton, D. T. Jones, Prediction and functional analysis
673 of native disorder in proteins from the three kingdoms of life. *J Mol Biol* **337**, 635-645 (2004).

674 6. S. M. King *et al.*, Brain cytoplasmic and flagellar outer arm dyneins share a highly conserved Mr
675 8,000 light chain. *The Journal of biological chemistry* **271**, 19358-19366 (1996).

676 7. G. J. Pazour, C. G. Wilkerson, G. B. Witman, A dynein light chain is essential for the retrograde
677 particle movement of intraflagellar transport (IFT). *J Cell Biol* **141**, 979-992 (1998).

678 8. E. Barbar, Dynein light chain LC8 is a dimerization hub essential in diverse protein networks.
679 *Biochemistry* **47**, 503-508 (2008).

680 9. N. Jespersen *et al.*, Systematic identification of recognition motifs for the hub protein LC8. *Life Sci*
681 *Alliance* **2** (2019).

682 10. D. Flemming, P. Stelter, E. Hurt, Utilizing the Dyn2 dimerization-zipper as a tool to probe NPC
683 structure and function. *Methods Cell Biol* **122**, 99-115 (2014).

684 11. M. Gaik *et al.*, Structural basis for assembly and function of the Nup82 complex in the nuclear pore
685 scaffold. *J Cell Biol* **208**, 283-297 (2015).

686 12. P. K. Singh *et al.*, Dynein light chain 1 induces assembly of large Bim complexes on mitochondria
687 that stabilize Mcl-1 and regulate apoptosis. *Genes Dev* **31**, 1754-1769 (2017).

688 13. L. Regue *et al.*, DYNLL/LC8 protein controls signal transduction through the Nek9/Nek6 signaling
689 module by regulating Nek6 binding to Nek9. *The Journal of biological chemistry* **286**, 18118-18129
690 (2011).

691 14. O. I. Kahn *et al.*, APC2 controls dendrite development by promoting microtubule dynamics. *Nat*
692 *Commun* **9**, 2773 (2018).

693 15. A. King *et al.*, Dynein light chain regulates adaptive and innate B cell development by distinctive
694 genetic mechanisms. *PLoS Genet* **13**, e1007010 (2017).

695 16. E. L. Eastwood *et al.*, Dimerisation of the PICTS complex via LC8/Cut-up drives co-transcriptional
696 transposon silencing in Drosophila. *eLife* **10** (2021).

697 17. S. A. Clark, N. Jespersen, C. Woodward, E. Barbar, Multivalent IDP assemblies: Unique properties
698 of LC8-associated, IDP duplex scaffolds. *FEBS Lett* **589**, 2543-2551 (2015).

699 18. E. Barbar, A. Nyarko, Polybivalency and disordered proteins in ordering macromolecular
700 assemblies. *Semin Cell Dev Biol* **37**, 20-25 (2015).

701 19. A. K. Dunsch *et al.*, Dynein light chain 1 and a spindle-associated adaptor promote dynein
702 asymmetry and spindle orientation. *J Cell Biol* **198**, 1039-1054 (2012).

703 20. S. Jurado *et al.*, ATM substrate Chk2-interacting Zn2+ finger (ASCIZ) Is a bi-functional
704 transcriptional activator and feedback sensor in the regulation of dynein light chain (DYNLL1)
705 expression. *The Journal of biological chemistry* **287**, 3156-3164 (2012).

706 21. K. Toropova *et al.*, Structure of the dynein-2 complex and its assembly with intraflagellar transport
707 trains. *Nat Struct Mol Biol* **26**, 823-829 (2019).

708 22. S. Clark *et al.*, Multivalency regulates activity in an intrinsically disordered transcription factor. *eLife*
709 **7** (2018).

710 23. P. N. Reardon *et al.*, The dynein light chain 8 (LC8) binds predominantly "in-register" to a
711 multivalent intrinsically disordered partner. *The Journal of biological chemistry* **295**, 4912-4922
712 (2020).

713 24. N. Palopoli *et al.*, Intrinsically Disordered Protein Ensembles Shape Evolutionary Rates Revealing
714 Conformational Patterns. *J Mol Biol* **433**, 166751 (2021).

715 25. M. Aznauryan *et al.*, Comprehensive structural and dynamical view of an unfolded protein from the
716 combination of single-molecule FRET, NMR, and SAXS. *Proc Natl Acad Sci U S A* **113**, E5389-
717 5398 (2016).

718 26. P. Stelter *et al.*, Molecular basis for the functional interaction of dynein light chain with the nuclear-
719 pore complex. *Nat Cell Biol* **9**, 788-796 (2007).

720 27. C. O. S. Sorzano *et al.*, Survey of the analysis of continuous conformational variability of biological
721 macromolecules by electron microscopy. *Acta Crystallogr F Struct Biol Commun* **75**, 19-32 (2019).

722 28. T. Nakane, D. Kimanius, E. Lindahl, S. H. Scheres, Characterisation of molecular motions in cryo-
723 EM single-particle data by multi-body refinement in RELION. *eLife* **7** (2018).

724 29. C. Zhang *et al.*, Analysis of discrete local variability and structural covariance in macromolecular
725 assemblies using Cryo-EM and focused classification. *Ultramicroscopy* **203**, 170-180 (2019).

726 30. A. Punjani, D. J. Fleet, 3D variability analysis: Resolving continuous flexibility and discrete
727 heterogeneity from single particle cryo-EM. *J Struct Biol* **213**, 107702 (2021).

728 31. E. D. Zhong, T. Bepler, B. Berger, J. H. Davis, CryoDRGN: reconstruction of heterogeneous cryo-
729 EM structures using neural networks. *Nat Methods* **18**, 176-185 (2021).

730 32. J. Frank, A. Ourmazd, Continuous changes in structure mapped by manifold embedding of single-
731 particle data in cryo-EM. *Methods* **100**, 61-67 (2016).

732 33. S. J. Ludtke, Single-Particle Refinement and Variability Analysis in EMAN2.1. *Methods Enzymol*
733 **579**, 159-189 (2016).

734 34. S. Clark, A. Nyarko, F. Lohr, P. A. Karplus, E. Barbar, The Anchored Flexibility Model in LC8 Motif
735 Recognition: Insights from the Chica Complex. *Biochemistry* **55**, 199-209 (2016).

736 35. G. Tang *et al.*, EMAN2: an extensible image processing suite for electron microscopy. *J Struct Biol*
737 **157**, 38-46 (2007).

738 36. F. D. Smith *et al.*, Intrinsic disorder within an AKAP-protein kinase A complex guides local substrate
739 phosphorylation. *eLife* **2**, e01319 (2013).

740 37. J. B. Myers *et al.*, The CaMKII holoenzyme structure in activation-competent conformations. *Nat*
741 *Commun* **8**, 15742 (2017).

742 38. O. R. Buonarati, A. P. Miller, S. J. Coultrap, K. U. Bayer, S. L. Reichow, Conserved and divergent
743 features of neuronal CaMKII holoenzyme structure, function and high-order assembly *BioRxiv*
744 https://doi.org/10.1101/2021.01.21.42764 (2021).

745 39. P. J. Flory, Spatial configuration of macromolecular chains. *Science* **188**, 1268-1276 (1975).

746 40. M. A. Kiskowski, J. F. Hancock, A. K. Kenworthy, On the use of Ripley's K-function and its
747 derivatives to analyze domain size. *Biophys J* **97**, 1095-1103 (2009).

748 41. A. Nyarko, Y. Song, J. Novacek, L. Zidek, E. Barbar, Multiple recognition motifs in nucleoporin
749 Nup159 provide a stable and rigid Nup159-Dyn2 assembly. *The Journal of biological chemistry*
750 **288**, 2614-2622 (2013).

751 42. C. Frattini *et al.*, TopBP1 assembles nuclear condensates to switch on ATR signaling. *Mol Cell* **81**,
752 1231-1245 e1238 (2021).

753 43. T. A. Tabish, M. R. Hamblin, Multivalent nanomedicines to treat COVID-19: A slow train coming.
754 *Nano Today* **35**, 100962 (2020).

755 44. J. L. Cuellar-Camacho *et al.*, Quantification of Multivalent Interactions between Sialic Acid and
756 Influenza A Virus Spike Proteins by Single-Molecule Force Spectroscopy. *J Am Chem Soc* **142**,
757 12181-12192 (2020).

758 45. P. Yang *et al.*, G3BP1 Is a Tunable Switch that Triggers Phase Separation to Assemble Stress
759 Granules. *Cell* **181**, 325-345 e328 (2020).

760 46. N. R. Hajizadeh *et al.*, Probing the Architecture of a Multi-PDZ Domain Protein: Structure of PDZK1
761 in Solution. *Structure* **26**, 1522-1533 e1525 (2018).

762 47. T. Shinohara *et al.*, Nonfilament-forming RecA dimer catalyzes homologous joint formation. *Nucleic*
763 *Acids Res* **46**, 10855-10869 (2018).

764 48. G. Nishide *et al.*, High-Speed Atomic Force Microscopy Reveals Spatiotemporal Dynamics of
765 Histone Protein H2A Involution by DNA Inchworming. *J Phys Chem Lett* **12**, 3837-3846 (2021).

766 49. M. Beck, E. Hurt, The nuclear pore complex: understanding its function through structural insight.
767 *Nature reviews. Molecular cell biology* **18**, 73-89 (2017).

768 50. T. Wagner *et al.*, SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-
769 EM. *Commun Biol* **2**, 218 (2019).

770 51. J. Zivanov *et al.*, New tools for automated high-resolution cryo-EM structure determination in
771 RELION-3. *eLife* **7** (2018).

772 52. F. W. Studier, Protein production by auto-induction in high density shaking cultures. *Protein Expr*
773 *Purif* **41**, 207-234 (2005).

774 53. P. Schuck, Size-distribution analysis of macromolecules by sedimentation velocity
775 ultracentrifugation and lamm equation modeling. *Biophys J* **78**, 1606-1619 (2000).

776 54. S. E. Harding, A. J. Rowe, J. C. Horton, Analytical Ultracentrifugation in Biochemistry and Polymer
777 Science. *Royal Society of Chemistry* (1992).

778 55. N. R. Voss, C. K. Yoshioka, M. Radermacher, C. S. Potter, B. Carragher, DoG Picker and TiltPicker:
779 software tools to facilitate particle selection in single particle electron microscopy. *J Struct Biol* **166**,
780 205-213 (2009).

781 56. P. H. Sneath, The application of computers to taxonomy. *J Gen Microbiol* **17**, 201-226 (1957).

782 57. J. Hall, P. A. Karplus, E. Barbar, Multivalency in the assembly of intrinsically disordered Dynein
783 intermediate chain. *The Journal of biological chemistry* **284**, 33115-33121 (2009).

