Automated and statistically corrected identification of flexible multivalent IDP-bound assemblies in electron micrographs

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

[1] Equal contributions

* Corresponding: Elisar.Barbar@oregonstate.edu (EB); reichow@pdx.edu (SLR); zuckermd@ohsu.edu (DMZ)

Abstract

Multivalent intrinsically disordered proteins (IDPs) bound to multiple protein ligands are found in numerous cellular systems. The 'beads-on-a-string' architecture that is common amongst such multivalent IDPs, consists of a highly flexible IDP "string" bound to multiple regulatory or scaffold protein "beads". The inherent conformational flexibility of the IDP, coupled with the potential compositional heterogeneity of ligand assemblies due to low binding affinities has made these systems difficult to characterize structurally. Electron microscopy (EM) has emerged as a powerful tool for structural characterization of heterogeneous protein complexes; however, in cases of continuum dynamics traditional "class averaging" effectively washes out the heterogeneity of primary interest. Furthermore, recently deployed methods in EM for characterizing such highly dynamic systems are not suitable for small proteins (e.g., < 50 kDa), due to a low signal-to-noise ratio. Here, we report automated analysis for a particular class of multivalent IDPs bound to ~20 kDa regulatory 'hub' proteins, which exhibit not only a multiplicity of bound species but also continuous conformational flexibility. The analysis (i) identifies oligomers and provides 'direct' counts of all species, (ii) statistically corrects the direct population counts for artifacts resulting from random proximity of unbound ligand 'beads', and (iii) provides conformational distributions for all species. We demonstrate our approach on a synthetic multivalent four-site IDP, which binds in a parallel duplex fashion to the ubiquitous hub protein, the LC8 homodimer. The duplex IDP architecture allows for potentially greater heterogeneity due to the possibility of off-register assemblies, which could in principle lead to runaway polymerization. We employ negative-stain EM (NSEM) because of its high contrast, which enabled direct visualization of individual LC8 homodimers for single particle analysis, although fundamentally our approach should be applicable to other 'beads-on-a-string'-like systems whenever there is sufficient contrast within the EM dataset. The automated analysis shows a heterogeneous population distribution of oligomeric species that are consistent with manually analyzed data. The statistical correction suggests that five-bead 'off-register' complexes identified in both automated and manual analysis, likely are four-bead oligomers extended by a randomly distributed free LC8 particle. Finally, significant conformational heterogeneity is resolved and characterized for the oligomeric assemblies that were not resolved by traditional 2D class averaging methods.

Introduction

Electron microscopy (EM), and particularly cryoEM, has emerged as a powerful tool for elucidating structure of large biomolecular complexes,[1-4] but highly flexible complexes that display a continuum of conformational states represent a significant challenge to EM that is in the early stages of being addressed in special cases.[5-6] In this study, we demonstrate methodology suited for the particularly challenging 'beads on a string' class of systems which exhibit both conformational and compositional heterogeneity. Our focus is on multivalent complexes consisting of intrinsically disordered protein (IDP) strands which form a duplex ladder-like assembly, reversibly cross-linked by the LC8 hub protein (DYNLL1) which forms the 'rungs' of the 'ladder'-like assembly (Figure 1A). Such LC8 duplexes have emerged as key structural players in cellular complexes ranging from the nuclear pore, to mitotic structures, to transcription machinery.[7-9] However, the inherent dynamical properties and transient formation of multiple oligomeric states, which are key aspects to their cellular function,[10-11] have stymied progress toward understanding the mechanistic details of how this class of protein facilitates such diverse functional roles.