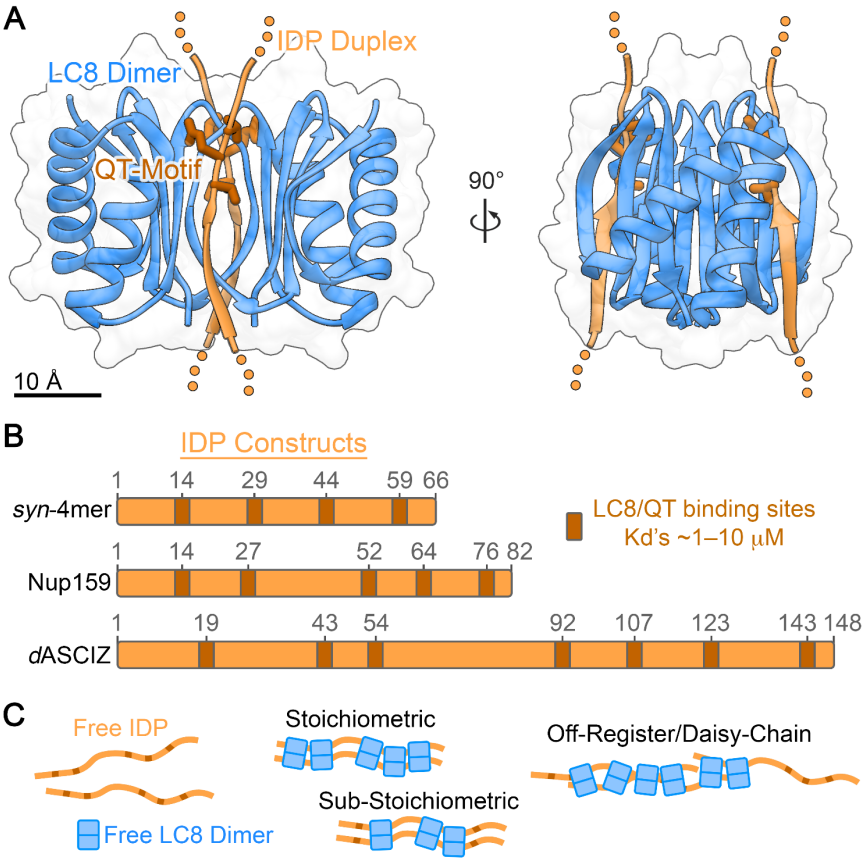
784

785

786 **FIGURES AND LEGENDS**

787

788 **FIGURE 1**



789

790 **Fig. 1: Overview of multivalent LC8/IDP systems.**

791 A) Crystallographic structure of the LC8 dimer (blue ribbon) bound to a duplex IDP (orange ribbon) (PDB
792 ID 3GLW; (57)). Each LC8 protomer binds to a single IDP through a characteristic QT-motif (dark orange
793 stick representation). Unresolved regions of the crystallized IDP construct are indicated by dotted line.

794

795 B) Schematic representation of the IDP constructs under investigation, corresponding to the synthetic four-
796 site IDP (*syn*-4mer), Nup159 and drosophila ASCIZ (*d*ASCIZ). LC8 binding QT-motifs are indicated in dark
797 orange. Sequence numbering for each IDP construct and QT-motif is indicated.

798

799 C) Illustration of various modes of assembly, ranging from *(left)* free IDP (orange) and free LC8 dimers
800 (blue), to *(middle)* sub-stoichiometric and stoichiometric assemblies of LC8 bound to IDPs in a duplex
801 fashion. In addition *(right)*, putative modes of off-register (or daisy-chain) assembly are illustrated, where
802 > 2 IDPs are linked together by LC8 dimers.

803

804

FIGURE 2

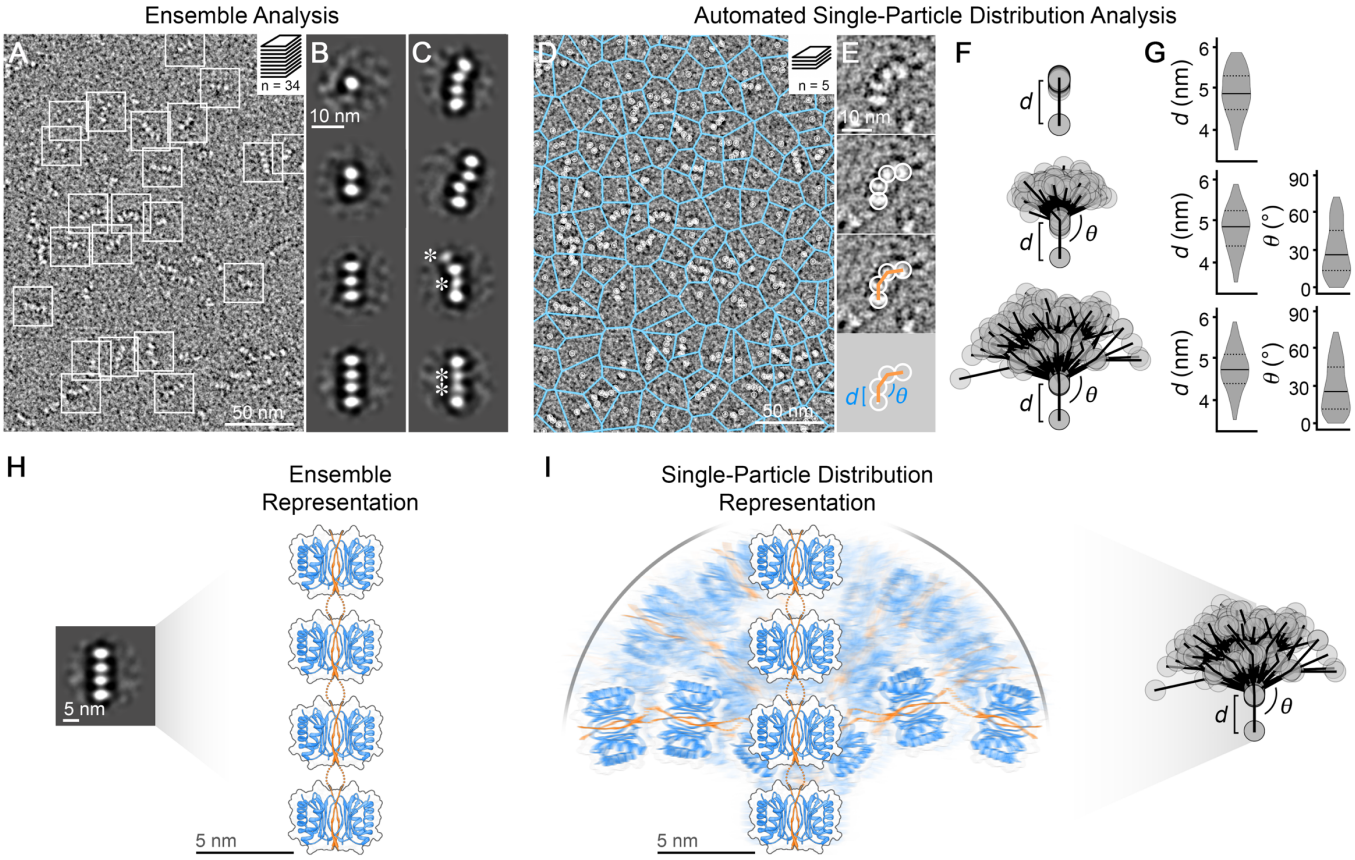


Fig. 2: Ensemble versus automated single-particle characterization of LC8/IDP oligomers.

A) Micrograph of negatively stained LC8/*syn*-4mer complexes. Representative complexes indicated by white box. Scale bar = 50 nm. *Inset*, indicates the number of micrographs collected for traditional ensemble analysis (n = 34)

B) Representative two-dimensional (2D) class-averages depict free LC8 *(top)* and a range of assembled LC8/*syn*-4mer complexes (2-mers to 4-mers) present in the image dataset that were well-resolved.

C) 2D class-averages of assembled LC8/*syn*-4mer complexes displaying varying degree of conformational heterogeneity. Asterisk indicate densities of LC8 that display blurred features that are less well-resolved, indicative of unresolved conformational/configurational heterogeneity. Scale bar = 10 nm in panels B and C.

D) Micrograph of negatively stained LC8/*syn*-4mer complexes as shown in panel A, with auto-picked LC8 densities highlighted as white circles and the single-linkage clusters indicated with blue lines, calculated as edges from a Voronoi partitioning of cluster centers. Scale bar = 50 nm. *Inset*, indicates the number of micrographs collected for automated single-particle distribution analysis (n = 5)

E) Zoom view from panel D, showing the automated classification and geometric analysis workflow. Individual LC8 dimers are selected in an automated fashion (white circles) and classified by our scoring function (represented by orange line). Geometric descriptions of the assigned oligomers are then extracted for analysis (*e.g.* LC8 to LC8 separation distances (d) and bend angles (Θ)). Scale bar = 10 nm.

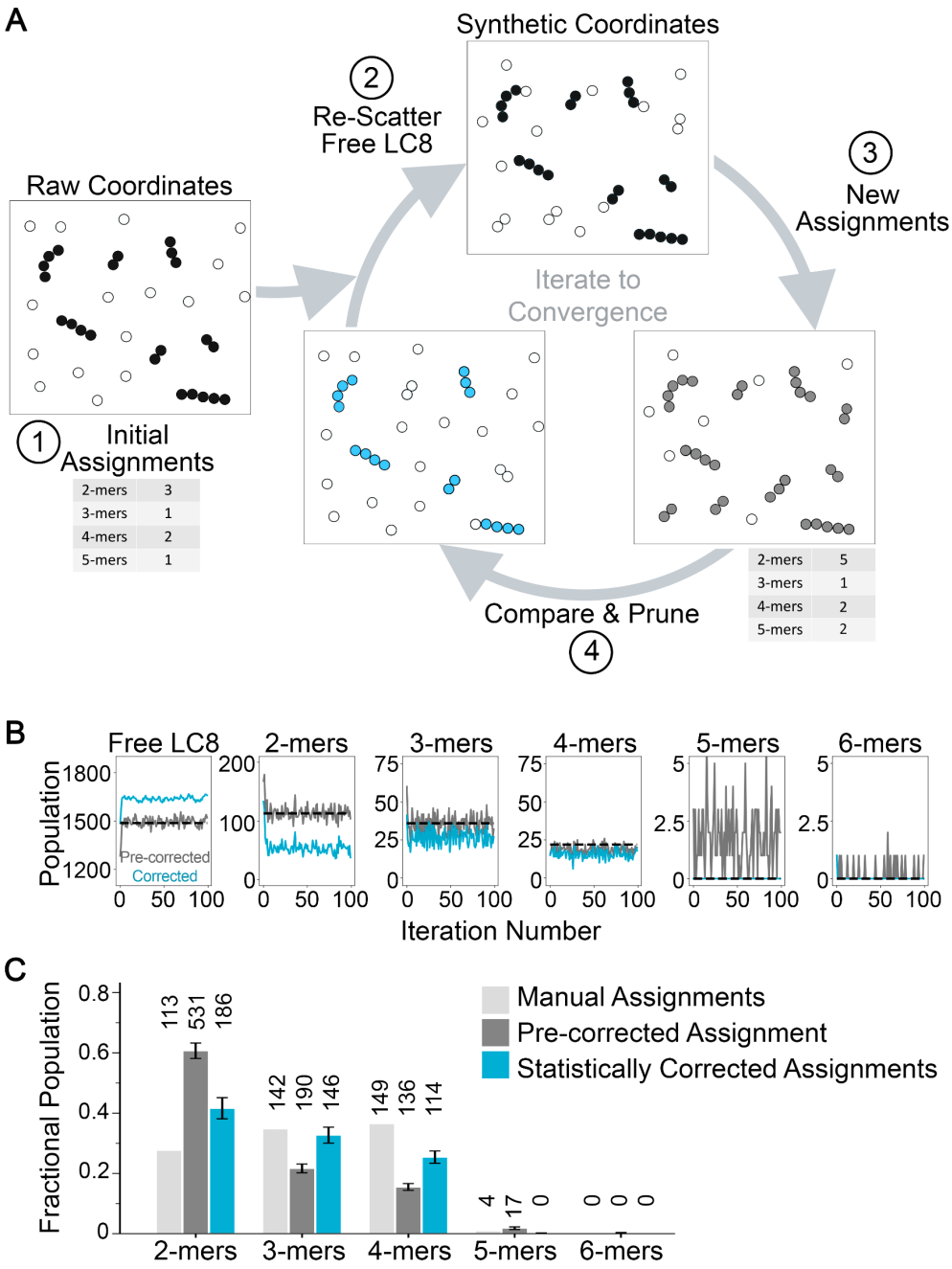
F) Illustrative representation of classified LC8/*syn*-4mer oligomers obtained by automated single-particle distribution analysis (n = 100 for each class of oligomer). Individual *n*-mers are aligned along the connection of the first two LC8 dimers (represented as grey circles). The conformational heterogeneity of identified oligomers is illustrated by the variability in separation distances (d) and bend angles (Θ).

G) Violin plots showing the distribution of separation distances (d) and bend angles (Θ) of all 2-mers, 3-mers, and 4-mers. The solid line indicates the median and the dotted lines indicate the corresponding first and third quartiles. A full table of statistics is reported in Supplemental Table 2.

H) Illustrative interpretation of *syn*-4mer duplex (orange ribbon) with four assembled LC8 dimers (blue) obtained by ensemble 2D classification methods.

I) Illustrative interpretation of *syn*-4mer duplex (orange ribbon) with four assembled LC8 dimers (blue) obtained by automated single-particle distribution analysis, demonstrating the wide spectrum of conformational states accessible by the IDP scaffold.

862 **FIGURE 3**



863

864 **Fig. 3: Self-consistent statistical correction of unbound LC8 particles**

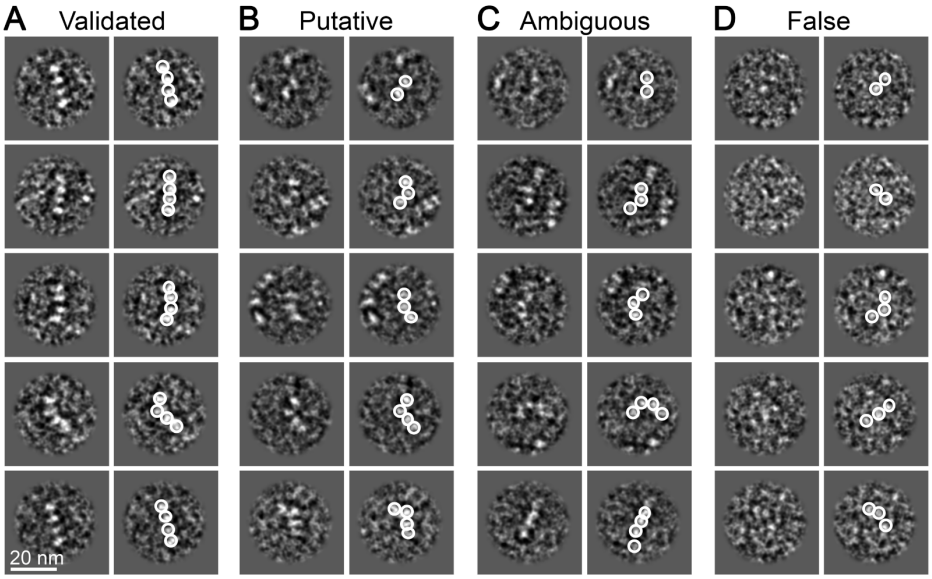
865 A) Initial oligomer assignments of the experimental micrograph are corrected to account for random
866 proximity effects resulting from free LC8 particles. Accepting at first the initial oligomer assignments (black
867 filled circles), the free LC8 population (white filled circles) is randomly rescattered to create a synthetic
868 micrograph, from which new oligomer populations are obtained (grey filled circles). Deviations between
869 pre-corrected and initial assignment populations are used to 'correct' the initial oligomer counts for the next
870 round of rescattering (cyan filled circles). The rescattering and correction procedure is then iterated until
871 self-consistency is obtained.

872

B) The pre-corrected (gray) and corrected (cyan) populations of oligomers and free LC8 as a function of iteration number during the statistical correction simulation of one example micrograph with LC8 dimers bound to *syn*-4mer. At every iteration, the pre-corrected populations are from the synthetic micrograph and the corrected populations are the putative oligomers from the experimental micrograph after pruning. The dashed black lines indicate the corresponding populations of initial oligomer assignments in the experimental micrograph, to which the gray lines are expected to converge. Note that, in this example, the significantly overestimated number of 2-mers (and the vastly underestimated number of free LC8) is corrected within the first 10-20 iteration cycles.

C) The fractional population distribution of LC8/*syn*-4mer oligomeric states obtained by manual inspection (light gray), by automated 'pre-corrected' single-particle distribution analysis (dark gray) and by statistical correction (blue). Total number of classified states are indicated above each bar. Error bars correspond to the effective standard deviation as described in Methods.

889 **FIGURE 4**

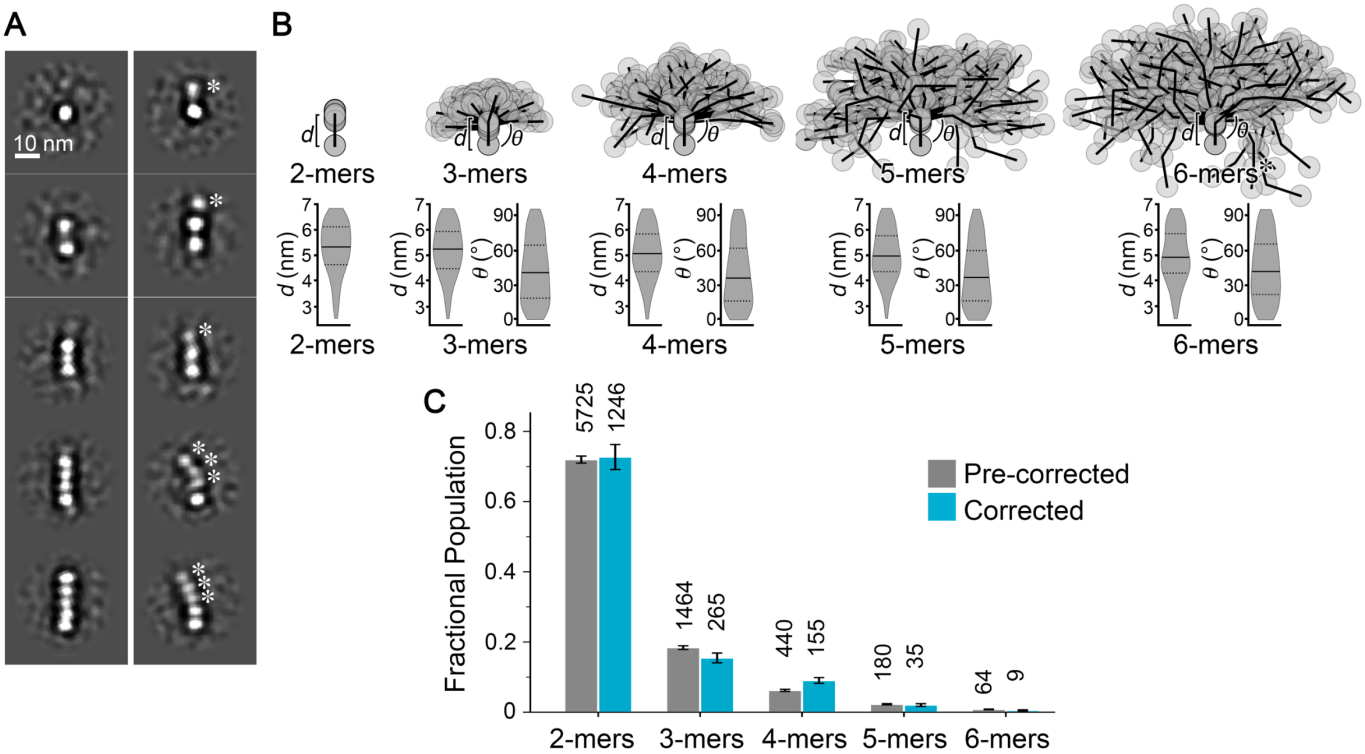


890

891 **Fig. 4: Validation of automated LC8/IDP oligomer assignments**

892 A) Representative images of validated LC8/*syn*-4mer oligomers that were assigned by both the
893 microscopists and the scoring function.
894
895 B) Representative images of putative complexes assigned by scoring function and deemed to be
896 acceptable by the microscopist upon re-evaluation.
897
898 C) Representative images of complexes assigned by scoring function and deemed to be too ambiguous
899 to confidently assign by the microscopists upon re-evaluation (*e.g.,* containing weak LC8 density and/or
900 neighboring LC8 densities that were not autopicked).
901
902 D) Representative images of complexes assigned by scoring function and deemed to be too falsely
903 assigned by the microscopist upon re-evaluation (*e.g.,* containing autopicked densities corresponding to
904 background carbon).
905
906 Scale bar for all panels = 20 nm. In all panels, raw images are shown in the left column and autopicked
907 results obtained by DoG Picker (55) shown by white circles.
908
909
910
911

912 **FIGURE 5**



913

914 **Fig. 5: Quantitative characterization and statistical correction of LC8/Nup159 complex assemblies.**
915 A) Representative 2D class-averages depicting, *(left column)* free LC8 and a range of assembled
916 LC8/Nup159 complexes ranging from 2-mers to 5-mers present in the image dataset that were well-
917 resolved, and *(right column)* complexes displaying varying degrees of conformational heterogeneity.
918 Asterisk indicate densities of LC8 that display blurred features that are less well-resolved, indicative of
919 unresolved conformational/configurational heterogeneity. Scale bar = 10 nm.