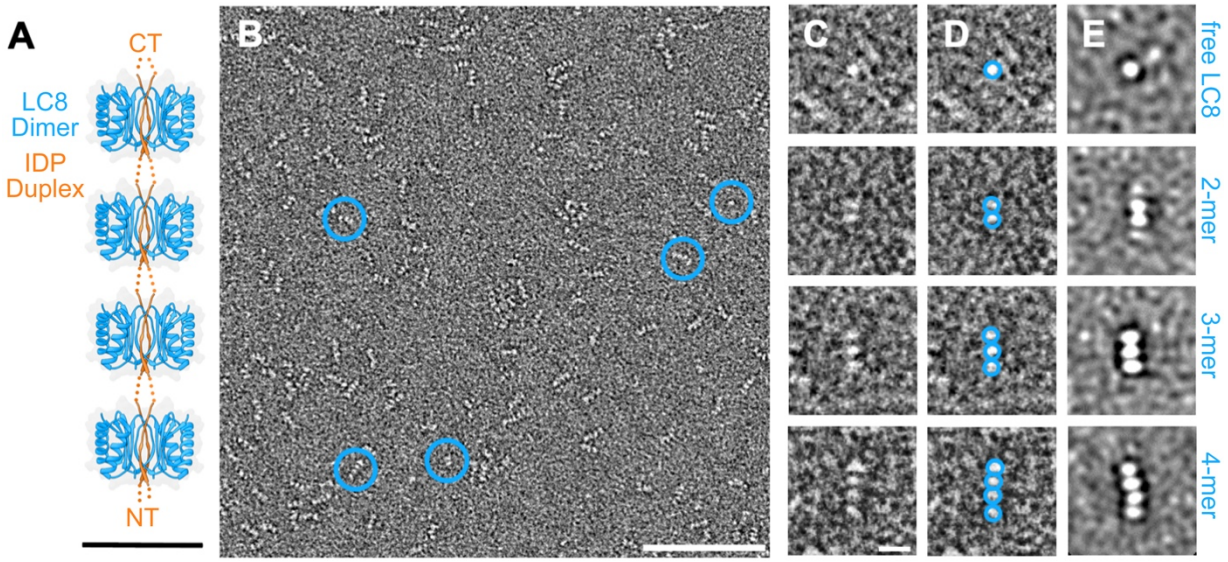

Figure 1. Model and representative EM data of complexes formed by the LC8 hub protein bound to intrinsically disordered peptides. (A) Model of the LC8 homodimer (blue) bound to an intrinsically disordered peptide (IDP, orange) in a parallel duplex fashion with four LC8 binding sites (PDB 3GLW). The N-termini (NT) and C-termini (CT) of the IDP are labeled, and the disordered linker regions of the peptide are represented by dotted lines. Scale bar = 5 nm. (B) Representative micrograph of negatively stained LC8 dimers (white puncta) in complex with a synthetic four-site intrinsically disordered peptide. The IDP is not visible. Representative complexes are circled, to indicate the heterogenous distribution of free and bound LC8-IDP complexes. Scale bar = 100 nm. (C,D) Selected complexes showcasing species containing between 1-4 LC8 dimers. Scale bar = 10 nm. In panel D, individual LC8 particles (LC8 dimers) have been circled in blue. (E) Conventional 2D classification of LC8 species. Scaled the same as panels C,D.

The combination of conformational and compositional heterogeneity of multivalent LC8-IDP duplexes – i.e., a continuum of shape fluctuations and differences in the number of LC8s per complex – rules out common EM analysis software, which predominantly rely on class averages that suppress conformational fluctuations by construction, and clustering/classification methods that presume the existence of discrete states.[12-14] Traditional 2D classification methods were shown explicitly to fail in our analysis of the 11-site LC8-ASCIZ duplex system, due to extreme conformational heterogeneity,[15] thus requiring painstaking manual curation of the EM image dataset. The commonly used single particle EM image processing software, RELION, has a 'multi-body' scheme,[16] but it requires establishing orientations for the individual 'bodies' which is not possible for the 20 kD LC8 dimers, which are far below the detection limit of current Cryo-EM methods and just at the limit of resolvability by negative stain EM. In principle, emerging methods of '3D variability analysis' may be applicable to the continuum flexibility displayed by multivalent duplex IDP systems, and we plan to test these in the future.[6, 17-20]

Here, we establish a fully automated analysis pipeline for inferring both species populations and conformational ensembles from single-particle analysis of negative-stain EM (NSEM) images, which completely bypasses traditional methods of particle averaging. Our computational approach builds on two principles: (i) simplicity and physical interpretability are advantageous; and (ii) analysis should be consistent with the underlying structural features of the specimen. Our scoring and classification of oligomer species builds on simple geometric and polymer principles, while our self-consistent statistical correction accounts for random ordering events which occur due to the presence of randomly distributed free LC8 particles, which occur due to the inherently weak binding affinity (K_d ~ 1uM) and can artifactually appear to form or extend oligomeric assemblies.

This process proceeds in two stages. For the first stage, oligomers are 'directly' identified from EM micrographs using a straightforward clustering and scoring approach detailed below, which relies on a minimum of training data. In brief, after the coordinates of individual LC8 particles (i.e., dimers) are autopicked using existing software,[21] their locations are clustered by a simple 'single-linkage' proximity rule: any two LC8s with centers closer than a threshold are in the same cluster. By construction, no oligomer can belong to more than one cluster, so we need only extract oligomers from one cluster at a time. Oligomers are classified from unbound free LC8 particles and the oligomeric states of bound LC8 particles is assigned (2mer, 3mer, 4mer, etc.) using geometric criteria – based on center-to-center distances and angles formed by three sequential particles (or beads) – trained from a few dozen hand-picked oligomers.

The preceding direct counting process, while it provides an unprecedented view of the ensemble of multivalent species and conformations, must be considered naive because it will inevitably include false-positive (FP) oligomers formed when free LC8s are contiguous with other LC8s or shorter, true-positive (TP) oligomers by random chance. This motivated the development of a second stage of analysis, involving an apparently novel statistical approach capable of estimating TP and FP populations in a self-consistent way. The approach, detailed below, synthetically replicates construction of the experimental micrograph using a two-step *conceptual* process, which applies equally to the experimental process itself: (i) placement of TP oligomers, followed by (ii) random distribution of the population of free LC8 particles. The two-step process and

subsequent analysis can be synthetically repeated with updated 'putative' TP populations – obtained by adjusting current TP estimates – until the direct counts measured from the synthetic micrograph match those observed in the original. The result is a self-consistent estimate of the TP population that has been corrected for artifactual influence of the abundant free LC8 population.

With a statistically corrected set of putative TP oligomers, we then generate large conformational ensembles describing the conformational sampling for each oligomer species, extracted from the coordinates of the original micrographs at the single-particle level. We are unaware of previous reports of an automated method for generating conformational ensembles of IDP complexes from EM data. These conformational ensembles can be analyzed for structural details about the oligomers, and lend themselves to further refinement of the geometric scoring process and also as references for future molecular dynamics simulations.