920

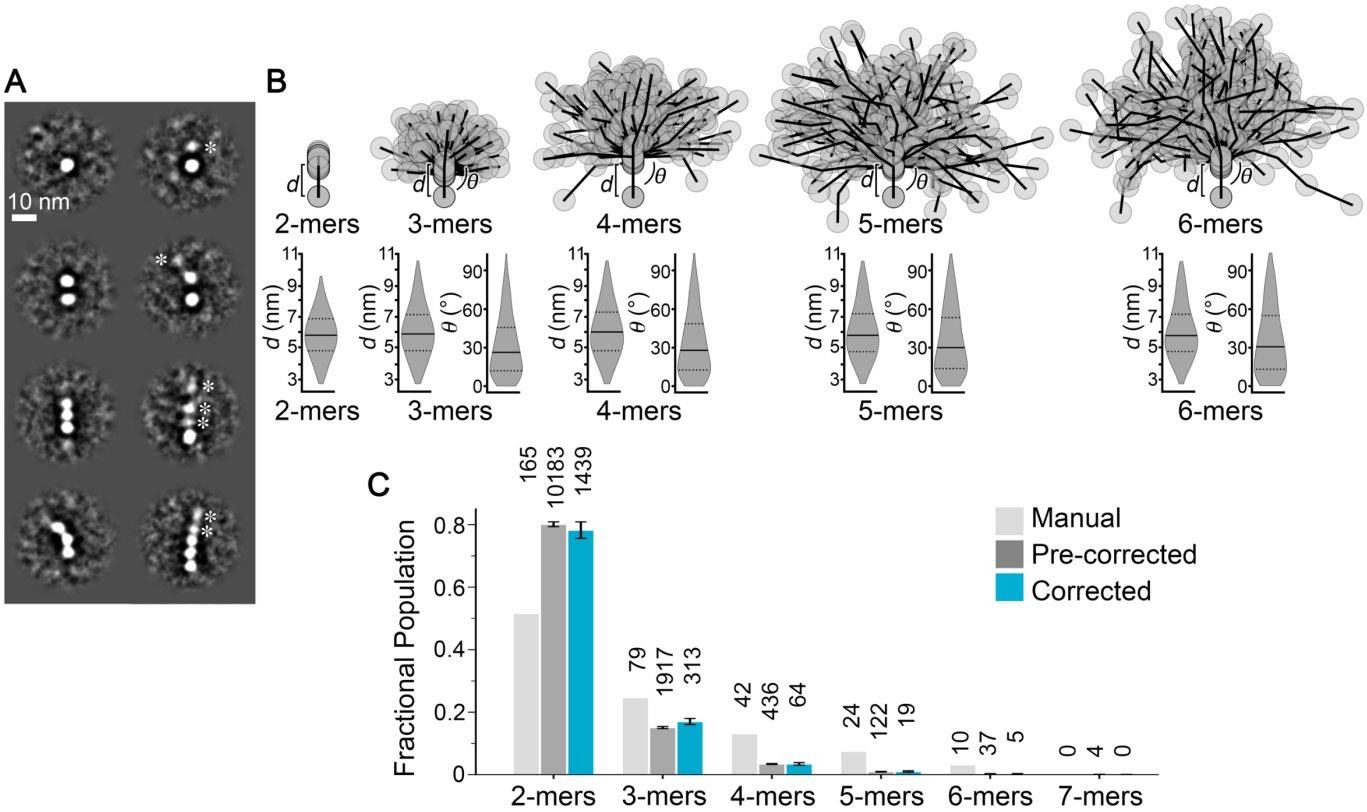
921 B) Illustrative representation of classified LC8/Nup159 oligomers obtained by automated single-particle
922 distribution analysis (n = 100 for each class of oligomer). Individual *n*-mers are aligned along the
923 connection of the first two LC8 dimers (represented as grey circles). The conformational heterogeneity is
924 illustrated by the violin plots showing the distribution of separation distances (d) and bend angles (Θ) of
925 all oligomers identified from a total of 30 micrographs. The solid line indicates the corresponding median
926 and the dotted lines indicate the corresponding first and third quartiles. A full table of statistics is reported
927 in Supplemental Table 2.

928

929 C) The fractional population distribution of oligomeric states obtained by automated single-particle
930 distribution analysis prior to correction (gray) and following statistical correction (blue). Total number of
931 classified states are indicated above each bar. Error bars correspond to the effective standard deviation
932 as described in Methods.

933

934 **FIGURE 6**



935

936 **Fig. 6: Quantitative characterization and statistical correction of LC8/*d*ASCIZ complex assemblies.**
937 A) Representative 2D class-averages depicting, *(left column)* free LC8 and a range of assembled
938 LC8/*d*ASCIZ complexes ranging from 2-mers to 3-mers present in the image dataset that were well-
939 resolved, and *(right column)* complexes displaying varying degree of heterogeneity. Asterisk indicate
940 densities of LC8 that display blurred features that are less well-resolved, indicative of unresolved
941 conformational/configurational heterogeneity. Scale bar = 10 nm.

942

943 B) Illustrative representation of classified LC8/*d*ASCIZ oligomers obtained by automated single-particle
944 distribution analysis (n = 100 for each class of oligomer, except for the 6-mer class which was limited to n
945 = 37). Individual *n*-mers are aligned along the connection of the first two LC8 dimers (represented as grey
946 circles). The conformational heterogeneity is illustrated by the violin plots showing the distribution of
947 separation distances (d) and bend angles (Θ) of all oligomers identified from a total of 302 micrographs.
948 The solid line indicates the corresponding median and the dotted lines indicate the corresponding first and
949 third quartiles. A full table of statistics is reported in Supplemental Table 2.

950

951 C) The fractional population distribution of oligomeric states obtained by manual inspection (light gray),
952 automated single-particle distribution analysis prior to correction (dark gray) and following statistical
953 correction (blue). Total number of classified states are indicated above each bar. Error bars correspond to

954 the effective standard deviation as described in Methods. The populations obtained by manual inspection
955 are based on a subset of only 50 micrographs.
956

957 **SUPPLEMENTAL INFORMATION**

958

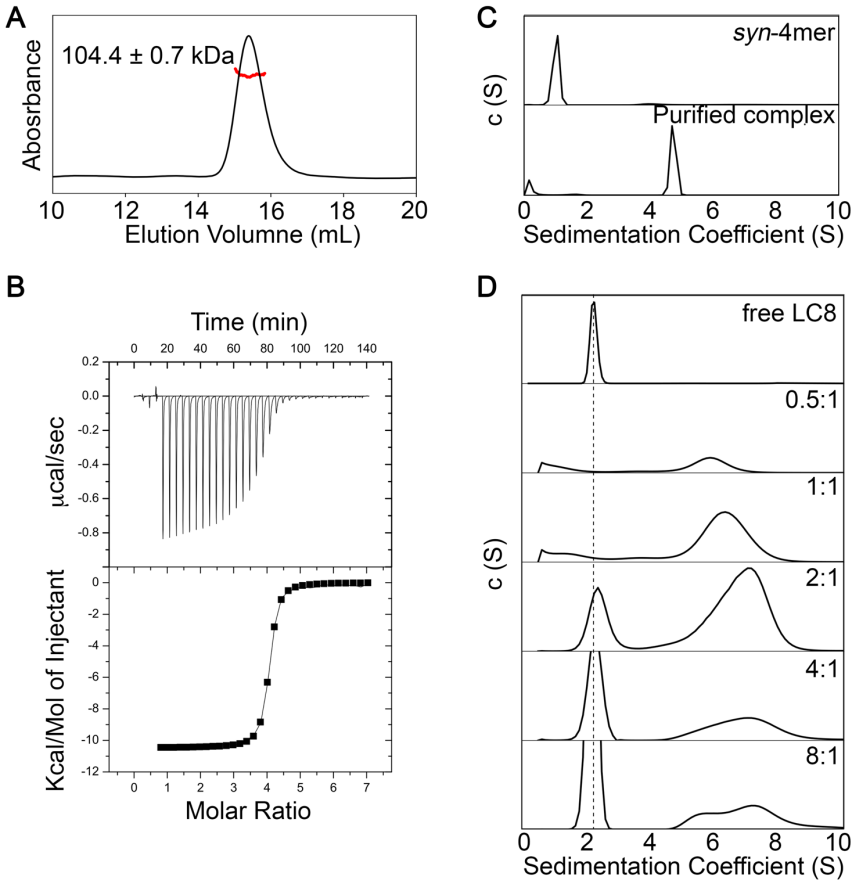
959 (SUPPLEMENTAL FIGURES, TABLES AND LEGENDS)

960

961 **SUPPLEMENTAL FIGURES AND LEGENDS**

962

963 **SUPPLEMENTAL FIGURE 1**



964

965 **Supplemental Fig. 1: Sedimentation velocity analytical ultracentrifugation (AUC) of LC8 complexes.**
966 A) SEC-MALS of *syn*-4mer in complex with LC8. Purified complex eluted as a single peak, with a mass of
967 104.4±0.7 kDa, within uncertainty of the expected mass for a 2:8 complex, 105.2 kDa.

968

969 B) Isotherm of binding between LC8 and the syn-4mer. The isotherm fits well to a simple binding model
970 with K_d = 36±3 nM, ΔH = 10.47±0.04 kcal/mol, and N=3.98±0.01. Model fit is shown as a line.

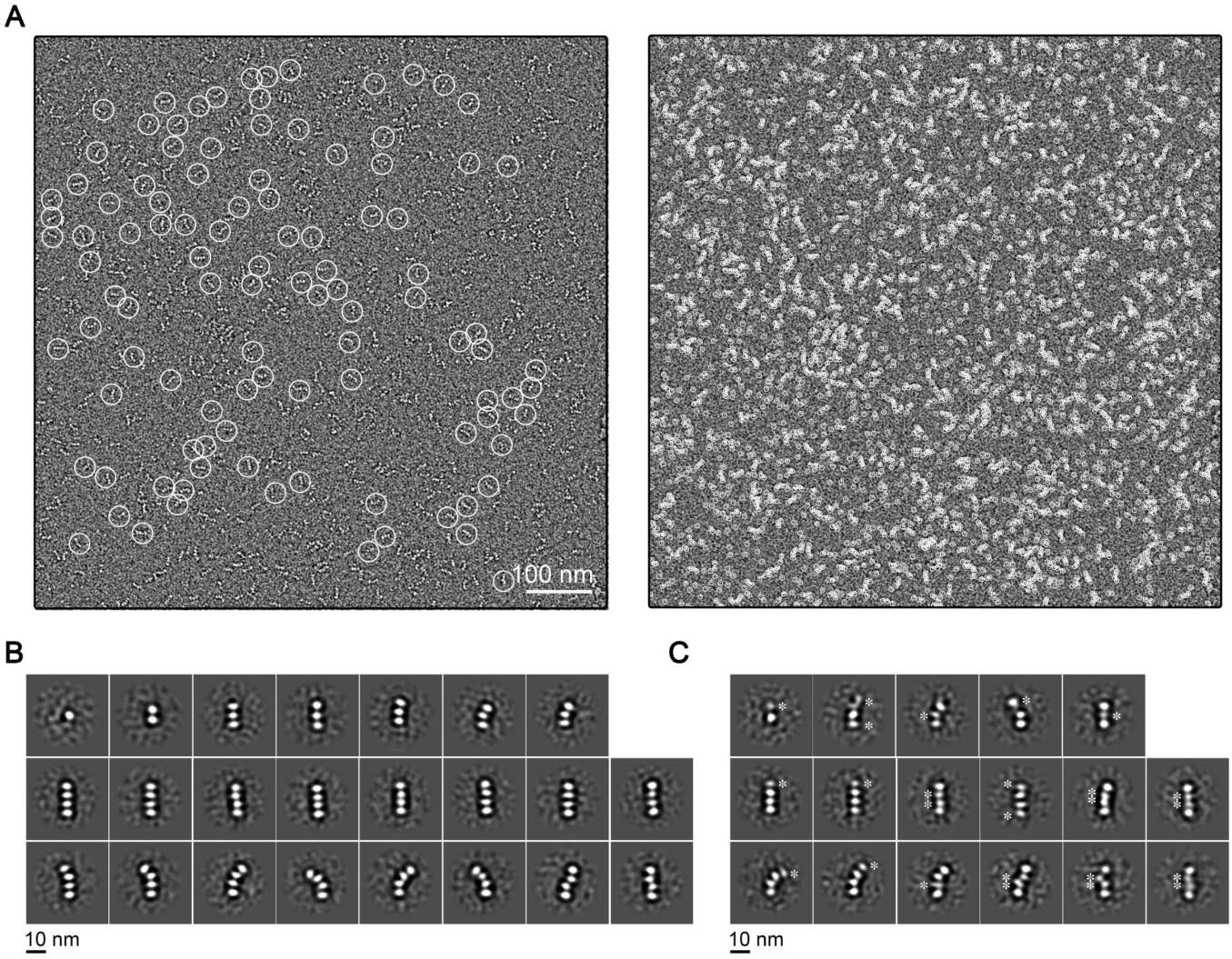
971

972 C) AUC data for the syn-4mer and size exclusion purified LC8/syn-4mer complex. A sharp peak at a
973 sedimentation coefficient of 4.7 S indicates a tight and homogeneous complex.

974

975 D) AUC data for LC8 and LC8/Nup159 complexes formed at increasing ratios of LC8. The dashed line is
976 centered on the LC8 peak. The multiple peaks in the 6-8 S for the complex indicates heterogeneity of the
977 complex and with an S value close to 8, it suggests a higher order assembly than a 5-mer and two Nup159
978 chains.
979

980 **SUPPLEMENTAL FIGURE 2**



981

982 **Supplemental Fig. 2: Single-particle EM analysis of LC8/*syn*-4mer.**

983 A) Representative micrograph of LC8/*syn*-4mer showing *(left)* manually selected oligomers and *(right)*
984 auto-picked LC8 densities using DoG Picker (55). Selected particles are indicated with white circles. Scale
985 bar = 100 nm.

986

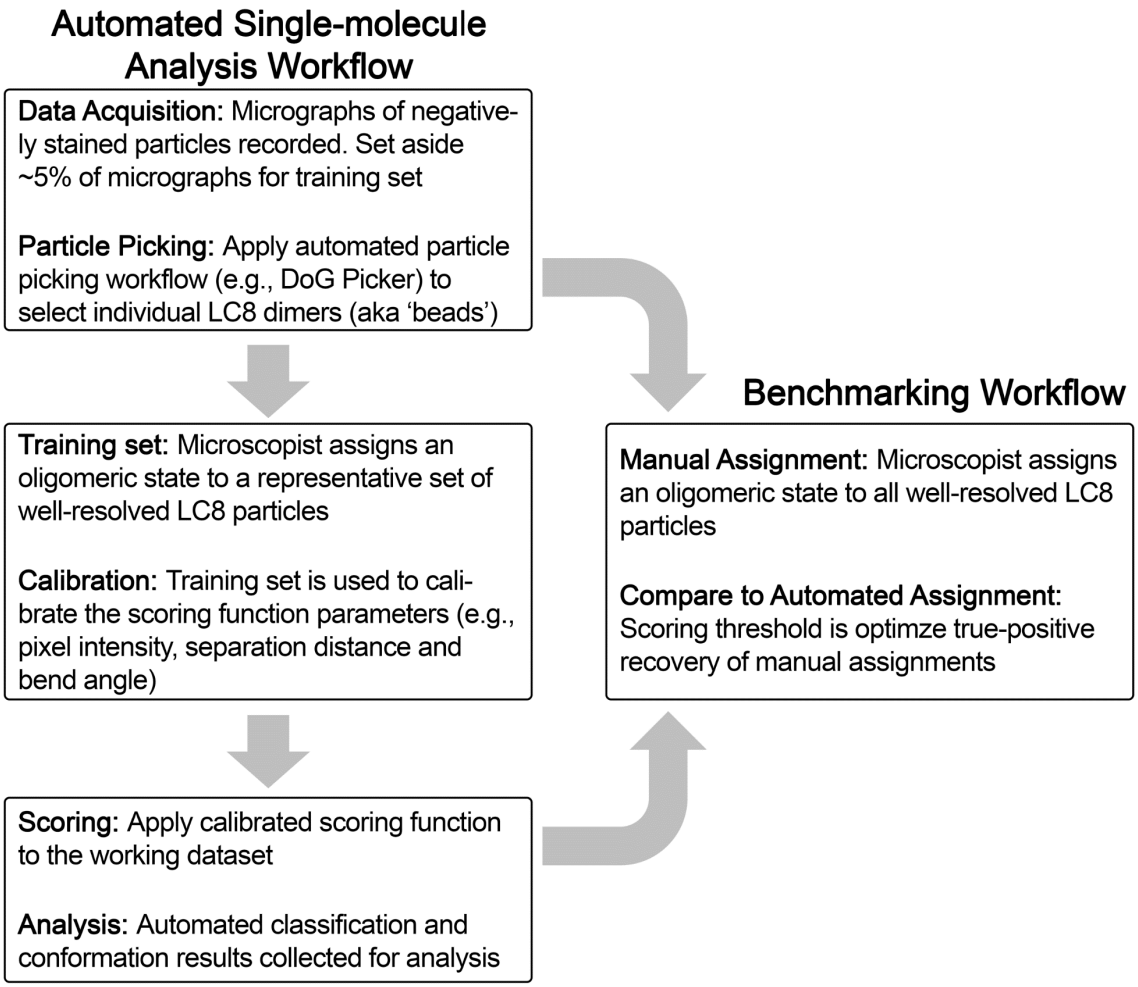
987 B) Expanded set of two-dimensional (2D) class-averages depicting free LC8 *(top left)* and a range of
988 assembled *syn*-4mer/LC8 complexes (2-mers to 4-mers) present in the image dataset that were well-
989 resolved.

990

991 C) Expanded set of 2D class-averages showing assembled *syn*-4mer/LC8 complexes displaying varying
992 degree of conformational heterogeneity. Asterisk indicate densities of LC8 that display blurred features
993 that are less well-resolved, indicative of unresolved conformational/configurational heterogeneity. Scale
994 bar = 10 nm in panels B and C.

995

996 **SUPPLEMENTAL FIGURE 3**



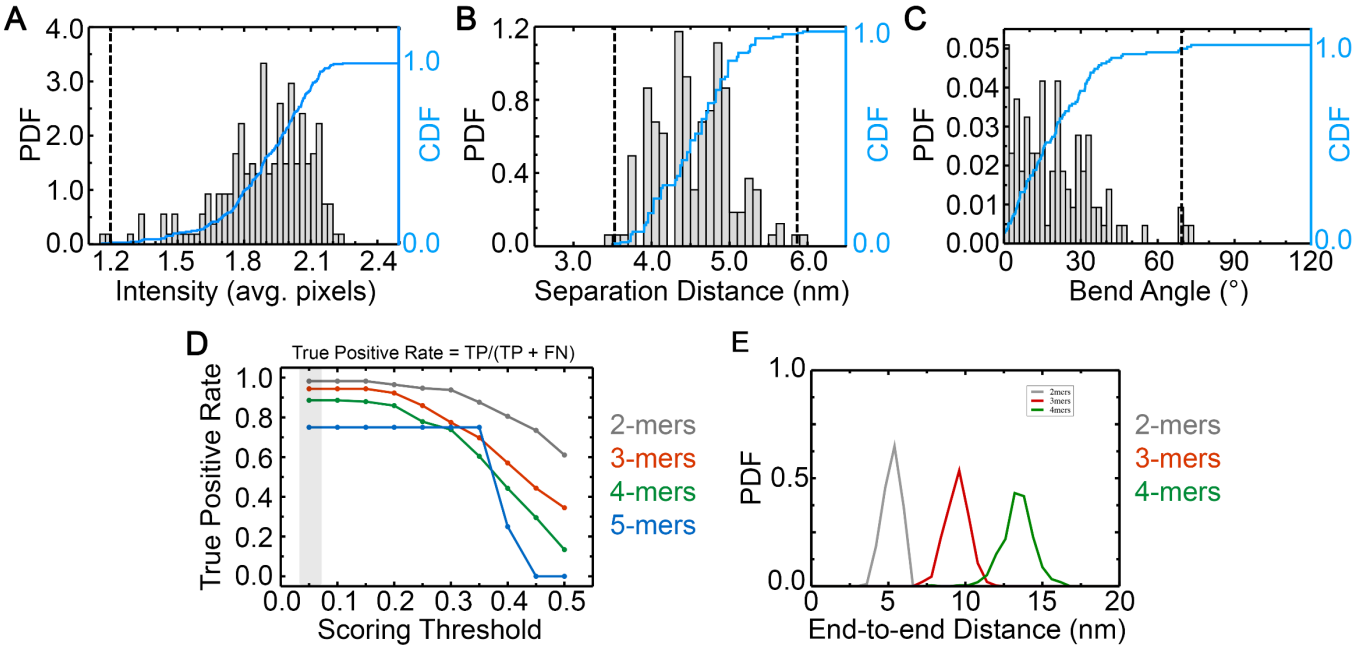
997

998 **Supplemental Fig. 3: Automated single-particle distribution analysis and benchmarking workflow.**
999 Diagram depicting the individual steps of the automated oligomer assignment and benchmarking
1000 procedures. The Analysis Workflow was applied to the LC8/*syn*-4mer, the LC8/Nup159, and the
1001 LC8/*d*ASCIZ datasets. The Benchmarking Workflow was applied to the LC8/*syn*-4mer system to validate
1002 the results.