The 'proof of principle' results below demonstrate that the analysis indeed can reliably select multivalent oligomers, as judged by comparison with a held-out test set of hand-scored oligomers. The self-consistent evaluation of TP oligomers appears to be unique in the field and enables comparison with, and potential validation of, other experimental measures including isothermal titration calorimetry and native mass spectroscopy. The unprecedented conformational ensembles described for this class of LC8-IDP complex that are generated for each species offer a wealth of data for further probing these important multivalent systems.

Methods

LC8-IDP complex preparation for EM

We designed a novel LC8-binding peptide using a series of 4 repeats of the amino acid sequence RKAIDAATQTE, taken from the tight-binding LC8 motif of the protein Chica (Uniprot Q9H4H8), spaced by uniform disordered linker sequences, totaling 4 identical motifs separated by 3 linkers (GSYGSRKAIDAATQTEPKETRKAIDAATQTEPKETRKAIDAATQTEPKETRKAIDAATQTEGSY GS). Flanking GSYGS sequences were added to the N and C termini of the constructs to allow for quantitation.

A gene sequence for the LC8-binding 4-mer was purchased as a block (integrated DNA technologies, Coralville, Iowa) and cloned into a pET24d expression vector with an N-terminal Hisx6 affinity tag and a tobacco etch virus protease cleavable site. LC8 from Drosophila melanogaster was also cloned into a pET24d vector with the same affinity tag and cleavable site. Proteins were expressed in ZYM-5052 autoinduction media at 37° C for 24 hr. Cells were harvested and both proteins were purified on a TALON resin, with the synthetic 4-mer purified under denaturing conditions. For LC8, the Hisx6 tag was cleaved by tobacco etch virus protease, and further purified in a reverse affinity chromatography step. LC8 and the 4-mer were further purified with a gel filtration step on a Superdex 75 column (GE Health). All proteins were stored at 4° C and used within one week of purification.

LC8 complex samples were prepared for electron microscopy studies by mixing excess of the purified LC8 with the synthetic IDP 4-mer and purifying the complexes by size-exclusion

chromatography (SEC; Superdex 200, in a buffer of 25 mM tris pH 7.5, 150 mM NaCl and 5 mM BME). Negative stain EM grids were prepared by diluting the LC8 complexes to a final concentration of 16 nM (presumed to be fully bound) in SEC buffer. A 3 µl drop of sample was applied to a glow-discharged continuous carbon coated EM specimen grid (400 mesh Cu grid, Ted Pella, Redding, CA). Excess protein was removed by blotting with filter paper and washing the grid two times with dilution buffer. The specimen was then stained with freshly prepared 0.75% (wt vol^{-1}) uranyl formate (SPI-Chem).

Electron microscopy

Negatively stained specimens were imaged on a 120 kV TEM (iCorr, FEI) at a nominal magnification of 49,000x at the specimen level. Digital micrographs were recorded on a 2K × 2K CCD camera (FEI Eagle) with a calibrated pixel size of 4.37 Å $pixel^{-1}$ and a defocus of 1.5 – 2 µm. A training dataset obtained from 4 micrographs was picked in an automated fashion to select the center of ~4 – 5 nm densities, corresponding to individual LC8 dimers, using DoG-picker[21] with settings for radius equal to 8 pixels and optimal thresholds ranging from 4.0 – 4.4, resulting in ~2000 – 3700 particle picks per micrograph with minimal contribution from background, assessed manually. Using these particles and referencing the micrograph for confirmation, a training set of 14,306 particles was generated. A separate validation set of 5 micrographs was prepared similarly, using DoG-picker, yielding a total of 17,245 particles.

For use in method development and validation studies, the training dataset was curated by the microscopist, who is familiar with the LC8-IDP structure (see Figure 1A) and the NSEM dataset,[15] to manually classify a representative set of LC8 oligomers as 2-mers, 3-mers, 4-mers, etc. To minimize ambiguity, the microscopist selected complexes that were well separated from neighboring particles on the micrograph (see Figure 1B-D). This procedure resulted in a curated set of 54 oligomers of varying valency (216 LC8 particles in total) that were used for calibration of our automated analysis workflow. For further comparative analysis, a traditional dataset of 817 putative LC8-IDP oligomers and free LC8 particles were manually selected from the training micrographs using EMAN2[13] (i.e., by selecting the center of mass of the putative complex), extracted with a box size of 128 pixels and processed using reference-free 2D classification methods (as shown in Figure 1E).

Automated identification and population counting of oligomers

The x,y coordinates obtained from the curated training data, described above, were used to calibrate the automated analysis by using inter-particle distances (center-to-center) and geometric angles (defined by coordinates of three adjacent particles) from the manually classified oligomers (see Figure S1).

To make the computations easily tractable, we used a 'divide and conquer' approach of clustering, followed by detailed geometric analysis (Figure 2). Single-linkage clustering of all LC8 coordinates from the auto-picked micrographs was first performed. In this clustering method, data points that are separated by less than a given distance threshold are grouped together, which is ideal for distinguishing sets of particles from each other that cannot form an oligomer based on the particle coordinates. The threshold was set to a value (6.5 nm) that is assumed to be larger

than the typical separation of neighboring LC8 binding sites on the IDP, as derived from the distance distribution of sequential particles of the oligomers in the curated training set (see Figure S1).

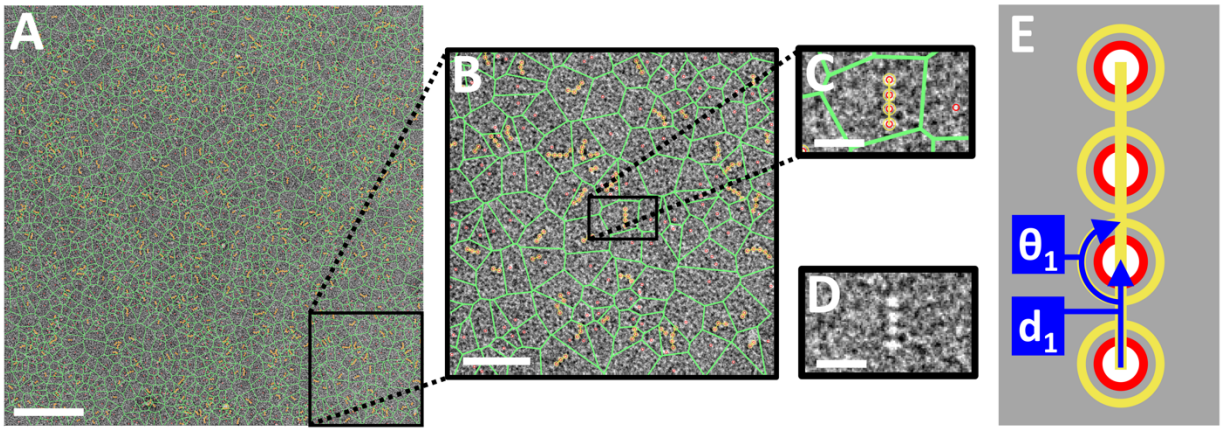

Figure 2. Automated oligomer assignment in negative stain electron micrographs. (A) A representative micrograph of negatively-stained LC8-IDP complexes. Individual LC8 particle picks are highlighted by red circles. Single-linkage clusters of particles are indicated by the partitioning in corresponding cells (green edges). Oligomers assigned by our automated analysis are highlighted by yellow circles around the corresponding particles and a connector line between them. Scale bar = 150 nm. (B) Zoom view of panel (A), to better illustrate the annotated micrograph. Scale bar = 50 nm. (C) A further magnification of the section shown in (B) to visualize one of the automatically assigned 4-mers. Scale bar = 15 nm. (D) The unannotated micrograph section shown in (C), showing the four neighboring LC8 particles. (E) A schematic representation of the assigned 4-mer shown in (C), which can be described by its three sequential inter-particle distances and two sequential inter-particle angles. One such distance (d_1) and one such angle (θ_1) are illustrated as blue arrows.

To obtain oligomer assignments from the clustered particles, a greedy algorithm was applied that takes into consideration every possible combination of particle sequences (or oligomeric states) within a cluster and scores them independently. As visualized in Figure 2E, the scoring algorithm is informed by the oligomer geometry, i.e., particle-to-particle separation (distance, d) and angles (θ) defined by three adjoining particles. The total score for any *n*-mer is the normalized sum over all of its sequential distance and angle log-probability scores:

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

where P_d and P_θ are the probability scores for the distance d or angle θ between two or three given particles, respectively, based on the scoring functions shown in Figure S1. The potential oligomers were then ranked by their length and their total score, thus giving preference to longer

assemblies over shorter ones, and the highest-ranked non-overlapping oligomers in each cluster were saved.

In order to prevent assignments in crowded, ambiguous regions of the micrograph, we filtered oligomer assignments by a distancing threshold, counting only those oligomers that were >9.0 nm from other LC8 particles whether oligomeric or free. This same threshold was applied to the manual curated dataset as well, to facilitate direct comparison. No additional filtering was imposed to prevent classification of longer oligomers – which does occur in 'naive' analysis – see Results.

Correcting oligomer populations through self-consistent statistical re-scoring

The accuracy of the oligomer prediction depends not only on the correct assignment of TP oligomers but also on the identification of artifactual structures that should not be counted – i.e., spurious FP oligomers resulting from random proximity of free LC8 particles not bound to any IDP. For example, such proximity would extend TP *n*-mers to be wrongly counted as (*n+1*)-mers. To provide an estimate for the actual number of the underlying TP oligomers, we iteratively simulated the experimental process of random placement of single LC8 particles and evaluated the degree of FP oligomer creation. As described below, the number of free LC8 particles and putative oligomer assignments may change between iterations, however the overall number of LC8 particles on the synthetic micrographs remains constant. The process is repeated until self-consistency with the direct oligomer populations is obtained – *i.e.,* putative TP oligomers and free LC8's are adjusted until synthetic micrographs yield the originally observed populations, using the naive scoring process described above.

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 original identification and scoring process. The single particles are 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 produces a synthetic micrograph that includes all predicted oligomers from the corresponding 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. This is because the random relocation of free LC8 particles 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). In this way, we simulate the possible artifacts that may also have occurred in the experimental specimen preparation process. 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 this 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 obtained from the synthetic micrographs, as compared to the original micrograph. To this end, we compared, at each iteration *i*, the direct count of each *n*-mer oligomer species (abbreviated by *n*) in the synthetic micrograph 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 pruned. Pruning 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 direct 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 original data.