1003

1004

1005 **SUPPLEMENTAL FIGURE 4**



1006

1007 **Supplemental Fig. 4: Scoring function calibration for the syn-4mer/LC8 complex.**

1008 A – C) Distributions of (A) LC8 particle intensities (n = 216), (B) LC8-to-LC8 distance separations (*d*) (n =
1009 162) and (C) bend angles (n = 108) between sequential LC8-LC8 'bond' vectors (θ), obtained from a
1010 training dataset of syn-4mer/LC8 micrographs. The probability density function (PDF, gray bars) is shown
1011 together with the cumulative distribution function (CDF, blue). Outlier values with CDF < 0.005 and/or >
1012 0.995 were excluded for calibration of the scoring function and given a score of zero, as indicated by the
1013 vertical dashed lines.

1014

1015 D) The True Positive Rate, defined as the fraction of manually assigned oligomers that were also
1016 automatically assigned, for 2-mers (grey), 3-mers (red), 4-mers (green), and 5-mers (blue). A smaller (more
1017 permissive) threshold value = 0.05 yielded the most accurate automatic assignments, and was selected
1018 for application to the full dataset (transparent grey).

1019

1020 E) Distribution of end-to-end distances, *i.e.,* the distance between terminal LC8s in an oligomer, based on
1021 all oligomers predicted for this system. Data displayed independently for 2-mers (gray, n= 1312), 3-mers
1022 (red, n= 485) and 4-mer (green, n= 359).

1023

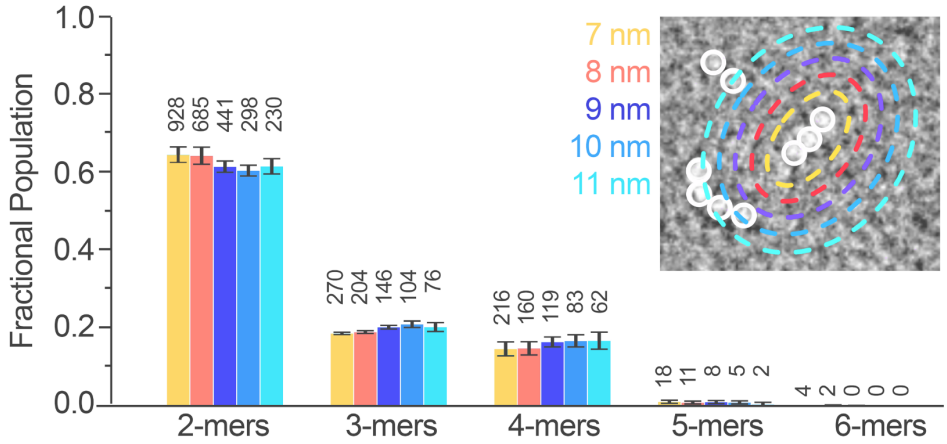
1024

1025

1026

1027

1028 **SUPPLEMENTAL FIGURE 5**



1029

1030 **Supplemental Fig. 5: Effects of distance thresholding on assigned oligomer populations.** The
1031 fractional population distribution of oligomeric states automatically assigned with different values for the
1032 nearest neighbor distance criterion, applied to a training set of LC8/*syn*-4mer particles. The numbers on
1033 top of bars indicate total population counts. *Inset,* zoom of representative region of an electron micrograph
1034 depicting an assigned oligomer in a crowded region of neighboring particles (white circles), with distance
1035 radii indicated as dotted circles. The closest distance from any oligomer in the reference oligomer (inset,
1036 center) to any other LC8 particle is considered.

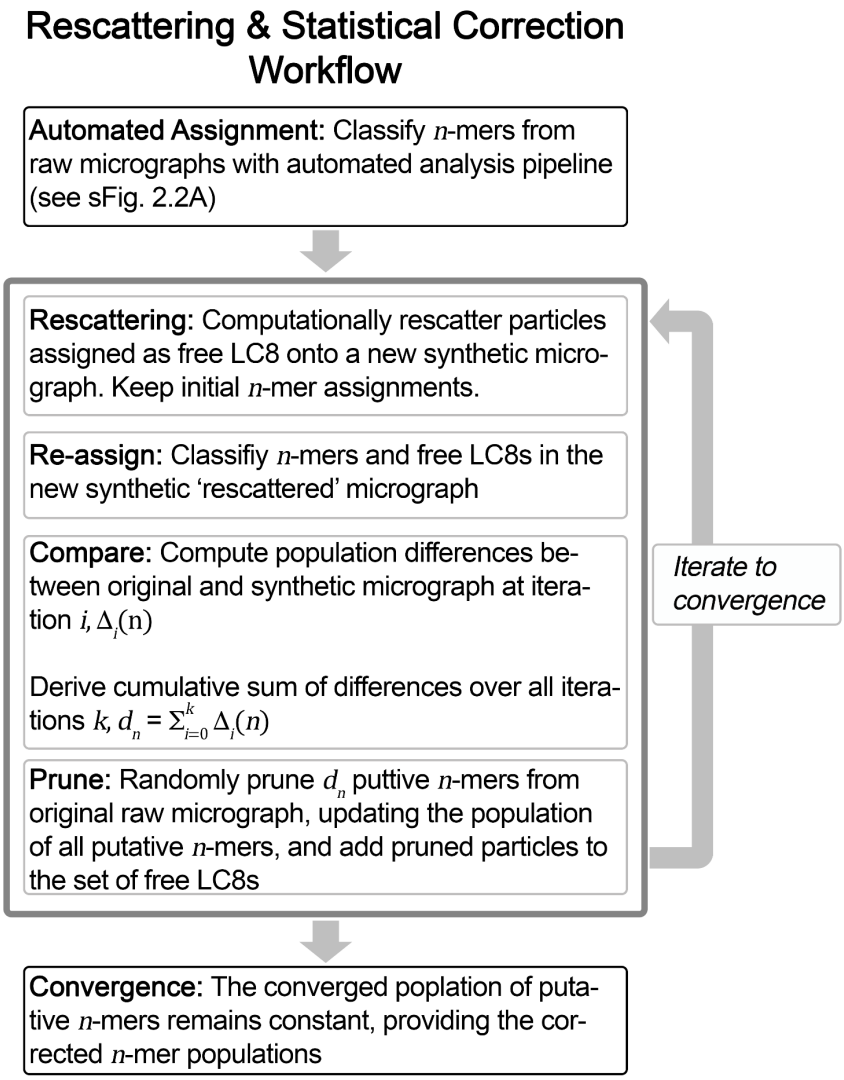
1037

1038

1039

1040

1041 **SUPPLEMENTAL FIGURE 6**



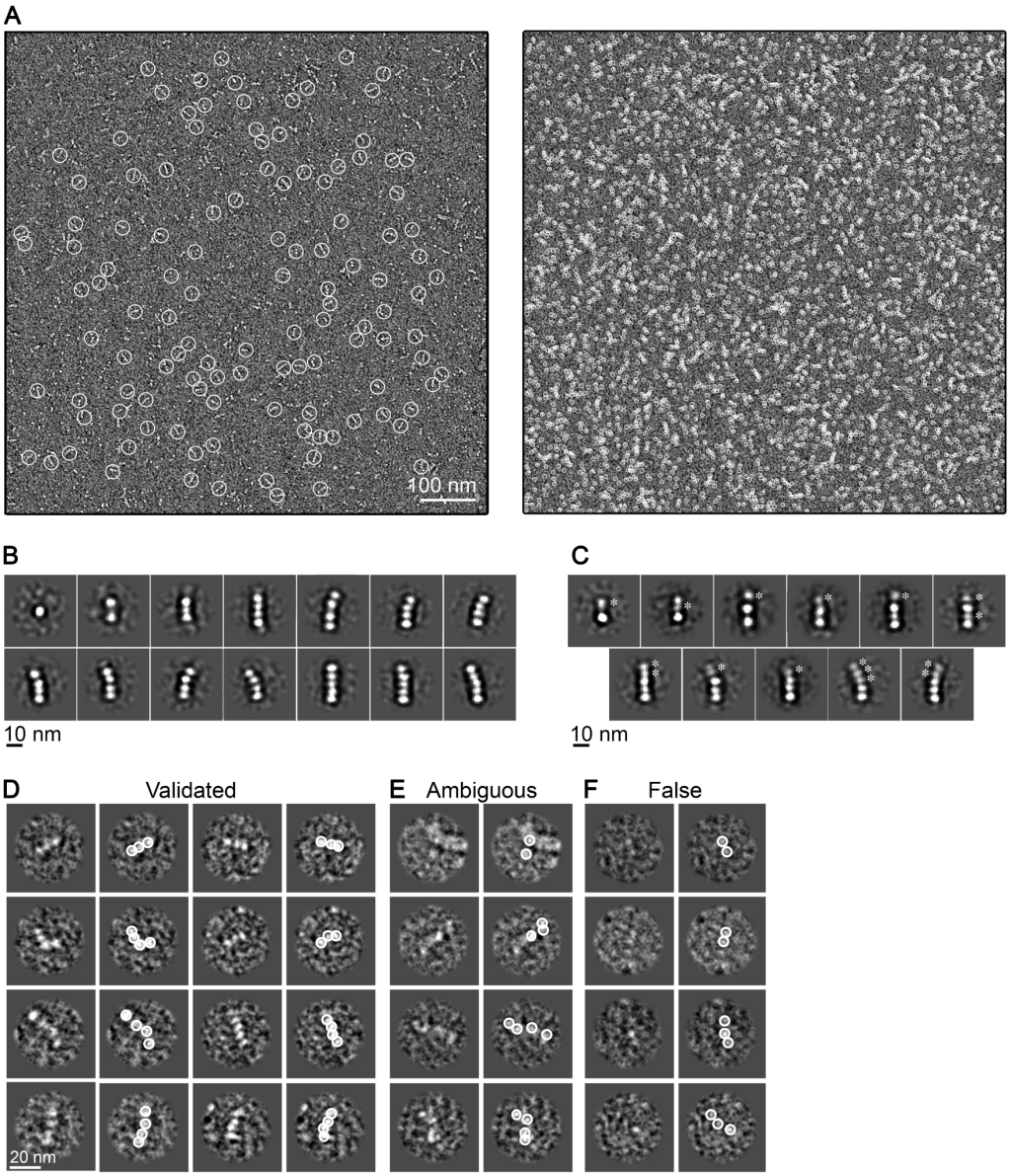
1042

1043 **Supplemental Fig. 6: Flow-chart of the re-scattering and statistical re-scoring protocol.**

1044 Diagram depicting the individual steps of the statistical correction procedure summarized in Fig. 3A.

1045

1046 **SUPPLEMENTAL FIGURE 7**



1047

1048 **Supplemental Fig. 7: Single-particle EM analysis of LC8/Nup159**

1049 A) Representative micrograph of LC8/Nup159 showing *(left)* manually selected oligomers and *(right)* auto-
1050 picked LC8 densities using DoG Picker (55). Selected particles are indicated with white circles. Scale bar
1051 = 100 nm.