Results

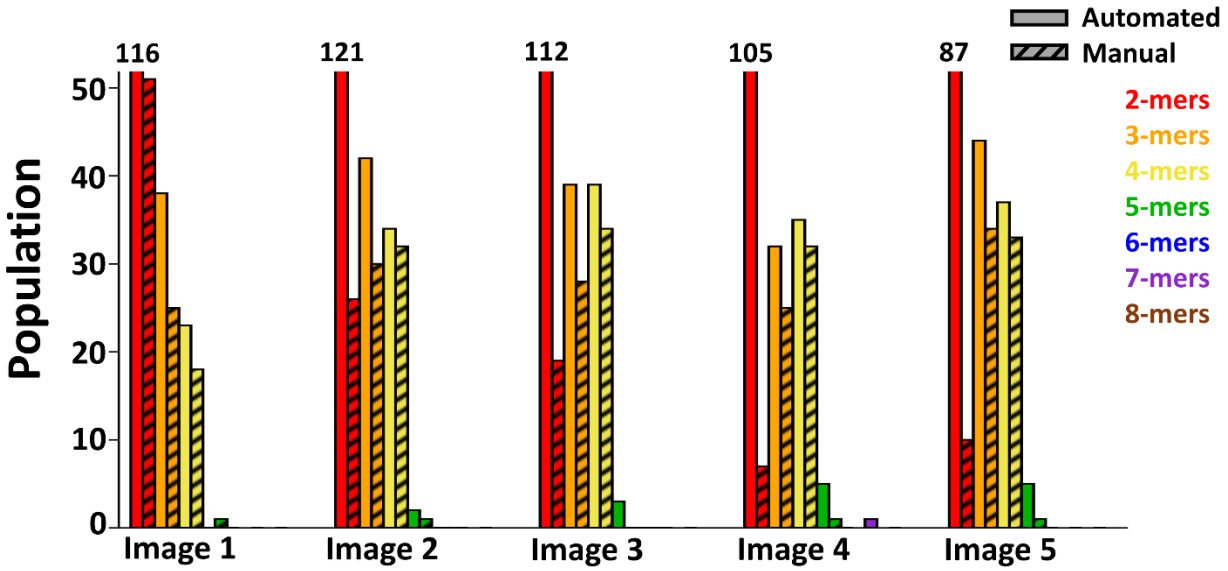

Figure 3. Histograms showing the population distribution of different oligomers, identified manually (hatched bars) or by the automated procedure (filled bars) in 5 different micrograph images. Numbers on top of bars show the population counts for the 2-mers.

Comparison of manual and automated oligomer populations. We performed a double-blind comparison of manually and automatically generated oligomer populations in the set of five test

micrographs (Figure 3). Almost all manually-labeled oligomers were detected by the fully automated scheme. In fact, when applying a 9 nm separation criterion to both populations, all 2-mers and >95% of all 3-mers and 4-mers assigned manually were found by automated analysis, over all test micrographs. However, this high sensitivity comes with a relatively low precision value, *i.e.,* the ratio of true-positive assignments over all automated assignments, due to the large number of false-positive (FP) picks. In particular, for 2-mers the precision is ~20%, whereas for 3-mers and 4-mers this value lies between 60 – 80%.

A heterogeneous population is found by both manual and automated methods (see Figure 3). Although the number of automatically assigned 2-mers far exceeds the manual count, the populations of other oligomers are comparable between the two methods. Importantly, the oligomer distributions highlight the compositional heterogeneity of the LC8-IDP binding, *i.e.,* different LC8 occupancies can occur in complexes formed with the synthetic peptide, under these experimental conditions (where LC8 concentrations are below the Kd). Of particular interest, a few 5-mers were assigned in all test micrographs, by both manual and automated procedures, suggesting the possibility of 'off-register' alignment of the two IDP strands opening up additional LC8 binding sites.[22] More detailed conformational analysis and comparison between manual and automated analyses is given below.

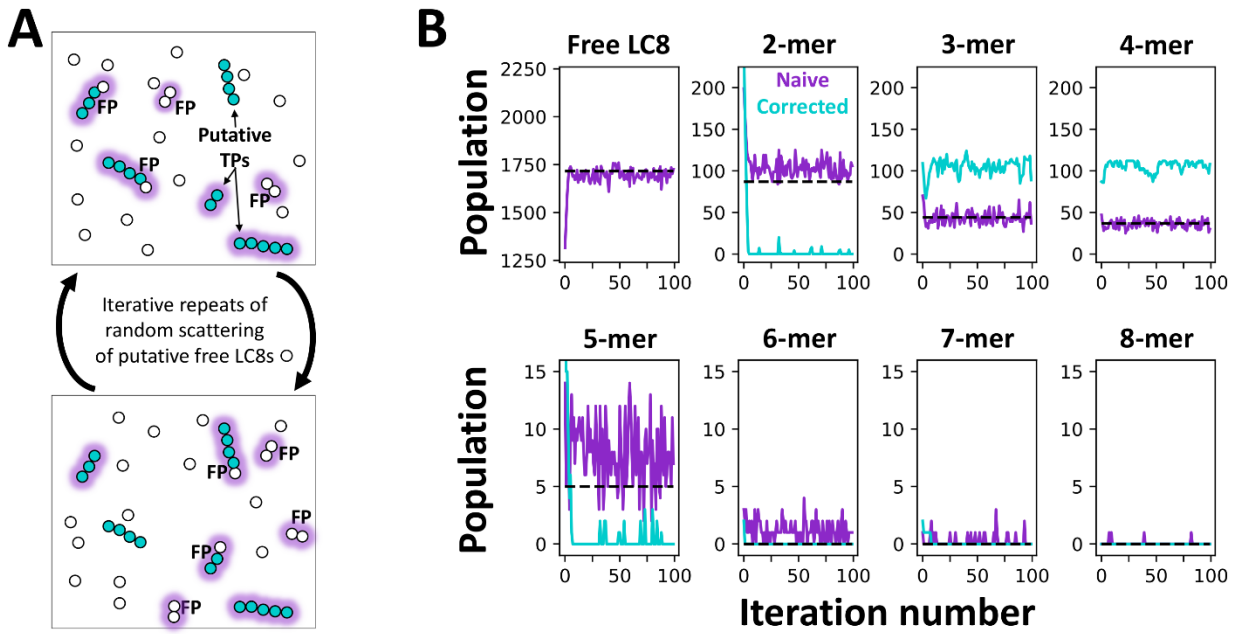

Figure 4. Self-consistent statistical re-scoring and population correction. (A) A schematic showing the assignment of oligomers (purple shadowing) in the experimental micrograph (top). These are categorized as putative true-positive *n*-mers (TP; cyan circles) or false positive (FP) *n*-mer assignments due to the random placement of free LC8 particles (white circles). The repeated pruning of oligomers and random relocating of free LC8 particles leads to synthetic micrographs (bottom) with different oligomer populations. (B) The naive (purple) and corrected (cyan) population counts of different oligomers and free LC8s as a function of iteration number. The set of results presented is for Image 5 in Figure 3. During the simulation the naive counts converge

to the corresponding experimental population count (dashed black line). All data shown are based on automated, not manual, analysis.

Self-consistent, statistical re-analysis. We applied the self-consistent statistical re-scoring strategy to the five test micrographs, which effectively re-distributes free LC8 particles and re-scores the 'synthetic' micrograph in an iterative fashion to obtain an estimate of the FP population (see Methods and Figure 4A). The output of this analysis resulted in a significantly smaller population count of 2-mers (negligible in Image 5, Figure 4B), and a slight increase in the counts of 3-mers and 4-mers, thus rectifying the "naive" oligomer assignment by reducing the effect of FP assignments (Figure 4B). Note that 9 nm distancing was *not* enforced in the corrected populations. On average, the corrected populations are ~30 2-mers and ~80 3-mers and 4-mers per micrograph. Notably, the small population of 5-mers drops to zero after this correction despite the lack of distance filtering, implying that the observed 5-mers in the naive oligomer assignment were most likely TP 4-mers that only appeared to be 5-mers because of the random proximity of a free LC8 particle.

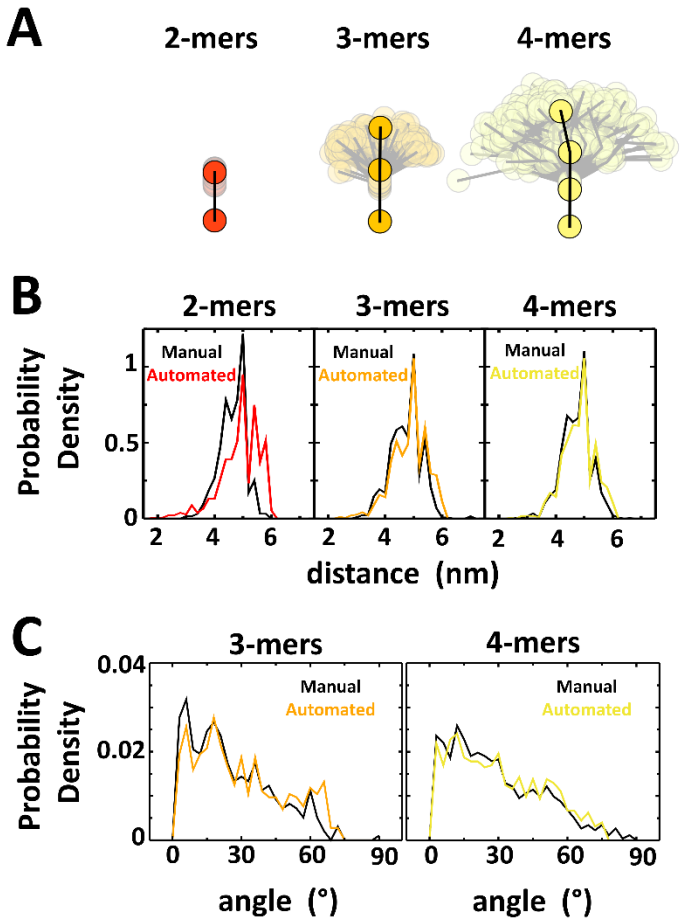

Figure 5. Characterization of LC8-IDP conformational ensembles. (A) Illustration of the conformational ensembles of the different LC8-IDP complexes, displaying the conformational flexibility present in each class. The distributions of LC8-to-LC8 separation distances (in nm) (B)

and geometric angles defined by three neighboring LC8s (C) present in the datasets obtained by manual (black line) and automated analysis (colored lines).