1052

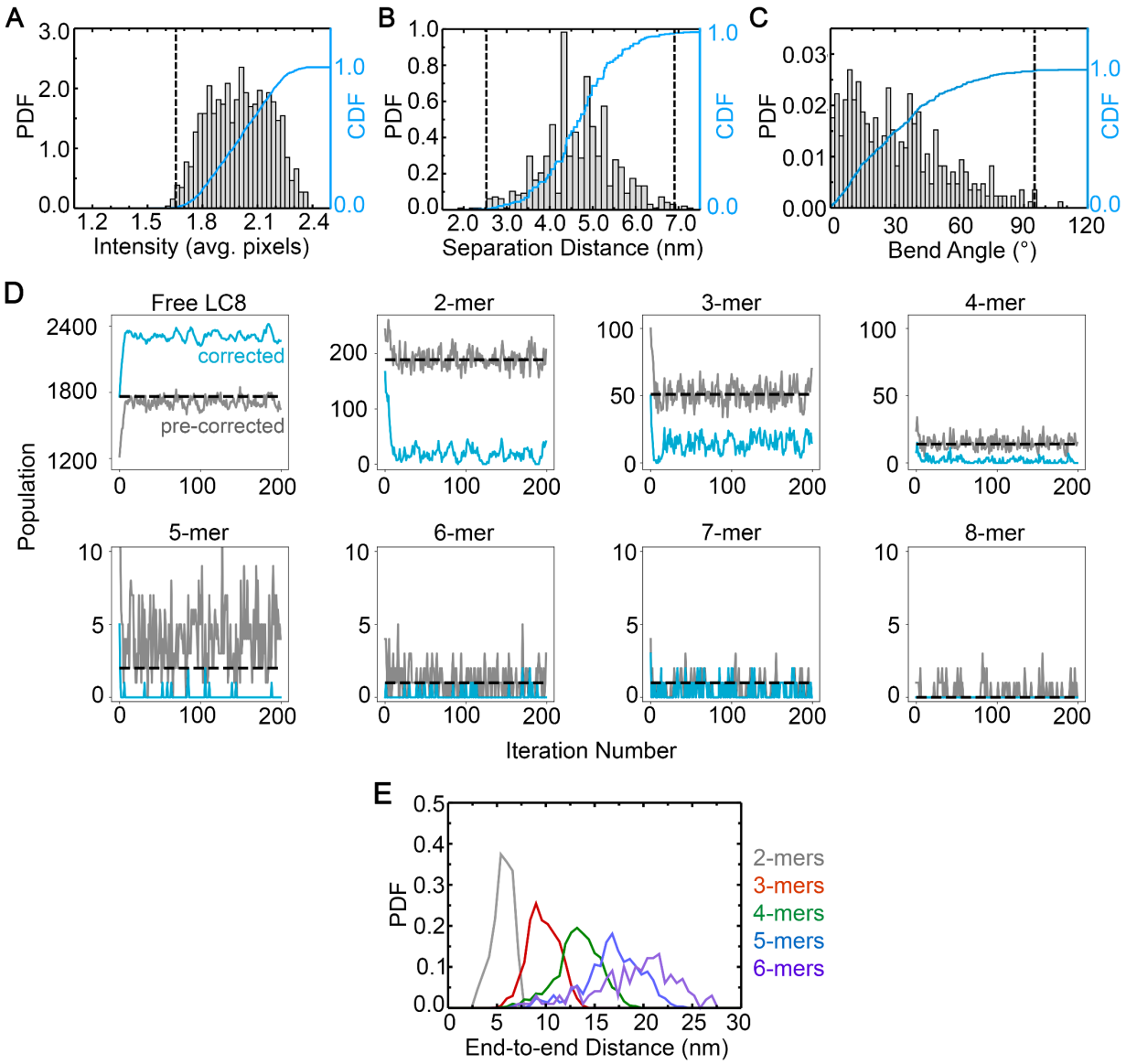
1053 B) Expanded set of representative two-dimensional (2D) class-averages depicting free LC8 (top left) and
1054 a range of assembled LC8/Nup159 complexes (2-mers to 5-mers) present in the image dataset that were
1055 well-resolved.

1056

1057 C) Expanded set of 2D class-averages showing assembled LC8/Nup159 complexes displaying varying
1058 degree of conformational heterogeneity. Asterisk indicate densities of LC8 that display blurred features

1059 that are less well-resolved, indicative of unresolved conformational/configurational heterogeneity. Scale
1060 bar = 10 nm in panels B and C.
1061
1062 D - F) Microscopist validation of automated single-particle assignments. D) Representative images of auto-
1063 assigned LC8/Nup159 oligomers that were deemed acceptable by the microscopists. E) Representative
1064 images of complexes assigned by scoring function and deemed to be too ambiguous to confidently assign
1065 by the microscopists upon evaluation (*e.g.,* containing weak LC8 density and/or neighboring LC8 densities
1066 that were not autopicked). F) Representative images of complexes assigned by the scoring function and
1067 deemed to be too falsely assigned by the microscopist upon re-evaluation (*e.g.,* containing one or more
1068 autopicked densities corresponding to background carbon).
1069
1070 For panels D – F, raw images shown in the left column and autopicked results obtained by DoG Picker
1071 shown by white circles. Scale bar = 20 nm.
1072
1073

1074 **SUPPLEMENTAL FIGURE 8**



1075

1076 **Supplemental Fig. 8: Scoring function calibration, statistical correction and end-to-end distance**
1077 **analysis of Nup159/LC8.**

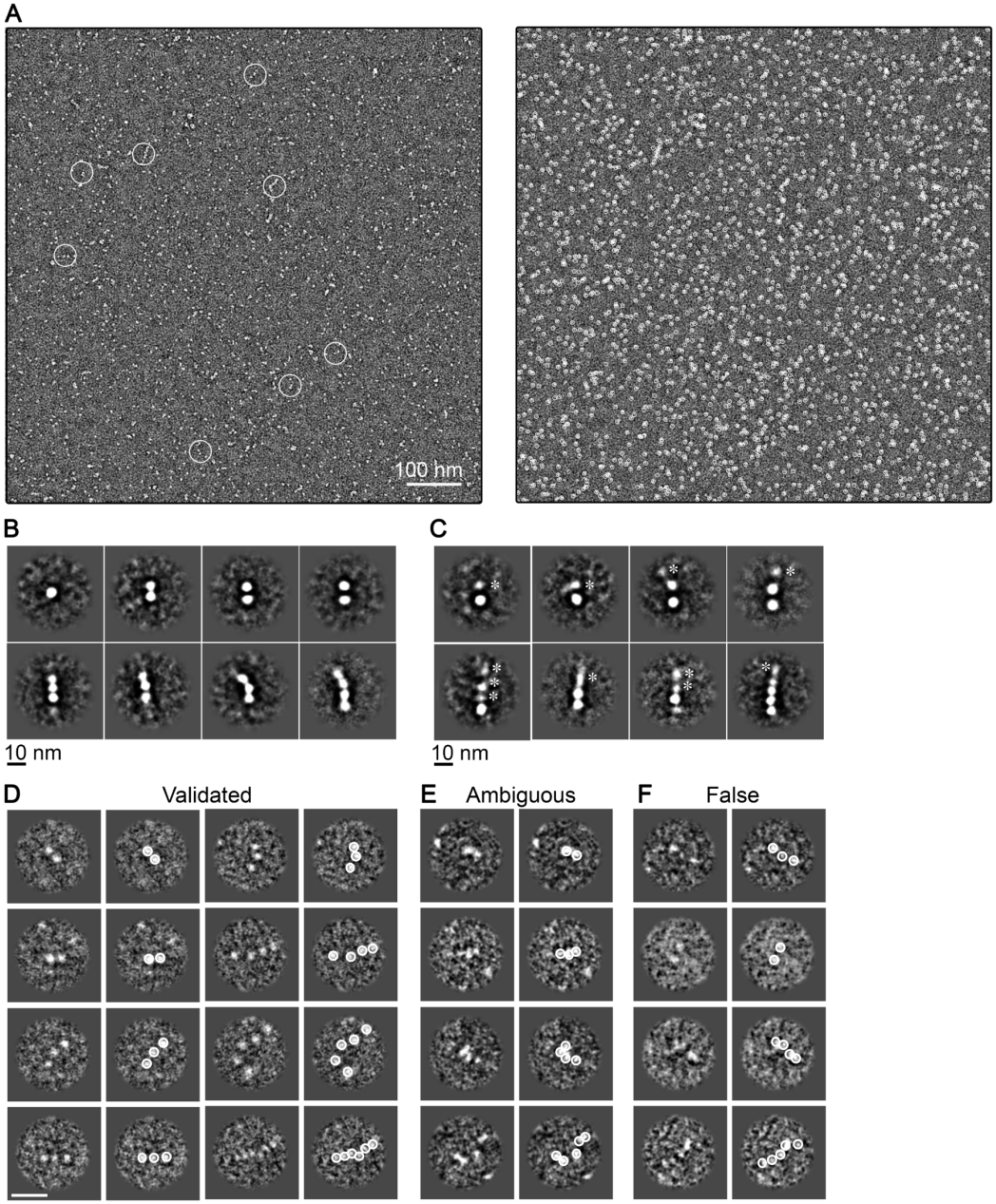
1078 A – C) Distributions of (A) LC8 particle intensities (n = 1037), (B) LC8-to-LC8 distance separations (n =
1079 732) (*d*) and (C) bend angles (n = 427) between sequential LC8-LC8 'bond' vectors (θ), obtained from a
1080 training dataset of Nup159/LC8 micrographs. The probability density function (PDF, gray bars) is shown
1081 together with the cumulative distribution function (CDF, blue). Outlier values with CDF < 0.005 and/or >
1082 0.995 were excluded for calibration of the scoring function and given a score of zero, as indicated by the
1083 vertical dashed lines.