Oligomer conformational ensembles. A significant benefit to our analysis is that the geometric parameters that define each oligomer can be extracted for conformational analysis (Figure 5). This analysis reveals the full conformational ensemble that is present in each of the classified oligomeric states. Visualization of the conformational states in this beads-on-a-string system reveals the significant conformational heterogeneity, particularly for 3-mer and 4-mer species (Figure 5A). Notably, the degree of conformational heterogeneity revealed by this approach is far greater (and more representative of the raw data) than what was provided by traditional 2D class averaging methods (compare to Figure 1E). Furthermore, detailed analysis of the distribution of neighboring particle distances (Figure 5B) and angles (Figure 5C), that can be readily extracted, shows that the flexibility of oligomers identified automatically is comparable to that of manually identified oligomers.

Discussion

The initial results presented here point to the unrealized capacity of negative-stain EM to reveal the single-particle fluctuations of highly flexible and heterogeneous beads-on-a-string systems, in a quantitative fashion. The data also point to the need for further study and methodological refinement.

A number of improvements to our analysis are available. In particular, more data can be exploited and more self-consistency can be enforced. For example, the current analysis does not account for false-negative single-particle picks (*e.g.,* only beads 1, 3, and 4 of a true 4-mer with skipped bead 2), but scoring functions for such cases can readily be constructed. Skipped-bead scoring could be cross-referenced to pixel-based micrograph analysis for confirmation. In a related vein, the self-consistent statistical analysis (Figure 4) could be used to refine the geometrical scoring (Figure S1): the scoring criteria could be self-consistently adjusted to match the best estimate for the set of true-positive oligomers.

Conclusions

We have fully automated the single-particle analysis of 'beads-on-a-string multivalent systems based on electron micrographs. Our approach is notable for (i) quantifying *compositional* heterogeneity – i.e., variation in the number of connected beads/particles; (ii) yielding full two-dimensional *conformational* ensembles of each species; and (iii) developing and implementing a statistical scheme for eliminating artifacts arising from random proximity of single beads to true oligomers. In this initial study on a synthetic system consisting of a four-binding-site IDP binding to the homodimeric LC8 protein, the automated approach performed well as compared to manually analyzed micrographs, capturing >90% of the manually annotated oligomers in a blind test. Remarkably, although five-bead oligomers are possible in principle via off-register binding[22] the self-consistent statistical analysis demonstrates that the appearance of such 5-mers by

NSEM, and under the experimental conditions described, likely are artifacts of random proximity of single LC8 particles.

Acknowledgements

We are grateful to the staff at the OHSU Multiscale Microscopy Core for assistance and training. RM and SLR are supported by the National Institutes of Health (R35-GM124779). We further acknowledge support from the NSF (grants MCB 1715823 and MCB 1617019).

References

1. Cheng, Y., Single-particle cryo-EM-How Did It Get Here and Where Will It Go *Science* **2018,** *361* (6405), 876-880.
2. Fernandez-Leiro, R.; Scheres, S. H. W., Unravelling biological macromolecules with cryo-electron microscopy. *Nature* **2016,** *537*, 339-346.
3. Nakane, T.; Kotecha, A.; Sente, A.; McMullan, G.; Masiulis, S.; Brown, P. M. G. E.; Grigoras, I. T.; Malinauskaite, L.; Malinauskas, T.; Miehling, J.; Yu, L.; Karia, D.; Pechnikova, E. V.; de Jong, E.; Keizer, J.; Bischoff, M.; McCormack, J.; Tiemeijer, P.; Hardwick, S. W.; Chirgadze, D. Y.; Murshudov, G.; Ariscescu, A. R.; Scheres, S. H. W., Single-particle cryo-EM at atomic resolution. *bioRxiv* **2020**, 2020.05.22.110189.
4. Yip, K. M.; Fischer, N.; Paknia, E.; Chari, A.; Stark, H., Breaking the next Cryo-EM resolution barrier – Atomic resolution determination of proteins! *bioRxiv* **2020**, 2020.05.21.106740.
5. Bonomi, M.; Vendruscolo, M., Determination of protein structural ensembles using cryo-electron microscopy. *Current Opinion in Structural Biology* **2019,** *56*, 37-45.
6. Sorzano, C. O. S.; Jimenez, A.; Mota, J.; Vilas, J. L.; Maluenda, D.; Martinez, M.; Ramirez-Aportela, E.; Majtner, T.; Segura, J.; Sanchez-Garcia, R.; Rancel, Y.; Del Cano, L.; Conesa, P.; Melero, R.; Jonic, S.; Vargas, J.; Cazals, F.; Freyberg, Z.; Krieger, J.; Bahar, I.; Marabini, R.; Carazo, J. M., Survey of the Analysis of Continuous Conformational Variability of Biological Macromolecules by Electron Microscopy *Acta crystallographica Section F Structural Biology Communications* **2019,** *75*, 19-32.
7. Stelter, P.; Kunze, R.; Flemming, D.; Höpfner, D.; Diepholz, M.; Philippsen, P.; Böttcher, B.; Hurt, E., Molecular basis for the functional interaction of dynein light chain with the nuclear-pore complex. *Nature Cell Biology* **2007,** *9*, 788-796.
8. Dunsch, A. K.; Hammond, D.; Lloyd, J.; Schermelleh, L.; Gruneberg, U.; Barr, F. A., Dynein Light Chain 1 and a Spindle-Associated Adaptor Promote Dynein Asymmetry and Spindle Orientation *The Journal of Cell Biology* **2012,** *198* (6), 1039-1054.
9. Rapali, P.; Garcia-Mayoral, M. F.; Martinez-Moreno, M.; Tarnok, K.; Schlett, K.; Albar, J. P.; Bruix, M.; Nyitray, L.; Rodriguez-Crespo, I., LC8 Dynein Light Chain (DYNLL1) Binds to the C-terminal Domain of ATM-interacting Protein (ATMIN/ASCIZ) and Regulates Its Subcellular Localization *Biochemical and Biophysical Research Communications* **2011,** *414* (3), 493-498.
10. Cortese, M. S.; Uversky, V. N.; Dunker, A. K., Intrinsic disorder in scaffold proteins: Getting more from less. *Progress in Biophysics & Molecular Biology* **2008,** *98* (1), 85-106.

11. Clark, S. A.; Jespersen, N.; Woodward, C.; Barbar, E., Multivalent IDP Assemblies: Unique Properties of LC8-associated, IDP Duplex Scaffolds *FEBS Letters* **2015,** *589* (19), 2543-2551.
12. Scheres, S. H. W., RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *Journal of Structural Biology* **2012,** *180* (3), 519-530.
13. Tang, G.; Peng, L.; Baldwin, R. P.; Mann, D. S.; Jiang, W.; Rees, I.; Ludtke, S. J., EMAN2: An extensible image processing suite for electron microscopy. *Journal of Structural Biology* **2007,** *157* (1), 38-46.
14. Grigorieff, N., Frealign: An Exploratory Tool for Single-Particle Cryo-EM. *Methods in Enzymology* **2016,** *579*, 191-226.
15. Clark, S.; Myers, J. B.; King, A.; Fiala, R.; Novacek, J.; Pearce, G.; Heierhorst, J.; Reichow, S. L.; Barbar, E. J., Multivalency regulates activity in an intrinsically disordered transcription factor. *eLife* **2018,** *7*, e36258.
16. Nakane, T.; Kimanius, D.; Lindahl, E.; Scheres, S. H. W., Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. *eLife* **2018,** *7*, e36861.
17. Punjani, A.; Fleet, D. J., 3D Variability Analysis: Directly resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM images. *bioRxiv* **2020**, 2020.04.08.032466v1.
18. Zhong, E. D.; Bepler, T.; Berger, B.; Davis, J. H., CryoDRGN: Reconstruction of heterogeneous structures from cryo-electron micrographs using neural networks. *bioRxiv* **2020**, 2020.03.27.003871.
19. Frank, J.; Ourmazd, A., Continuous changes in structure mapped by manifold embedding of single-particle data in cryo-EM. *Methods* **2016,** *100* (1), 61-67.
20. Ludtke, S. J., Single-Particle Refinement and Variability Analysis in EMAN2.1 *Methods in Enzymology* **2016,** *579*, 159-189.
21. Voss, N. R.; Yoshioka, C. K.; Radermacher, M.; Potter, C. S.; Carragher, B., DoG Picker and TiltPicker: Software Tools to Facilitate Particle Selection in Single Particle Electron Microscopy *Journal of Structural Biology* **2009,** *166* (2), 205-213.
22. Reardon, P. N.; Jara, K. A.; Rolland, A. D.; Smith, D. A.; Hoang, H. T. M.; Prell, J. S.; Barbar, E. J., The Dynein Light Chain 8 (LC8) Binds Predominantly "In-Register" to a Multivalent Intrinsically Disordered Partner *The Journal of Biological Chemistry* **2020,** *295* (15), 4912-4922.

Supplemental Information

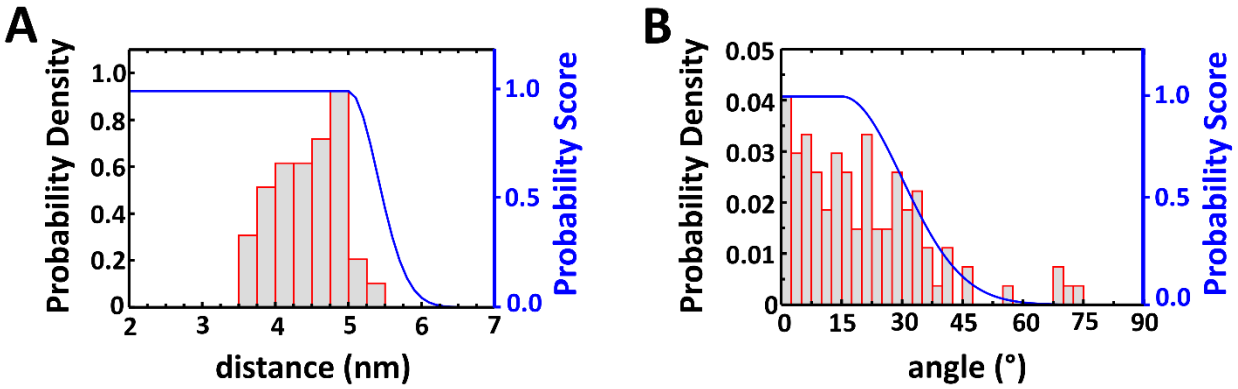

Figure S1. Structural features of ”handpicked” oligomers. The distance (A) and angle (B) distributions of sequential particles from 54 oligomers that were manually selected with great

confidence on the raw micrographs are shown as gray bars. The continuous blue line is the corresponding probability score used for the scoring of each distance and angle between neighboring particles in an oligomer.