1084

1085 D) The pre-corrected (gray) and corrected (cyan) populations of oligomers and free LC8 as a function of
1086 iteration number during the statistical correction simulation of one example micrograph. The dashed black

1087 lines indicate the corresponding populations of initial oligomer assignments in the experimental
1088 micrograph, to which the gray lines converge.
1089
1090 E) Distribution of end-to-end distances, *i.e.,* the distance between terminal LC8s in an oligomer, based on
1091 all oligomers predicted for this system. Data displayed independently for 2-mers (gray, n= 9786), 3-mers
1092 (red, n= 3083), 4-mer (green, n= 1196), 5-mers (blue, n= 489), and 6-mers (purple, n= 165).
1093
1094
1095

1096 **SUPPLEMENTAL FIGURE 9**



1097

1098 **Supplemental Fig. 9: Single-particle EM analysis of LC8/*d*ASCIZ.**

1099 A) Representative micrograph of LC8/*d*ASCIZ showing *(left)* manually selected oligomers and *(right)* auto-
1100 picked LC8 densities using DoG Picker (55). Selected particles are indicated with white circles. Scale bar
1101 = 100 nm.

1102

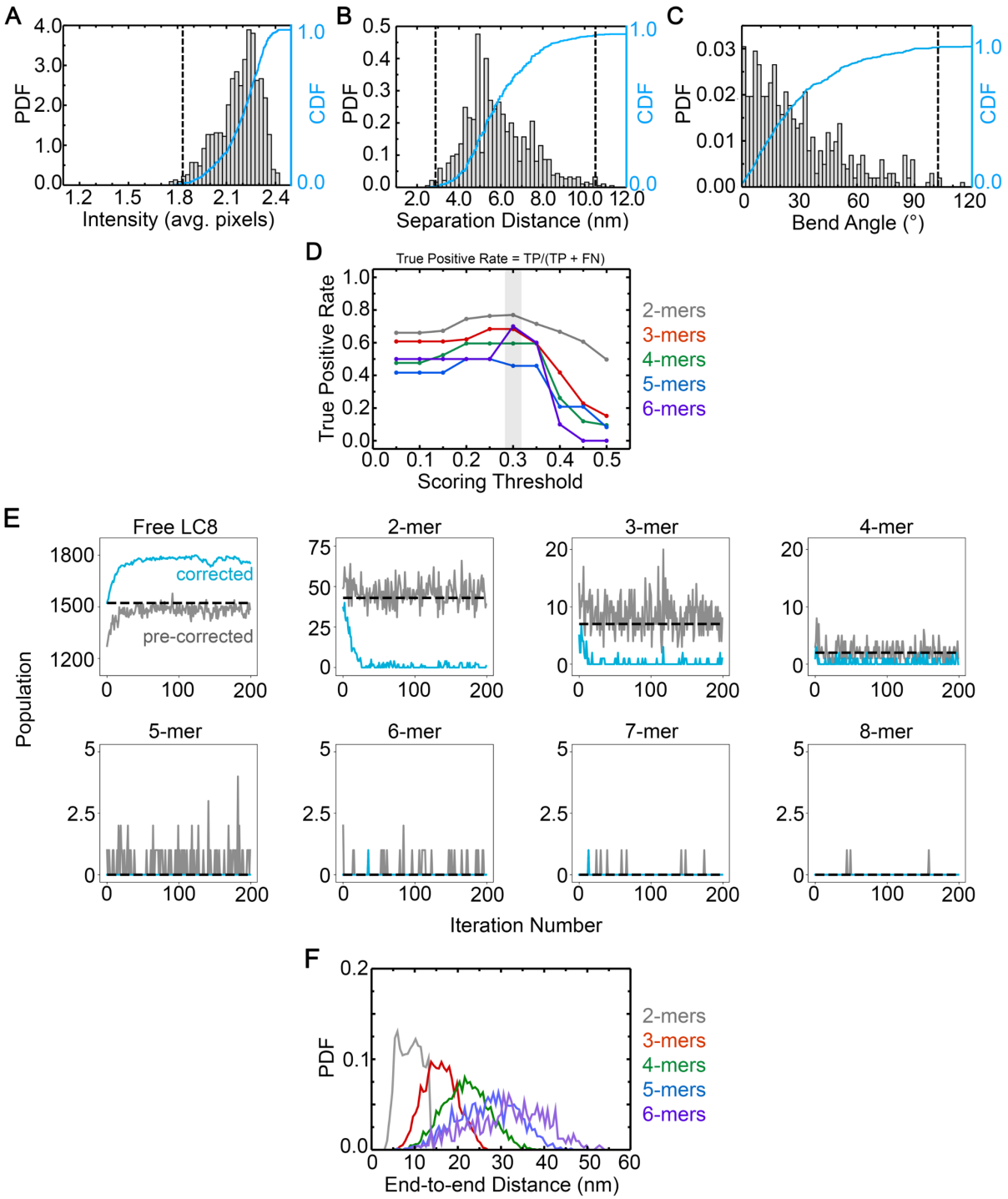
1103 B) Expanded set of 2D class-averages depicting free LC8 (top left) and a range of assembled *d*ASCIZ/LC8
1104 complexes (2-mers to 4-mers) present in the image dataset that were well-resolved.

1105

1106 C) Expanded set of 2D class-averages showing assembled *d*ASCIZ/LC8 complexes displaying varying
1107 degree of conformational heterogeneity. Asterisk indicate densities of LC8 that display blurred features

1108 that are less well-resolved, indicative of unresolved conformational/configurational heterogeneity. Scale
1109 bar = 10 nm in panels B and C.
1110
1111 D - F) Microscopist validation of automated single-particle assignments. D) Representative images of auto-
1112 assigned LC8/*d*ASCIZ oligomers that were deemed acceptable by the microscopists. E) Representative
1113 images of complexes assigned by scoring function and deemed to be too ambiguous to confidently assign
1114 by the microscopists upon evaluation (*e.g.,* containing weak LC8 density and/or neighboring LC8 densities
1115 that were not autopicked). F) Representative images of complexes assigned by the scoring function and
1116 deemed to be too falsely assigned by the microscopist upon re-evaluation (*e.g.,* containing one or more
1117 autopicked densities corresponding to background carbon).
1118
1119 For panels D – F, raw images shown in the left column and autopicked results obtained by DoG Picker
1120 shown by white circles. Scale bar = 20 nm.
1121
1122

1123 **SUPPLEMENTAL FIGURE 10**



1124

1125 **Supplemental Fig. 10: Scoring function calibration, statistical correction and end-to-end distance**
1126 **analysis of *d*ASCIZ/LC8.**

1127 A – C) Distributions of (A) LC8 particle intensities (n = 1360), (B) LC8-to-LC8 distance separations (*d*) (n
1128 = 925) and (C) bend angles between sequential LC8-LC8 'bond' vectors (θ) (n = 509), obtained from a
1129 training dataset of syn-4mer/LC8 micrographs. The probability density function (PDF, gray bars) is shown
1130 together with the cumulative distribution function (CDF, blue). Outlier values with CDF < 0.005 and/or >

1131 0.995 were excluded for calibration of the scoring function and given a score of zero, as indicated by the
1132 vertical dashed lines.
1133
1134 D) The True Positive Rate, i.e., the fraction of manually assigned oligomers that were also automatically
1135 assigned, for 2-mers (gray), 3-mers (red), 4-mers (green), 5-mers (blue), and 6-mers (purple). A threshold
1136 value of 0.3 yields the most accurate automatic assignments and was selected for application to the full
1137 dataset (transparent grey).
1138
1139 E) The pre-corrected (gray) and corrected (cyan) populations of oligomers and free LC8 as a function of
1140 iteration number during the statistical correction simulation of one example micrograph. The dashed black
1141 lines indicate the corresponding populations of initial oligomer assignments in the experimental
1142 micrograph, to which the gray lines converge.
1143
1144 F) Distribution of end-to-end distances, *i.e.,* the distance between terminal LC8s in an oligomer, based on
1145 all oligomers predicted for this system. Data displayed independently for 2-mers (grey, n = 27823), 3-mers
1146 (red, n = 6136), 4-mer (green, n = 1591), 5-mers (blue, n = 530), and 6-mers (purple, n = 215).
1147

1148 **SUPPLEMENTAL TABLES**

1149

1150 **Supplemental Table 1: Overview of EM data collection and parameters.**

	syn-4mer	Nup159	*d*ASCIZ
# micrographs	30	104	305
Defocus range (μm)	1.9–4.0	1.6–4.0	2.0–3.5
Pixel size (Å/pixel)	4.37	4.37	4.37
Magnification	49,000	49,000	49,000
Microscope power (kV)	120	120	120
Picked complexes	4151	5875	2434
Picked LC8 domains	14,333	246,328	557,134
Dog-pick thresholds	4; 1000	4; 1000	4; 1000
Dog-pick radius (pixels)	8	8	8

1151

1152

1153 **Supplemental Table 2: Overview of statistics obtained by singe particle distribution analysis.**

IDP System	Putative oligomers	Avg. bead distance (nm)	Std. dev. of bead distance (nm)	Avg. bead angles (˚)	Std. dev of bead angles (˚)
syn-4mer					
2-mer	1312	4.91	0.55	-	-
3-mer	485	4.79	0.53	29.53	20.23
4-mer	359	4.73	0.51	29.71	19.80
Nup159					
2-mer	9786	5.28	1.01	-	-
3-mer	3083	5.19	0.99	42.46	27.09
4-mer	1196	5.07	1.00	40.11	27.56
5-mer	489	5.03	0.99	39.70	26.42
6-mer	165	5.00	1.04	43.68	26.95
*d*ASCIZ					
2-mer	27823	5.89	1.42	-	-
3-mer	6136	6.12	1.65	31.05	23.19
4-mer	1591	6.20	1.69	32.66	24.22
5-mer	530	6.11	1.72	34.99	25.06
6-mer	215	6.13	1.72	35.37	26.24
7-mer	-	-	-	-	-

1154

1155

1156

1157

1158

1159 **SUPPLEMENTAL MOVIE LEGEND**

1160

1161 **Supplemental Movie 1**

1162 **Single-particle conformational heterogeneity in the *syn*-4mer/LC8 complex.** Montage of selected syn-
1163 4mer/LC8 particles aligned and oriented to display the conformational heterogeneity observed captured
1164 from the raw image dataset by the automated assignment procedure. Individual LC8 dimers are highlighted
1165 by a white.

1166

1167 **Supplemental Movie 2**

1168 **Annotated conformational heterogeneity in the *syn*-4mer/LC8 complex.** Annotated version of syn-
1169 4mer/LC8 particles aligned and oriented as in Figure 2F. Individual LC8 dimers are indicated by a grey
1170 circles.

1171

1172